The Handbook of Environmental Chemistry 102
Series Editors: Damià Barceló · Andrey G. Kostianoy

Andrea Scozzari · Steve Mounce
Dawei Han · Francesco Soldovieri
Dimitri Solomatine *Editors*

ICT for Smart Water Systems: Measurements and Data Science

Springer

The Handbook of Environmental Chemistry

Volume 102

Founding Editor: Otto Hutzinger

Series Editors: Damià Barceló • Andrey G. Kostianoy

Editorial Board Members:
Jacob de Boer, Philippe Garrigues, Ji-Dong Gu,
Kevin C. Jones, Thomas P. Knepper, Abdelazim M. Negm,
Alice Newton, Duc Long Nghiem, Sergi Garcia-Segura

In over three decades, *The Handbook of Environmental Chemistry* has established itself as the premier reference source, providing sound and solid knowledge about environmental topics from a chemical perspective. Written by leading experts with practical experience in the field, the series continues to be essential reading for environmental scientists as well as for environmental managers and decision-makers in industry, government, agencies and public-interest groups.

Two distinguished Series Editors, internationally renowned volume editors as well as a prestigious Editorial Board safeguard publication of volumes according to high scientific standards.

Presenting a wide spectrum of viewpoints and approaches in topical volumes, the scope of the series covers topics such as

- local and global changes of natural environment and climate
- anthropogenic impact on the environment
- water, air and soil pollution
- remediation and waste characterization
- environmental contaminants
- biogeochemistry and geoecology
- chemical reactions and processes
- chemical and biological transformations as well as physical transport of chemicals in the environment
- environmental modeling

A particular focus of the series lies on methodological advances in environmental analytical chemistry.

The Handbook of Envir onmental Chemistry is available both in print and online via http://link.springer.com/bookseries/698. Articles are published online as soon as they have been reviewed and approved for publication.

Meeting the needs of the scientific community, publication of volumes in subseries has been discontinued to achieve a broader scope for the series as a whole.

ICT for Smart Water Systems: Measurements and Data Science

Volume Editors: Andrea Scozzari · Steve Mounce ·
Dawei Han · Francesco Soldovieri ·
Dimitri Solomatine

With contributions by

C. Agudelo-Vera · L. Alfonso · R. Battaglia · D. Bouziotas · E. Brill ·
M. Castro-Gama · B. Della Ventura · O. Grievson · T. Koster ·
M. Mauro · M. Mazzoleni · S. R. Mounce · S. Mvulirwenande ·
N. Nanding · M. A. Rico-Ramirez · M. Romano · D.P. Solomatine ·
J. van den Broeke · R. Velotta · U. Wehn · L. Zhuo

Springer

Editors
Andrea Scozzari
National Research Council of Italy
Institute of Information Science and
Technologies (CNR-ISTI)
Pisa, Italy

Steve Mounce
Department of Civil and Structural
Engineering
University of Sheffield
Sheffield, UK

Dawei Han
Department of Civil Engineering
University of Bristol
Bristol, UK

Francesco Soldovieri
National Research Council of Italy
Institute for Electromagnetic Sensing
of the Environment (CNR-IREA)
Napoli, Italy

Dimitri Solomatine
Hydroinformatics Chair, IHE Institute for
Water Education and Water Resources Section
Delft University of Technology
Delft, The Netherlands

ISSN 1867-979X ISSN 1616-864X (electronic)
The Handbook of Environmental Chemistry
ISBN 978-3-030-61972-5 ISBN 978-3-030-61973-2 (eBook)
https://doi.org/10.1007/978-3-030-61973-2

© Springer Nature Switzerland AG 2021
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Series Editors

Prof. Dr. Damià Barceló

Department of Environmental Chemistry
IDAEA-CSIC
C/Jordi Girona 18–26
08034 Barcelona, Spain
and
Catalan Institute for Water Research (ICRA)
H20 Building
Scientific and Technological Park of the University of Girona
Emili Grahit, 101
17003 Girona, Spain
dbcqam@cid.csic.es

Prof. Dr. Andrey G. Kostianoy

Shirshov Institute of Oceanology
Russian Academy of Sciences
36, Nakhimovsky Pr.
117997 Moscow, Russia
and
S.Yu. Witte Moscow University
Moscow, Russia
kostianoy@gmail.com

Editorial Board Members

Prof. Dr. Jacob de Boer

VU University Amsterdam, Amsterdam, The Netherlands

Prof. Dr. Philippe Garrigues

Université de Bordeaux, Talence Cedex, France

Prof. Dr. Ji-Dong Gu

Guangdong Technion-Israel Institute of Technology, Shantou, Guangdong, China

Prof. Dr. Kevin C. Jones

Lancaster University, Lancaster, UK

Prof. Dr. Thomas P. Knepper

Hochschule Fresenius, Idstein, Hessen, Germany

Prof. Dr. Abdelazim M. Negm

Zagazig University, Zagazig, Egypt

Prof. Dr. Alice Newton

University of Algarve, Faro, Portugal

Prof. Dr. Duc Long Nghiem

University of Technology Sydney, Broadway, NSW, Australia

Prof. Dr. Sergi Garcia-Segura

Arizona State University, Tempe, AZ, USA

Series Preface

With remarkable vision, Prof. Otto Hutzinger initiated The Handbook of Environmental Chemistry in 1980 and became the founding Editor-in-Chief. At that time, environmental chemistry was an emerging field, aiming at a complete description of the Earth′s environment, encompassing the physical, chemical, biological, and geological transformations of chemical substances occurring on a local as well as a global scale. Environmental chemistry was intended to provide an account of the impact of man′s activities on the natural environment by describing observed changes.

While a considerable amount of knowledge has been accumulated over the last four decades, as reflected in the more than 150 volumes of The Handbook of Environmental Chemistry, there are still many scientific and policy challenges ahead due to the complexity and interdisciplinary nature of the field. The series will therefore continue to provide compilations of current knowledge. Contributions are written by leading experts with practical experience in their fields. The Handbook of Environmental Chemistry grows with the increases in our scientific understanding, and provides a valuable source not only for scientists but also for environmental managers and decision-makers. Today, the series covers a broad range of environmental topics from a chemical perspective, including methodological advances in environmental analytical chemistry.

In recent years, there has been a growing tendency to include subject matter of societal relevance in the broad view of environmental chemistry. Topics include life cycle analysis, environmental management, sustainable development, and socio-economic, legal and even political problems, among others. While these topics are of great importance for the development and acceptance of The Handbook of Environmental Chemistry, the publisher and Editors-in-Chief have decided to keep the handbook essentially a source of information on “hard sciences” with a particular emphasis on chemistry, but also covering biology, geology, hydrology and engineering as applied to environmental sciences.

The volumes of the series are written at an advanced level, addressing the needs of both researchers and graduate students, as well as of people outside the field of "pure" chemistry, including those in industry, business, government, research establishments, and public interest groups. It would be very satisfying to see these volumes used as a basis for graduate courses in environmental chemistry. With its high standards of scientific quality and clarity, The Handbook of Environmental Chemistry provides a solid basis from which scientists can share their knowledge on the different aspects of environmental problems, presenting a wide spectrum of viewpoints and approaches.

The Handbook of Environmental Chemistry is available both in print and online via www.springerlink.com/content/110354/. Articles are published online as soon as they have been approved for publication. Authors, Volume Editors and Editors-in-Chief are rewarded by the broad acceptance of The Handbook of Environmental Chemistry by the scientific community, from whom suggestions for new topics to the Editors-in-Chief are always very welcome.

Damià Barceló
Andrey G. Kostianoy
Series Editors

Preface

The well-publicised and growing concerns about worldwide freshwater resources for human consumption impose increasing attention and focus on the monitoring, assessment, and protection of these resources. Indeed, information and communication technologies (ICT) are currently playing a key role in the observation of water systems, both for what are regarded as man-made infrastructures and for those that are regarded as being water in its natural form.

The wide context covered by ICT implies a very heterogeneous technological framework, which involves several multidisciplinary aspects, and also encompassing several application fields. Without seeking to be exhaustive, one may mention the development of new observational approaches, direct sensing techniques, sensor networking architectures, data processing and analysis methods, and integration with large data systems. These are just a few examples of the several thematic areas that can be individuated. In addition, such wide multidisciplinary coverage embraces many of today′s hot topics, such as crowdsourced data collection, the internet of things (IoT), and the consequent management and analysis of big data.

Today, ′smart cities′ and ′smart water networks′ are cutting-edge topics in the technical literature concerning water systems. The practical objectives of smart water networks (and of digitalisation in general) are essentially driven by the demand for increased efficiency of whole systems, as a response to increased consumption scenarios, uncertain climate change, and the relating pressure on the higher quality portion of freshwater resources destined for human consumption.

The chapters of this book focus on new perspectives for the monitoring, assessment, and control of water systems, offering an updated survey of recent advances in tools and concepts originating from the ICT sector applied in the ′smart water′ context. The aim of this book is to present a portrait of up-to-date observational techniques, data processing approaches, and sensing technologies for water, giving further particular attention to the implication of multiple data science aspects, e.g., data analytics, cloud computing, and machine learning.

Within this framework, the chapter by Mounce [1] investigates the opportunities offered by data science in the context of smart water utilities. He discusses the role played by digitalisation for smart water networks, analysing several aspects of IoT, artificial intelligence, cloud computing, blockchain, and other new technologies. The chapter explores relevant issues connected with data science applications to the water industry. The first obstacle to the adoption of digital technologies is related to the extraction of useful information from big datasets, being that the water industry is generally considered as ′data-rich but information poor′. Thus, collected data are typically underused and data analytics are still not generally perceived as valuable, in the road map to more efficient networks based on information and knowledge. Currently, available computing power permits the implementation of data-driven modelling and deep learning techniques for prediction and classification purposes. The author foresees strong possibilities offered by deep artificial neural networks, based on their excellent unsupervised feature extraction capabilities. Big datasets concerning water quality generated by multiparametric sensor systems are given as an example of relatively undeveloped sectors for data analytics, offering opportunities for further developments, which are discussed also in other chapters of this book [3, 5]. Finally, the chapter provides reference and overview of case studies, demonstrating the kind of applications that are candidates to be more commonplace in the near future.

As pressure increases on water resources, there is a growing emphasis for water service providers to minimise the loss from leakage. Optimal sensor placement in water distribution systems for leak/burst detection and localisation is a well-established and very productive research field. Its primary focus is to minimise the cost of a proposed sensor network infrastructure while maximising the capability to detect and localise leaks and bursts through the analysis of the collected data. Romano [2] provides a systematic review of previous work covering relevant articles published over the last decade aiming at rationalising the work carried out in this field. The chapter presents a synthesis and analysis of the relevant published works to: (1) provide insight and awareness of differing arguments, theories, and approaches; (2) highlight their capabilities and limitations; (3) identify the state of the art in their development. The chapter also provides insight and awareness of differing approaches that have been proposed to tackle specific issues encountered by researchers when developing their proposed techniques such as model and measurement uncertainties. Trends and gaps in the current research and future research directions are identified and discussed, and a number of considerations to promote further developments in this important field of research are presented. This comprehensive chapter can serve as a useful reference resource for researchers and practitioners involved in sensor network design for leak/burst detection and localisation methodologies and in the development/adoption of these techniques.

During recent decades, the role of data as a vital resource that enhances decision-making and which supports efficient systems operation has become evident, with a growing number of water supply companies viewing data as a key organisational aspect that has to be properly managed, instead of an operational side-product. At the same time, drinking water systems have increased in complexity and feature smarter elements, which in turn leads to a data-richer operational environment. Castro-Gama et al. [3] address this challenging context and the often-overlooked factor of ensuring high data quality and preventing errors in data streams. The chapter provides a bird′s eye view of data validation in the drinking water industry of The Netherlands towards better data quality control policies, by providing insights on (raw) data validation in two problem types, one of water quantity and one of water quality. The chapter concentrates on a specific aspect of the overall data quality control chain, which deals with faulty data detection and isolation. Furthermore, of interest here are errors in the measurements, because sensing and human data editing processes lead to raw data distortion in the form of, e.g., drift, bias, precision degradation, or sensor failure. The focus lies on data validation to determine faulty data and the identification techniques, without expanding further on the decision-making process regarding accepting or rejecting faulty data. The authors present the results of surveys conducted with four water companies, a literature review on faulty data detection techniques, and then propose a data quality control approach using simple techniques. Case study results are presented including data validation for one water company. Best practices and issues arising from these examples regarding data quality control by water utilities are identified, as well as recommendations for future research and application of faulty detection techniques in the Dutch drinking water sector.

Monitoring wastewater has always been a challenge. Wastewater systems can vary in both size and complexity ranging from small and simple rural catchments to large and complex urban conurbations. In the wastewater collection network, historically, there has been a lack of permanent wastewater monitoring because of the propensity for fouling and the complications of monitoring both gravity and pressurised networks. In engineering and operational terms, the wastewater network has also been treated as a separate entity to the wastewater treatment works, which is, in reality, part of the same system. The wastewater treatment works tend to be much better monitored depending upon the size of the works. However, this monitoring has been very much based upon single system instrument-based control systems (e.g., a dissolved oxygen control system for an activated sludge plant). Grievson [4] presents a more holistic systematic approach, which is based upon the philosophy of the resource factory and treating the outputs from the wastewater treatment works as a product. The chapter looks at the different elements of the system as a whole and looks at the philosophy of operation that a smart system would put in place and the measurement and control needs required. The future of both the wastewater network and the wastewater treatment works will be a much more holistic approach bringing the network and the treatment works together and treating it as a single system. In this way, rather than operating the wastewater treatment system for process control with the aim of protecting the water environment, it can also be operated for resource recovery and energy efficiency with a much wider environmental benefit. This chapter provides valuable background for researchers and practitioners interested in smart wastewater networks (including opportunities and barriers) and smart wastewater treatment (including preliminary, primary/secondary treatment and sludge and resources recovery processes).

Online monitoring of several common parameters of water quality is used in water distribution systems, in order to ensure its safety for drinking and sanitation. The most common parameters are free chlorine, turbidity, and pH. A water quality event will typically result when one or more parameter values reach abnormal levels. Detection of water quality events before customers are affected is paramount to prevent possible public health impacts and potential regulatory action. Several general methods have been suggested in the past for identifying and classifying such water quality events from multiple parameters. These methods include supervised methods such as regression or regression trees and methods that make use of unsupervised learning such as clustering. The chapter by Brill [5] presents and demonstrates the utilisation of radial basis function as a tool for detection and classification of abnormal events in water quality. The methodology is based on calibration of a radial basis function using historical true events classified by human experts. The aim of the process is the selection of parameters that ensure zero false-negative events. The chapter continues to describe the main method of using radial basis function and then compares four different kernel functions, which are used for implementing the radial basis function. The case study part of the chapter illustrates actual analysis of real-world data (obtained from a monitoring station located in a large city) as well as an illustrative example (data originating from a laboratory rig). The chapter concludes with some practical advice on how kernel functions should be selected for this task and will be of value to practitioners implementing their own water quality alert systems.

The first five chapters of this book show a general portrait of the many implications of data science with the water industry, in particular for what regards specific monitoring demands and also, more in general, when dealing with big datasets of heterogeneous parameters. The potentialities and the opportunities offered by the information and communication technologies (ICTs) for improving the management of water are globally recognised. However, ICT solutions are not well exploited in developing countries and for solving this issue a partnership approach, based on open innovation, is necessary. Mvulirwenande and Wehn [6] show how ICT-focused water innovation partnerships (ICT-WIPs) can play a relevant role in building the capabilities of developing countries to implement smart water systems. In particular, this study demonstrates that ICT-WIPs allow a variety of stakeholders in the water sector (such as municipalities, ministries, large utilities, and regulatory agencies) to work together and increase the awareness about the potential of smart water systems. At the same time, the innovation partnership approach promotes the culture of mutual learning, thus allowing partners to strengthen each other′s innovation competences relating to smart water systems in developing countries. This is particularly true when, as in the case of the ICT-WIPs analysed in this study, partners from foreign countries need the knowledge and experience of local partners (e.g., about specific problems concerning the water systems, existing solutions and their weaknesses, possible additional local risks). Finally, the nature of the ICT-focused water innovations analysed in this study leads to the insight that fostering smart water systems in developing countries requires to rethink not only the technologies themselves but also the business models around them.

In times of IoT and ′social sensing′, also the observation of hydrological contexts may take benefit from low cost and ′pervasive′ sensing systems, opening the possibility to novel approaches for capturing relevant information. In this frame, one of the main issues to solve is the availability of heterogeneous and intermittent observations. The chapter by Mazzoleni et al. [7] describes novel methods for optimally assimilating such observations into hydrological models, focusing on the particular application of flood prediction. The aim of this chapter is to explore numerical approaches for integrating crowdsourced observations from static social sensors within hydrological and hydrodynamic modelling framework to improve flood prediction. The distinctive characteristic of such heterogeneous observations is their varying lifespan and their spatial distribution, which make more complex the implementation of standard model updating techniques. This chapter applies different innovative assimilation techniques within two case studies, where synthetic flow observations are generated to represent the different intermittency and accuracy scenarios of the crowdsourced observations. It was found that crowdsourced observations can significantly improve flood prediction if integrated into hydrological and hydraulic models. Moreover, a network of low-cost static social sensors can actually complement traditional networks of static physical sensors, for the purpose of improving flood forecasting accuracy.

Precipitation is a key hydrological process in the water cycle, whose observation is increasingly required for modern water and environmental management. Conventional precipitation measurements by rain gauges cannot provide sufficient spatial and temporal coverage for many hydrological applications, such as urban drainage system modelling. Weather radar is a remote sensing instrument that has been increasingly used to estimate precipitation for a variety of hydrological and meteorological applications, including real-time flood forecasting, severe weather monitoring and warning, and short-term precipitation forecasting. Weather radar provides unique observations of precipitating systems at fine spatial and temporal resolutions. The potential benefit of using radar rainfall in hydrology is huge, but practical hydrological applications of weather radar have been limited by the inherent uncertainties and errors in radar rainfall estimates. Uncertainties in radar rainfall estimates can lead to large errors in their applications, so radar rainfall measurements must be corrected before the data are used quantitatively. Nanding and Rico-Ramirez [8] have introduced the latest advances in the measurement and forecasting of precipitation with weather radar. The common uncertainty sources include radar hardware calibration, echoes due to non-meteorological origin, attenuation, variations in the vertical profile of reflectivity, and variations of raindrop size distribution. The techniques for adjusting radar rainfall with rain gauge measurements are described. Precipitation forecasting (called ′nowcasting′) using weather radar is valuable in its applications in real-time flood forecasting.

Accurate soil moisture information is critically important for hydrological applications such as water resources management and hydrological modelling. This is because soil moisture is an important element in the ecosystem and hydrological cycle, regulating evapotranspiration, precipitation infiltration, and overland flow. In contrast with in situ instruments, modern satellite remote sensing has shown a huge potential for providing soil moisture measurements at a large scale. However, its effective utilisation in the practical projects still needs comprehensive research. Zhuo [9] has introduced the advances and potential issues in the current application of satellite soil moisture observations in hydrological modelling. The key issues include soil moisture measuring methods, hydrological evaluation of satellite soil moisture, error distribution modelling of soil moisture measurements, and the need for new hydrological soil moisture product development. It has been found that hydrological application of soil moisture data requires the data relevant to hydrology. In order to meet the requirement, two important research tasks are needed: the first is to carry out comprehensive assessments of satellite soil moisture observations for hydrological modelling, not merely based on evaluations against point-based in situ measurements; the second is that a soil moisture product (e.g., soil moisture deficit) directly applicable to hydrological modelling should be developed. Only fully accomplishing these two steps will push forward the utilisation of satellite soil moisture in hydrological modelling to a greater extent.

There is an increasing demand for automatic chemometric solutions for water quality monitoring, the main requirements being their autonomous operation, low cost, and low maintenance. Today, there is a range of optical sensor technologies that are capable to perform most analytical tasks and are characterised by full solid-state, no need for reagents, and capability to withstand harsh working conditions, as it is needed when the measurement points are outside protected monitoring locations (e.g., treatment plants) and relevant parameters have to be acquired directly in the external environment. van den Broeke and Koster [10] introduce a selection of optical sensing technologies, which can provide valuable information on the quality of water. The measurement techniques that they describe are suitable for process monitoring and control applications, as well as for early-warning systems. The chapter explores the basic principles of radiative transfer at the foundation of spectroscopic methods and the fundamentals of the signal processing for the extraction of chemical information from the acquired signals. The survey of sensing methods covers the absorption spectrometry in the UV/Vis spectral region, illustrating both selective measurements of specific substances (e.g., BTEX, nitrate, and nitrite) and more generic features, like the colour, the amount of total suspended solids, and the ′sum organic parameters′, which are recognised as excellent overall water quality indicators. The chapter also covers basic aspects and applications of fluorescence spectroscopy and infrared spectroscopy in the NIR (near-infrared) domain. Further methods, which are considered as promising and are also already marketable, are described in this survey: Raman and laser-induced breakdown spectroscopy, refractive index measurements, and image analysis. Despite the fact that optical methods are based on mature technologies that have a solid physical background, they are still not much used by the industry and still have a big potential to deliver. Optically based methods are very attractive candidates to perform automatic online field measurements, both for selective parameters, i.e., in a way equivalent to traditional analytical methods, and for overall water quality assessments, i.e., for change-detection and early warning purposes.

In addition to the online measurement of chemical parameters, the monitoring of possible organic threats to the quality of water is also of utmost importance, and technologies for the direct sensing of such contaminants based on biosensing techniques are currently under development. Della Ventura et al. [11] describe a very attractive sensing technique for the in situ monitoring of organic contaminants based on quartz crystal microbalance (QCM) devices. The chapter introduces the basic theory of QCMs, the detection scheme and its practical embodiment (the electronic interface and signal analysis) for an appropriate extraction of the needed information. The peculiarity of the proposed approach consists in the functionalisation of the QCM gold surface, obtained by immobilising a ′recognition layer′ of antibodies on the surface of the crystal, which is the key aspect for the sensitivity and specificity of QCM-based immunosensors. The chapter, after introducing the theory and modelling of the QCM working mechanism, analyses the response of a quartz crystal resonator in contact with a liquid sample. Finally, the surface functionalisation and the detection scheme are discussed, with regard to particular bacteriological contaminants like Escherichia coli and pesticides like parathion. This promising sensing approach offers very high selectivity, real-time measurement capability, and high sensitivity, thanks to the fact that a QCM is capable of measuring mass changes as small as a fraction of a monolayer of atoms, as the authors report.

This book is intended for a wide audience of readers, such as postgraduates, researchers, and stakeholders at various levels. It is also intended for those experts who want to widen their purview to adjacent fields of expertise and do not necessarily have an ICT or hydroinformatics background.

Without aiming to be exhaustive, the present volume seeks to be a selective survey of novel measurement technologies and data analysis approaches for water systems.

Chapters of this book have been peer-reviewed by two reviewers per chapter. In some cases, more than one review round was needed. Reviewers have been selected partly internal and partly external to the book project. The editors are very indebted to the reviewers for their excellent and thorough contribution to the overall quality of the book.

Here we acknowledge the reviewers external to the book project:

- Romeo Bernini (CNR-IREA, Italy)
- Alessandro Nottola (Technoprobe S.p.A., Italy)
- Dleen Al-Shrafany (Salahaddin University, Iraq)
- Andreja Jonoski (IHE Delft Institute for Water Education, Netherlands)
- Antonio Annis (University for Foreigners of Perugia, Italy)
- Henrik Madsen (Danish Hydraulic Institute, Denmark)
- Qiang Dai (Nanjing Normal University, China)
- Peter van Thienen (KWR Water Research Institute, Netherlands)
- Ina Vertommen (KWR Water Research Institute, Netherlands)

- John Machell (University of Sheffield, UK)
- Franco Dinelli (CNR-INO, Italy)

In addition, we are grateful to Prof. Andrey Kostianoy (co-editor-in-chief of the book series) for having supported this project, and all the editorial staff within Springer, who have always been available and supportive at all times.

Finally, our sincere thanks go to the contributors of this volume, who have always shown a great level of engagement since the beginning of the book project.

Pisa, Italy — Andrea Scozzari
Sheffield, UK — Steve Mounce
Bristol, UK — Dawei Han
Napoli, Italy — Francesco Soldovieri
Delft, The Netherlands — Dimitri Solomatine

References

1. Mounce SR (2020) Data science trends and opportunities for smart water utilities. In: Han D, Mounce S, Scozzari A, Soldovieri F, Solomatine D (eds) ICT for smart water systems: measurements and data science, Hdb Env Chem. Springer Nature Switzerland AG. https://doi.org/10.1007/698_2020_482
2. Romano M (2019) Review of techniques for optimal placement of pressure and flow sensors for leak/burst detection and localisation in water distribution systems. In: Han D, Mounce S, Scozzari A, Soldovieri F, Solomatine D (eds) ICT for smart water systems: measurements and data science, Hdb Env Chem. Springer Nature Switzerland AG, https://doi.org/10.1007/698_2019_405
3. Castro-Gama M, Agudelo-Vera C, Bouziotas D (2020) A bird′s eye view of data validation in the drinking water industry of The Netherlands. In: Han D, Mounce S, Scozzari A, Soldovieri F, Solomatine D (eds) ICT for smart water systems: measurements and data science, Hdb Env Chem. Springer Nature Switzerland AG, https://doi.org/10.1007/698_2020_609
4. Grievson O (2019) Monitoring and controlling a smarter wastewater treatment system: a UK perspective. In: Han D, Mounce S, Scozzari A, Soldovieri F, Solomatine D (eds) ICT for smart water systems: measurements and data science, Hdb Env Chem. Springer Nature Switzerland AG, https://doi.org/10.1007/698_2019_418
5. Brill E (2019) Using radial basis function for water quality events detection. In: Han D, Mounce S, Scozzari A, Soldovieri F, Solomatine D (eds) ICT for smart water systems: measurements and data science, Hdb Env Chem. Springer Nature Switzerland AG, https://doi.org/10.1007/698_2019_424
6. Mvulirwenande S, Wehn U (2019) Promoting smart water systems in developing countries through innovation partnerships: evidence from VIA water-supported projects in Africa. In: Han D, Mounce S, Scozzari A, Soldovieri F, Solomatine D (eds) ICT for smart water systems: measurements and data

science, Hdb Env Chem. Springer Nature Switzerland AG, https://doi.org/10.1007/698_2019_422

7. Mazzoleni M, Alfonso L, Solomatine DP (2019) Exploring assimilation of crowdsourcing observations into flood models. In: Han D, Mounce S, Scozzari A, Soldovieri F, Solomatine D (eds) ICT for smart water systems: measurements and data science, Hdb Env Chem. Springer Nature Switzerland AG, https://doi.org/10.1007/698_2019_403
8. Nanding N, Rico-Ramirez MA (2019) Precipitation measurement with weather radars. In: Han D, Mounce S, Scozzari A, Soldovieri F, Solomatine D (eds) ICT for smart water systems: measurements and data science, Hdb Env Chem. Springer Nature Switzerland AG, https://doi.org/10.1007/698_2019_404
9. Zhuo L (2019) Satellite remote sensing of soil moisture for hydrological applications – a review of issues to be solved. In: Han D, Mounce S, Scozzari A, Soldovieri F, Solomatine D (eds) ICT for smart water systems: measurements and data science, Hdb Env Chem. Springer Nature Switzerland AG, https://doi.org/10.1007/698_2019_394
10. Van den Broeke J, Koster T (2019) Spectroscopic methods for online water quality monitoring. In: Han D, Mounce S, Scozzari A, Soldovieri F, Solomatine D (eds) ICT for smart water systems: measurements and data science, Hdb Env Chem. Springer Nature Switzerland AG, https://doi.org/10.1007/698_2019_391
11. Della Ventura B, Mauro M, Battaglia R, Velotta R (2019) Quartz crystal microbalance sensors: new tools for the assessment of organic threats to the quality of water. In: Han D, Mounce S, Scozzari A, Soldovieri F, Solomatine D (eds) ICT for smart water systems: measurements and data science, Hdb Env Chem. Springer Nature Switzerland AG, https://doi.org/10.1007/698_2019_390

Contents

Data Science Trends and Opportunities for Smart Water Utilities

Stephen R. Mounce

Contents

Abstract We are witnessing an industry change which is transitioning to a more intelligent (or smarter) water network. In the UK a 5-year planning period and investment cycle called the Asset Management Plan (AMP) is the regulatory mechanism. This process is used to manage a water utility′s infrastructure and other assets to deliver an agreed standard of service. The challenge of AMP 6 and 7 (to 2025) and beyond is to maximise efficiency by moving from reactive to proactive management. This can be achieved by using data, information and (where possible) control of the system. The more intelligence that is captured, the more that can be learned and understood about the network and subsequently be predicted. Extra data provides new opportunities for asset maintenance and event analytics. Data science is an emerging discipline which combines analysis,

S. R. Mounce (*)
Department of Civil and Structural Engineering, University of Sheffield, Sheffield, UK

Mounce HydroSmart, Beverley, UK
e-mail: s.r.mounce@sheffield.ac.uk

Andrea Scozzari, Steve Mounce, Dawei Han, Francesco Soldovieri, and Dimitri Solomatine (eds.), ICT for Smart Water Systems: Measurements and Data Science, Hdb Env Chem (2021) 102: 1–26, https://doi.org/10.1007/698_2020_482,
© Springer Nature Switzerland AG 2020, Published online: 22 June 2020

programming and business knowledge and uses new and advanced techniques and technologies to work with complex data. The water sector needs to address the issue of ′big data′ and obtaining ′signal from the noise′. Primarily, the focus is on data to action by the application of data science.

The role of digitalisation for smart water networks is covered in this chapter, exploring issues of IoT, artificial intelligence, blockchain and other novel technologies. Reference to case studies demonstrates the type of applications which will become increasingly common place. Some recommendations based on future possibilities and opportunities are proposed.

Keywords Blockchain, Data science, IoT, Machine learning, Smart networks

1 Introduction

Population growth, urbanisation, industrialisation and climate change are placing increasing pressure on water resources. The water-energy nexus is a term being used to describe the complex linkages and dependencies among water, energy and food security, and this is of vital importance for the twenty-first century. At the current pace of growth and consumption, water scarcity has the potential to grind food and energy supply chains to a halt impacting on economic growth. A paradigm shift in water industry systems is required, by considering water and waste treatment holistically across all sectors using flexible and responsive processes to meet rapidly changing global challenges, minimising the impact on the environment, and not compromising on public health and quality standards. The importance of such a transition to a circular water industry by 2050 is becoming increasingly apparent [1].

This chapter will focus on the UK water sector and thus form the context for discussion of relevant technologies. Over time, water supply networks have evolved into extremely large (over 300,000 km of each of supply and sewerage pipes in the UK), complex, interconnected systems of pipes, storage reservoirs, pumps, valves and other assets that are required to hold and deliver enough water to meet all the drinking, hygiene, washing, gardening and recreational requirements of modern society as well as much of the industrial demands and for disposal of wastewater as well as surface drainage. Operation and control of water networks [2] is the responsibility of water service providers (WSP) who are continually challenged to be ever more efficient and to improve year-on-year as demanded by regulators and investors. In the UK, over £90 billion has been invested in upgrading water industry assets since privatisation in 1989, and there is now a shift in focus to the aspiration for intelligent and proactively managed water networks. In 2018 it was announced that over £50 billion would be invested over the next 5 years, increasing spending 13% above current levels.

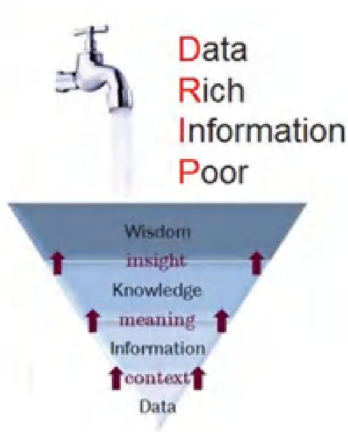
1.1 Data Rich Information Poor: Fixing the DRIP

The water industry has often been perceived as being ′data rich but information poor (DRIP)′ – see Fig. 1. Knowledge comes from understanding the information about a subject and then using it to make decisions, form judgments/opinions or make predictions. Data is the basis for acquiring information, and information is the basis for further deriving knowledge. Water utilities are actually really interested in information and knowledge, not raw data. DRIP relates to the lack of interpretation of data that is generated by instrumentation deployed in the highly complex systems. The DRIP needs to be fixed for the ageing infrastructure of water networks to make them resilient to the combined and interacting challenges of climate change, population growth, urbanisation and energy costs.

These gaps have partly to do with the necessarily conservative nature of the sector, and with the economic drivers affecting the water industry, and reflect the demands of its regulators and consumers. One of the most significant barriers to adoption of digital technologies in the water sector is the relative perceived lack of value; even though water is an essential element of life, oil or gas is more ′valuable′ in the marketplace in economic terms (yet there are few industries as critical to humans as water).

WSPs are struggling to archive data or to transform the data effectively into knowledge with which to enable operational control. It has been estimated that water utilities in the UK only use 10% of the data they collect [3]. Accurate recent figures are difficult to obtain, but evidently the amount of data being collected has exploded in the last decade, and its usage has not kept pace. The quality of this data is not only variable from one water utility to the next; it is also very variable depending on the nature of the data and the purpose for which it is being collected. It has traditionally been difficult to justify efforts to improve data quality in the water industry because, although seen as of interest, investment in asset improvements takes priority. It is challenging to put a price on the value of improved

Fig. 1 Fixing the DRIP



data quality and/or increased data collection and even online instrumentation. However, regulation in the UK is encouraging this with companies in the future seeking a basis for their asset planning with the analysis of data from, or directly related to, their operations. Where good quality data is available, accessible and well maintained, the ability of the distribution engineer to monitor, evaluate and make good decisions with regard to the operation and maintenance of the network is greatly enhanced. However this is rare, and there is a need to aid the decision-making process to make the best possible use of the data that is available.

Datasets currently maintained by water companies include historic and updated asset records, discrete water quality sampling and associated laboratory analysis and continuous/online (typically hydraulic only) data collection from an ever increasing telemetry footprint. Companies also keep records of customer contacts of various types (such as complaints about leakage or water discoloration); however such data is very variable in nature and information richness as it depends on non-standardised customer behaviour. Good procedures for formatting and using that data should then be implemented by WSPs. In order, the key data used in the UK to inform decision-making processes are hydraulic meter data, customer contacts, water quality sample data and analysis and network data such as asset records, burst records and pipe samples. Currently interventions are often responsive to customer contacts leading to a reactive management that does not necessarily deal with the underlying issues. An interim position currently exists, where water companies have substantial databases but lack the connectivity and deployed methods to extract the full potential benefit. There is a wealth of data at works/ reservoir outlets but currently less intelligence on what is happening in the distribution network. Customers should not be acting as ′surrogate telemetry points′, which is often the case (e.g. to help with leakage detection or water quality issues). Monitoring and control systems play a role in the daily operation and maintenance of water supply and sewerage facilities.

One of the main issues in the water industry is the lack of linking together of data as required for identifying/solving problems occurring at system interfaces. Current practice is for the collection and maintenance of data in disparate corporate systems, severely limiting the potential for deeper understanding. Utility databases containing data in an unprocessed format do not lend themselves to analysis to establish relationships or trends. The databases are generally on separate platforms, in differing formats, with non-uniform IDs and contain many unpopulated or utility-specific fields. Often there has been too much focus on technology and IT infrastructure and not enough on improving data quality and data integration and making best use of existing corporate data resources. An example of this has been the UK water industry gradual move to enterprise resource planning solutions (with platforms such as SAP and MAXIMO) whereby all of the water utilities′ business information is coordinated into a single environment and similarly to enterprise infrastructure for management of real-time data and events (platforms such as OSI-PI). Large implementations of these platforms can take (many) years for large companies, and customisation to allow practical use can substantially increase implementation times [4]. Hence, many legacy systems are not yet migrated to such new systems. And when data is migrated to such systems, integration and techniques to derive information are at often overlooked. Better use of data needs to be made by bringing datasets together, particularly in the development and use of metadata models. Essentially, data that is in many silos needs to be located and pulled together and (sometimes) missing data accounted for before sophisticated analysis can be achieved. Most large utilities today have an EAM (enterprise asset management) or CMMS (computerised maintenance management system) in place for network operational processes. In the future, utilities will be able to move beyond time-based to condition-based maintenance, so by adopting the ability to understand the effective age of their assets and then forecasting potential failures, they will be able to identify and schedule improvements in life extension maintenance activities as well as strategically plan for their replacement in their long-term asset plan.

With advances in data manipulation (such as the use of metadata and format interoperability) and analysis systems, in particular the integration of GIS information with data mining methodologies, it is now possible to explore relationships between data in increasingly sophisticated ways. If data is available, is of good enough quality and can be linked and associated with other data, it can, in principle, be used for a multitude of applications such as business analytics, problem ′hotspot′ mapping, operational assessment and investment modelling.

1.2 Big Data and Analytics Opportunities

The availability and affordability of varying forms of sensing, smart systems, data storage and transmission technology means water utilities are becoming able to collect more data than ever before. This information revolution era opens up hereto unseen possibilities in the creation of tools for future engineering application. Water utility databases are currently growing rapidly and will continue to do so. Globally, IBM [5] estimates that 2.5 quintillion bytes of data per day are being collected. In fact, more than 50% of the world′s data was created last year, but less than 0.5% was analysed or used. Experts predict that 40 zettabytes of data will be in existence by 2020. Collecting more data doesn′t necessarily result in better information or knowledge. But, substantial datasets do offer a potential way to tackle traditional issues via development and application of novel data-driven analysis. Big data has been compared to being like an iceberg where most of the value to be unlocked is still hidden under the surface.

Machine learning (ML) relates with the study, design and development of the algorithms that give computers the capability to learn without being explicitly programmed. Problems usually need describing in features to use ML. Machine learning and big data can be used to ′learn′ how the system operates and reacts to events. The combination of big data and machine learning will eventually result in a breakthrough in the integration and analysis of heterogeneous data types as is already occurring in other more developed industries. Big data analytics lies at the heart of the ability to derive actionable value from an array of structured and increasingly unstructured, text based or sensory data and execute or automate the next best action based on predictive and prescriptive data science.

The current nature of water utility network data is that it remains sparse (e.g. not all locations are sampled) and typically is not linked across functions (e.g. water quality data is not linked to hydraulic model data). Maximising the quality of data (its usefulness) requires consideration of a chain of processes and manipulation, e.g. data source, collection, storage and the anticipated data end use. Machine learning or data-driven analyses, which map inputs to outputs without attempting to accurately model underlying processes, can potentially yield useful understanding, such as determination of dominant variables and empirical relationships, and therefore have been used for many different environmental and water quality applications. The required blend of foresight and experience means a move towards ′big data′ solutions, and so-called business intelligence (turning an organisation′s data into patterns that help make intelligent business decisions) in water utilities must be somewhat iterative and will require significant development time. In the future, engineers will access data, tools and analysis in real time and collaborate on diagnostic decisions relating to the condition of a remote monitored asset. This could enable network engineering decisions to be increasingly evidence-based with associated provenance to allow the reasoning behind decisions to be evaluated for compliance with statutory regulation and to create a knowledge repository to form a basis for future decisions. In an industry where customer perception of service (and statutory obligations) is dependent on very complex, distributed, non-linear dynamical water networks with a consequent high degree of uncertainty, such knowledge-based engineering represents a key advantage in commercial terms (i.e. company efficiency) and in the ability to serve the wider society.

A vision for this integrated future is illustrated in Fig. 2. At the rate at which data and our ability to analyse it are growing in society in general, it is reasonable to expect that most UK companies will be using the impact of big data analytics in the next 5 years [6]. It should however be noted that this is unlikely to be the case in many other countries, with a more gradual trickle down of technology transfer occurring over time.

Ultimately developments will result in data exploration tools for non-ICT specialists reducing the cost of and ability for high-fidelity visualisation of data for enabling human interpretation. The exploration and analysis of data using visualisation techniques is a powerful approach for conveying potential hypotheses and exploring correlations, due to the fact that vision plays an important role in human cognition. One example of how data-driven techniques can be used for data visualisation is the use of self-organising maps (SOM), a form of unsupervised artificial neural networks (ANNs), as has been used for analysing water data in various UK industry research projects [7]. The application of SOMs has been demonstrated in water distribution system data mining for microbiological and physico-chemical data at laboratory scale [8] and in the field [9]; in clustering of water quality, hydraulic modelling and asset data for a single water supply zone [10];

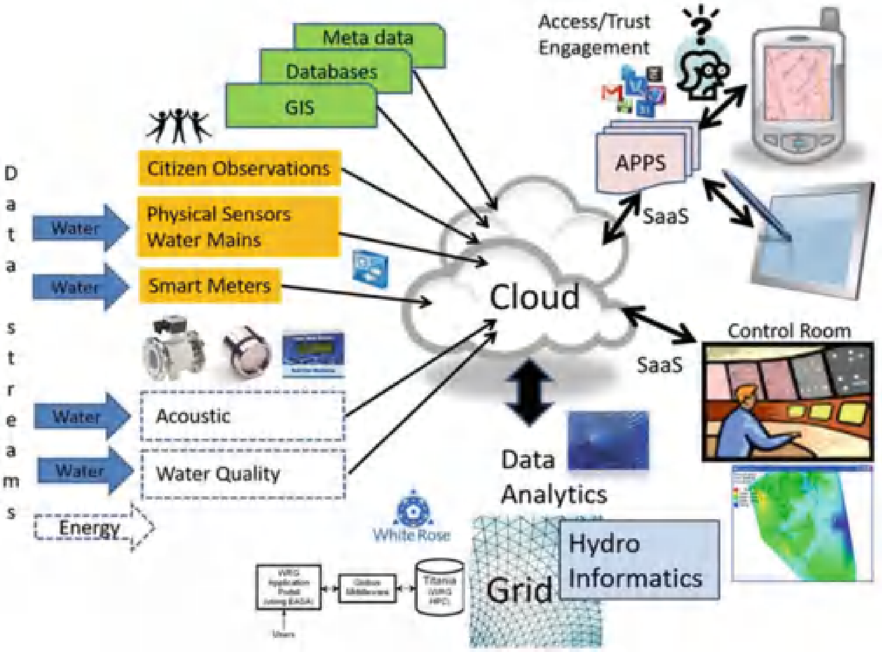


Fig. 2 Vision of an integrated future

in interpreting the risk of iron failures [11]; in relating water quality and age in drinking water [12]; in exploring the rate of discolouration material accumulation in drinking water [13]; and for geospatial burst behaviour [14]. The SOM is an unsupervised ANN model which resembles the way biological brain maps spatially order their responses by modelling those self-organising and adaptive learning features of the brain [15]. Trained vectors are positioned on a regular low-dimensional grid in a spatially ordered fashion hence facilitating improved visualisation, readily enabling presentation and interpretation. SOMs are noise tolerant; this property is highly desirable when sparse data are used, and thus there are many potential applications in the water industry. The SOM (see example in Fig. 3) contains colour-coded hexagons that summarise all of the component planes that represent individual variables. Each hexagonal cell represents individual neurons, which are the mathematical linkages between the input and output layers. In the component planes for individual variables, the colouring or shading corresponds to actual numerical values for the input variables that are referenced in the scale bars adjacent to each plot. Blue shades show low values, and red corresponds to high values. Visual inspection and comparison of the component planes allow examination of how variables vary against each other. Figure 3 shows SOM analysis of data from a water quality sensor monitoring in a test loop facility [16] for a 28-day period measuring water quality parameters every minute. While absolute accuracy was uncertain, the way in which these parameters are related in a time invariant fashion can be displayed by the component planes of

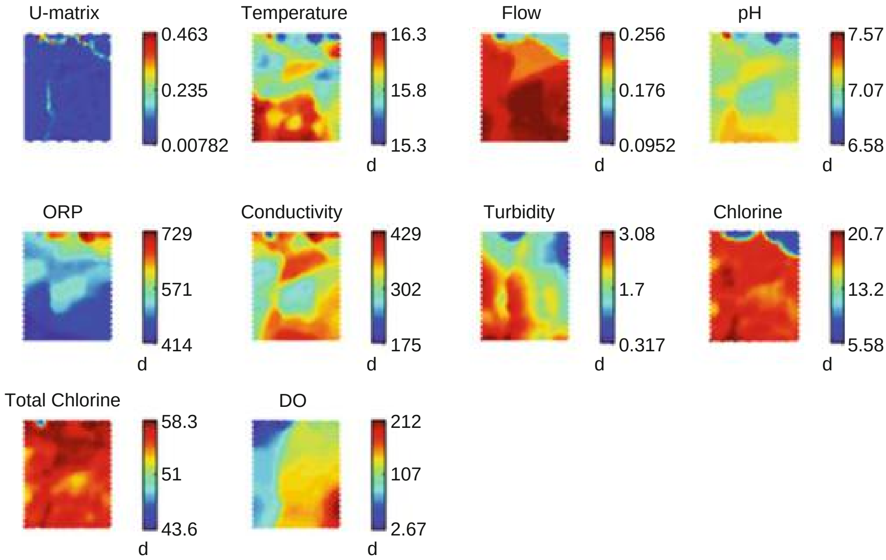


Fig. 3 Example SOM for water quality monitoring

a SOM (total 39,945 measurement vectors on a 37 $\times$ 27 map). In particular, the relationships between flow, oxidation-reduction potential (ORP) and chlorine can be seen as similarly shaped areas of colour in the same sections of the map (e.g. upper right corner). The use of data analysis techniques like SOMs, which have a good resilience to sparse data, can provide value where other techniques would likely fail.

Other approaches to data-driven visualisation/mapping and cluster analysis include using k-means (typically with Euclidean distance), hierarchical clustering, distribution models (such as the expectation-maximisation algorithm), fuzzy clustering and density-based models, e.g. DBSCAN [17], Sammon′s projection [18] and t-SNE [19].

1.3 Hydroinformatics

Hydroinformatics has emerged over the last decade to become a recognised and established field of independent research within the hydrological sciences. Hydroinformatics is concerned with the development and hydrological application of mathematical modelling, information technology, data science (e.g. data mining and knowledge discovery, big data and deep learning techniques) and computational intelligence tools. It provides the computer-based decision support systems that are now becoming increasingly prevalent for use by consulting engineers, water service providers and government agencies to implement solutions such as smart networks.

Data-driven modelling seeks to provide a mapping between the inputs and outputs of a given system, with little prior process knowledge, and is emerging as an attractive option for prediction and classification in water systems. Over the next decade, advancements in the general progresses of ICT (hard and soft) such as in data analytics, accelerated computing power and mega-networking (already becoming available) will help solve the frustrating fragmentation of scientific information in the field of water resources. The increase of CPU power (massive parallel computing, cloud computing, etc.) extends the possibilities of numerical and data-driven models and of 3-D/augmented reality displays, and Web 2.0/3.0 opens up access to information sources to millions of new users. New developments and products in the fields of micro-sensors, alternative power supply and wireless telecoms all revolutionise the whole domain of real-time monitoring and consequently real-time management.

Deep learning [20] is when big data intersects with machine learning. Techniques allow the tackling of problems that exceed human understanding. Deep learning′s important innovation is to have neural nets learn categories incrementally, attempting to model lower-level categories (like letters) before attempting to acquire higher-level categories (like words). Deep learning excels at this sort of problem and has transformed the application of AI in the last decade. Their usage has been directly facilitated by big data on top of algorithmic breakthroughs. Next-generation formulations of deep artificial neural networks allow for the direct transition from data to action. Such state-of-the-art algorithms include unsupervised feature extraction as part of the data-driven learning. Deep learning technology has the potential to leverage the power of big data and help develop an insight-driven culture. Deep learning has been applied to visualising high-dimensional smart water meter data [21]. T-SNE [19] is a technique that can be used for human-intuitive (two-dimensional) visualisation of high-dimensional data. The parametric version of t-SNE uses deep neural networks.

2 Data Analytics: Prediction

Predictive analytics involves using the patterns of past behaviour to predict behaviour in the future. Predictive analytics encompasses a variety of statistical techniques from predictive modelling, machine learning and data mining that analyse current and historical facts to make predictions about future (or unknown) events. Predicting future values of time series is one such application useful in many water resource domains. Since there is a set of temporal ordered observations for which (and working on the assumption that) there exist serial correlations along the series, previous observations can be used to predict future values. The task is essentially one of function approximation, i.e. to approximate the underlying continuous valued function producing the time series.

3 Data Analytics: Classification

There are other ways in which analytics can be used than for prediction, from understanding customers to optimising a business process. In machine learning and statistics, classification is the problem of identifying which category a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known (hence described as supervised learning). An example would be assigning a given pipe to a high-risk or low-risk category for water quality issues based on asset characteristics (pipe material, age, diameter, etc.). Classification can be viewed as a special case of function approximation using some type of discriminant for the decision.

3.1 Internet of Things and Edge Computing

Internet of Things (IoT) objects and sensors can be connected to the Internet (via the cloud) giving rise to the concept of ′smartness′ and the development of ′smart cities′ and ′smart water′. By definition they have an IP address. It seems clear that in the future, the whole water sector is going to be completely penetrated by ICT and Internet-like technologies. It is estimated that 50 billion devices in all industries will be connected in this way by 2020 (with an estimated 6.1 billion smartphones) and as many as 75 billion by 2025. In a decade, tens or even hundreds of petabytes of data may be routinely available. The sensing of data that could not be gathered in the past and collecting them on IoT platforms is expected to create new value. As these technological capabilities advance, so does the ability to collect information from remote devices and correlate that information across diverse systems. Standards will be required to manage this level of interconnectivity which may ultimately lead to the unification of information management for the industry. An infrastructure that can connect the monitoring and control systems to an IoT platform allows the effective use of the operational information the systems hold and help achieve near real-time situational awareness. Demands for solutions and tools will become more urgent to meet the aspiration for intelligent water networks, proactively managed through access to timely information.

In recent years, the UK water industry has been making increasing use of advances in sensor technology for monitoring parameters of water systems to identify performance shortfalls in order to improve asset management and hence provide better customer service, value and regulatory performance. For example, in supply systems sensors for flow and pressure have become more widely used, especially on trunk mains and at district metered area (DMA) level, to facilitate zone-based asset management. The monitoring of sewerage systems has not progressed as far; however there is an increasing interest and deployment of instrumentation, for example, for CSO (combined sewer overflows) level measurement and pump station flows. Water quality data (parameters such as water temperature, turbidity, conductivity, colour, pH, etc.) is not currently collected in a very consistent and automated manner across networks, and these sensor technologies (and their price) have moved little in the last decade. New spectroscopic methods are now coming into use (see chapter "Spectroscopic Methods for Online Water Quality Monitoring"). Water quality measurements can help to identify discolouration events and monitor chlorine residuals (see chapter "Using Radial Basis Function for Water Quality Events Detection"). As the technology improves and the whole life cost of ownership falls, water quality instruments could be used for the provision of operational data and become as widespread as for flow metering. Water quality data analytics are relatively undeveloped because there are limited significant deployments of real-time water quality monitoring within networks [22]. Future technologies promise enhanced sensors for natural biological/biochemical markers. Interpretation and analysis across multiple online parameters is expected to provide deeper understanding of water supply system state and asset condition.

The proliferation and diminishing costs of automated data transfer, such as by GPRS, 4G/5G, Wi-Fi, LoRa/LoRaWAN and Sigfox systems, are allowing all types of recorded data to be transferred from many disparate points on the networks. New developments and products in the fields of micro-sensors, alternative power supply and wireless telecoms all revolutionise the whole domain of real-time monitoring and consequently real-time management. It is easy to anticipate that the environment may quite soon be teeming with tens of thousands of small, low-power, wireless sensors. Each of these devices will produce a stream of data, and those streams will need to be monitored and combined to detect interesting changes in the environment. The emergent properties of data from networks of simple, low-cost sensors will be increasingly fruitfully explored.

IoT will be dominated in the coming years by the growth of edge computing systems as roll-outs of the technology in the field become more complex and larger scale. Edge computing represents cutting-edge hardware and software co-located at IoT endpoints that make the technology far more efficient, scalable, secure and manageable. The presence of on-board CPUs can form basic analytical functions (e.g. which data to send to customer′s smartphone apps, which data to send to the WSP and when it should be sent), combined with on-board data storage that would really make devices smart, and not just a measurement and communication device.

It will also make IoT solutions much smarter by enabling the deployment of machine learning and AI capabilities much nearer the point-of-use that will enrich and optimise what is possible. Edge computing will enable AI applications in scenarios where processing is better performed locally. Advances in sensors and low-power computing architectures will enable edge computing with high-performance, real-time and increasingly complex AI solutions. This has promise to yield huge cost savings for remote locations with limited or expensive Internet connectivity and to reduce the overhead of centralised processing. Traditionally, in IoT, the predictive analytics have been done by analysing the data in the cloud. However, that may not always be possible with large amounts of streaming data arriving from the edge devices. Making edge devices smarter will thus be critical. The prediction rules may be discovered a priori in the cloud, and these lightweight rules can then be deployed on the edge devices.

4 Cloud Computing and Condition Monitoring

Utility companies have been exploring more advanced architectures for real-time performance management such as data sources going via a WAN (wide area network) to the database system and then a telemetry portal allowing access on demand, e.g. through a web browser. Future operational applications will allow data to be processed locally or remotely on the cloud to provide exception alert generation or other information involving more than one signal using peer-to-peer technologies. Water utilities have been slow to utilise cloud hosting and services due to perceived concerns about security. Security should not be a reason to not adopt cloud-based solutions – if the correct platforms are being leveraged. Placing workloads in the cloud does not require a security trade-off, with enterprises actually benefitting now from the security built into the cloud, often resulting in fewer security incidents than when using traditional data centres.

By 2020, at least a third of all data will pass through the cloud. Cloud technology can provide a catalyst for innovation in business and a transformation of traditional ways of operating. Users can run services on data over the cloud, moving processor-intensive operations away from the desktop into disparate remote locations. A benefit of cloud hosting is ′time to value′. By having infrastructure available on demand, new innovations can be developed, tested and launched much faster than in comparison to deploying traditional IT infrastructure. It is also easier to cope with spikes in demand or unplanned growth. Further, with cloud-hosted desktops, lost or compromised hardware no longer poses the same security threat, as data is not on the device. This extra layer of protection ensures that critical data is only stored inside the data centre (where it is more easily managed, protected and recovered). Cloud native applications are designed to be self-healing, able to seamlessly adapt to loss of infrastructure components with no human intervention required.

Cloud computing offers an opportunity to make the results of condition monitoring readily available to a range of stakeholders responsible for the maintenance of an asset. Data from sensors distributed across one or more assets at one or more sites are uploaded to the cloud compute resource for continual analysis. Users can then run services on data over the cloud, using the computational and data processing power of grids and moving processor-intensive operations away from the desktop into disparate remote locations. A shared environment allows engineers to access data, tools and analysis in real time and to collaborate on diagnostic decisions relating to the condition of a remote monitored asset.

5 Digitalisation

The ′Fourth Industrial Revolution′ and ′The Second Machine Age′ are used to describe current efforts to digitise sectors, from government to healthcare to education to manufacturing to business services. Digitisation is transforming the world we live in with social, political and economic consequences. In recent years we have witnessed the rise of social media, powerful smartphone proliferation and new e-commerce businesses/e-services. The negative mirror image has included disappearance of traditional businesses, tech giant monopolies, unfettered data harvesting and fake news. Digitalization poses potential disruption to work and workers but if new technologies and new data are integrated successfully by reimagining business processes, great opportunities. These new technologies have the potential to deliver significant outcomes in the water sector. In a 2017 report, the World Economic Forum identified a $100 trillion opportunity by 2025 for both industry and society through the adoption of these technologies [23].

Digital water is about setting the foundation for utilities to begin applying data science and related technologies (such as IoT, AI, blockchain and augmented reality) to amplify the power of data to optimise real-world decision-making across water networks. Digitalization has the potential to take water utilities into the twenty-first century where almost everything can be measured. Digital twin solutions involve the creation of a virtual model of the real world and can be used to help understand the impact of incidents and prioritise the appropriate response actions. Gartner predicts that by 2021, half of the large industrial companies will use digital twins, resulting in a 10% improvement in effectiveness. Digital twins drive the business impact of IoT by providing a powerful way to monitor and control assets and processes. These virtual representations of water systems will enable situational awareness and/or near real-time hydraulic and quality monitoring, which has great potential to solve many of the business challenges faced by the industry such as improving efficiency, resilience, predictive maintenance and/or driven down operational costs. Figure 4 illustrates how water sits within a larger digital landscape. It is important that there is a communication strategy in place to bring together all ′digital′ stakeholders, including consumers, utility employees, regulators, environmentalists, etc., in the digital transformation. Most stakeholders and the communities they serve can understand the benefits of digitization. Water users are already profoundly connected with digital technologies permeating our daily lives in ways that were unimaginable several years ago. Digitisation has accelerated the collection and dissemination of actionable information to all stakeholder groups including customers. Early engagement by bringing these stakeholders into the discussions as early as possible encourages more collaboration and joint consideration around the ultimate vision. Customer′s expectations around sustainability are driving behavioural changes in traditional utility practices. Some consumers already participate in water conservation, and they will be able to do that more and more as utilities digitise, making smarter decisions about how they use and reuse their water using new sources of information such as smart meters. Smart-sensing technology, social media, utility web-monitoring portals and apps (quantified-self/digital customer), gamification and AI chatbots allow water consumers greater access to information and improved rates of engagement. At the same time, consumers are also coming up with innovations and hence driving the change rather than acting as a recipient only.

Fig. 4 The larger digitalization landscape that water companies operate within: data sources, issues and actors

6 Blockchain, Data Sharing and Web 3.0

Approximately 80% of data is closed (organisational) or personal – large volumes remain untapped behind firewalls. Secondary or tertiary uses of this data are not realised or known. The organisations that own closed data and the direct interactions between consumers and producers are essentially best positioned to dominate the market and create new business models that will lead to a successful digital transformation of a utility. How is it possible to unlock water sector proprietary data in faster, better and more trusted ways? The water sector has historically had trust issues with proprietary data. However, the potential benefits of open data (data that anyone can access, use and share) need to be considered, specifically by opening up previously closed data to data innovators. Data can be best considered as an ′e-infrastructure′ to facilitate this. Water companies are heavily reliant on data, but there are often concerns over data quality. A move to more standardised information representation for interoperability is needed for the future, resulting in independence from any particular computer software system and leading to standardised interfaces, technical information models and convertibility to XML and Web 3.0 formats. Web 3.0 refers to a semantic web, which is a

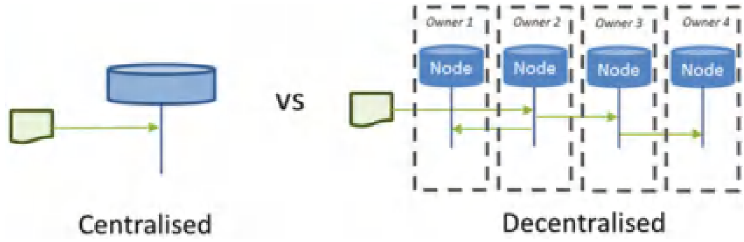


Fig. 5 Blockchain (distributed ledger) decentralised architecture

web where all information is categorised and stored in such a way that a computer (perhaps through AI) can understand it as well as a human. Machine to machine communication in Web 3.0 will be key to a step change in optimising performance with distributed middleware infrastructure allowing receiving systems to communicate with downstream applications. The new focus for Web 3.0 is user centric: decentralisation, privacy and security.

Blockchain is a shared, secure, distributed ledger which is ideal for applications that require trust and transparency. Figure 5 illustrates how there are many database copies (hence many owners) which result in resilience to both technical and organisational failure. The potential for blockchain to disrupt a wide range of industries has been well publicised, from Fintech to energy, governance, healthcare, supply chain, creative industries and many others. Originating in the cryptocurrency space, Bitcoin was introduced as a peer-to-peer version of electronic money [24]. Features include proof of work (cryptographic hash power to secure the network and mine new tokens), digital scarcity, immutability and censorship resistance. The technology is, in the most basic form, a digital record distributed across the Internet and allowing two users to conduct a transaction without the need for a third party or centralised intermediary (such as a bank, VISA or PayPal). Blockchain can provide a mathematically secured (using cryptography), verifiable and traceable database or a ′ledger′ of transactions (hence also referred to as distributed ledger technology (DLT)). Blockchain ledgers are tamper-proof and safe from malicious actors because the data does not exist in any single location. Innovators and regulators continue to believe that tokens (virtual utility or security digital tokens) can be foundational to Web 3.0 infrastructure and represent the opportunity for new business models including the convergence with other technologies including AI and IoT.

Blockchain technology could be used to validate transactions, ensure trust (hence security) and reduce costs for IoT. It can potentially be employed to trace billions of connected devices and process microtransactions (machine to machine) between them. However, large and busy networks can potentially suffer from substantial latency. Propagation issues may leave significant parts of the network out of the consensus loop (out of date). Another bottleneck can be transaction backlogs or block size issues leaving nodes unable to participate in consensus due to local issues. Practical blockchain implementations need to address these risks through simplification and compromise (such as through the consensus mechanism). Criticisms regarding Bitcoin′s usage of increasing amounts of electricity (and hence water) have been recently levelled, particularly in light of a desire to reduce carbon footprints [25]. However, a blockchain only needs substantial power if trying to secure something valuable on a public permission-less platform. A private permissioned blockchain requires very little proof of work security. This type of blockchain relies on user access control (for trust) and consensus and governance being achieved by other methods (such as proof of stake), and these platforms are being implemented by large corporations for a wide range of applications.

Blockchain technology could alleviate utility fears about the security and confidentiality of data by creating transparent supply chains and provide analysts with better datasets to use in their analyses. Data mining, analytics and AI depend crucially on the quality and quantity of data so this has been a bottleneck. Blockchain could help with sharing of data among utilities. They are often reluctant to contribute data to joint databases, often citing confidentiality and commercial sensitivity as the key reason. With blockchain technology, it is also possible for industries, consumers, households, and water managers to retrieve valuable information about water quantity and quality and help make informed decisions. This data empowers users to make better decisions about water usage and conservation, once regulatory constructs exist (with near-term potential in water billing and supply management). Other use cases include:

- Peer-to-peer trading. Blockchain can deliver a financial platform for decentralised water treatment and management in local communities in a similar manner as solar panel energy trading. Civic Ledger has used blockchain to improve water trading in Australia, allowing smaller irrigators to enter the market by reducing intermediaries. This company has also been investigating how a blockchain micro trading platform could open up rainwater trading with residents near Melbourne.
- Smart contracts and settlements. A water treatment technology company Origin Clear has developed a system called ′Waterchain′ to finance water treatment by embedding smart contracts and a cryptocurrency for payments on a decentralised water funding platform.
- Water rights trading – such as trading water credits and tracking shortfall and surplus with a digital twin.
- Buying and selling energy.
- Cybersecurity.
- Capital raising.

7 Smart Networks

Smart cities need to bring together hard infrastructure, social capital including local and community institutions and technologies to fuel sustainable economic development and provide an attractive environment for all [26]. The concept of a smart city within the context of water means using technologies for optimising water resources and waste treatment, monitoring and controlling water and providing real-time information to help water companies and households manage their water better. Smart water networks have been described as layered architecture, beginning with the sensing-and-control layer through data collection and data management and ending with the data fusion-and-analytics layer [27]. Although the technology components for smart water cities are available, the route to application is uncertain. The main hurdles are lack of integrated and open solutions, difficulty to comply with user and integration requirements, lack of clear and validated business cases for solutions, lack of business intelligence awareness and lack of political and regulatory support.

The quantity and complexity of sensor and environmental data is growing at an increasing rate, while the demands for new solutions and tools to utilise and interpret this data are likewise growing due to financial and regulatory pressures. The phrase ′big data′ may then become a reality for the water sector particularly on the customer side, since when smart metering becomes more prevalent a huge amount of data will be collected. If the UK goes to a point where the entire water industry is universally metered with smart metering, there will be approximately 25 million water meters for customers. Organising, managing and supporting such massive ICT network infrastructure, however, are substantial technical challenges. This data could be used, in conjunction with mapping software and hydraulic models to map consumption in DMAs where there are spikes in usage.

As demand for clean water increases with population growth in the coming decades and supply remains stagnant or shrinks due to climate change, solutions to manage and minimise leaks will become increasingly critical. Many water utilities are struggling to measure and locate leaks in their distribution networks beyond the economic level of leakage, and there is a drive to efficiency by implementing leak-reducing solutions. Leakage results in wasted energy costs (such as spent pumping water), water treatment costs (energy and chemicals), misdirected repair activities, regulatory penalization and environmental damage to city infrastructure.

Smart water networks offer the potential to identify leaks early, thus reducing the amount of water that is wasted and saving utilities money. These solutions include the use of flow and pressure sensors to gather data, analyse the data using algorithms to detect patterns that could reveal a leak in the network and provide real-time data on the location of a leak. A condition monitoring approach for smart networks can allow the early detection of potential faults in assets. The integration of real-time analytics can facilitate rapid determination (i.e. before customers are impacted) of abnormal flow events.

A number of approaches from the fields of artificial intelligence and statistics have been applied for detecting abnormality in WDSs from time series data. Alert systems that convert flow and pressure sensor data into usable information in the form of timely alerts (event detection systems) have been developed with a focus on burst detection to help with the issue of leakage reduction. Analysis systems need to provide useful classifications of system status, events and conditions and not provide an onerous number of alerts or alarms to system operators who will otherwise ignore warnings hence compromising the value of the information.

Most of these systems are for detecting leaks/bursts at district metered area (DMA) level. DMAs are designed to be hydraulically isolated areas that are generally permanent in the system. Automated online analysis systems have the potential to be a useful tool for real-time identification of small- to medium-sized bursts. Their use promotes a more proactive approach to leakage management, with awareness of leakage incidences soon after they occur and before the customer is seriously impacted. Such systems make it feasible to identify and hence find and fix leaks that would previously have become background leakage, providing the potential to reduce the so-called Economic Level of Leakage. Mounce et al. [28] review approaches for event detection in WDS measured time series data, with a focus on data-driven methodologies for leak detection. Wu and Liu [29] also provide a more extensive summary of data-driven approaches and their performance. Mounce and Boxall [30] describe an online system pilot implemented with a UK water company using an ANN and fuzzy logic system for detection of leaks/bursts at DMA level. This AI system was not reliant on any special hardware or network configuration and produces intelligent ′smart alarms′. The system was subsequently commercialised as FlowSure (Servelec Technologies Ltd) demonstrating how academic research can have real-world impact.

Smart water network technologies have the potential to deliver an improved service to customers and cost-effective performance improvements for the water industry. Sensor technology and the ′big data′ they generate combined with advanced ML techniques are providing new opportunities for deeper understanding of WDS. SmartWater4Europe (SW4EU) was a European FP7 demonstration project. Four demo sites comprise solutions for leakage control, water quality management and energy optimisation incorporating sensors, data processing, modelling and analytics technologies. The UK demo site (TWIST) in reading investigated how emerging technologies can be used to create a SWN with near real-time notification of performance to enable proactive management and intervention (particularly for leakage). Technologies utilised during the pilot have included installation of flow instruments through full-bore hydrants, instruments capable of high-resolution monitoring (thus enabling the identification of pressure transients), AMR customer smart meters as well as traditional sensors. Three network leakage algorithms to detect leakage and other abnormalities (as soon as they occur) in the water network were tested and assessed: AURA BED alerts as described in [28, 31], dynamic bandwidth monitoring (DBM) and Netbase envelopes (developed by Crowder Consulting). Having a standard approach to test different algorithms allows an objective comparison of their effectiveness.

Increasing amounts of SWN data are only of real business value if this valuable resource is ultimately used to inform and support decision-making, i.e. data to information to insight to action. Projects such as SW4E allow the exploration, at demo and full WDS pilot scale, of deploying multiple smart network technologies, both hardware and software, and the multiplicative synergy between them. The deployment of a smart water network has its own challenges such as large network data stores, false positives, limited analytical capability, pipe location and condition, failure prediction, meter coverage, response to failures, etc.

Dealing with these challenges at a small scale, before increasing the scope and area, enables the understanding of the best way to do it and allows us to assess the risks and benefits. Successful demonstration of the return on investment business case will ultimately allow full-scale roll-out of smart DMAs. Successful demonstration of the return on investment business case will ultimately allow full-scale roll-out of smart DMAs.

The full possibilities of smart network technologies to deliver improved service to customers and cost-effective performance improvements in the water industry are yet to be realised. Smart metering technologies need to be able to support decisions at both the household and utility levels [32]. Sensor technology and the ′big data′ they generate combined with advanced machine learning techniques are providing exciting new opportunities for new scales of understanding of WDS. With increased DMA flow and pressure measurements comes the possibility of leak localisation at the sub-DMA level [33, 34]. AMR data expands information availability even further and has the potential to be used for customer profiling at the WSP side, allowing urban water planning and management based on consumer types and for informing customers on their water end-use patterns. Nguyen et al. [35] present a methodology using hidden Markov model and dynamic time warping algorithm techniques disaggregating customer data into its end-use categories, including for rapidly alerting consumers of occurring leak events. By utilising AMR for demand forecasting, the possibility for reducing costs for treatment, storage and distribution arises such as through optimisation of pump scheduling. Candelieri et al. [36] present a data-driven, fully adaptive self-learning algorithm for short-term water demand forecasting utilising AMR data. In the future, water and energy use could be more efficiently managed through smart meter adoption and changing the 24-h diurnal demand profile.

8 Use of Smart Meter Data

There is more to smart water metering than accurate billing, but the business case for investing in smart meters and automatic meter reading is often complex. The technological development and digitalization of the water industry show no sign of slowing down, but the right tools can help water utilities decrease the complexity of their daily work and transform their meter data into valuable knowledge.

The concept of a smart city places citizens at the centre of services within a city. This involves bringing together hard infrastructure, social capital including local skills and community institutions and technologies to fuel sustainable economic development and provide an attractive environment for all. For the water sector, this means using technologies for optimising water resources and waste treatment, monitoring and controlling water and providing real-time information to help water companies and households manage their water better. The increasing use of smart water metering technologies for monitoring networks in real time is providing water utilities with an evergrowing amount of data on their business operations and infrastructure. Advanced metering technologies coupled with informatics create an opportunity to form digital multiutility service providers [37]. Such metering devices embrace two distinct technologies: meters that record water usage and communication systems that can store and transmit real-time water use information [38]. The ideal approach for their smart city application is installing smart water meters at the property boundary in conjunction with intelligent end-use pattern recognition algorithms either in-built into the meter software or within a processing module at the utilities data centre. However, such an end goal requires the ability to analyse collected data without human interaction and manual reclassification, and this is non-trivial.

Increasing amounts of smart network data are now being collected by WSPs; however the data is only of real business value if this valuable resource is ultimately used to inform and support decision-making. The full range of uses for these observations is only beginning to be realised and exploited. Recent work has explored the use and analysis of such data. It has been argued that using actual observed data, demand profiles can be calculated to provide more accurate representations of high-granularity historical data, with potential applications in real-time leakage detection, customer profiling and the provision of network modelling demand patterns [39]. Time series clustering is an active area of research, with the major issues being high dimensionality, temporal order and noise [40]. McKennaa et al. [41] investigated employing Gaussian mixture models (GMMs) as the basis set for representing demand patterns using a dataset of hourly demand readings spanning a 6-month study period, for 85 service connections within a single DMA. While there was no customer information available for the dataset, it was hypothesised after applying k-means that evidence of patterns found may represent both residential and commercial customers. Garcia et al. [42] demonstrated the potential use of k-means for clustering AMR data based on shape. Hadoop and Spark were used in a big data context to provide an unsupervised classification of the demand patterns from smart meters, with hourly interval feature vectors of a weekly profile for 51,117 smart meters over a 1-year period (approximately 317 million observed readings). Nine distinctive clusters were identified. However, no additional information (including customer type) was available other than demands. In Mounce et al. [43], a case study of approximately 250 million readings is presented, using a workflow for cleaning and preprocessing AMR data and then clustering average daily demand patterns using the k-means ++ algorithm with a correlation distance metric. Three natural clusters in the data (confirmed by using silhouette plots) were found to correspond strongly to a residential and commercial composition based on customer type which was used for post-analysis. Further, a classification approach was also presented, comparing five classification models with K-fold cross-validation, in order to classify into residential and commercial customers. When using an ensemble of RUSBoosted decision trees (for a fivefold), the overall accuracy was 91.3% (TPR 92% for residential and 84% for commercial) confirming dominant patterns of usage.

Potential application areas for further work as more smart data becomes routinely available include:

- Data mining: understanding how businesses and households use water and whether and where unique patterns in this use exist is essential for proactive management (including weekday vs. weekend analysis).
- Applying customer segmentation based on consumption data for customer loading and variable water pricing in a similar manner to energy (some industrial customers already have tariffs based on time of day usage).
- For leak detection activities based on detecting pattern changes (deviation from cluster centroids/distributions); data-driven models of demand could also help identify atypical customers or unusual changes in consumption.
- Filling missing data for audits/regulatory purposes using cluster centroids/typical usage perhaps allowing volumetric usage and flow profiles to be estimated for unmetered customers.
- Allowing a WSP to forecast the demand of a client or the overall DMA and hence improved network operations.
- More accurate demand profiles for hydraulic modelling.

9 Discussion of Technology Adoption and Recommendations

Water is considered a ′right′ in many parts of the developed world which have grown accustomed to clean drinking water and sanitary facility provision. Many of the world′s leading water companies have been around for decades, even centuries. They enjoy high prestige, low staff turnover and healthy margins. Few see threats from competitors poaching metered customers who are tethered to miles of buried pipe infrastructure. Regulation can also be a barrier to innovation. As natural monopolies, the incumbent water utilities generally feel safe and insulated from competition. Digitisation was once seen as a luxury. However, innovators are disrupting the old business models, and there is little place anymore for complacency with threats such as decentralised and distributed technology arising. Disruptive innovation need not be a zero-sum game in which only one side emerges victorious. There is no reason why water utilities cannot learn from insurgents, engage with new thinking and embrace innovation to update their business models to deliver new solutions that benefit all.

A further issue is that the water sector is not generally perceived as a ′cool′ industry, partly due to it not being at the forefront of the technology adoption curve. In contrast, new digital technologies are a hot topic particularly as machine learning and AI begin to proliferate into industrial application. This fact means that a career in the water industry generally isn′t a top priority for data scientist professionals and attracting the right type of talent can be difficult. Startups in the water sector can rarely compete financially with Google for hiring coders, graphic designers and tech engineers. Nor can potential return on investment compete with unicorns like Facebook, Netflix or Uber. However, the sector needs to avoid being a slow adopter of data science and consequently should consider investing decisively in this direction. Solutions include:

- Some key recommendations to drive the digital agenda more generally are as follows:
  - First, secure executive buy-in, and then devise a digital strategy with an action plan (and stick to it).
  - Second, build the technological foundation by ensuring the basics are in place to support future growth.
  - Third, focus on business priorities, and communicate quick wins to tie the investment in digital to outcomes that support the strategy (immediate pay back is often needed in order to gain approval for implementation).
- Promote in-house expertise and roles (e.g. data scientist) hence embedding personnel within water companies. Leveraging existing toolboxes (particularly for machine learning) and workflows is becoming easier so that the barrier to entry has now been now lowered (PhD qualifications were usually required in the past for AI).
- Adopting open-source programming languages (such as Python) and tools can facilitate collaboration and shared development. Software produced in water engineering research has previously often been bespoke and stand alone, and more thought needs to be directed at reusability, sustainability and maintainability. For example, university research project software is usually developed in isolation in languages such as MATLAB and may only persist through one or two generations of researchers (such as passed down from an academic to a PhD student).
- In the UK, centres for doctoral training like STREAM and WISE embed a doctoral engineer in water companies helping such transfer.
- Other collaborations between industry, SMEs and research establishments in projects (e.g. through European H2020 or UK Innovate funding) encourage knowledge transfer. Data dives and hackathons provide forums for building teams of collaborators.

Some private sector water utilities are already leaders in the digital space and active in sharing their early successes with leveraging connected devices, IoT and machine learning. This digital re-imagination of the sector will enable a broad spectrum of outcomes from improved efficiencies to optimised asset management across providers and potentially new business models such as consolidated multiutility retail operations. While there is an increase of digital adoption in water, the sector still lags behind other industries in integrating new, smart technologies. Water utilities can benefit from the lessons learned in other sectors (such as energy) and the established best practices and network infrastructures. As technology evolves, the price for smart devices decreases, the functionality increases, and consequently piggybacking on other sectors′ lead can result in an accelerated adoption rate for realisation of benefits in the water ecosystem.

Cybersecurity is a growing concern today, and the risk is increasing. Historically, WSP control systems were not designed with security in mind, and while this alone doesn′t make them vulnerable, considerations must be made when digitising an existing system with older applications and tools. Increasingly, there are threats around the critical control systems, especially those that control water flows, so treatment works and other key infrastructure are potential security risks.

10 Summary

In a game-changing period of rapid technological transformation, smart networks are at the forefront of investment plans for UK water companies as part of a progression to a circular economy [6]. Technological advancements allow water companies to gather more information about their networks and assets than ever before. Extra data provides new opportunities for asset maintenance and event analytics. Digital water will bring automation and connect the sector to the Internet of Things. Technologies like data analytics, cloud computing, augmented intelligence and blockchain provide new capabilities to analyse, automate, correct in real time, predict and minimise risks. Edge computing will help make IoT roll-outs more integral and core to the way businesses work in coming years.

Water utilities are investing in new technologies that improve customer service, enhance efficiency and drive resilience. A smart water network is a fully integrated set of products, solutions and systems that enable water utilities to remotely and continuously monitor and diagnose problems, pre-emptively prioritise and manage maintenance issues and remotely control and optimise all aspects of the water distribution network using data-driven insights.

When considering artificial intelligence, cloud and the latest sensor technology from the standpoint of a water utility, many are still struggling with paper reports, so there remains a long road ahead. The digitalisation of water is no longer optional. It takes an entire ecosystem to digitise a utility, including innovation partners and supply chain providers. By ′digitising the utility′, metered account holders can grow more involved and act on valuable knowledge on their usage, costs and conservation strategies.

With AI becoming more prevalent across industries, there is a growing need to make it broadly available, accessible and applicable to engineers and scientists with varying specialisations. Engineers, not just data scientists, will drive the experimentation and adoption of AI in water industry applications. Complexity of larger datasets, embedded applications and bigger development teams will drive solution providers towards interoperability, greater collaboration, reduced reliance on IT departments and higher productivity workflows.

References

1. Caffoor I (2010) Perspective: a vision of a low-carbon water sector in 2050. Proc Inst Civil Eng-Eng Sustain 163(1):9–14. https://doi.org/10.1680/ensu.2010.163.1.9
2. Boxall JB, Machell J, Dewis N, Gedman K and Saul A (2011) In: Dragan A. Savic and John K. Banyard (eds) Operation, maintenance and performance in water distribution systems, ICE, London. ISBN: 978-0-7277-4112-7.
3. SWIG (Sensors for Water Interest Group) (2007) ′SCADA systems for future management of assets′ 24 Jan 2007
4. Bingi P, Sharma MK, Godla JK (1999) Critical issues affecting an ERP implementation. Inf Syst Manage 16(3):7
5. IBM (2014) Web ref. Retrieved from 30 Nov 2014
6. UKWIR (2019) Smart metering and smart networks for leakage management report. UK Water Industry Research, London – ISBN: 978-1-84057-880-5
7. Conroy P, De Rosa S (2015) Framework for expenditure decision making: development of service forecasting approaches. UK Water Industry Research Limited, London
8. Mounce SR, Douterelo I, Sharpe R, Boxall JB (2012) A bio-hydroinformatics application of self-organizing map neural networks for assessing microbial and physico-chemical water quality in distribution systems, Proceedings of 10th international conference on hydroinformatics, Hamburg, Germany
9. Mounce SR, Jensen HJ, Boxall JB, Biggs CA (2012) Data mining T-RFLP profiles from water system sampling using self-organising maps. Proceedings of 8th international conference on natural computation and the 9th proceedings of international conference on fuzzy systems and knowledge discovery (IEEE), vol 1, pp 219–223.
10. Mounce SR, Sharpe R, Speight V, Holden B, Boxall JB (2014) Knowledge discovery from large disparate corporate databases using self-organising maps to help ensure supply of high quality potable water. Proceedings of 11th international conference on hydroinformatics, New York, USA
11. Ellis K, Mounce SR, Edwards J, Speight V, Jakomis N, Boxall JB (2015) Interpreting and estimating the risk of iron failures, Computing and Control for the Water Industry (CCWI2015). Procedia Eng 119:299–308
12. Blokker EJM, Furnass WR, Machell J, Mounce SR, Schaap PG, Boxall JB (2016) Relating water quality and age in drinking water distribution systems using self-organising maps. Environments 3(2):10. https://doi.org/10.3390/environments3020010
13. Mounce SR, Blokker EJM, Husband SP, Furnass WR, Schaap PG, Boxall JB (2016) Multivariate data mining for estimating the rate of discoloration material accumulation in drinking water distribution systems. IWA J Hydroinf 18(1):96–114
14. Mackey T, Mounce SR, Boxall JB, Cashman A (2017) A multivariate geospatial data-driven approach to descriptive modelling of burst behaviour in a small island context. Proceedings of computing and control for the water industry (CCWI2017), Sheffield, UK
15. Kohonen T (1990) The self-organizing map. Proc IEEE 78(9):1464–1480. http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html
16. Sharpe RL, Smith CJ, Biggs CA and Boxall JB, (2010) Pilot scale laboratory investigations into the impact of steady state conditioning hydraulics on potable water discolouration. Proceedings of 11th annual international symposium on water distribution systems analysis, Arizona, USA, September
17. Ester M, Kriegel HP, Sander J, Xu X (1996) Simoudis E, Han J, Fayyad UM (eds) A density-based algorithm for discovering clusters in large spatial databases with noise. AAAI Press, Palo Alto, pp 226–231
18. Sammon JW (1969) A nonlinear mapping for data structure analysis. IEEE Trans Comput C-18(5):401–409

19. van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605
20. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444
21. Mounce SR (2017) Visualising smart water meter dataset clustering with parametric t-distributed stochastic neighbour embedding. Proceedings of 13th international conference on natural computation, fuzzy systems and knowledge discovery (IEEE). Guilin, China. vol 1, pp. 1940–1945. ISBN: 978-1-5386-2165-3
22. UKWIR (2013) Cost Benefit Analysis of Ubiquitous Data Collection in Water Distribution – CBA Scenarios. 13/DW/12/2 – ISBN: 1 84057 692 8
23. World Economic Forum (2017). http://reports.weforum.org/digital-transformation/introducing-the-digital-transformation-initiative/
24. Nakamoto S (2009) Bitcoin: a peer-to-peer electronic cash system. Cryptography Mailing list at. https://metzdowd.com.
25. IWA (Source) (2018) How Bitcoin′s footprint is impacting water use. https://www.thesourcemagazine.org/how-bitcoins-footprint-is-impacting-water-use/. Accessed 18 Mar 2019
26. Bull R, Azennoud M (2016) Smart citizens for smart cities: participating in the future. Proc Inst Civil Eng-Energy 169(3):93–101. https://doi.org/10.1680/jener.15.00030
27. Cahn A (2014) An overview of smart water networks. J Am Water Works Ass 106(7):68–74
28. Mounce SR, Mounce RB, Jackson T, Austin J, Boxall JB (2014) Associative neural networks for pattern matching and novelty detection in water distribution system time series data. IWA J Hydroinf 16(3):617–632
29. Wu Y, Liu S (2017) A review of data-driven approaches for burst detection in water distribution systems. Urban Water J 14(9):972–983. https://doi.org/10.1080/1573062X.2017.1279191
30. Mounce, S. R. and Boxall, J. B., (2011). Chapter 9: online monitoring and detection, water loss reduction. Bentley Institute Press, Exton Ed. Zheng Wu, ISBN: 978-1-934493-08-3.
31. Mounce SR, Fargus A, Weeks M, Young J, Jackson T, Goya E, Boxall JB (2017) Online advanced uncertain reasoning architecture with binomial event discriminator system for novelty detection in smart water networks, Proceedings of computing and control for the water industry (CCWI2017), Sheffield, UK, 5th to 7th September 2017
32. Savić D, Vamvakeridou-Lyroudia L, Kapelan Z (2014) Smart meters, smart water, smart societies: the iWIDGET project. Procedia Eng 89:1105–1112
33. Farley B, Mounce SR, Boxall JB (2013) Development and field validation of a burst localization methodology. ASCE J Water Resour Plann Manage 139(6):604–613
34. Romano M, Kapelan Z, Savic DA (2013) Geostatistical techniques for approximate location of pipe burst events in water distribution systems. J Hydroinf 15(3):634–651
35. Nguyen KA, Stewart RA, Zhang H (2013) An intelligent pattern recognition model to automate the categorisation of residential water end-use events. J Environ Modell Software 47:108–127
36. Candelieri A, Soldi D, Archetti F (2015) Short-term forecasting of hourly water consumption by using automatic metering readers data. Proceedings of CCWI 2015, Leicester, UK, 2nd–4th September
37. Stewart RA, Nguyen K, Beal C, Zhang H, Sahin O, Bertone E, Vieira AS (2018) Integrated intelligent water-energy metering systems and informatics: visioning a digital multi-utility service provider. Environ Model Software 105:94–117. https://doi.org/10.1016/j.envsoft.2018.03.006
38. Stewart RA, Willis R, Giurco D, Panuwatwanich K, Capati G (2010) Web-based knowledge management system: linking smart metering to the future of urban water planning. Aust Planner 47:66e74
39. Vitorinoa D, Loureirob D, Alegreb H, Coelhob S, Mamadeb A (2014) In defense of the demand pattern, a software approach. Procedia Eng. 89:982–989
40. Rani S, Sikka G (2012) Recent techniques of clustering of time series data: a survey. Int J Comput Appl 52(15):1–9

41. McKennaa SA, Fuscoa F, Ecka BJ (2014) Water demand pattern classification from smart meter data. Procedia Eng 70(2014):1121–1130
42. García D, Gonzalez D, Quevedo J, Puig V, Saludes J (2015) Water demand estimation and outlier detection from smart meter data using classification and Big Data methods. In: Proceedings of 2nd IWA new developments in IT and water conference, Rotterdam, February 2015
43. Mounce SR, Furnass WR, Goya E, Hawkins M, Boxall JB (2016) Clustering and classification of aggregated smart meter data to better understand how demand patterns relate to customer type, Proceedings of computing and control for the water industry (CCWI2016), Amsterdam, Netherlands, 7th to 9th November 2016

Review of Techniques for Optimal Placement of Pressure and Flow Sensors for Leak/Burst Detection and Localisation in Water Distribution Systems

Michele Romano

Contents

Abstract Optimal sensor placement in water distribution systems (WDSs) for leak/burst detection and localisation is a well-established and very productive research field. Its primary focus is to minimise the cost of a proposed sensor network infrastructure while maximising the capability to detect and localise leaks and bursts through the analysis of the collected data. This chapter reviews in a systematic manner relevant articles published over the last decade aiming at rationalising the work carried out in this field. It presents a synthesis and analysis of the relevant published works to (1) provide insight and awareness of differing arguments, theories and approaches, (2) highlight their capabilities and limitations and (3) identify the state of the art in their development. It also provides insight and awareness of differing approaches that have been proposed to tackle specific issues encountered

M. Romano (*)
United Utilities Water Limited, Warrington, UK
e-mail: Michele.Romano@uuplc.co.uk

Andrea Scozzari, Steve Mounce, Dawei Han, Francesco Soldovieri, and Dimitri Solomatine (eds.), ICT for Smart Water Systems: Measurements and Data Science, Hdb Env Chem (2021) 102: 27–64, https://doi.org/10.1007/698_2019_405,

© Springer Nature Switzerland AG 2019, Published online: 18 September 2019

by researchers when developing their proposed techniques such as model and measurements uncertainties. Trends and gaps in the current research and future research directions are identified and discussed in this chapter, and a number of considerations to promote further developments in this important field of research are presented. The desired outcome of this chapter is to serve as a useful resource for researchers and practitioners involved in sensor network design for leak/burst detection and localisation and in the development/adoption of leak/burst detection and localisation techniques.

Keywords Leak and burst detection and localisation, Literature review, Pressure and flow sensors, Sensor network design

1 Introduction

The problem of leak and pipe burst events in water distribution systems (WDSs) is a compelling issue for water companies worldwide. Leak and pipe burst events not only cause economic losses to water companies [1] but also represent an environmental issue (i.e. waste of water and energy) and a potential risk to public health [2]. Furthermore, they have a negative impact on water companies′ operational performance, customer service and reputation. Currently, a wide range of leak/burst event detection and location techniques exists that are based on various principles [3–7]. However, none is ideal, and the number of techniques currently practised by water companies is limited. In many cases, pipe bursts are brought to the attention of a water company only when someone calls in to report a visible event. Water companies embracing modern leakage management technologies devote considerable manpower and resources to proactive detecting and localising leaks and pipe bursts by utilising techniques that make use of highly specialised hardware equipment (e.g. leak noise correlators, acoustic sensors mounted on inline pipeline inspection gauges, ground penetrating radars, etc.). Despite some of these techniques being the most accurate ones used today [6], they are also costly, labour-intensive and slow to run. Consequently, much research has been focused on finding inexpensive (i.e. numerical) techniques that can help the water companies significantly reducing the leaks/bursts′ lifecycle by making them aware of the occurrence of these events much faster and guiding the water company personnel straight to the problem areas.

In the above scenario (and bearing in mind that in the last decade the importance of a proactive approach to network management and near real-time assets monitoring have become apparent as water companies have had to deal with tightening regulatory and budgetary constraints), it is clear that instrumentation and analytics can play a vital role in addressing the aforementioned issues. In the UK, and as recommended internationally by the International Water Association (IWA), WDSs are divided into District Metered Areas (DMAs), which may consist of approximately 300–5,000 properties. The current UK industry practice is to install a flow sensor (and sometimes a pressure sensor as well) at the inlet of a DMA and any outlet (e.g. to another DMA or to a large industrial user) and a (supplementary) pressure sensor at the so-called critical monitoring point within the DMA (i.e. the point located either at the point of highest elevation or alternatively at a location farthest away from the inlet). With recent improvements in sensor technology and communication technologies (such as GSM, GPRS and, more recently, LORAWAN, Sigfox, NB-IoT and 5G), data can now be transferred via wireless systems, and batteries last much longer, meaning that sensors can be placed in less accessible areas and data from these devices can be received in near real time (e.g. every 15 min). Furthermore, it is becoming more feasible to deploy larger numbers of instruments per DMA, as the cost of both pressure and flow instrumentation (and their maintenance) has been reduced considerably. As a result, a vast amount of pressure and flow data originating from the many DMAs that typically form a UK WDS is now frequently available and expected to quickly grow over time. This data can give insights into the operation and current/future status of water networks and support many water loss-related activities, such as estimating background leakage levels, establishing and maintaining hydraulic models of water systems and detecting and localising new leaks and bursts as they occur. With regard to the latter, data-driven techniques utilising machine learning and advanced statistical tools have been developed that automatically manage and analyse in an on-line fashion increasing numbers of near real-time data streams aiming at enabling the detection and (in certain instances) the approximate location of leaks, bursts and other similar network events (e.g. [8–18]). These techniques can complement traditional leak/burst localisation methods such as acoustic surveys, which can then be used for accurately determining the exact leak/burst position (i.e. pinpointing). The value of the information that can be derived through analysis of sensor data and hence the success of the aforementioned methods (especially for localisation), however, is critically linked to the number and types of sensors deployed and their locations. As previously mentioned, it is envisaged that in the near future higher numbers of sensors (especially pressure, for their lower cost and easier installation and maintenance when compared to flow sensors) will be used to monitor WDSs. However, due to the financial constraints placed on water companies, the costs of increased instrumentation in WDSs (both capital and ongoing maintenance) must be weighed against the operational and other cost savings which can be made by improving network operations and management. It is therefore desirable to limit the number of additional instruments to be deployed by selecting the optimal number and location of sensors in a DMA.

This chapter provides a critical state-of-the-art literature review on the subject of optimal sensor placement in WDSs for leak/burst detection and localisation. It provides details of a number of existing sensor macro-location design methodologies intended to facilitate the efficient collection of relevant measurements in WDSs for that specific purpose. Generally speaking, the optimal placement of a limited number of sensors within a WDS (′ideal′ locations at which measurements of selected quantities should be taken) is a necessary step in the application of intelligent and cost-effective monitoring for current and prospective WDSs. The definition of an "optimised" sensor network is dependent on the intended purpose of the sampling scheme and the resulting sensor data. Design methodologies in the literature are typically catered towards one of a number of distinct agendas, and the field is consequently segmented into a range of subsidiary groups – i.e. methods to determine optimal placement schemes for effective contaminant detection (e.g. [19–25]), methods to determine optimal placement schemes for model calibration (both hydraulic, e.g. [26], and water quality, e.g. [27]) and methods for leak and burst detection, each formulation of which may be largely irrelevant outside of its own context. Although a large amount of methods that consider how to identify the optimal placement of both pressure and flow instrumentation within WDSs at the DMA level for leak/burst detection and localisation can be found in the sampling design literature, a comprehensive review of their capabilities, limitations and other aspects important for assessing the potential of these techniques to be beneficially utilised by water companies has not yet been presented in any review paper.

As only a limited number of sensors can be installed in WDSs due to budget constraints and since improper selection of their location may seriously hamper leak/burst detection and localisation performance, the development of optimal sensor placement strategies has become an important research issue in recent years. This chapter aims at rationalising the relevant published works in the field and is organised as follows. After this introduction, Sect. 2 presents a synthesis and analysis of the relevant published works aimed at (1) providing insight and awareness of differing arguments, theories and approaches, (2) highlighting their capabilities and limitations and (3) identifying the state of the art in their development. Section 3 focuses on specific issues encountered when developing techniques for the optimal placement of sensors for leak/burst detection and localisation that researchers in this field have tried to address (e.g. model and measurements uncertainties, simultaneous use of pressure and flow sensors, etc.) and provides insight and awareness of differing approaches that have been proposed in those contexts. Section 4 presents considerations regarding, inter alia, the potential of the proposed techniques to help water companies minimising the leaks/bursts′ runtime by effectively detecting and localising these events as they occur in a DMA and the gaps in the current research. Finally, a summary of the chapter containing the main conclusions and highlighting the key considerations made is given in Sect. 5 in order to promote further developments in this important field of research.

The desired outcome of this chapter is to serve as a useful resource for researchers and practitioners involved in sensor network design for leak/burst detection and localisation and in the development/adoption of leak/burst detection and localisation techniques.

2 Synthesis and Analysis of Optimal Sensor Placement Techniques for Leak/Burst Detection and Localisation

The amount of information on a leak/burst event occurrence in a pressure or flow signal from a DMA is a function of the number and types of sensors and their locations, as well as the DMA structure and event location, among the others. As such, measurements at some locations can include more information regarding an event than measurements at other locations. The main aim of optimal sensor placement for leaks/burst detection and localisation techniques is therefore to place the minimum number of sensors in a DMA to capture the event "effects" no matter where in a DMA the event occurs and then effectively use this information to provide detection alarms and accurately identify the approximate event′s location.

Model-based leak/burst detection and localisation techniques, using pressure and flow measurements and hydraulic models of WDSs, have been studied for approximately two decades, since the paper by Pudar and Liggett [28], which formulates the leak detection and localisation problem as an indirect (see [29]) least-squares parameters estimation problem. However, the estimation of the parameters describing a WDS model is a difficult task since these models are non-linear. This said, with the papers by Farley et al. [30] and Pérez et al. [31], the last decade has seen a large number of papers published on this subject that attempt to use direct (see [29]) methods to solve the leak/burst detection and localisation problem. Almost all these studies work by running multiple hydraulic model simulations of various leak/burst scenarios and then evaluating the sensitivity of different monitoring points to the imposed ′fault′ conditions. Because of this, many of these studies are inspired by the model-based fault diagnosis theory (see, e.g. [32]), the main objectives of which are maximising fault detectability (i.e. ability to identify a fault occurrence in a system) and fault isolability (i.e. ability to distinguish between two possible fault occurrences – as, if the effects of different faults are similar, they may result in similar sensors′ measurements). However, different approaches have also been proposed.

Farley et al. [30] introduced an approach which simulates, for an idealised 24 h period, leaks/bursts at all possible locations in a DMA (i.e. as an emitter at all the nodes of the relevant hydraulic model) and subsequently builds a matrix with rows corresponding to the possible leak/burst points and column corresponding to possible monitoring points. Each element of this matrix contains the sum, $\sum X^2$ (over the 24 h period), of instantaneous chi-squared values computed as $X^2 = \frac{(P_{lc} - P_n)^2}{P_n}$, where P_{lc} is the simulated pressure recorded under leak/burst conditions and P_n is the simulated pressure recorded under normal conditions. A threshold computed as the mean of the matrix is then applied to each value in the matrix to map those values to zeros (i.e. leak/burst undetected) and ones (i.e. leak/burst detected) by simply checking whether the specific $\sum X^2$ is less or greater than the threshold, respectively. The column with the largest number of ones is considered as the most sensitive sensor location. The authors expanded this methodology to determine the best combination of two sensor locations by using a complete enumeration procedure of all possible pairings of locations. They tested their method on two UK DMAs with different geometries assuming a perfect model and no measurements noise. Through comparison with the leak/burst detection performance of already installed instrumentation (according to UK standard practices), they demonstrated that the optimal location(s) identified using their method enable detecting a higher percentage of simulated leak/burst events. The authors noted however that the threshold selection is cause of concern as, if obtained from the simulation of large leaks/bursts, smaller events may not be detected. Of particular importance is the fact that, in a later study [33], the authors conducted a set of field trials to evaluate their approach. These field trials simulated five different leak/burst events through the opening of fire hydrants within a selected DMA. By installing pressure instrumentation at different locations in the DMA, an understanding of how accurately the model methodology can determine sensitivity of instrument location was obtained. Indeed, the results showed that pressure instrumentation location is crucial to sensitivity and that their modelling methodology was able to predict instrument location sensitivity to leak/burst events reasonably well.

Farley et al. [11, 34] built on the work carried out in [30, 33] and proposed to search the sensitivity matrix to achieve selective sensitivity to events in different network areas. By doing this, their approach enabled providing useful leak/burst localisation information by subdividing a DMA in smaller detection zones. The main differences from the work presented in Farley et al. [30, 33] are that a genetic algorithm (GA – see, e.g. [35]) is used to improve the search efficiency when identifying the best sensor locations and an uncertainty band is applied to either side of the threshold to (somehow) account for a certain degree of model/measurements uncertainty. The single objective function of the GA search aims to identify combinations of instruments that provide an even division of the DMA and to minimise the number of nodes within the penalty zone (i.e. a zone whereby a response within the uncertainty band is produced at one or more instruments). The authors presented results from field tests using hydrant flushing to simulate leak/burst events in real DMAs. The field tests′ results demonstrated the practical applicability of the method, showing that by combining quantification of differential sensitivities with event detection techniques for data analysis (i.e. [12, 13]), events can effectively be localised using a small number of instruments (i.e. taking into account existing instrumentation and one or two additional pressure sensors). However, it was noted that the effectiveness of the localisation method was dependent on where in a DMA a leak/burst event occurs (e.g. event near a DMA inlet are likely to be missed) and, most importantly, that the method only works if all the considered instruments are working and the event detections from all the sensors′ data agree with what the model expects to happen (i.e. an incorrect or uncertain detection even at a single sensor location would cause the method to indicate an incorrect leak/burst search area).

Pérez et al. [31] proposed a sensor placement method conceptually similar to the one presented in Farley et al. [30]. This method is based on computing and analysing the differences (i.e. residuals) between the pressure measurements at the sensor locations following a leak/burst and their estimations obtained using a hydraulic model. The basic idea is that the values of these residuals for a particular leak/burst can be seen as that leak/burst signature. The calculated residuals are then evaluated against a threshold (that may be selected to take into account the measurement noise and model uncertainty – [36]). If a residual violates the threshold (for a given time window, in the general case) then, the leak/burst isolation process is initiated. The isolation process is based on comparing the residuals against the leak/burst sensitivity matrix, $S(k)$, that contains the effect of each possible leak/burst on the available pressure measurements at the sensor locations. A mathematical representation of the sensitivity matrix is shown in Eq. (1) [29]:

$$S(k) = \begin{bmatrix} \frac{p_1^{f_1}(k) - p_1(k)}{f_1} & \cdots & \frac{p_1^{f_m}(k) - p_1(k)}{f_m} \\ \vdots & \ddots & \vdots \\ \frac{p_n^{f_1}(k) - p_n(k)}{f_1} & \cdots & \frac{p_n^{f_m}(k) - p_n(k)}{f_m} \end{bmatrix} \tag{1}$$

where $p_i^{f_j}(k)$ is the pressure of sensor i at the time instant k when a leak/burst with constant flow, f_j, is present at node j, m is the number of nodes in the network (possible leak/burst locations – if leaks and burst are assumed as occurring at nodes), n is the number of sensors in the network and $p_i(k)$ represents the pressure of sensor i at the time instant k without the presence of a leak/burst in the network. The candidate leaks/bursts are those whose effect matches the best (when compared using some metric) with the observed residuals. In this study the authors proposed to normalise (i.e. divide each row by the maximum value of that row that corresponds to the leak/burst most important for that node) and then binarise the sensitivity matrix in order to be used as a leak/burst signature matrix (i.e. set of all the leak/burst signatures). The threshold used to evaluate the residuals and to binarise the sensitivity matrix was identified by looking at the trade-off between the number of unique signatures present in the binarised sensitivity matrix and the number of leaks/bursts with the same signature. The authors used a GA-based optimisation to find the location of sensors that minimises the maximum number of possible leaks/bursts with a particular signature (i.e. maximise isolability) for a pre-specified number of sensors. This method was tested using the hydraulic model of a real WDS, Placa del Diamant, in the Barcelona WDS. Here the authors simulated leaks/bursts as a single, constant demand that can appear in any node, considered a single time step k, assumed the availability of a perfect hydraulic model (e.g. no model uncertainty) and did not account for measurements noise. Subsequently, however, Pérez et al. [37] repeated their experiments considering multiple time steps (i.e. a particular time step during the night-time period for 15 days – they introduced a voting mechanism to then assign a leak/burst to a particular group of nodes) and uncertainty in the nodal demands. The authors found that, independently by the presence (or absence) of the added uncertainty, their approach required to recalculate new sensitivity matrices every day – because these matrices are strongly affected by the changing boundary conditions and total consumption. Furthermore, they found that localisation performance strongly decreases when nodal demand uncertainty is introduced, in addition to different sensor placement configurations being identified as “optimal” despite using the same number of sensors.

One of the main issues in the work presented in Pérez et al. [31, 37] is the threshold selection. Furthermore, even with an optimal threshold selection, binarising the sensitivity matrix leads to a loss of information [38]. Therefore, aiming at circumventing these issues, Casillas et al. [39] formulated the optimal sensor placement problem as an integer optimisation problem based on projections from a non-binarised leak/burst sensitivity matrix solved with a semi-exhaustive search or a GA. The projection-based method used in this study (i.e. angle method) is based on evaluating the angle between the vector of the “actual” residuals and every column (i.e. possible leak/burst nodes) of the leak/burst sensitivity matrix. The “actual” leak/burst node is then identified by looking at the column (sensitivity matrix vector) that presents the smallest angle with the residual vector. This method was first proposed by Casillas et al. [29] for the sole purpose of leaks/bursts localisation. It was then compared in that study and in Casillas et al. [9, 10] against other ways of using the leak/burst sensitivity matrix to isolate/localise a leak/burst (including the binarisation method proposed by [31] and the correlation method presented in [38] and in [40]), and, through tests on small synthetic networks and on a real-life network (i.e. Nova Icaria, in the Barcelona WDS), it was found to offer better localisation performance than the other tested methods. With specific regard to the method used for solving the integer optimisation problem, the authors evaluated the performance of a semi-exhaustive search, which uses a lazy evaluation mechanisms to reduce the computation cost by discarding potential sensor configurations as soon as it is found that they cannot be candidates for the optimum solution, against the performance of a GA on the Hanoi network (see [41]) and on a relatively small real-life network in Limassol, Cyprus. They found that the semi-exhaustive search would not scale up well to bigger networks, whereas the GA allowed the finding of good near-optimal solutions in a computationally efficient manner. Bearing all this in mind, it is important to stress that Casillas et al. [39] also proposed improving the robustness of their sensor placement methodology by (1) carrying out a time horizon analysis (which, by performing an extended-horizon analysis of pressure sensitivities and residuals and then looking at the mean projection, can reduce the sensitivity to demand changes and noise in the measurements observed when using methods that consider a time instant evaluation only – see, e.g. [9, 10]), (2) using a distance-based scoring during the optimisation process (which, by accounting for the topological distance between the “actual” leak/burst node and the node indicated by the projection-based method, attempts to retain more information than the traditional binary scoring process would in the case of leaks/bursts incorrectly localised – as all the incorrectly localised leaks/bursts are treated in the same way), (3) incorporating sets of sensitivities and residuals in their evaluation function that are computed considering different leak/burst sizes and (4) adding noise to the model pressures before computing the residuals to simulate measurements noise. Through comparison of the results obtained on the Limassol network with and without considering the proposed improvements, the authors found that leak magnitude changes were impacting the resulting optimal sensor placement found in the case of no improvements, requiring a post-treatment analysis to tackle such a problem, whereas considering the improvements enabled them to avoid any post-treatment analysis.

Sarrate et al. [42] proposed a sensor placement method based on an extension of the work done, although not focusing on WDSs specifically, in Rosich et al. [43]. This method takes into account maximum diagnosability (i.e. leak/burst isolability and detectability maximisation) specifications for a specific number of sensors to be installed. The strategy is based on the structural model of a WDS. A structural model is a coarse model description, based on a graph representation of the analytical model structure whereby only the relationship between variables and equations is taken into account, while the mathematical expression of this relationship is neglected. Because of this an efficient graph-based method (i.e. depth-first branch and bound search algorithm) was applied to solve the sensor placement problem. Bearing this in mind, it is important to stress that, due to the coarse nature of a structural model, the diagnosis performance obtained using such a model cannot be guaranteed for the real WDS. Considering only a small subset of nodes as potential sensor locations, the authors applied their method to a DMA in the Barcelona WDS and demonstrated the feasibility of their approach. However, in a later study [44], the author stated that, because of the size and the complexity of the optimal sensor placement problem in real-life WDSs, the applicability of the method proposed in Sarrate et al. [42] is limited to small-/medium-sized networks. Therefore, they attempted to reduce the size and complexity of the problem by combining their structural model-based method with clustering techniques. Clustering techniques enable the unsupervised classification of patterns (observations, data items or feature vectors) into groups (clusters) and have been used to solve various problems in different domains [45]. Specifically, a k-means clustering technique (see – e.g. [46]) was used in that study as a pre-processing step to reduce the number of candidate sensor locations before solving the sensor placement problem proposed by Sarrate et al. [42]. Aiming at grouping together nodes that respond in a similar manner to leak/burst events, the authors built a fault sensitivity matrix as done in Pérez et al. [31]. However, they did not binarise that matrix but used the cosine distance on the residuals for the k-means algorithm. As a result, the number of candidate sensor locations to be used in the depth-first branch and bound search algorithm was reduced by selecting only one candidate sensor location (i.e. the nearest to the cluster centroid) from each cluster. In this study, the authors tested their method on the same DMA used in Sarrate et al. [42], simulated leaks/bursts as a single, constant demand that can appear at selected (in order to limit problem complexity) nodes, assumed the availability of a perfect hydraulic model and did not account for measurements noise. Of particular note in this study is the fact that the authors stated that although it might seem appealing (in order to reduce computational efforts) to skip the branch and bound step and directly apply the clustering step to obtain the final sensor configuration, such an approach may lead to suboptimal results as “only a reduced set of directional residuals (the primary residuals) are represented in the fault sensitivity matrix according to the simulation method used”. That statement was then validated in a following study by the authors

[47] where, however, it was also noted that results from the direct application of the clustering step were not too far from the global optimum.

Bearing in mind the above, in order to overcome the aforementioned intrinsic limitation of their structural model-based strategy, Sarrate et al. [47] proposed a further approach based entirely on analysis of the leak/burst sensitivity matrix. In this study, projections are calculated from the sensitivity matrix, a “leak locatability” index (to be maximised for the specific number of sensors to be installed) is introduced, and a two-step hybrid methodology that combines clustering techniques (the evidential c-means algorithm was used in that study – see [48]) with an exhaustive search is utilised to search for the optimal sensor configuration. Again, the authors tested their method on the same DMA used in Sarrate et al. [42] and conducted their experiments with settings very similar to those used in that study. It was found that their further approach enables solving the optimal sensor placement problem in a reasonable time. However, the authors noted that despite the exhaustive search approach providing an optimal result, “optimality” of this result over the set of original candidate pressure locations is strongly dependent on the performance of the clustering algorithm.

Wu and Song [49] developed a pressure sensor placement method that maximises the number of leak/burst events that can be detected for a given number of sensors by performing the following two steps. Firstly, a Monte Carlo method (see – e.g. [50]) is used to generate a large number of random events with different magnitudes and that may occur at a single location or at two locations simultaneously. In this step, the simulated nodal pressures are compared with the baseline condition, and residuals are stored in a matrix. Then a binary matrix is obtained from the residuals matrix using the sensors′ accuracy (from manufacturer′s specifications) as the threshold. That is to say, an event is considered to be detected as long as a pressure change is greater than the pressure sensor accuracy. In the second step, the pressure sensor locations are optimised using a GA in the Darwin optimization framework [51] for a given number of sensors, so that the optimised sensor locations are able to cover the maximum number of leak/burst events. The authors tested their method on two real-life networks considering a perfect model.

Hagos et al. [52] presented a method that attempts to mitigate the “arbitrary threshold” selection issue present in many of the previous studies that convert the sensitivity matrix to a binary matrix by promoting the use of statistical process control tools. Specifically, the use of Shewhart control charts [53] and of the Western Electric Company detection rules [54] was proposed in that study as the authors deemed this detection approach more statistically robust, in addition to enabling up to eight most recent past measurements rather than a single/current value/measurement. The method focuses on the placement of pressure and flow sensors independently, makes use of linear programming for the optimisation (binary integer programming problem solved by the general reduced gradient non-linear solver – see [55]) and was demonstrated on a modified Austin network (see [56]). In this study, the authors looked into the issue of false alarms in determining sensor placements′ detection effectiveness and made use of the average detection time as a secondary (given placements with the same detection effectiveness, the placement with a shorter average detection time is more favourable) detection efficiency indicator. With specific regard to the issue of false alarms, the authors found that maximising the rate of correct event detections and minimising the rate of false alarms are contradictory goals and the best detection locations are not likely to be the best locations for minimising the rate of false alarms. Furthermore, they found that, as the number of sensors in the DMA increases, both the rate of correct event detections and the rate of false alarms increase.

Huang et al. [57] developed a clustering-based pressure sensor placement method for pipe burst detection. They used a fuzzy self-organising map neural network (see – e.g. [58]) due to its capability to classify the inputs without knowing the number of clusters in advance and, hence, with the capability of enabling to determine the optimal number of pressure sensors required. In this method, the nodes in a WDS/DMA are grouped according to their similarity in responding to the change of node demands due to leaks/bursts. A small real-world network was used to demonstrate the effectiveness of the method. Here, the authors simulated leaks/bursts as a single, constant demand, considered a single time step (i.e. the hour of maximum daily water consumption), assumed the availability of a perfect hydraulic model and did not account for measurements noise. Because of the limited verification of the methodology carried out in this study, it is difficult to assess the value of the proposed method. Bearing this in mind, it is also important to stress that the authors stated that setting the parameters of the self-organising map neural network properly is not a trivial task and further investigations into this issue are required if this method is to be used by water companies.

Candelieri et al. [59] proposed a method that makes use of (1) a graph-based, spectral clustering procedure (see – e.g. [60]) of similar variations in pressure and flow induced by leaks/bursts simulated using a hydraulic model and (2) support vector machines classification (see – e.g. [61]) to learn the relationship between the variations in pressure and flow at the deployed sensor locations and the most probable set of pipes affected by a leak/burst (i.e. to learn to approximate the non-linear mapping performed by the spectral clustering procedure and estimate the most probable cluster which an "actual" vector of variations in pressure and flow would belong to). They run several leak/burst scenarios by varying leak/burst location and magnitude, assuming the availability of a perfect model and perfect sensors′ measurements. They proposed to use a "localisation index" measure [62] and a novel "quality of localisation" measure to evaluate the quality of the identified clusters. The authors looked at the simultaneous deployment of pressure and flow sensors by introducing a simple measure of cost (i.e. the cost of a flow sensor is ten times the cost of a pressure sensor). They demonstrated the capabilities of their method by applying it to the study of the optimal sensor locations for a real-life DMA in Timisoara, Romania.

Boatwright et al. [63] proposed a novel combined sensor placement – leak/burst localisation methodology based upon a spatially constrained version of the inverse distance weighted geospatial interpolation technique (see [64]) that aims at ensuring that optimal sensor locations (with respect to the leak/burst localisation technique used) are selected. The proposed methodology makes use of the

GALAXY multi-objective evolutionary algorithm [65] to identify the optimal location of pressure sensors in the DMA given a specified number of sensors. Similarly to the work presented in Farley et al. [30], the first step for solving the optimal sensor placement problem involves hydraulic modelling of leaks/bursts at all nodes and building a matrix containing instantaneous chi-squared values (as only a single time step was considered in this study). These chi-squared values are then used for building various interpolation surfaces during the optimisation step, which aims at maximising (using an objective function also based on the spatially constrained inverse distance weighted interpolation technique and a threshold that defines the leak/burst search area on an interpolation surface) the localisation performance of each configuration of sensors for every leak/burst being modelled. After determining the optimal sensors configuration by looking at the results of the optimisation step, the spatially constrained inverse distance weighted interpolation technique is used again to calculate the approximate location of an "actual" leak/burst occurring in a DMA (once a leak/burst has been identified or is suspected) based on the "actual" pressures measured at the sensor locations. The authors considered a perfect model and perfect sensors′ measurements and tested their method on a small synthetic network from the literature, the Bakryan benchmark WDS (see [66]). Despite the limited testing/validation of this method, it is worth highlighting one of the potential benefits of the approach proposed in this work, namely, the use of spatially constrained geostatistical techniques. Generally speaking, geostatistical techniques have the potential to limit the number of instruments which are deployed in a DMA as they can estimate the values of parameters at locations which are not measured based on the measurements from nearby sensors and, hence, to enable higher leak/burst localisation performance to be achieved for a given number of sensors. The use of geostatistical techniques for leak/burst localisation was already proposed by Romano et al. [15] with encouraging results. However, the use of a spatially constrained version of the inverse distance weighted interpolation technique proposed in Boatwright et al. [63] enables the overcoming of the obvious limitation of using the Euclidean distance instead of the pipe length between the estimation locations and the instrument locations (i.e. not accounting for the actual network layout of a DMA).

3 Considerations on Specific Issues Encountered When Developing Optimal Sensor Placement Techniques for Leak/Burst Detection and Localisation

In this section several issues that have been considered by researchers when developing sensor placement methodologies for leak/burst detection and localisation are presented together with details of relevant research works that have aimed at addressing these issues.

This section is organised as follows. Firstly, the issue of model uncertainties and sensitivity to the leak/burst size assumed for hydraulic simulations is considered in Sect. 3.1. Then, Sect. 3.2 focuses on the issue of uncertainties in the sensors′ measurements. Once this is done, Sect. 3.3 deals with the issues of sensor/communication failures in sensor networks. Section 3.4 examines the topic of the simultaneous use of pressure and flow sensors. Finally, Sect. 3.5 focuses on the issue of accounting for risk when developing optimal sensor placement techniques for leak/burst detection and localisation.

3.1 Model Uncertainties and Sensitivity to the Leak/Burst Size Assumed for Hydraulic Simulations

Almost all the sensor placement algorithms for leak/burst detection and localisation in a DMA rely on modelling a large number of leak/burst scenarios. A number of reliable, readily available hydraulic solver packages exist (e.g. EPANET (see [67]); PICCOLO (see [68]); AQUIS (see [69]); WaterGEMS (see [70]); OOPNET (see [71]); etc.), which allow leaks/bursts to be modelled relatively easily. Many of the sensor placement studies for leak/burst detection and localisation found in the literature assume a perfect model (i.e. that reflects reality at all times). Such a perfect model is assumed to contain up-to-date estimates of nodal demands, background (i.e. not burst type) leaks, pipe friction factors, statuses and characteristics of valves, pumps and other devices and any other model parameter/input values (e.g. heads in service reservoirs) that may affect its predictions of network pressures and flows. However, it is well known that a perfect model does not exist. In this context, demand allocation in a hydraulic model, which requires a good characterisation of consumers, is considered as one of the most critical issues. In addition to all this, leaks and bursts in WDSs have a stochastic nature. The size, location, timing and nature/type of a leak/burst event are generally unknown. However, a nominal leak/burst size is assumed in many of the sensor placement methodologies that can be found in the literature. This section presents a selection of studies that have attempted to deal with these issues.

Blesa et al. [72] studied the robustness of the methodology introduced by Sarrate et al. [47] against sensitivity matrix uncertainties by taking into account different leak/burst magnitudes on the one hand and several operating points (although only inflows variations were considered in this study) on the other hand. The authors introduced a “robustness percentage” index, which is based on the “leak locatability” index (see [47]), to assess the robustness of the selected sensor placement methodology. Additionally, they made use of an extended sensitivity matrix that considers all possible leak/burst scenarios and operating point scenarios in their clustering analysis to reduce the number of candidate sensor locations. The authors illustrated their robustness studies by means of a simple synthetic network (note, however, that the clustering analysis was not deemed necessary there) and the same DMA in the

Barcelona WDS used in previous studies by the authors (i.e. [42, 44, 47]). The main result was that the identified sensor positions are relatively insensitive to the size of the leaks/bursts. However, variation of the "leak locatability" index can be significant when different operating point scenarios are considered. Bearing this in mind, aiming at accounting for this variation and ensure robust performance, Blesa et al. [73] extended the optimal sensor placement method by Sarrate et al. [47] by formulating a multi-objective optimisation strategy to place sensors. This strategy has the following objectives: (1) to maximise the mean "leak locatability" index and (2) to maximise the worst "leak locatability" index. Optimisation was carried out by using their two-step hybrid methodology that combines clustering techniques (note that the extended sensitivity matrix considered there encompass all possible operating point scenarios only) with an exhaustive search procedure, resulting in an approximation of the entire Pareto front. The authors utilised again the simple synthetic network (without clustering analysis) and the DMA in the Barcelona WDS used in Blesa et al. [72] to test their strategy. Through comparison of the results achieved in the Barcelona DMA with and without the use of the proposed robust sensor placement methodology, the authors demonstrated that not to account for different operating point scenarios leads to solutions that are not Pareto optimal.

As mentioned in Sect. 2, Casillas et al. [39] attempted to mitigate the effect of model uncertainties and of the unknown leak/burst size by incorporating in their method an extended-horizon analysis of pressure sensitivities/residuals and by considering sets of sensitivities and residuals computed using different leak/burst sizes. However, in Casillas et al. [74, 75], the authors proposed a different sensor placement method inspired by the leak signature space-based leak/burst localisation technique presented in Casillas et al. [76]. The leak signature space analysis enables a specific signature to be associated to each leak/burst location that is minimally affected by the leak/burst size. It considers a linear model approximation of the relationship between pressure residuals and leaks/bursts to perform a transformation that allows representing leak/burst locations by means of points in the leak signature space that are not dependent on the leaks/bursts magnitude. The authors introduced the concept of a domain of influence for a particular leak signature and solved the sensor placement optimisation problem by attempting to minimise the overlapping between domains of influence considering the signatures of all network nodes. A time horizon analysis was also considered by looking at the mean number of overlaps along the time horizon analysed. A GA and a particle swarm optimisation (i.e. PSO – see [77]) algorithm were separately used to perform the optimisation. The capabilities of the proposed methodology (i.e. efficiency, in terms of the percentage of leaks correctly localised) were evaluated on the same two networks, Hanoi and Limassol, considered in a previous study by the authors (i.e. [39]), assuming perfect models but accounting for measurements uncertainties by adding Gaussian white noise. The result obtained demonstrated that efficiencies of 100% and up to about 85% could be achieved using a small number of sensors (up to 4 and 3, respectively) on the Hanoi and Limassol networks, respectively. They also emphasised, similarly to what is found in Casillas et al. [39], the benefits of the time horizon analysis. Worth of note, here, are also the results from the GA/PSO comparison that the authors detailed in these studies. They found that, generally speaking, PSO works faster than the GA (being very effective for small networks or few sensors) but, when the problem complexity increases (e.g. when more sensors are considered), the GA tends to find placements with higher efficiency. In this regard, the authors observed that PSO may tend to be trapped in a local suboptimum, probably because it has memory of past successes and therefore tends to explore around those recorded configurations; whereas when it is necessary to leap from one region of the search space to a distant other region, crossover operations like those in a GA are probably more effective. Finally, the authors stressed that, although relatively small networks were used in their studies, trying to find optimal placements for larger numbers of sensors than those detailed in their papers would be prohibitive in terms of computational time required to obtain a solution.

Steffelbauer and Fuchs-Hanusch [78] extended the work by Steffelbauer et al. [79] in which the effect of demand uncertainty on modelled predictions of pressure was incorporated in the optimal sensor placement problem (solved by using the method proposed by Casillas et al. [9, 10] but adapted in order to penalise potential sensor locations with high uncertainties) by using Monte Carlo simulation to calculate pressures for multiple realisations of nodal demands. In Steffelbauer and Fuchs-Hanusch [78], the authors solved the problem for different numbers of sensors ranging from two to ten (the study by Steffelbauer et al. [79] was limited to four sensors) taking into account different strengths of uncertainties. One of the main findings was that incorporating uncertainties leads to very different optimal placements than without uncertainties. Indeed, without uncertainties the algorithm tended to place sensors in regions with high demand uncertainties spread over the whole system. With high strength uncertainty, on the other hand, the sensors tended to be clustered "too much" in regions with low demand uncertainties, thus indicating that points which are sensitive to leaks/bursts are also likely to be points which are most sensitive to demand variations and, hence, not ideal locations to place sensors at. Worth of note in this study is also the fact that the authors derived a cost-benefit function to describe the relation between the number of sensors and the leak/burst localisation quality. The main reason for this was to provide water companies with a methodology to answer the question of how many sensors are needed to identify a specific number of leak/burst scenarios correctly. They found that the simple cost-benefit function they derived follows a power law. That is to say, for a linear improvement of the localisation quality, the number of sensors has to double. Furthermore they observed that the power law behaviour still applies even if demand uncertainties are accounted for. The only difference to simulations without uncertainties is that the localisation quality for a placement with a particular number of sensors decreases as the strength of the uncertainty increases.

A further interesting investigation into the issue of demand uncertainty can be found in Puleo et al. [80]. In this study, the authors proposed an "identifiability analysis" [81] method that makes use of the Fisher information matrix to select points that are sensitive to leaks/bursts and also provide less correlated measurements under uncertain demands. They performed Monte Carlo simulations whereby demand was randomly drawn from a normal distribution and, through limited tests on a small synthetic network (i.e. Apulian – see [82]), found that their method was not affected by the demand uncertainties.

3.2 Measurement Uncertainties

Pressure and flow sensors are the primary devices to monitor WDSs, and data coming from these devices can potentially enable timely and reliable leak/burst event detection and localisation. However, these sensors are subject to measurement errors associated with any measuring device. Differences between measured and expected data are at the core of many optimal sensor placement techniques. As uncertainties in using measured values due to the possible range of errors for these devices exist, the difference between measured and expected data must exceed the measurement error to be considered an “anomaly”. In this context, “anomalies” caused by small leaks/bursts may be difficult to identify. Furthermore, as leaks/bursts may occur at any location in a network, a leak/burst occurring farther away from a sensor may result in small variations in the signal recorded by that sensor. In view of all this, it is important to investigate if sensor placements obtained using assumptions of perfect sensors′ measurements would perform suboptimally when implemented in real-life WDSs.

A small number of studies found in the literature have considered the sensors′ accuracy as a key component of their methodology. For example, Wu and Song [49] and Forconi et al. [83] used the sensors′ accuracy as a threshold to discriminate between detections and non-detections. A few other studies have attempted to build robustness to measurements uncertainties by accounting for noisy measurements in their optimal sensor placement frameworks (e.g. [39, 75]). The work by Raei et al. [84], on the other hand, attempted to investigate this issue as its primary aim. The authors proposed to solve the sensor placement problem by using a multi-objective optimisation framework. They explored the effect of measurements uncertainty on the selection of sensor locations by identifying alternative non-dominated fronts for different values of sensor accuracy and then selected the final sensor placement from those non-dominated fronts. The sensor placement problem formulation presented in this study is based on sensitivity to leaks/bursts that are simulated at all potential nodes in a network (note that the absolute error is used in this study) and solved using the Non-dominated Sorting Genetic Algorithm-II (NSGA-II – see [85, 86]) to explore trade-offs between the minimisation of the number of sensors to be deployed and the detection time (i.e. from leak/burst start time to the time at which one sensor out of the set of sensors registers a pressure difference that is larger than an error threshold) objectives. The authors tested their approach on the C-town (see [87]) synthetic network considering a perfect model. The result obtained showed that the detection times increase as the sensor accuracy decreases. However, the sensor uncertainties did not seem to greatly affect the placement of the sensors. Worth of note in this study is also the fact that the authors proposed to simulate leaks/bursts that start at different times during the day by discretising the demand patterns into four clusters. Bearing in mind that the authors stressed that this discretisation was only necessary to mitigate the computational burden that would be faced if leaks/bursts were allowed to start at every time step of a hydraulic simulation, their attempt to account for more realistic leak/burst modelling assumptions is valuable and highlights a further source of uncertainty that has been somehow neglected by optimal sensor placement studies.

3.3 Sensor and Communication Failures

Sensor networks are exposed to failure conditions, such as sensor malfunctions and communication system failures. In the current hyperconnected world, for example, cyberattacks are now a major risk for sensor and communication malfunctions/failures [88]. Therefore, a sensor network′s robustness should be considered for the reliable provision of informative sensor data. A sensor network′s robustness should be considered at the design stage because the overall information gain and, hence, the effectiveness of the sensor network should be assessed as a whole. Despite the information gain varies for different locations, information gains from data collected at some locations can compensate for those at other locations [89].

The vast majority of optimal sensor placement methods that can be found in the literature have assumed that all sensors perform without any failure. However, this assumption is not realistic and may result in the design of a sensor network that performs poorly when the network is partially impaired (e.g. a sensor fails).

In the above context, Jung and Kim [89] proposed a leak/burst detection approach similar to that proposed by Hagos et al. [52] but that builds on that work by (1) using the NSGA-II for multi-objective optimal sensor placement and, most significantly, by (2) introducing a further criterion in that optimisation, namely, the maximisation of the robustness of a sensor network given a predefined number of sensors. The authors defined the sensor network′s robustness as its ability to consistently provide quality data in the event of sensor failure. Individual sensor failures were considered in that study, and a coefficient of variation of the rate of correct detections in the event of a sensor failure was used to assess the variation in performances of the subsets of a given set of sensors. The authors tested this method on the same synthetic network used in Hagos et al. [52], performed experiments with very similar settings and considered a similar number of pressure and flow sensors to be independently deployed. By accounting for robustness of a sensor network, quite different sensor placements were proposed, thereby confirming that a sensor network′s robustness should be considered at the sensor network′s design stage.

3.4 Use of Flow Sensors

While pressure sensors are still cheaper than flow sensors, the price difference has considerably lowered over the past years. For example, collecting flow data is now possible via insertion sensors using through bore hydrants (which, however, typically have lower accuracy than the full bore electromagnetic flow sensors normally used at a DMA inlet). Through the use of insertion sensors, the costs for excavation, pipe cut-out, installation of valves, backfilling and pavement work and the potential need to temporary decommission parts of the WDS can be avoided. Bearing this in mind, flow and pressure sensors in WDS work differently. Flow measurements are sensitive to all downstream changes, while pressure measurements are sensitive to additional head loss on the flow route to them – thus generally more sensitive to events local to the instrument, both up and down stream of the instrument. Flow data is also generally via pulse counting systems providing an average value over a time period (e.g. 15 min) generating smoothed data with good confidence, while pressure data is generally an instantaneous value including noise and variability [14, 90]. Therefore, using additional flow instrumentation should hypothetically improve the performance of optimal sensor placement methods that only use additional pressure sensors. However, in the literature there has been less analysis of the simultaneous optimisation of the locations of both pressure and flow sensors for leak/burst event detection and localisation.

In the above context, worth of mention is the work by Imschoot et al. [91]. The authors utilised an approach very similar to that presented in Farley et al. [30, 33] for event detection and in Farley et al. [11, 34] for achieving selective sensitivity. However, they incorporated data from not only pressure but also flow sensors to detect and localise leaks/bursts. The authors populated two sensitivity matrices, one for flow and one for pressure, used an absolute error rather than a chi-squared formulation (as the latter is not applicable to simulated flow measurements that could be null or negative) and considered a more conservative (than using an uncertainty band) safety factor that simply shifts the threshold (the mean of the values in each sensitivity matrix) used to binarise the matrices to a higher limit. They then performed a complete enumeration search of these matrices using a fitness function that aims at finding optimal solutions for the placement of one or two additional sensors that results in similarly sized subdivided areas. The authors tested their method on two UK DMAs assuming a perfect model and no measurements uncertainty and found that (as a general tendency) placing optimal flow sensors plus the inlet flow sensor seems to provide better results than the flow sensor at the DMA inlet with optimally placed pressure sensors.

Findings similar to those reported by Imschoot et al. [91] have also been recently presented in Raei et al. [92] whereby the authors observed that, despite the use of pressure sensors having clear benefits in improving leak/burst detection rates, the impact of pressure sensors in improving those rates diminishes quickly as the number of flow sensors increases. Overall, these initial findings seem to suggest that further development of sensor placement methods that attempt to simultaneously use pressure and flow sensors for leak/burst detection and, most importantly, localisation is needed. However, despite what stated at the beginning of this section with regard to costs, these methods could/should be further developed to also account for the differences in costs and budget constraints. Bearing this in mind, accounting for differences in costs has been attempted in studies such as those presented in Candelieri et al. [59] and Jung and Kim [93], for example, but a more thorough analysis framework for dealing with these issues would be beneficial to water companies.

3.5 Accounting for Risk

A potential drawback of all the optimal sensor placement approaches reviewed so far is that they tend to treat all leaks/bursts in the network equally – i.e. without considering the potential impact they may have on customers, for example. In real-life circumstances, a water company may decide to favour sensor placements that ensure quick detection and localisation of events that may have a major impact on nearby customers (e.g. cause local road or property damage) and especially if the customers in question are sensitive/critical (e.g. hospitals).

In Forconi et al. [83], three different risk-based functions were used to derive optimal placements of a given number of sensors in a WDS: a simple function based on likelihood of leak/burst non-detection and two other risk-based functions, where impact and exposure/vulnerability are combined with the leak/burst detection likelihood. The impact is measured by the effects of a leak/burst occurrence on the demands (i.e. volume of undelivered water), while the exposure/vulnerability is measured by the intrinsic importance of the elements that can be damaged (by assigning higher weights to certain nodes). This method therefore enables to take into account social, economic and/or safety considerations. The results obtained showed that accounting for risk can lead to significantly different sensor placements. In this context, the methodology proposed in this study can represent a useful tool for the WDS′s managers for placing sensors in the network in order to not only detect and localise leaks/bursts but to also comply with hydraulic, social and economic requirements.

Venkateswaran et al. [94] presented a good example of work that focus on refining the means of estimating the likelihood and impact components of risk (i.e. one of the most important issues in risk-based approaches). In that study, the authors proposed an approach to model and quantify the real-world impact of a leak/burst event on a community using various geospatial, infrastructural and societal factors. Specifically, they modelled the vulnerability of a community to flooding by simulating the propagation of water from a leak/burst along the surrounding terrain using a hydrodynamic flood simulation algorithm. They also partitioned the community into regions (driven by flood maps, which depend on the terrain) and determined the relative criticality of these regions by assigning scores based on the population density as well as the critical infrastructure (e.g. healthcare, transportation, government facilities, educational, etc.) present within each region. The sensor placement algorithms they developed, however, are greedy type (i.e. algorithms that solve the problem by placing one sensor and find the next sensor position through incorporation of the previous one), which have been shown to be likely to fail in finding optimal sensor placements [26, 78].

4 Discussion

Based on the literature review carried out in the previous two sections, it is possible to state that the various optimal sensor placement techniques that have been proposed by researchers have many differences but also similarities. Some studies have focused on leak/burst detection only, while others have considered both leak/burst detection and localisation. The optimal sensor placement problem has been formulated in a number of different ways, and the proposed solutions to the problem have involved the use of different tools such as different hydraulic solvers and different optimisation algorithms. Even when the same hydraulic solver is used, different modelling approaches have been taken by researchers such as accounting for the pressure-driven behaviour of a network or not, performing extended period or single period simulations and simulating the occurrence of leaks/bursts by using additional demands at nodes or emitters at nodes (or on pipes). Furthermore, some studies have attempted to deal with one or more sources of uncertainty such as demand and measurements uncertainty, while others have assumed the availability of a perfect model and measurements, among other things. All the proposed techniques have been tested and demonstrated on one or more case study networks. The characteristics of such case studies vary widely from small synthetic benchmark networks to real-life DMAs in various parts of the world. Tests and demonstrations of the proposed techniques have often involved numerical experiments only, but in some cases field tests have also been carried out. Bearing this in mind, Tables 1, 2, 3 and 4 summarise the main characteristics of a number of selected publications that have been reviewed in Sects. 2 and 3.

By scrutinising these tables and in the light of the literature review carried out in Sects. 2 and 3, a number of considerations regarding, inter alia, the state of the art of optimal sensor placement techniques, the potential of these techniques to help water companies minimising the leaks/bursts′ runtime by effectively detecting and localising these events as they occur in a DMA and the gaps in the current research can be made. These considerations are detailed below.

Notwithstanding the individual contributions to the body of knowledge in the field made by studies that have focused on optimal placement of pressure sensors for leak/burst detection only, it is possible to observe that such studies have limited value for water companies when considering the aim of minimising the leaks/bursts′ runtime. Indeed, studies such as Hagos et al. [52] noted that the majority of optimally located pressure sensors in a network tend to detect the same set of leaks/bursts and, thus, they provide little information on where a leak/burst may be located.

Table 1 Main characteristics of selected publications – primary objective (i.e. leak/burst detection or leak/burst detection and localisation) and problem formulation/solution

	Detection/ localisation	Problem formulation/solution
Farley et al. [30]	Detection	Threshold applied to the chi-squared matrix + complete enumeration (only two additional pressure sensors considered)
Farley et al. [11]	Detection and localisation	Threshold and uncertainty band applied to the chi-squared matrix + optimisation using a GA
Pérez et al. [31]	Detection and localisation	Threshold applied to the leak/burst sensitivity matrix + optimisation using a GA
Casillas et al. [39]	Detection and localisation	Angle method to analyse the leak/burst sensitivity matrix + optimisation using a semi-exhaustive search/ GA
Sarrate et al. [42]	Detection and localisation	Structural model of a DMA + depth-first branch and bound search algorithm
Sarrate et al. [44]	Detection and localisation	Structural model of a DMA + k-means clustering + depth-first branch and bound search algorithm
Wu and Song [49]	Detection	Threshold based on the accuracy of sensors + optimisation using a GA in the Darwin optimization framework
Hagos et al. [52]	Detection	Statistical process control + linear programming for the optimisation
Huang et al. [57]	Detection	Influence coefficient matrix + fuzzy self-organising map neural network clustering
Candelieri et al. [59]	Detection and localisation	Spectral clustering + support vector machines classification
Boatwright et al. [63]	Detection and localisation	Spatially constrained version of the inverse distance weighted geospatial interpolation technique + optimisation using the GALAXY multi-objective evolutionary algorithm
Blesa et al. [73]	Detection and localisation	Projections calculated from the extended sensitivity matrix + two-step hybrid methodology combining evidential c-means clustering algorithm and an exhaustive search
Casillas et al. [75]	Detection and localisation	Leak signature space method + optimisation using a GA/PSO
Steffelbauer and Fuchs-Hanusch [78]	Detection and localisation	Angle method between the general sensitivity of potential measurement points with respect to all possible leak scenarios and residual vectors + optimisation using a GA
Puleo et al. [80]	Detection	Identifiability analysis
Raei et al. [84]	Detection	Threshold based on the accuracy of sensors and ranking applied to the absolute error matrix + optimisation using the Non-dominated Sorting Genetic Algorithm-II
Forconi et al. [83]	Detection	Threshold based on the accuracy of sensors + ranking using the Max-Sum method [95]

Table 2 Main characteristics of selected publications – hydraulic solver used, demand-/pressure-driven formulation, use of single/extended period simulations and choice of leak/burst simulation method

	Hydraulic solver used	Demand-/ pressure-driven	Single/ extended period	Leak/burst simulation method
Farley et al. [30]	AQUIS	Not mentioned[a]	Extended period	Emitter at nodes
Farley et al. [11]	AQUIS	Not mentioned[a]	Extended period	Emitter at nodes
Pérez et al. [31]	PICCOLO	Demand-driven	Single period	Additional demand at nodes
Casillas et al. [39]	EPANET	Not mentioned[a]	Extended period	Emitter at nodes
Sarrate et al. [42]	N/A	N/A	N/A	Not mentioned[b]
Sarrate et al. [44]	EPANET	Not mentioned[a]	Single period	Not mentioned[b]
Wu and Song [49]	WaterGEMS	Not mentioned[a]	Not mentioned[c]	Emitter at nodes
Hagos et al. [52]	EPANET	Not mentioned[a]	Extended period	Emitter at nodes
Huang et al. [57]	Not mentioned	Not mentioned[a]	Single period	Additional demand at nodes
Candelieri et al. [59]	EPANET	Pressure-driven [96]	Not mentioned[c]	Emitter on pipes [96]
Boatwright et al. [63]	EPANET	Demand-driven	Single period	Emitter at nodes
Blesa et al. [73]	EPANET	Not mentioned[a]	Not mentioned[c]	Emitter at nodes
Casillas et al. [75]	EPANET	Not mentioned[a]	Extended period	Emitter at nodes
Steffelbauer and Fuchs-Hanusch [78]	OOPNET	Not mentioned[a]	Not mentioned[c]	Emitter at nodes
Puleo et al. [80]	EPANET	Not mentioned[a]	Not mentioned[c]	Emitter in the middle of pipes
Raei et al. [84]	EPANET	Demand-driven	Extended period	Emitter at nodes
Forconi et al. [83]	EPANET	Pressure-driven	Extended period	Emitter in the middle of pipes

[a] Assumed demand-driven
[b] Assumed emitter at nodes
[c] Assumed single period

Furthermore, Farley et al. [11] noted that if pressure sensors were to be used solely to detect leak/burst events, the issue of false alarms would be of concern. In that study, the authors observed that false alarms are more common when using pressure time-series values for detection, as pressure fluctuates much more than the flow in the system. Similar/related observations have been made by a number of other researchers. For example, as already mentioned in Sects. 2 and 3, Hagos et al. [52] observed that the best pressure sensor locations for detection are not likely to be the

Table 3 Main characteristics of selected publications – accounting for model uncertainty, measurements uncertainty and choice of using a single leak/burst size or multiple leak/burst sizes in the proposed optimal sensor placement frameworks

	Model uncertainty	Measurements uncertainty	Leak/burst magnitude – single/multiple
Farley et al. [30]	No	No	Multiple – but one at a time
Farley et al. [11]	No	No	Single
Pérez et al. [31]	No	No	Single
Casillas et al. [39]	No	Yes – incorporated in the placement method	Multiple – incorporated in the placement method
Sarrate et al. [42]	No	No	Single
Sarrate et al. [44]	No	No	Single
Wu and Song [49]	No	No	Multiple – performed Monte Carlo simulations
Hagos et al. [52]	Yes – demand uncertainty; introduced as random noise	No	Multiple – considered the emitter discharge coefficient as a random variable
Huang et al. [57]	No	No	Single
Candelieri et al. [59]	No	No	Multiple – varying in a given range
Boatwright et al. [63]	No	No	Single
Blesa et al. [73]	Yes – inflows variations; incorporated in the extended sensitivity matrix	No	Multiple – varying in a given range
Casillas et al. [75]	No	Yes – incorporated in the placement method	Multiple – incorporated in the placement method
Steffelbauer and Fuchs-Hanusch [78]	Yes – demand uncertainty; performed Monte Carlo simulations and used 4 different strength of uncertainty	No	Single
Puleo et al. [80]	Yes – demand uncertainty; performed Monte Carlo simulations	No	Single
Raei et al. [84]	No	Yes – incorporated in the placement method	Multiple – varying in a given range
Forconi et al. [83]	No	No	Single

Table 4 Main characteristics of selected publications – details of the case study network(s) and methodology validation through field trials

	Case study network(s)	Field validation
Farley et al. [30]	2 × real-life: Dendritic UK DMA – 260 nodes Looped UK DMA – 86 nodes	No
Farley et al. [11]	14 × real-life: UK DMAs – 204–1,091 nodes; ~6.3–36 km of pipes	Yes – simulated bursts in 3 DMAs
Pérez et al. [31]	Real-life: Placa del Diamant, Barcelona WDS, Spain – 1,600 nodes; ~41 km of pipes	No
Casillas et al. [39]	Synthetic: Hanoi, Vietnam – 31 nodes; 34 pipes Real-life: Limassol, Cyprus – 197 nodes; 239 pipes	No
Sarrate et al. [42]	Real-life: DMA, Barcelona WDS, Spain – 883 nodes (31 considered as possible sensor locations); 927 pipes; ~17.4 km of pipes	No
Sarrate et al. [44]	Real-life: DMA, Barcelona WDS, Spain – 883 nodes (311 considered as possible sensor locations, 31 clusters); 927 pipes; ~17.4 km of pipes	No
Wu and Song [49]	2 × real-life: UK DMA – 1,321 pipes United Arab Emirates – 86 pipes	No
Hagos et al. [52]	Synthetic: modified Austin – 125 nodes; 90 pipes	No
Huang et al. [57]	Real-life: DMA – 77 nodes; 108 pipes	No
Candelieri et al. [59]	Real-life: Timisoara, Romania – 335 nodes; ~4 km of pipes	No
Boatwright et al. [63]	Synthetic: Bakryan – 35 nodes; 58 pipes; ~102 km of pipes	No
Blesa et al. [73]	Synthetic: small benchmark – 12 nodes; 17 pipes; ~102 km of pipes Real-life: DMA, Barcelona WDS, Spain – 883 nodes (311 considered as possible sensor locations, 25 clusters); 927 pipes; ~17.4 km of pipes	No
Casillas et al. [75]	Synthetic: Hanoi, Vietnam – 31 nodes; 34 pipes Real-life: Limassol, Cyprus – 197 nodes; 239 pipes	No
Steffelbauer and Fuchs-Hanusch [78]	Real-life: DMA – 392 nodes; 452 pipes; ~37 km of pipes	No
Puleo et al. [80]	Synthetic: Apulian – 23 nodes; 34 pipes	No
Raei et al. [84]	Synthetic: C-town – 388 nodes; 429 pipes	No
Forconi et al. [83]	Real-life: E023 DMA, Harrogate, Yorkshire, UK – 448 nodes (291 considered as possible sensor locations); 468 pipes; ~16 km of pipes	No

best locations for minimising the rate of false alarms and that, as the number of pressure sensors in the DMA increases, the rate of false alarms increases, thus exacerbating the problem; Steffelbauer and Fuchs-Hanusch [78] stated that pressure sensor locations that are sensitive to leaks/bursts are also likely to be locations that are most sensitive to demand variations and, hence, not ideal locations to place sensors at. Mounce et al. [14] also reported that flow signals are much more reliable for leak/burst event detection than pressure signals. In view of this, it is envisaged that future optimal sensor placement studies should focus on simultaneously considering the possibility of detecting and, most importantly, localising leaks/bursts. In this scenario, flow measurements (usually available at the inlet of DMAs already) can be used first to determine detection, and the pressure instruments can then be used to determine location in addition to provide further confidence in the detection alarms (see – e.g. [16]) and to provide information useful for root-cause identification (e.g. a flow increase and a simultaneous pressure decrease can indicate a leak/burst in a DMA, whereas a simultaneous flow and pressure increase can indicate a different issue such as a pressure reducing valve failure).

With regard to the problem formulation and with specific focus on the use of hydraulic simulation packages, it is possible to state that using hydraulic models to simulate a large number of leak/burst scenarios and then (somehow) analysing the differences between the simulated pressures under leak/burst conditions and the simulated pressures recorded under normal conditions are common practices among researchers and, possibly, the only way forward. Methods that have attempted to avoid using hydraulic models such as the structural model-based approach proposed by Sarrate et al. [42, 44] have intrinsic limitations (see Sect. 2) that make their use difficult for effectively solving the optimal sensor placement for leak/burst detection and localisation problem. Therefore, it is clear that numerical models are instrumental to the future of cost-effective monitoring of WDSs for leak/burst detection and localisation purposes. Unfortunately, the numerous sources of uncertainty associated with such an approach remain a key concern. Temporarily ignoring these issues here together with issues related to increasing the complexity of the problem formulation (and, hence, the computational burden), as they will be discussed in further detail below, it is envisaged that more realistic modelling practices should be taken into consideration during the development of optimal sensor placement methodologies in the future. For example, it may be beneficial to use pressure-driven modelling rather than demand-driven modelling as leaks and bursts may induce pressure-deficient conditions in a network under certain circumstances. Additionally, better leak/burst localisation performance may be achieved by more realistically simulating leaks and bursts, which may occur at any point along the pipe (and not at nodes, as commonly done) and start at any time during the day.

With specific focus on the analysis of the differences between the simulated pressures under leak/burst conditions and the simulated pressures recorded under normal conditions, it is possible to observe that the development of different approaches for performing this particular task has attracted the attention of a large number of researchers. Generally speaking, binarisation of the residuals/sensitivity matrix (e.g. [11, 30, 31, 49]) has been recognised as leading to a loss of information [38]; therefore methods that make full use of the hydraulic simulation results (e.g. [39, 63, 73, 75, 78]) should be preferred. Many of the latter methods have been developed with the aim of addressing issues related to model, measurements and leak/burst size uncertainties, and they are very valuable for future research. The main findings from these studies have shown that different operating point scenarios and demand uncertainties may significantly affect the performance of "optimal" sensor placements if these factors are not carefully accounted for in the sensor placement methodologies. On the other hand, measurements uncertainties and uncertainties related to the leak/burst size have a much lesser impact on the optimal sensor placements. Having said this, it is envisaged that the influence of other model uncertainties such as uncertain pipe friction factors and other model-reality divergences on optimal sensor locations should be explored further.

With specific focus on the methods used to solve the optimal sensor placement problem, it can be observed that the optimal sensor placement problem has been often solved through optimisation. Complete enumeration techniques (see – e.g. [30]) or semi-exhaustive search routines utilising a lazy evaluation mechanisms to reduce the computational cost (see – e.g. [39]) have clearly shown not to scale up well as the number of sensors to be deployed, the size of the studied network and the complexity of the problem formulation, among the others, increase. GA, on the other hand, has shown the potential to be efficiently used to solve carefully formulated problems in small- to medium-sized networks. They have also been shown to outperform algorithms such as PSO in terms of quality of the sensor placements obtained (see [74, 75]). However, given the extremely large solution spaces which are typically present when real-life networks and assumptions are considered, their limitations in terms of the computational time required to obtain a solution (by not just exploring a small part of the total solution space) have also been highlighted by several researchers (see – e.g. [74, 75, 78]). Notwithstanding the fact that several researchers have attempted to reduce the size of the solution spaces/complexity of the optimisation problem using clustering techniques (as discussed in further detail below), it is envisaged that experiments with other, potentially more efficient, optimisation techniques should be carried out. Having said this, another relevant issue highlighted by researchers is the need for accurate tuning of the GA algorithms′ parameters. Therefore, the use of algorithms that are able to automatically adjust their hyper-parameters such as the hybrid GALAXY multi-objective evolutionary algorithm used by Boatwright et al. [63] may be beneficial. In addition to all this, investigations into the possibility of using parallel computing in a multi-core processor framework and high performance computing should be carried out to effectively enable considering real-life networks and assumptions and obtain a solution in a reasonable (bearing in mind that identifying an optimal sensor placement is a task that, in general, needs to be only carried out at the sensor network′s design stage) time.

With regard to the use of clustering algorithms and as briefly anticipated in the previous paragraph, several researchers have experimented with such methods with the aim to reduce the size of the solution space/complexity of the optimisation problem (e.g. [44, 73]). Valuable contributions were demonstrated in this respect, and, thus, further investigations into the potential of such methods to help solve the optimal sensor placement problem in an efficient and effective way should be carried out. Bearing this in mind, clustering algorithms have also been proposed by some researchers (e.g. [57, 59]) for solving the sensor placement problem on their own (i.e. without coupling clustering algorithms with GAs/semi-exhaustive search routines/etc.). Such an approach has been criticised by Sarrate et al. [44, 47] who argue that it may lead to suboptimal results. However, it should not be possible to deprecate the use of this approach based on that critique alone. Having said this, a further concern relevant to the use of clustering algorithms on their own may be the fact that the majority of optimally located pressure sensors in a network tend to detect the same set of leaks/bursts (i.e. the observation by [52], already mentioned above). In this regard, it could be argued that clustering algorithms, if used on their own, may struggle to provide information useful for enabling efficient leak/burst localisation.

Some of the optimal sensor placement methods that can be found in the literature utilise a different method for determining the instrumentation locations and for localising leaks/bursts (e.g. [11, 34, 91]). It is clear that this approach implies that the resulting sensor placements will not be optimised for the chosen method of leak/burst localisation. Therefore, it is envisaged that tightly coupled optimal sensor placement and leak/burst localisation frameworks should be developed by researchers in the future. Leak/burst localisation and sensor placement should be considered together since the best placement depends on the method that is used to localise the potential leaks/bursts and the efficiency of the leak/burst localisation depends on the sensor placement.

Much greater attention should be paid in the future to the issue of sensor/communication failures as this is of critical importance for the effectiveness of optimal sensor placements for leak/burst detection and localisation. In Sect. 2, it was noted that the use of the method proposed by Farley et al. [11, 34] would make the task of correctly localising a leak/burst impossible if a single sensor is not working or data are not timely received. Bearing this in mind, similar considerations could be made for the majority of the reviewed optimal sensor placement methods for leak/burst detection and localisation as they have been developed under the unrealistic assumption that all the sensors perform without any failure at all times. In this context, the methodology proposed by Boatwright et al. [63] may offer a potentially appealing way to mitigate the issue under scrutiny. Indeed, geostatistical interpolation techniques are less reliant on the availability of data from all the optimally deployed sensors when performing leak/burst localisation than methods that are based on assessing the similarities between the observed residuals and the results of hydraulic simulations, for example. Based on similar arguments, it may be possible to state that the use of geostatistical interpolation techniques could also be beneficial for mitigating some of the issues that arise because of model and measurements uncertainties. Indeed, a resulting interpolation surface created by using observed pressure measurements (which provides inferred values of pressure at every point in a network) attempts to mimic the results of a hydraulic simulation but without the reliance on an accurate hydraulic model/good measurements fed into a hydraulic model.

Methods for the optimal placement of pressure and flow sensors simultaneously should also be the focus of further research and development in the future. This is because early studies (e.g. [91]) have indicated that using additional flow instrumentation can improve the leak/burst detection and localisation performance of optimal sensor placement methods that only use additional pressure sensors. Due to the higher costs associated with obtaining flow measurements, however, these methods could/should also account for the differences in costs and budget constraints and balance all this with the additional benefits that could be realised.

It is envisaged that optimal sensor placement methods should also account for risk as, while detecting and localising all leaks/bursts is important, not all leaks/ bursts are equally impactful. Nowadays, the resulting potential unplanned interruptions to the water supply and the damaging consequences of the leak/burst events are tolerated to a lesser extent, and water companies are increasingly judged by the public (and the regulatory agencies alike, where applicable) based on how well (or otherwise) they manage contingency situations. In this context, future work on this subject should focus on further refining the means of estimating the likelihood and impact components of risk.

The literature review carried out in this chapter has highlighted that different methods, the inclusion of different objectives in similar methods and even slightly changing specific settings within the same method (e.g. incorporating different strengths of uncertainty) lead to significantly different sensor placements. Despite measures for the assessment of performance being found in the literature, the majority of these measures are tailored to the particular method being proposed. All this makes the task of assessing whether a sensor placement is better that another sensor placement almost impossible. Bearing this in mind, optimal sensor placement research is pressingly in need for field trials and validation, which are the only way to understand the real value and practicality of a proposed approach. In the relevant literature, only the studies by Farley et al. [11, 33, 34] and Fuchs-Hanusch and Steffelbauer [97] report the results of field trials and validations. The main findings from the field trials carried out in the studies by Farley et al. [11, 33, 34] have been detailed in Sect. 2. These findings demonstrated the practical applicability of the methods proposed in those studies. On the other hand, in Fuchs-Hanusch and Steffelbauer [97], a comparison of several methods including the methods proposed by Pérez et al. [31], Casillas et al. [39] and Steffelbauer and Fuchs-Hanusch [78] was carried out by opening fire hydrants to simulate different leak/burst scenarios in a real network and then assessing the leak/burst localisation capabilities of the different methods by calculating the distance between the suggested leak/burst locations and the opened fire hydrants. The results from the limited tests carried out in that study showed that for different leak/burst positions, different sensor sets, mainly those with sensors close to the leak/burst position, led to the best performance. These quite disappointing findings cast a shadow on the real value of the various "optimal" sensor placement methods that have been proposed so far and demonstrated using numerical simulations only, therefore stressing even more the need for any future optimal sensor placement study to be thoroughly field validated in real-life networks. It is envisaged that, as a bare minimum, future optimal sensor placement studies should include an assessment of their underlying capabilities using a set of common quantitative metrics which may include/take inspiration by those recently proposed by Qi et al. [98]. In this context, the use by researchers in the field of a common set of benchmark models that cover a range of network layouts/sizes/etc. and a common set of leak/burst scenarios could also be beneficial.

5 Conclusions

Given that the deployment of an increased number of pressure and flow sensors in a single DMA is becoming more common and affordable, the automated analysis of multiple pressure and flow signals to provide useful information for efficient and timely leak/burst detection and localisation is of paramount interest to water companies. Pressure and flow measurements at some locations, however, can include more information regarding an event than measurements at other locations. Furthermore, only a limited number of sensors can be installed in a DMA due to budget constraints. In this context, a number of methodologies have been developed in the last decade that aim at identifying the optimal placement of a small number of pressure and flow sensors to capture the leak/burst effect no matter where in a DMA the leak/burst occurs and then effectively use this information to provide reliable detection alarms and accurately identify the approximate leak/burst event′s location. A comprehensive review of these methodologies has been carried out and presented in this chapter. After the introduction in Sect. 1, a synthesis and analysis of relevant published work have been presented in Sect. 2. Then, specific issues encountered by researchers when developing optimal placement of sensors for leak/burst detection techniques have been discussed in Sect. 3 together with different approaches proposed to solve these issues. Finally, Sect. 4 has presented considerations regarding, inter alia, the state of the art of optimal sensor placement techniques, the potential of the reviewed techniques to benefit water companies and the current research gaps. All this has enabled us drawing a number of conclusions, the most notable of which is perhaps that the optimal sensor placement for leak/burst detection and localisation problem remains unresolved despite the many efforts by researchers in the field. This fact is exacerbated by the almost complete lack of field tests and validation of the proposed techniques on real-life networks to enable assessing their true value and practicality for beneficial use by water companies and the fact that comparing the effectiveness of the different proposed approaches remains an almost impossible task. The other important conclusions that can be drawn from the literature review carried out in this chapter can be summarised as follows:

- Future optimal placement of pressure sensors studies should focus on simultaneously considering the possibility of detecting and, most importantly, localising leaks/bursts.
- Hydraulic models are instrumental to the future development of cost-effective sensor placement techniques for leak/burst detection and localisation purposes. Adoption of more realistic hydraulic modelling practices such as considering pressure-driven modelling, for example, is envisaged.
- Sensor placement methods that make full use of the hydraulic simulation results, as opposed to methods that involve binarisation of the residuals/sensitivity matrix, should be preferred. Further development of these methods to more effectively deal with the various sources of uncertainties is envisaged, especially for model-reality divergences that have not been considered by researchers so far.

- Research into the use of other (as opposed to the techniques used so far), potentially more efficient, optimisation techniques to solve the sensor placement for leak/burst detection and localisation problem should be carried out. This research should favourably look into algorithms that are able to automatically adjust their hyper-parameters and into the possibility of using parallel and high performance computing.
- Further investigations into the potential of using clustering algorithms coupled with optimisation techniques (as opposed to clustering algorithms used on their own) to reduce the size of the solution space/complexity of the sensor placement problem should be carried out.
- Tightly coupled optimal sensor placement and leak/burst localisation frameworks should be developed by researchers in the future as an optimal placement depends on the method that is used to localise the potential leaks/bursts and the efficiency of the leak/burst localisation depends on the sensor placement.
- Much greater attention should be paid by researchers in the future to the issue of sensor/communication failures as this is of critical importance for the effectiveness of optimal sensor placements for leak/burst detection and localisation. Research into the use of artificial intelligence-type and (geo)statistical techniques with the potential to mitigate this issue and issues related to the use of imperfect hydraulic models should also be carried out.
- Methods for the optimal placement of pressure and flow sensors simultaneously should also be the focus of further research and development in the future as using additional flow instrumentation can improve the leak/burst detection and localisation performance. In this context, the developed methods should carefully account for cost-benefit considerations as, because of the higher costs associated with obtaining flow measurements, including such considerations becomes even more important than it currently is.
- Future optimal sensor placement methods should also account for risk in order to be of even more value to water companies. Further research into refining the means of estimating the likelihood and impact components of risk should be carried out.
- Future optimal sensor placement studies should strive to incorporate results from field demonstrations as this is the only way to ultimately assess the actual capabilities of a proposed approach. Where this is not possible, future studies should at least include an assessment of their underlying capabilities using a set of common quantitative metrics, benchmark models and leak/burst scenarios.

Although optimal sampling design for leak/burst detection and localisation has been the focus of this review, researchers and practitioners interested in this topic should also look at macro-location of sensors in the wider context of WDSs management (i.e. look at optimal sampling design techniques developed for the numerous other optimisation agendas such as detection of contamination events).

Indeed, elements of research conducted for other sampling design purposes are conceptually applicable to frameworks aimed specifically at optimal sampling design for leak/burst detection and localisation. Most notably, the optimisation approaches and algorithms developed for (or merely implemented in) the wider sampling design literature are distinct from the objectives to which they are applied.

In addition to all of the above, it is also worth highlighting the fact that with the rise of easy-to-use and low-cost sensing devices, Internet of Things (IoT) technologies and edge analytics an increase in the density of heterogeneous sensors deployed in WDSs may be expected in the near future. In this scenario, pressure and flow devices will be part of a much wider network of sensors. Therefore, considerations regarding issues that have been the focus of research in the broader wireless sensor networks′ literature, such as reliable communications, efficient routing protocols, power management and computation/communication overhead, to mention just a few, will need to be accounted for when developing optimal sensor placement techniques for leak/burst detection and localisation in WDSs [99].

Finally, it must be noted that in this review the requirements for additional instrumentation have been looked at in the context of WDSs subdivided in DMAs. The rationale for this is that many of the optimal sampling design techniques for leak/burst detection and localisation found in the literature have been developed and tested under the “DMAs existence” assumption. This, in turn, is probably due to the fact that DMAs are seen as ground zero data-wise on the water companies′ journey towards operating smart water networks [100]. As evidenced by the fact that, over the past two decades, a number of technology vendors have aligned their business models based on the data flows captured or technologies required by a DMA approach, many water consultancies have implemented their DMA-dependent water balance methodologies around the world, and major industrial players have tailored products for managing sectorised networks [100]. Having said all this, however, the “DMA model” (led by UK water companies) only proliferates in Europe while gradually been adopted in countries such as Singapore, Chile, Brazil and Australia. The “non-DMA model”, on the other hand, represents the bulk of the world′s WDSs, and it proliferates in the USA as well as Germany and most of the developing world [100]. From the “non-DMA model” perspective, it may not be cost-effective or even practical to subdivide a WDS into DMAs. Therefore, the development of further optimal sampling design techniques for leak/burst detection and localisation should bear the global situation in mind and be pursued independently by the existence (or otherwise) of DMAs, which does not seem to be a strict requirement for this task. Furthermore, it is worth mentioning that this task may perhaps also be facilitated by the fact that, in recent years, the Virtual DMAs (V-DMAs) concept has started to come to fruition as data from a suite of technologies such as insertion, ultrasonic and acoustic flow metering and smart customer meters has started to be leveraged on a larger scale.

References

1. Colombo AF, Karney BW (2002) Energy and costs of leaky pipes: toward comprehensive picture. J Water Resour Plan Manag 128(6):441–450
2. Sadiq R, Kleiner Y, Rajani B (2006) Estimating risk of contaminant intrusion in distribution networks using Dempster-Shafer theory of evidence. Civ Eng Environ Syst 23(3):129–141
3. Colombo AF, Lee PJ, Karney BW (2009) A selective literature review of transient-based leak detection methods. J Hydro Environ Res 2(4):212–227
4. Li R, Huang H, Xin K, Tao T (2015) A review of methods for burst/leakage detection and location in water distribution systems. Water Sci Technol Water Supply 15(3):429–441
5. Mutikanga HE, Sharma SK, Vairavamoorthy K (2013) Methods and tools for managing losses in water distribution systems. J Water Resour Plan Manag 139(2):166–174
6. Puust R, Kapelan Z, Savić DA, Koppel T (2010) A review of methods for leakage management in pipe networks. Urban Water J 7(25):25–45
7. Wu Y, Liu S (2017) A review of data-driven approaches for burst detection in water distribution systems. Urban Water J 14(9):972–983
8. Boatwright S, Romano M, Mounce S, Woodward K, Boxall J (2016) Approximate location of leaks and bursts within a district metered area using statistical process control and geostatistical techniques. Proceedings of the 14th international conference on computing and control for the water industry, Amsterdam, The Netherlands
9. Casillas MV, Garza-Castanon LE, Puig V (2013) Extended-horizon analysis of pressure sensitivities for leak detection in water distribution networks: application to the Barcelona network. Proceedings of the 2013 European control conference, Zurich, Switzerland
10. Casillas MV, Garza-Castanon LE, Puig V (2013) Model-based leak detection and location in water distribution networks considering an extended-horizon analysis of pressure sensitivities. J Hydroinf 16(3):649–670
11. Farley B, Mounce SR, Boxall JB (2013) Development and field validation of a burst localisation methodology. J Water Resour Plan Manag 139(6):604–613
12. Mounce SR, Boxall JB, Machell J (2010) Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows. J Water Resour Plan Manag 136(3):309–318
13. Mounce SR, Farley B, Mounce RB, Boxall JB (2010) Field testing of optimal sensor placement and data analysis methodologies for burst detection and location in an urban water network. Proceedings of the 9th international conference on hydroinformatics, Tianjin, China
14. Mounce SR, Mounce RB, Boxall JB (2011) Novelty detection for time series data analysis in water distribution systems using support vector machines. J Hydroinf 13(4):672–686
15. Romano M, Kapelan Z, Savic DA (2013) Geostatistical techniques for approximate location of pipe burst events in water distribution systems. J Hydroinf 15(3):634–651
16. Romano M, Kapelan Z, Savic DA (2014) Automated detection of pipe bursts and other events in water distribution systems. J Water Resour Plan Manag 140(4):457–467
17. Romano M, Woodward K, Kapelan Z (2016) Statistical process control based system for approximate location of pipe bursts and leaks in water distribution systems. Proceedings of the 18th water distribution systems analysis conference, Cartagena de Indias, Colombia
18. Van Thienen P (2013) A method for quantitative discrimination in flow pattern evolution of water distribution supply areas with interpretation in terms of demand and leakage. J Hydroinf 15(1):86–102
19. Berry J, Carr RD, Hart WE, Leung VJ, Phillips CA, Watson J-P (2009) Designing contamination warning systems for municipal water networks using imperfect sensors. J Water Resour Plan Manag 135(4):253–263

20. Hart WE, Murray R (2010) Review of sensor placement strategies for contamination warning systems in drinking water distribution systems. J Water Resour Plan Manag 136(6):611–619
21. Kessler A, Ostfeld A, Sinai G (1998) Detecting accidental contaminations in municipal water networks. J Water Resour Plan Manag 124(4):192–198
22. Kumar A, Kansal ML, Arora G, Ostfeld A, Kessler A (1999) Detecting accidental contaminations in municipal water networks. J Water Resour Plan Manag 125(5):308–310
23. Ostfeld A, Salomons E (2004) Optimal layout of early warning detection stations for water distribution systems security. J Water Resour Plan Manag 130(5):377–385
24. Ostfeld A, Uber JG, Salomons E, Berry JW, Hart WE, Phillips CA, Watson J-P, Dorini G, Jonkergouw P, Kapelan Z, di Pierro F, Khu S-T, Savic DA, Eliades D, Polycarpou M, Ghimire SR, Barkdoll BD, Gueli R, Huang JJ, McBean EA, James W, Krause A, Leskovec J, Isovitsch S, Xu J, Guestrin C, VanBriesen J, Small M, Fischbeck P, Preis A, Propato M, Piller O, Trachtman GB, Wu ZY, Walski T (2008) The Battle of the Water Sensor Networks (BWSN): a design challenge for engineers and algorithms. J Water Resour Plan Manag 134(6):556–568
25. Shastri Y, Diwekar U (2006) Sensor placement in water networks: a stochastic programming approach. J Water Resour Plan Manag 132(3):192–203
26. Kapelan Z, Savic DA, Walters GA (2005) Optimal sampling design methodologies for water distribution model calibration. J Hydraul Eng 131(3):190–200
27. Wu ZY, Roshani E (2014) Sensor placement optimization for water quality model calibration. Proceedings of the world environmental and water resources congress, Portland, USA
28. Pudar RS, Liggett JA (1992) Leaks in pipe networks. J Hydraul Eng 118(7):1031–1046
29. Casillas MV, Garza-Castanon LE, Puig V (2012) Extended-horizon analysis of pressure sensitivities for leak detection in water distribution networks. Proceedings of the 8th IFAC symposium on fault detection, supervision and safety of technical processes, safeprocess, Mexico City, Mexico
30. Farley B, Boxall JB, Mounce SR (2008) Optimal locations of pressure meter for burst detection. Proceedings of the 10th water distribution systems analysis conference, Kruger National Park, South Africa
31. Pérez R, Puig V, Pascual J, Peralta A, Landeros E, Jordanas L (2009) Pressure sensor distribution for leak detection in Barcelona water distribution network. Water Sci Technol Water Supply 9(6):715–721
32. Gertler JJ (1998) Fault detection and diagnosis in engineering systems. Marcel Dekker, New York
33. Farley B, Mounce SR, Boxall JB (2010) Field testing of an optimal sensor placement methodology for event detection in an urban water distribution network. Urban Water J 7 (6):345–356
34. Farley B, Mounce SR, Boxall JB (2010) Field validation of ′optimal′ instrumentation methodology for burst/leak detection and location. Proceedings of the 12th water distribution systems analysis conference, Tucson, USA
35. Goldberg DE (1989) Genetic algorithms in search optimization and machine learning. Addison-Wesley, Boston
36. Blanke M, Kinnaert M, Lunze J, Staroswiecki M (2006) Diagnosis and fault-tolerant control2nd edn. Springer, Berlin
37. Pérez R, Puig V, Pascual J, Quevedo J, Landeros E, Peralta A (2010) Leakage isolation using pressure sensitivity analysis in water distribution networks: application to the Barcelona case study. Proceedings of the 12th IFAC symposium on large-scale systems: theory and applications, Villeneuve-d′Ascq, France
38. Quevedo J, Cugueró-Escofet MA, Pérez R, Nejjari F, Puig V, Mirats JM (2011) Leakage location in water distribution networks based on correlation measurement of pressure sensors. Proceedings of the 8th international water association symposium on system analysis and integrated assessment, San Sebastian, Spain

39. Casillas MV, Puig V, Garza-Castanon LE, Rosich A (2013) Optimal sensor placement for leak location in water distribution networks using genetic algorithms. Sensors 13 (11):14984–15005
40. Pérez R, Quevedo J, Puig V, Nejjari F, Cugueró-Escofet MA, Sanz G, Mirats JM (2011) Leakage isolation in water distribution networks: a comparative study of two methodologies on a real case study. Proceedings of the 19th mediterranean conference on control and automation, Corfu, Greece
41. Fujiwara O, Khang DB (1990) A two-phase decomposition method for optimal design of looped water distribution networks. Water Resour Res 26(4):539–549
42. Sarrate R, Nejjari F, Rosich A (2012) Sensor placement for fault diagnosis performance maximization in distribution networks. Proceedings of the 20th mediterranean conference on control and automation, Barcelona, Spain
43. Rosich A, Yassine AA, Ploix S (2010) Efficient optimal sensor placement for structural model based diagnosis. Proceedings of the 21st international workshop on principles of diagnosis, Portland, USA
44. Sarrate R, Blesa J, Nejjari F, Quevedo J (2014) Sensor placement for leak detection and location in water distribution networks. Water Sci Technol Water Supply 14(5):795–803
45. Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. ACM Comput Surv 31 (3):264–323
46. MacQueen J (1967) Some methods for classification and analysis of multivariate observations. Proceedings of the 5th Berkeley symposium on mathematical statistics and probability, volume 1: statistics, 281–297, University of California Press, Berkeley, USA
47. Sarrate R, Blesa J, Nejjari F (2014) Clustering techniques applied to sensor placement for leak detection and location in water distribution networks. Proceedings of the 22nd mediterranean conference on control and automation, Palermo, Italy
48. Masson M, Denoeux T (2008) ECM: an evidential version of the fuzzy c-means algorithm. Pattern Recogn 41(4):1384–1397
49. Wu ZY, Song Y (2012) Optimising pressure logger placement for leakage detection and model calibration. Proceedings of the 14th water distribution systems analysis conference, Adelaide, Australia
50. Pearl J (1987) Evidential reasoning using stochastic simulation of causal models. Artif Intell 32(2):245–258
51. Wu ZY, Mi T, Wang Q, Butala S (2011) Generalized framework for high performance infrastructure system optimization. Proceedings of the 11th international conference on computing and control for the water industry, Exeter, UK
52. Hagos M, Jung D, Lansey KE (2016) Optimal meter placement for pipe burst detection in water distribution systems. J Hydroinf 18(4):741–756
53. Shewhart W (1931) Economic control of quality of manufactured product. Van Nostrand Reinhold, New York
54. Western Electric Company (1958) Statistical quality control handbook2nd edn. AT&T Technologies, Indianapolis
55. Abadie J (1970) Application of the GRG method to optimal control problems. In: Abadie J (ed) Integer and nonlinear programming. North-Holland Publishing, Amsterdam, pp 191–211
56. Brion L, Mays L (1991) Methodology for optimal operation of pumping stations in water distribution systems. J Hydraul Eng 117(11):1551–1569
57. Huang H, Tao T, Xin K (2012) Optimal pressure meters placement for bursts detection based on SOM. Proceedings of the 14th water distribution systems analysis conference, Adelaide, Australia
58. Kohonen T (1990) The self-organizing map. Proc IEEE 78(9):1464–1480
59. Candelieri A, Soldi D, Archetti F (2015) Cost-effective sensors placement and leak localization – the Neptun pilot of the ICeWater project. J Water Supply Res Technol AQUA 64 (5):567–582

60. Luxburg U (2007) A tutorial on spectral clustering. Stat Comput 17(4):395–416
61. Vapnik V (1998) Statistical learning theory. Wiley, New York
62. Candelieri A, Conti D, Archetti F (2013) A graph based analysis of leak localization in urban water networks. Proceedings of the 12th international conference on computing and control for the water industry, Perugia, Italy
63. Boatwright S, Romano M, Mounce S, Woodward K, Boxall J (2018) Optimal sensor placement and leak/burst localisation in a water distribution system using spatially-constrained inverse-distance weighted interpolation. Proceedings of the 13th international conference on hydroinformatics, Palermo, Italy
64. Wing SR, Bowman MHE, Smith F, Rutger SM (2004) Analysis of biodiversity patterns and management decision making processes to support stewardship of marine resources and biodiversity in Fiordland – a case study. Report 2 of 3, Technical report. Report to the Ministry for the Environment, New Zealand
65. Wang Q, Savic DA, Kapelan Z (2017) GALAXY: a new hybrid MOEA for the optimal design of water distribution systems. Water Resour Res 53(3):1997–2015
66. Lee S, Lee S (2001) Genetic algorithms for optimal augmentation of water distribution networks. J Korea Water Resour Assoc 34(5):567–575
67. Rossman L (2000) EPANET user′s manual. U.S. Environmental Protection Agency, Cincinnati
68. PICCOLO (1999) Systems reference manual. http://www.safege.com/piccolo/. Accessed Feb 2019
69. AQUIS (2012) User guide. https://www.schneider-electric.com/en/download/document/Aquis+Set+Up+Guide/. Accessed Feb 2019
70. WaterGEMS (2019) User guide. https://communities.bentley.com/products/hydraulics___hydrology/w/hydraulics_and_hydrology__wiki/11933/learning-resource-guide-for-new-users-of-bentley-watergems-and-watercad. Accessed Feb 2019
71. Steffelbauer D, Fuchs-Hanusch D (2015) OOPNET: an object-oriented EPANET in Python. Proceedings of the 13th international conference on computing and control for the water industry, Leicester, UK
72. Blesa J, Nejjari F, Sarrate R (2014) Robustness analysis of sensor placement for leak detection and location under uncertain operating conditions. Proceedings of the 16th water distribution systems analysis conference, Bari, Italy
73. Blesa J, Nejjari F, Sarrate R (2016) Robust sensor placement for leak location: analysis and design. J Hydroinf 18(1):136–148
74. Casillas MV, Garza-Castanon LE, Puig V (2015) Sensor placement for leak location in water distribution networks using the leak signature space. Proceedings of the 9th IFAC symposium on fault detection, supervision and safety of technical processes, safeprocess, Paris, France
75. Casillas MV, Garza-Castanon LE, Puig V (2015) Optimal sensor placement for leak location in water distribution networks using evolutionary algorithms. Water 7(11):6496–6515
76. Casillas MV, Garza-Castanon LE, Puig V, Vargas-Martinez A (2014) Linear model representation for robust leak location in water distribution networks. Proceedings of the 16th Congreso Latinoamericano de Control Automatico, Cancún Quintana Roo, Mexico
77. Kennedy J, Eberhart R (1995) Particle swarm optimization. Proceedings of the IEEE international conference on neural networks, Perth, Australia
78. Steffelbauer D, Fuchs-Hanusch D (2016) Efficient sensor placement for leak localization considering uncertainties. Water Resour Manag 30(14):5517–5533
79. Steffelbauer D, Neumayer M, Günther M, Fuchs-Hanusch D (2014) Sensor placement and leakage localization considering demand uncertainties. Proceedings of the 16th water distribution systems analysis conference, Bari, Italy

80. Puleo V, Freni G, La Loggia G (2018) Pressure sensors positioning for leakages detection under uncertain demands. Proceedings of the 13th international conference on hydroinformatics, Palermo, Italy
81. Freni G, Mannina G, Viviani G (2011) Assessment of the integrated urban water quality model complexity through identifiability analysis. Water Res 45(1):37–50
82. Giustolisi O, Savic D, Kapelan Z (2008) Pressure-driven demand and leakage simulation for water distribution networks. J Hydraul Eng 134(5):626–635
83. Forconi E, Kapelan Z, Ferrante M, Mahmoud H, Capponi C (2017) Risk based sensor placement methods for burst/leak detection in water distribution systems. Water Sci Technol Water Supply 17(6):1663–1672
84. Raei E, Shafiee ME, Nikoo MR, Berglund E (2018) Placing an ensemble of pressure sensors for leak detection in water distribution networks under measurement uncertainty. J Hydroinf 21(2):223–239
85. Deb K (2000) Multi-objective optimization using evolutionary algorithm. Wiley, Chichester
86. Deb K, Pratap A, Agarwal S, Meyarivan TA (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 6(2):182–197
87. Marchi A, Salomons E, Ostfeld A, Kapelan Z, Simpson A, Zecchin A, Maier H, Wu Z, Elsayed S, Song Y, Walski T, Stokes C, Wu W, Dandy G, Alvisi S, Creaco E, Franchini M, Saldarriaga J, Páez D, Hernández D, Bohórquez J, Bent R, Coffrin C, Judi D, McPherson T, van Hentenryck P, Matos J, Monteiro A, Matias N, Yoo D, Lee H, Kim J, Iglesias-Rey P, Martínez-Solano F, Mora-Meliá D, Ribelles-Aguilar J, Guidolin M, Fu G, Reed P, Wang Q, Liu H, McClymont K, Johns M, Keedwell E, Kandiah V, Jasper M, Drake K, Shafiee E, Barandouzi M, Berglund A, Brill D, Mahinthakumar G, Ranjithan R, Zechman EM, Morley MS, Tricarico C, Tolson BA, Khedr A, Asadzadeh M (2014) The battle of the water networks II (BWN-II). J Water Resour Plan Manag 140:7
88. Taormina R, Galelli S, Tippenhauer NO, Salomons E, Ostfeld A (2017) Characterizing cyber-physical attacks on water distribution systems. J Water Resour Plan Manag 143:5
89. Jung D, Kim JH (2017) Robust meter network for water distribution pipe burst detection. Water 9(11):820
90. Mounce SR, Mounce RB, Boxall JB (2012) Identifying sampling interval for event detection in water distribution networks. J Water Resour Plan Manag 138(2):187–191
91. Imschoot DPO, Furnass WR, Mounce SR, Boxall JB (2016) Flow-pressure sensor placement optimisation for pipe burst localisation in a water distribution network. Proceedings of the 14th international conference on computing and control for the water industry, Amsterdam, The Netherlands
92. Raei E, Nikoo MR, Pourshahabi S, Sadegh M (2018) Optimal joint deployment of flow and pressure sensors for leak identification in water distribution networks. Urban Water J 15 (9):837–846
93. Jung D, Kim JH (2018) Using mechanical reliability in multiobjective optimal meter placement for pipe burst detection. J Water Resour Plan Manag 144(7):1–12
94. Venkateswaran P, Han Q, Eguchi RT, Venkatasubramanian N (2018) Impact driven sensor placement for leak detection in community water networks. Proceedings of the 9th ACM/IEEE international conference on cyber-physical systems, Porto, Portugal
95. Bush CA, Uber JG (1998) Sampling design methods for water distribution model calibration. J Water Resour Plan Manag 124(6):334–344
96. Candelieri A, Messina E (2012) Sectorization and analytical leaks localizations in the H_2O leak project: clustering-based services for supporting water distribution networks management. Environ Eng Manag J 11(5):953–962
97. Fuchs-Hanusch D, Steffelbauer D (2016) Real-world comparison of sensor placement algorithms for leakage localisation. Proceedings of the 18th water distribution systems analysis conference, Cartagena de Indias, Colombia

98. Qi Z, Zheng F, Guo D, Maier HR, Zhang T, Yu T, Shao Y (2018) Better understanding of the capacity of pressure sensor systems to detect pipe burst within water distribution networks. J Water Resour Plan Manag 144:7
99. Shahra EQ, Wu W, Romano M (2019) Considerations on the deployment of heterogeneous IoT devices for smart water networks. Proceedings of the 16th IEEE international conference on ubiquitous intelligence and computing, Leicester, UK
100. Bluefield Research (2017) Europe smart water: market forecasts and utility strategies, 2017–2025, focus report. https://www.bluefieldresearch.com/research/europe-smart-water/. Accessed Feb 2019

A Bird′s-Eye View of Data Validation in the Drinking Water Industry of the Netherlands

Mario Castro-Gama, Claudia Agudelo-Vera, and Dimitrios Bouziotas

Contents

M. Castro-Gama (*), C. Agudelo-Vera, and D. Bouziotas
KWR Water Research Institute, Nieuwegein, Netherlands
e-mail: Mario.Castro-Gama@kwrwater.nl

Andrea Scozzari, Steve Mounce, Dawei Han, Francesco Soldovieri, and Dimitri Solomatine (eds.), ICT for Smart Water Systems: Measurements and Data Science, Hdb Env Chem (2021) 102: 65–110, https://doi.org/10.1007/698_2020_609,
© Springer Nature Switzerland AG 2020, Published online: 24 July 2020

Abstract During the last decades, the role of data as a vital resource that enhances decision-making and supports efficient systems operation has become evident, with a growing number of companies viewing data as a key organizational aspects that has to be properly managed, instead of an operational side-product. At the same time, drinking water systems increase in complexity and feature smart sensors, which in turn leads to data-richer operation environments for the water services. Given this challenging context, the often-overlooked factor of ensuring high data quality and preventing errors in data streams becomes increasingly important. In this chapter the current data validation techniques, challenges and best practices of the Dutch drinking water companies is presented.

Keywords Anomaly detection, Best practices, Data quality control, Hydroinformatics, Netherlands

1 Introduction

1.1 Background

Despite the emerging need for holistic, efficient data management policies, implementing a proper Data Quality Control (DQC) strategy is generally a non-trivial task, as the protocols and techniques used are process- and context-dependent. For the water sector, protocols to standardize data acquisition and analysis are being developed for different parts of the water cycle, targeting the data streams of specific processes. For example, management frameworks in the context of urban hydrology and sewer systems have been developed [1, 2], as well as initiatives for European ocean and sea data management [3].

In the Netherlands, since 2012 a protocol for data quality is being developed by KWR Water Research Institute and the Netherlands Organisation for Applied Scientific Research (TNO) for registration of groundwater levels and hydraulic heads [4–6]. A consequence of such protocol is the development of a validation tool named Menyantes 1 [7] for which a new version of this software, named HydroMonitor, is currently under development. In the drinking water sector, a uniform failure registration database (USTORE) has been used by eight Dutch drinking water companies [8]. This initiative started in 2001, and it has been an ongoing process of continuous improvement, with both the complexity of the registration and data requirements increasing over the years. In 2017, a protocol to guarantee data quality was included in the PCD (Praktijkcode Drinkwater[1]) no. 9: 'Uniform failure registration'. In 2018, this PCD was released [8].

Within these protocols, one of the core ways of improving data quality is by performing data validation. Data validation or, in other words, fault detection and isolation (FDI) refers to the identification and handling of anomalies and outliers in

[1] https://www.praktijkcodesdrinkwater.nl/

data that cannot be explained by the underlying physical rules of the measured system.[2] These anomalies, otherwise known as errors, can be further distinguished in three types [10]:

1. Measurement errors (e.g. failure of data registration, maintenance problems, drifts, bias, strong gradients, lack of redundancy, problems of coherence at both local and global scale, duplication of data)
2. Human errors (e.g. sensor placement, sensor settings, faulty/inadequate calibration, unit conversions, round-off and data conversion errors)
3. Any occurrence of unexpected processes, modifications and events in the monitored urban water systems, either controlled or uncontrolled (i.e. pipe bursts, flooded pump station, maintenance of a filter at a treatment plant)

Untreated or mismanaged data strongly impacts the service operation, as it propagates deeper [11] into the decision-making process and leads to erroneous or ill-informed decisions regarding system operation, organizational mistrust, reduced service efficiency and, eventually, customer dissatisfaction [12]. The detection and identification of the aforementioned errors can be carried out with a variety of methods that include threshold, data-driven and model-based approaches, as further discussed in Sect. 3.

1.2 Scope and Approach

This chapter aims at providing a bird′s-eye view of data validation in the drinking water industry of the Netherlands towards better Data Quality Control (DQC) policies in the drinking water sector, by providing insights on (raw) data validation in two problem types, one of water quantity and one of water quality. The focus of this chapter is on a specific aspect of the overall DQC chain, which deals with faulty data detection and isolation (FDI). Furthermore, of interest are errors in the measurements, because sensing and human data editing process lead to raw data distortion in the form of, e.g. drift, bias, precision degradation or sensor failure [13]. Mapping this focal point to the typology of errors seen in Sect. 1.1, it becomes evident that this chapter focuses only on errors of type (i) and type (ii), i.e. measurement and human errors. Moreover, the focus lies on data validation to determine faulty data and the identification techniques, without expanding further on the decision-making process regarding to accepting or rejecting the faulty data.

As a first step, in order to identify the needs of the industry and its current practice, an inventory of current applications regarding DQC within the water companies was conducted. Visits or interviews with four water companies took

[2]Given this definition, any outliers or anomalies in data owing to natural rare and/or extreme events, including very low probability cases such as black swans [9], should not be considered as faulty data due to errors that have to be corrected.

place during the period January–May 2018, as well as surveys using questionnaires that were sent to representatives of drinking water companies that are part of the Dutch joint research (BTO Bedrijfstakonderzoek in Dutch) [14].

Secondly, to gain insight on different approaches on data validation, a literature review on faulty data detection techniques was performed, resulting in an overview of available techniques that are relevant for the drinking water companies. This overview differentiates between simple and complex techniques, and it also includes the range of applications of each one. Based on the insight gained by the literature review and the data gathered from the water companies, a data quality control is proposed using simple techniques.

Based on the findings of the previous steps, an application in three cases focusing on two types of problems in drinking water follows. The two problem types:

i. The detection of anomalies in volume flow rate, as an example of data validation in water quantity
ii. Anomaly detection in datasets of temperature, turbidity and pH, as an example of validation in water quality

The analysis of the case studies was performed in close cooperation with the water companies. In this chapter, only data validation for one company, Company A, is presented.

Finally, using the information collected from all previous steps, best practices and issues regarding DQC by the water utilities are identified, as well as recommendations for future application of faulty detection techniques, along with ideas for future research in the field of DQC in the Dutch drinking water sector.

1.3 Outline

A brief overview of the contents of the following sections is provided in this paragraph. In §2, the necessary foundations and theoretical background in DQC is defined, and an overview of current experiences of the drinking water companies is provided. Having set the foundations, §3 contains the literature review on faulty data detection techniques. This overview leads to a selection of techniques directly applicable to Dutch water utilities, in Sect. 4, the form of a flowchart, to implement simple techniques for DQC is presented.

In §4 are described the data obtained from the drinking water companies for different studied cases and their results, derived from the application of simple techniques following the proposed data validation. At the end of §4, the best practices and issues found during the implementation of the cases is summarized. Following the analysis, §5 contains the discussion and recommendations that highlight the need for future research. Finally, §6 describes the main conclusions drawn from each case study, as well as general conclusions drawn from the application of the methodology.

2 Data Quality Control: Theory and Current Practice

2.1 Principles of Data Quality Control

Data is now considered one of the fundamental pieces of the daily operation across many services [15]. Over the recent decades, rapid technological changes have transformed multiple service fields into data-rich environments, where decision-makers are increasingly called to evaluate and decide based on data. Smarter and more frequent metering [16, 17], along with advances in hardware, editing technologies and new data analysis techniques have reshaped decision-making from an empirical to an increasingly data-driven process [18]. Furthermore, the role of data is foreseen to grow, with the inclusion of technologies such as cloud-based systems and big data analytics in the systems analysis and, eventually, the decision-making culture, thus causing a paradigm shift in the value of information, the nature of expertise and, eventually, the practice of management and decision-making itself [19, 20]. This paradigm shift is also occurring in the drinking water industry, as drinking water networks become smarter, more networked and more complex [16, 17], thus providing increasingly data-rich inputs to the operators and the decision-makers.

The elevated role of data in decision-making leads to a pressing need for more efficient data quality services, as poor data quality leads to ill-informed operational decisions and, thus, less reliable systems and higher customer dissatisfaction [12]. Moreover, data can be considered the foundation of knowledge creation that leads to knowledge (Fig. 1), as time scales shift from the operational collection of

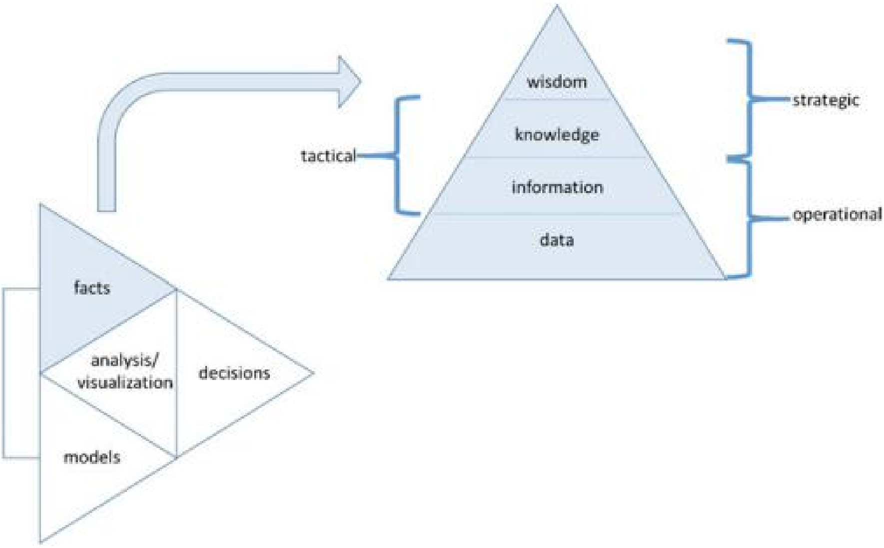


Fig. 1 Overview of the components feeding the decision-making process

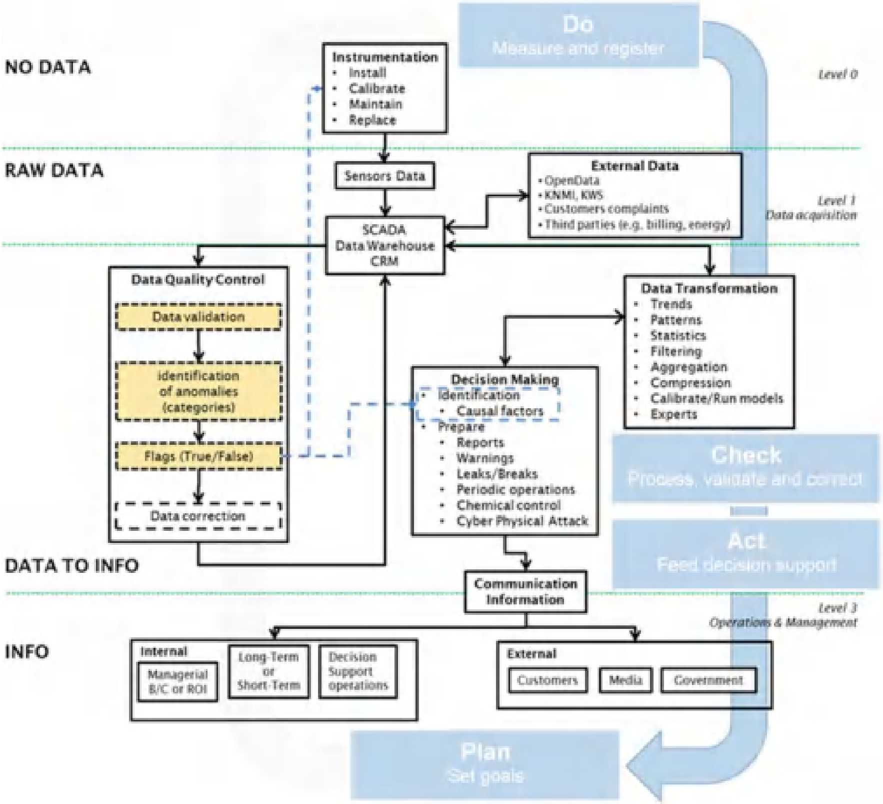


Fig. 2 The PDCA approach for data quality improvement, data to information at water utilities. Adapted from [22, 23]

(raw) bytes to information analyses at tactical level and, finally, strategic interpretation of the analytical results that provides knowledge and wisdom to management groups. It follows, as a result, that the water companies have acknowledged that good data provides a basis for good decision-making.[3] As such, the policies to control data quality serve a fundamental function to the transformation from data to wisdom [21] for water utilities, along with the broader frameworks that extend data applications for decision-making provided by the concept of hydroinformatics [14].

As in other product, process and service cycles in organizations, ensuring data of good quality requires an encompassing framework of continuous quality improvement, which can be defined as a framework for Data Quality Control (DQC). To design such a framework, classic quality improvement methodologies can be employed, such as the Plan-Do-Check-Act (PDCA) approach (Fig. 2) [22, 24, 25], which can be used to describe the continuous improvement of measurement systems

[3]Minutes, Hydroinformatics platform 12 October 2017

and their data products as a cyclic process. The PDCA approach can be viewed as a proactive framework which continuously monitors and registers data, checks their integrity, acts upon the checked datasets to feed information-based decision-making and plans strategies, including proposition of improvements on the sensing/monitoring system which in turn closes the loop [26, 27].

2.2 The Role of Validation in Data Quality Control

Within the DQC context, validation techniques play a key role in connecting the wealth of information obtained by raw data acquisition with decision-making and planning. The acquired data (i.e. the result of a "Do" step in a PDCA cycle, see Fig. 2) needs to be checked against errors and, in case faults are detected, needs to be corrected before feeding any decision-making process (i.e. the steps of "Act" and "Plan"). To complete this transition, a "Check" step is needed, which is better known in information analysis as Data Validation [23].

As seen in Fig. 3, data validation can be further distinguished in three steps: Collection, Detection and Correction. Data collection refers to the process of gathering data through data streams from each sensing device to a database, otherwise known as a data warehouse. The step that follows is the detection of a subset of data which could be deemed faulty. Detection techniques have to ensure that they can safely distinguish between actual faulty data and data which appears doubtful but its deviation could be attributed to something else than an error (Fig. 4). As a last step, the data confirmed to be faulty need to be corrected (e.g. empty values filled, outliers corrected based on other close values, etc.) before the data can be interpreted further and used as a basis for decision-making. This stepwise process of identifying and

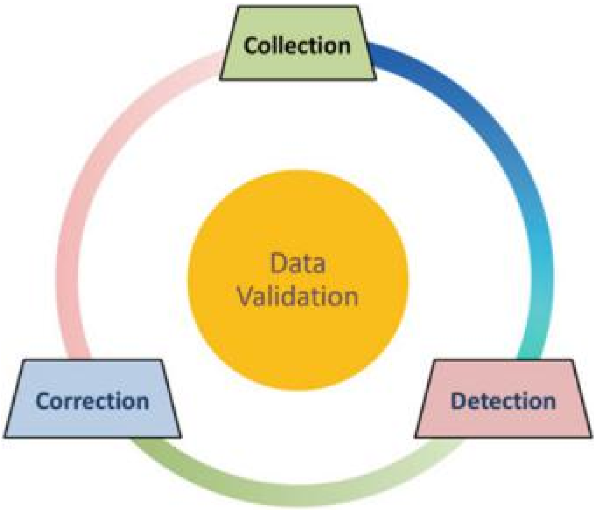


Fig. 3 The three steps comprising identification of faulty data

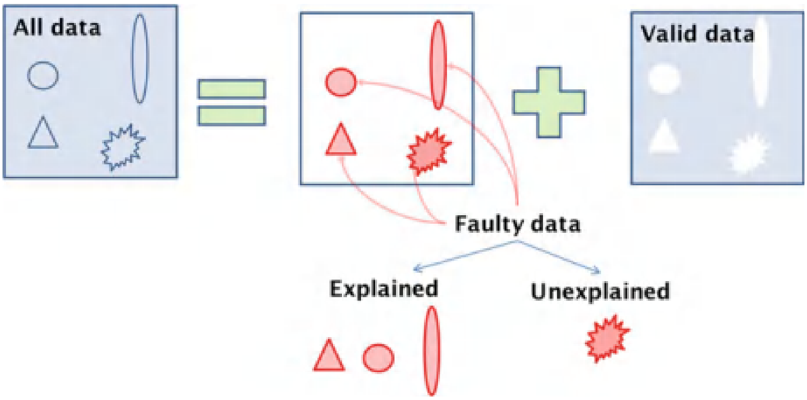


Fig. 4 The data set which is target of validation

correcting faulty data is also known in literature as fault detection and isolation (FDI) [28].

Primarily, the goal of data validation lies in identifying and extracting the subset of data which may be considered faulty (Fig. 4), i.e. not representing a valid measurement of reality, due to a measurement or human error [10]. From the likely faulty subset of data, some data represent occurrences of irregular/unexpected processes in the system (i.e. pipe bursts, catastrophes, maintenance downtime etc.). These data constitute a third type of human error that lies beyond the scope of this study, as explained in Sect. 1.2. The focal point of this study lies, therefore, in techniques that can be used to detect the subset of faulty data whose faultiness can be explained and attributed to measurement or human errors, i.e. the first two types of errors seen in Sect. 1.1. At the same time, the detection process has to ensure that irregular but non-faulty data are not classified as faulty. For instance, outliers owing to extreme events and even unprecedented events such as black swans [9] belong to the valid data subgroup and should not be classified as faulty data.

As a core process in DQC, data validation is not a new concept and has been developed heavily in DQC platforms [29], relying largely on algorithms and mathematical techniques of faulty data detection. However, automating the entire process of data validation is not realistic [30], and expert judgment is still required to cross-validate the results produced by mathematical methods.

2.3 Data Quality Control in the Context of Drinking Water

The concepts on DQC and validation described in §2.1 can be applied to any data stream production environment [31], including course drinking water (DW). In that case, data describing the status of the DW network (i.e. samples of physical variables such as water quantity, quality, water level, pressure head, etc.) are acquired by sensing devices from multiple points within the production, transport and distribution chain and typically stored in a central repository.

To demonstrate this, Fig. 2 presents the data-to-information workflow typically seen in the context of drinking water. Elements from the PDCA approach, as analysed in §2.1, have been mapped, focusing on the Do-Check-Act-Plan parts of the loop that describes the pathway from data to information (and, eventually at the plan stage, knowledge). One may observe three distinct levels: acquisition of data (level 1), followed by transformation and quality control (level 2) and finally dissemination of the information produced by data (level 3). In the data acquisition level, data is coming from sensors (possibly in real time) or can be fed from periodical manual checks, such as local visits, regular sampling, etc. Such data can be considered raw data, which means that they are stored as obtained by the sensors.

After acquisition, a common workflow inside a data warehouse is to upscale fine-scaled data through aggregation or averaging, in order to produce metrics and time series at intervals meaningful to management or to identify extreme or periodic events [32] and causal factors [1]. Prior to this step, raw data need to be cleaned of errors belonging to the two types explained in Sect. 2.1. In the context of water utilities, these errors could be due to sensor failure (maintenance problems, bias, de-calibration, communication failure, physical damage due to catastrophes, etc.), due to human mistakes (incorrect installation of measuring equipment, e.g. sensor settings, unit conversions, not using the validation protocol issued by the manufacturer or forgetting registering information) and due to unexpected processes, phenomena and events in the monitored urban water system (electrical power outage, failure of a pump).

Due to the numerous processes involved, data validation is not trivial but depends on:

- The type of variable monitored
- The overall measurement and sensor/monitoring network conditions and more specifically:
  - The degree of system complexity

    Larger systems may require larger sensor networks in this way more variables are measured simultaneously.

    Sensors located far away from each other may be correlated or measure completely different patterns with delays.
  - The operational age of the sensor/monitoring network, which is translated in the time length of available data
  - The type and technology sensors/equipment used

    Precision, accuracy, type of measurement, uncertainty of measurement
- The characteristics of the phenomenon being captured and more specifically:
  - The type of problem (leakage detection, water balance closure, water quality, etc.)

– The way data is represented (i.e. real numbers as pressures and flows, binary as pump switches, categorical as status of data provided by most systems),
– Data resolution (i.e. both temporal or spatial)
– The method/technique used for validation (see Sect. 3 "Literature Review on Faulty Data Detection Techniques for Water Utilities")

  A single technique cannot be used for all instances.
  The time spent between techniques can vary between pre- and post-processing.

– The metrics used (i.e. some variables such as pH, temperature and turbidity are based on a sensor calibration made through laboratory tests)

• The user and objective

  – Data may be used for real-time control (RTC) or offline historical analysis.
  – Some methods may be used by data warehouse administrators, while for water accounts managers as end users only performance indicators or data aggregation as post-process are relevant.

2.4 Current Implementation of Data Validation Techniques by the Drinking Water Companies

Drinking water companies own and manage extensive systems (with several facilities), which are continually monitored at different points, e.g. production, transport and distribution. For example, Company A has approximately 73,000 variables in total, measured every second. Currently every company is dealing with data quality issues. Due to the exponential growth of data and the specific characteristics of each variable, these cannot be easily manually validated.

Additionally, time series (TS) are becoming increasingly necessary for modelling, such us hydraulic, risks and decision models.

Other emerging drivers are stricter laws and regulations for the definition of standardization of data models and protocols. For example, the European INSPIRE directive[4] defines the technical guidelines for data specification of Infrastructure and its spatial information. Such initiative is currently an invitation for standardization moving forward (which may facilitate exchange of information), rather than a mandatory application for future implementations for the drinking water utilities.

Despite these drivers, there are still several barriers to validate the data. The volume of real-time information has become so extensive that validation of all the variables by a human becomes unrealistic. To deal with it, in some cases data

[4] Commission Regulation (EU) No 1312/2014 of 10 December 2014 amending Regulation (EU) No 1089/2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data services

validation is limited to aggregated data (e.g. daily water use in a supply area). In other cases, software tools are built to screen and flag the data which is identified as suspicious.

Currently, it is not possible to validate all the variables. In general, the most important data is validated, by a mix of manual and automatized routines. One of the companies introduced the concept of a data diet, which implies a profound consideration of which data have to be measured, in which kind of time interval they have to be stored and which of them have to be validated, before starting to generate data. For some datasets, it is not really needed to develop a high level of validation. For those datasets where validation is essential, we should look into the possibility of correlation between different variables, measure all these variables and use correlation techniques (data science, statistics, models) for the validation.

The techniques used by the water companies to validate the data include:

- Manual validation (expert judgment)
- Visual comparison
- Control of measuring range, plausibility, data types
- Cross-correlation, statistical methods and models
- Combining own data with validated external sources

Examples of current practices on data validation are:

1. Filling missing data in the records of produced water using registered energy use and relation between energy use and produced m^3 of water.
2. Determining missing year of installation of the pipes using the age of the buildings of the area. Although these methods are not exact, they help to improve the quality of the datasets. To the question regarding which platforms drinking water companies use to store and process the data, each company has its own (customized) systems. Some examples are shown in Table 1.

Table 1 Overview of tools per utility

Utility	Systems
A	FEWS, Aspen and Midas
B	A MS SQL server database
C	PI (real-time information assets), SAP (context information of the assets) sample manager (information regarding water quality) and SCADA (events and all process information)
D	PGIM (database van ABB 800x a process automatization) and own data warehouse (Microsoft SQL)
E	PA (PIMS) and SAP
F	GIS (ESRI) own information system (accent) and SAP SharePoint
G	Data warehouse and Infor PGIM
H	Oracle Data warehouse, SQL, Excel, MS Power and BI ARCGIS

2.5 Data Validation: Experiences of a Front Runner – Company D

One of the drinking water company, which is identified as one of the front runners in relation to automatization of routines for data validation, collects a lot of measurement data in PI (from OsiSoft), but still it only validates just a small percentage of all the data.

Company D has a system to validate water volume flows, and it validates the daily volume flow at measurement points on the boundaries of the DMAs (about 150 locations per day). The validation consists of checking if the difference between meter readings at the beginning and end of the day is equal to the sum of the analogue readings during the day. If that does not turn out to be correct, the user of the system is assisted in correcting the daily quantity for instance by showing typical values/ranges for this type of day and the historical values of the last 7 or 14 days. Validating always consists of two steps: (1) check whether the data to be validated is plausible; if not, (2) correct the data. Individual large customers ($>10 \times 10^3$k m^3/y, approx. 600 units) are validated on a monthly basis. Meter readings of each month are compared with the previous month. It is also checked whether the difference between both meter readings is equal to the sum of hourly values (which are collected to determine peak rates).

Within the data validation process, the responsibilities are well defined: An employee (from the control centre) validates the measurement points in the net and checks that the validation actions are carried out by production sites (if they appear to be necessary). An employee from the industrial water department validates large customers ($>$100,000 m^3/y) and all the industrial water customers. An employee of the customer contact centre validates customers with drinking water consumption between 10,000 and 100,000 m^3/y. The operators on site validate the outgoing flows of production sites, plus the waste water flows and the incoming water flows. The validation rules are also reported. Despite some companies having automation for data validation, this task still represents a lot of work and needs constant attention. It has become increasingly clear that not only the quality of the sensor and the data logger but also a good interface to PI are very important aspects within the process to validate the data.

3 Literature Review on Faulty Data Detection Techniques for Water Utilities

3.1 Background

The detection of anomalies corresponds to the first line of defence against faulty data, as it allows near real-time identification of individual values or sets of values that are erroneously measured or that deviate from what is regularly measured in the system (e.g. a water distribution network or a treatment plant).

Detection is performed by developing a validation technique for the object that generates data (e.g. a sensor) at the time it generates the data, with various techniques having been developed for this purpose in literature [3, 33, 34]. In typical operational cases, data validation is carried out manually by expert judgment using both analytic and visualization tools. The issue with that approach is that with current data streams only a small amount of data can be validated by operatives from the utilities [2], and as evidenced by several authors, there is always a human bias in the decision making. This human bias is particularly important while determining whether data corresponds to anomalous/irregular behaviour (which can be explained, e.g. due to extremes or failure of network parts, such as leakages and pipe bursts) or faulty data (unexplained, e.g. due to sensor faults) [35], as explained in Fig. 4. In this text, faulty data corresponds to data belonging to the two error categories explained in Sect. 2.1., i.e. measurement and human errors only.

An inventory of faulty data validation techniques is presented in §3.2, based on a literature review on the subject. In general, there are sequential steps for data validation corresponding to:

(a) Input variable selection, which consists of the selection of a subset of interest from the data warehouse that have to be validated [36].
(b) Pre-processing, which includes a number of statistical and modelling techniques that help to identify anomalies, either by statistical analyses or by contrasting real data to modelled data equivalents.
(c) Anomaly and faulty data detection, which can be performed with either simple or advanced (statistical) methods.

The diagram of Fig. 5 also provides an indication of the amount of data required (Data arrow) and the amount of time (Time arrow) for each technique. Evidently, more advanced techniques require more data and time. In pre-processing techniques, the use of models or meta-models may drastically increase the data and computational requirements, but it may lead to a significant reduction of the uncertainty in faulty data identification.

3.2 Faulty Data Detection Techniques

Faulty data detection techniques are generally classifiers which divide the data in two classes (correct and faulty/doubtful). While some detection techniques have been applied for generic problems [29, 37, 38], water research has also developed domain-specific techniques for sewer systems [36], geo-hydrological systems [4, 6, 7, 33, 39], water quality sensing [40], automatic or real-time data validation in urban systems [2] and specific problems such as the determination of leakages as anomalous data in water supply systems [41]. As depicted in Fig. 5, these techniques are divided based on their complexity to two main categories: simple tests and statistical tests.

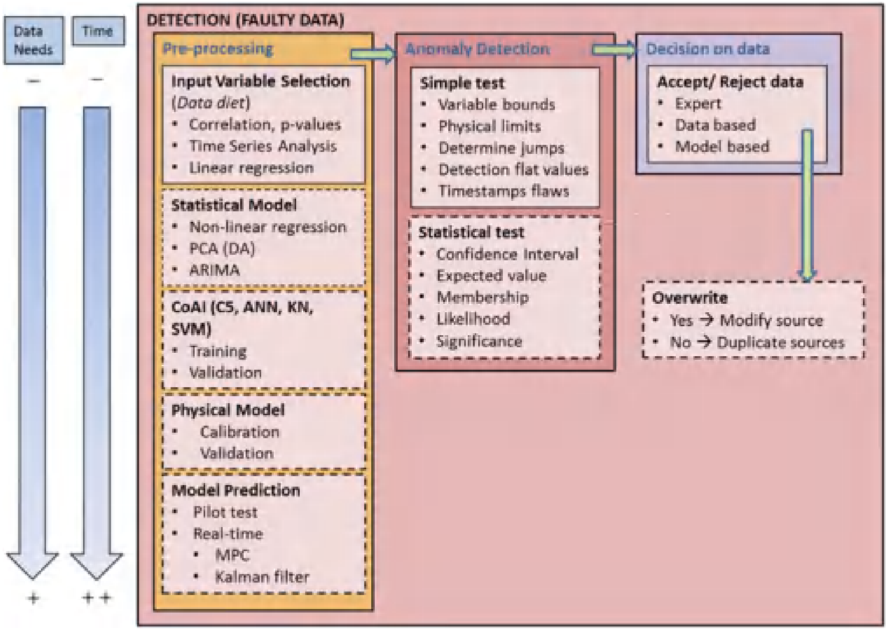


Fig. 5 Inventory of faulty data detection techniques. Solid lines represent the focal subject of this study

3.2.1 Simple Tests

These techniques offer an array of simple methods that can be used to determine if data follows regular patterns or is contained within certain thresholds. Examples of these techniques include (Fig. 6):

(a) Special value detection (zero, flagged values), where faulty data is identified based on a pre-set special value, which acts as a flag. This typically happens when a sensor fails to register a valid value (e.g. due to power failure or maintenance downtime) and logs data such as ′Null′, ′-9999′ or, in some cases, 0, which can be then easily tracked.

(b) Flat line detection, where groups of consecutive constant (zero or non-zero) data are identified and flagged as suspicious. Flat line detection can be used for multiple purposes, as a tool for the detection of gaps in the data or for determining constant values which tend to be very rare inside a water distribution network (WDN), especially if the time series are characterized by periodic patterns. The definition of the time window and duration for a flat value test is problem-specific, requiring fine-tuning for each variable and each sensor [40].

(c) Boundary detection (minimum and maximum threshold detection), where data that exceeds certain (pre-set) minimum and/or maximum thresholds is flagged as suspicious. In water systems, this analysis is often based on geometric, hydraulic

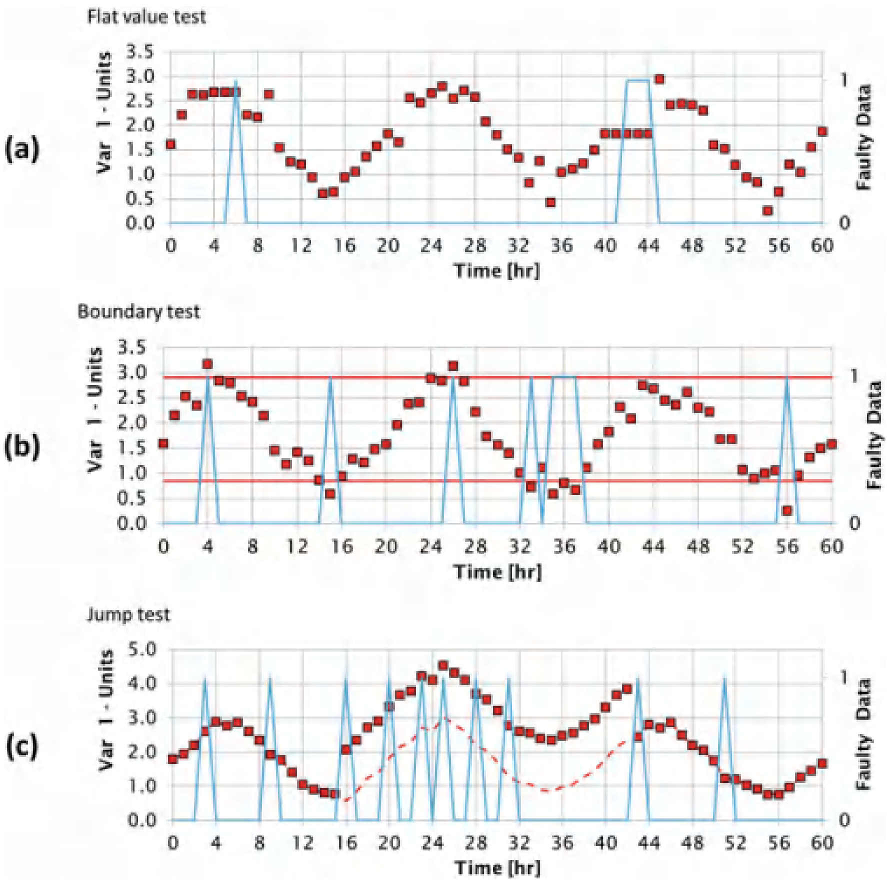


Fig. 6 Examples of simple testing for data validation, with measured samples in red, boundaries as red lines, and marks of faulty data in cyan blue. Faulty data 1, no flag 0. Panel (a) flat line detection, panel (b) min-max boundary testing, panel (c) jump-leap detection

and data quality constraints. For instance, tanks have a limited capacity, and valid water level values have to be below that; likewise, pressure values in a network are bounded by physical (and pipe strength) constraints. These tests can be set by setting appropriate thresholds on data before storage to the database or data warehouses, based on expert judgment, e.g. from both system operators and data managers. Thresholds can be also based on historical operational values, for instance, when deriving the operational temperature range of a wastewater treatment tank.

(d) Jump or leap detection (change of variance), where faulty data are identified based on leaps or jumps in signal data. As part of these techniques, any subsets that deviate from a general trend or (seasonal or diurnal) pattern are flagged as suspicious. These techniques are effective at identifying systematic errors, e.g. due to sensor calibration faults.

Evidently, different tests are suitable for specific water system assets, e.g. there is no point having a flat line test for a pump that gets periodically switched on or off. Besides these simple test types, other variations are also possible depending on the context, such as the use of slope/gradient tests or sensor-specific drift techniques; these methods fall beyond the scope of this study.

3.2.2 Statistical Tests

Beyond simple testing for faulty data detection, statistical analyses offer a formal framework to probabilistically detect faults, e.g. as outliers of given distributions. The statistical tests that exist in literature are given in the following sub-sections.

Comparison of Flow Pattern Distributions (CFPD)

Time series such as inflows and consumption patterns of a WDN generally follow patterns on a diurnal, weekly and seasonal scale. Variations in consumptions due to season change are common (i.e. winter, summer) [42], as well as due to specific events in a short time window [43]. In these cases, a comparison of flow pattern distributions (CFPD) can be performed for the detection of anomalies, for instance, based on the identification and interpretation of features in CFPD block diagrams. Feature analysis and techniques for the automated screening of data with seasonal statistics can be used to measure deviations for the expected value and infer the existence of faulty data [44].

Spatial Deviation

In some cases, variables exhibit correlation in space. In these cases, a complete set of techniques for faulty data identification can be explored with the use of geo-statistical techniques, such as Kriging, in order to estimate deviation between the measured value and the estimated [45]. These techniques have been mostly applied in hydrology and fall outside of the scope of this chapter.

Extreme Value Checks

Another type of statistical approach is to perform extreme value analyses (EVA), where the probability distributions of sensor variables are inferred from observed values. Once a distribution is fitted, the quartiles of the data can be obtained and a hypothesis test can be performed to identify values which fall outside the statistical boundaries. If the test has statistical significance, the value can be registered as an outlier, implying that is considered as faulty data. This approach can supplement the CFPD method described above.

Regression Analysis, Correlation Checks and Principal Component Analysis (PCA)

Statistical regression encompasses a large number of techniques that aim at fitting a line or curve to a data and using it as a basis to determine outliers and – potentially – faulty data points, as well as to predict/infer missing data, based on a subset of ′good′ data points [46–49]. Such a regression model in principle includes two (2) components: (a) predictors or explanatory variables, which form the basis of the prediction and (b) response variables, which can be predicted based on the explanatory variables.[5] Multivariate regression models can be also employed in case there are multiple explanatory variables for the water production. Techniques when data are sampled in unevenly distributed intervals also exist [50] and can be of use to water utilities, where some variables are stored irregularly at specific events and subsequently interpolated prior to storage in databases/data warehouses. On the assessment side, multiple metrics[6] to assess the successful adjustment of the regression line or curve exist [51], quantifying whether the model represents the response variable as a function of the explanatory variables accurately. Once the regression model is fitted, any large error between an actual data point and the estimation can be flagged as a faulty data point.

The main issue of relevance to regression models for data validation is to properly select the explanatory variables for a certain response variable. This issue is related with the concept of data diet (Fig. 5), in the sense of looking for a correlation between different variables by (1) deciding which correlations are possible by using physical and domain knowledge (expertise of drinking water) and (2) employing statistical correlation techniques used by data scientists.[7] The techniques used for the selection of explanatory variables are known as input variable selection (IVS) and

[5] For example, one may be interested in the relation between the monthly water production [m^3] of a utility and the total energy consumption [kWh] used for treatment, transmission and distribution. If there is a missing or suspect faulty data in water production, then this value can be estimated based on the total energy consumption of the utility (explanatory variable).

[6] Examples include the Root Mean Square Error (RMSE), Coefficient of Determination (R^2), Nash-Sutcliffe Efficiency (NSE), Kling Gupta Efficiency (KGE). For a successful regression model, RMSE should have low values (close to 0.0), while for efficiency metrics a value near 1.0 is optimal.

[7] Domain knowledge is always relevant in the water sector, as a SCADA, water data warehouse or database can exceed a thousand variables ($n_v 1000$) and thus a combination of response variables that exceeds $n_{dd} = \frac{n_v(n_v-1)}{2}$ 500:000 values.

have broader applications in engineering and environmental sciences [48, 52, 53]. Such techniques apply correlation and stepwise selection of explanatory variables to perform the identification of significant dependencies and are based on two main techniques:

(a) Correlation analysis, which explores auto- and cross-correlation structures based on time series analysis [54]. A multitude of techniques exists in both stochastic and deterministic time series analysis [55] and analytical models exist that include long- scale dependency [56, 57], persistence [58–60] and more sophisticated models such as ARIMA/ARIMAX [61–63]. In water quality analysis, correlation has been extensively applied for open and pressurized flow when data validation is required [64, 65].
(b) Principal component analysis (PCA), as its name states, has been applied mainly for determining variables which may project the data into its principal components. The data is transformed using an orthogonal transformation and then converted into a set of variables uncorrelated among themselves, which in turn denominates the principal components. Particular applications for PCA in the water sector are to support the identification of consumption patterns and leakage detection [66].

3.2.3 Data-Driven Models

Another approach that features methods heavily relying on the dataset itself is called data-driven modelling (DDM) [67, 68]. These methods originate from the computer science fields of computational- or artificial-intelligence (CoAI) and machine learning (ML) and are used for data exploration, data mining [69] and also classification. The latter case aims at building classifier models[8] based on a large number of independent (predictor) variables and is of use to data validation. Data-driven techniques, among others, include:

- Decision trees (DecT), which can be considered the simplest technique to perform classification of large datasets [70], seeing use in the prediction of leakages and breaks for WDN models [71]
- Support vector machines (SVM), which are non-linear regression algorithms over multi-dimensional input spaces that have been used for leakages and demand pattern identification [72–74]
- Artificial neural networks (ANN), which are machine learning algorithms inspired by biological neural networks and used for classification. In the water sector, ANN have been applied for the identification of losses and leakages in

[8]Classifier models (or classifiers) output a categorical variable (e.g. with values '1' or '0', that can be flags for a data validation problem), based on a (large) number of input variables, that can be for instance geophysical or water system time series.

distribution networks and supply systems [41, 75] but also as surrogate models of complex networks [76, 77], to simplify simulation and save computational time

The drawbacks of these data-driven approaches is that they require a large amount of data for training (and, ideally, validation) and that they are considered, in an operational sense, ′black-box′ estimators, with the human actors having limited insight of their internal structure and functionality.

3.2.4 Physical Models

Physical modelling is able to provide a reference set of (modelled) system measurements that can be used as a basis to perform faulty data identification. For instance, if a WDN model (e.g. in EPANET, Infoworks or WaterGEMS) of the system is available and properly calibrated, it is possible to simulate the behaviour of the network and then extract the data on pressures, given the proper drivers (e.g., current levels in tanks and reservoirs, demand forecast) are known. It is then possible to use that simulation output as a comparison basis with sensor data; any deviation from what is modelled might then be flagged as (potentially) faulty data. Deviations of the reality from the model could indicate for instance an increase of losses due to pipe breaks or background leakages [78] or in some cases faulty sensor behaviour (e.g. due to service downtime).

The trade-off for utilities with setting up a physical model lies between model reliability and computational cost. The more reliable the model is, the larger the effort to keep models up-to-date and properly calibrated, which is a (continuous) cost for the utility. The reduction of complexity and computational time can be curbed with techniques such as network skeletonization [79, 80], hydraulic simplification [81] and topological aggregation of serial pipes [82]. Another approach to reduce computational time lies in the use of surrogate modelling techniques [83]. Even at a higher computational cost, Model Predictive and Data Assimilation approaches are beginning to be developed for uncertainty reduction [43, 84, 85] and as a consequence may be used to identify data anomalies.

3.2.5 Knowledge-Based Techniques

Regardless of the mathematical, statistical or modelling techniques that assist faulty data detection, knowledge and expert-based judgment offers invaluable insight to validation and allows the operators to reach decisions on whether flagged data is actually faulty or not. A strategy for knowledge-based evaluation might include:

- Periodically checking the status of sensor or asset, where log files and metadata are checked to evaluate whether a particular sensor or asset is operating well and is well-calibrated. These checks also include periodic maintenance, such as (re-) calibration in sensors, as these are expected to have a reduction in their reliability and accuracy over time [86].

- Checking the duration between sensor maintenance and anticipating operational downtime periods.
- Maintaining and consulting a repository with data from past network failure events, such as a Pipe Failure Data (PFD) repository, which is of use to update physically based models and clarify certain anomalies which otherwise would be identified as faulty data [87]. Standardization and integration of PFD repositories among water companies helps streamline this task [88, 89].
- Combining expert-based judgment with ad hoc validation tools built for a specific part and function of the water system [7].

4 Application of Data Validation

4.1 Overview of the Data and Selection of the Techniques

In order to apply and test diverse data validation techniques, two problems in the field of drinking water distribution were identified:

- Anomaly detection in volume flow rates and energy use
- Anomaly detection in datasets of temperature, turbidity and pH.

In Table 2, the overview of the cases and the data validation techniques used during the 2018 survey are presented where only the proposed data validation applied to Company A′s data is discussed in this chapter.

The data provided by the water utilities contain several differences. In terms of variables, water quality data tends to be more homogeneous. Data resolution was also variable and depends on the type of registration for each utility. It can vary in minutes, quarters (15 min), events (when significant change occurs) and pulses.

Timestamp registration is also very heterogeneous across utilities, dates can contain summer and winter time as number or other indicator (+1.00 or +0.00), and some information is consistently absent on the same timestamps, most likely due to data communication. In this regard, at 00:00, Company A contains no data, indicating that the data transfer is most likely to occur at this time at night. Data from one utility was provided as a Last In, First Out (LIFO) format (reverse dates) for some variables, while the same utility provided a more common First In, First Out (FIFO) format.

Length of time series was also very variable as it was not possible to obtain more than a few months of data for Company B. Some sensors have recently started to send live data. Due to the high variability in data types and content, it was not possible to perform a one size fits all analysis of data validation and more specifically of faulty data detection, so the focus given to specific datasets is variable to present a broader set of analyses within the funding research instrument BTO.[9] In each

[9]BTO: Bedrijftaak Onderzoek. It corresponds to research within the consortium of the ten drinking water companies from the Netherlands and one Belgian drinking water company.

Table 2 Overview of data and techniques used for validation

Company	Datasets	Remarks	Action	Current validation	Techniques
Company A	Water quality WWTP I production	One location: Temperature, pH, turbidity. Company A has a built up system that labels the data with different flags	Compare identification of anomalies with own system and with registration of maintenance activities or reported incidents	Data is validated automatically by the system	Simple test confusion matrix
	Flow City A Pump stations	Large-scale City A. 5 pumping stations. Company A has a built up system that labels the data with different flags		Data is Validated automatically by the system	Data aggregation regression Knowledge based
	Energy City A Energy provider	Large-scale City A. 5 pumping stations. Data comes from a third party. No flags available for this data		Data from a third party not validated by Company A	Data aggregation regression Knowledge based

specific case study, the data used and the test applied are presented. Taking advantage of feedback sessions, it was possible to implement additional expert knowledge in the determination of faulty data of two utilities. Only the analysis for Company A will be discussed in this chapter.

4.2 Data and Proposed Data Validation for Company A

The utility currently monitors 73,000 variables (simultaneously) for drinking water and 23,000 for wastewater. It is intended here to compare current Company A′s data validation with a proposed data validation. For comparison purposes, the proposed data validation is composed of simple tests. The four tests are:

- Verification of boundaries
- Verification of timestamps
- Verification of flat values
- Verification of jumps

Data was collected from two sources the Pumping Stations (PS) for the whole system and Water Quality (WQ) data at Waste Water Treatment Plant (WWTP) I and WWTP II. Data from the pumping stations was analysed only for the identification of faulty data of timestamps and consistency current DQC. Data from WQ corresponds a total of four time series and three different variables (i.e. temperature (1), pH (1) and turbidity (2)), from their filtration processes.

Data collection corresponds to time series in the period between 1 January 2016 and 31 December 2017. After a visit to the facilities of Company A, it was verified that it is also possible to fetch data directly from their data warehouse. Values interpolated at different resolutions can be obtained. However, for the time being, a resolution of 1 min for all variables was selected for further analysis, with the exception of energy use which is provided by the energy provider at a 15 min resolution.

Data was provided as CSV files. Data contains four fields (columns), (A) the sensor ID, (B) the timestamp, (C) the measured value and (D) a status of signal′s health, established by the system. Such pre-screening DQC is split in four different categories as flags: (1) Good data, (2) Faulty data, (3) Dubious data and (4) Out of range. It was not possible to determine the specific rules which drive the definition of different categories as they are automatically triggered by the DQC system of Company A.

4.3 Results Obtained

4.3.1 Water Quality Data at WWTP I and II

For each time series, the corresponding number of flags identified in the data is presented in Table 3. It is evident that there are a limited number of flags identified by the system. This is indeed a cumbersome task for Company A as to our knowledge more than 73,000 variables are updated every minute by their system only for drinking water.

Subsequently the proposed data validation has been applied. The confusion matrices (Table 4) present the comparison between the observed faulty data in the current data validation, and the ones identified by applying the proposed data validation (KWR). If any timestamp sample is identified as faulty data by any of the simple tests proposed, then data is considered faulty. There are 3 possibilities:

- When both DQC schemes agree in the identification a Yes-Yes coincidence is identified. This can be understood as a validation of Company A′s validation. For

Table 3 Number of flags present in water quality data from Company A (% = percentage of total number of timestamps)

Treatment plant	Acronym	pH		Temperature		Turbidity	
		(–)	%	(C)	%	(%)	%
WWTP I	L01	25	0.00	33	0.00	30	0.00
WWTP II	L02	N/A	–	N/A	–	67	0.01

N/A not available

Table 4 Confusion matrices of water quality data for company A′S and proposed DQC

pH		KWR		
		Yes	No	Total
Company A	Yes	25	9	34
	No	254	-	
	Total	279		

Temperature		KWR		
		Yes	No	Total
Company A	Yes	29	4	33
	No	1250	-	
	Total	1279		

Turbidity II		KWR		
		Yes	No	Total
Company A	Yes	61	6	67
	No	187	-	
	Total	248		

Turbidity I		KWR		
		Yes	No	Total
Company A	Yes	20	10	30
	No	39	-	
	Total	59		

The significance of the colors is linked between this table and Figs. 7, 8, 9 and 10. The blue color and yellow color are represented both in this table and in Figs. 7, 8, 9 and 10. This means that the cases represented in the figures correspond to specific cases of Yes–No and No–Yes which are presented in this table

a perfect agreement among the two analyses a Yes-Yes cell must contain all faulty data flags for both DQC schemes.

- In the case that the current Company A′s DQC is unable to identify a faulty data compared to the one proposed, a No-Yes coincidence is identified.
- In the case that the proposed method is able to identify faulty data, while the current validation of Company A is not able to do it, a Yes-No coincidence is identified.

It is also of interest that for all variables the number of faulty data identified with the proposed validation is larger for this analysis than the number of flags obtained with the current one of Company A. There are two possible reasons for this. Either the proposed data validation is more compact and sensitive to faulty data or there is a need to improve the current data validation of the utility. In the first case, this is a disadvantage for the operatives, as this will translate to a large number of verifications required. For example, temperature data in WWTP I confirms more than 1,200 flags for a verification during a 2-year period, or almost two flags per day. As it is now, data validation for Company A is already cumbersome, so the possibility of performing such task for more than 73.000 variables seems impossible to come to reality. A calibration process of the parameters for each of the variables is required to be performed on individual basis. In the second case, it is possible that the proposed data validation has indeed identified additional faulty data. Although this may sound controversial, some examples are presented to discuss the reliability of DQC of Company A.

Figure 7 presents the time series for pH at WWTP I. The range of variability of pH is quite small due to the need by Company A to keep its magnitude within a narrow band. However, there are some spikes present in the data which are identified both by Company A′s system and the proposed data validation.

In Fig. 8, time series of temperature in WWTP I is presented. In this case most of Company A′s system flags are captured by the simple data validation proposed, and indeed three time windows in which the possibility of faulty data were identified are presented. Such time windows are centred in 2016 around April 18th, May 22nd and December 29th.

The case of turbidity is presented in Fig. 9 (in logarithmic scale). Here the most relevant feature for data validation is that the time series presents jumps at different periods. Such jumps (drifts or changes in variance) occur during 2016, around July 1st and in 2017, around 3rd of April and 9th of August. This can be due to a modification in the operational conditions of the treatment plant, given that data corresponds to the filtration system of the treatment plants. Without further information it was not possible to elaborate a hypothesis on this change of behaviour.

A duplicate analysis for the same variable, this time at WWTP II, (see Fig. 10, vertical axis in logarithmic scale) shows that Company A′s flag system tends to allow higher values of turbidity as normal events. An example is the spike in 2016, during April 1st. This could have an operational reasoning; however this behaviour is not identified in the logbooks provided by the utility.

On the other hand, there are some time windows in which the simple tests identified plateau values registered in the raw data, while Company A′s system was not able to do so (see Fig. 10). Such time windows are identified in 2016 around December 22nd and in 2017 around September 26th.

It was possible to identify most of the faulty data, without previous knowledge of the system rules. However, in some cases with the simple tests, some additional “likely” faulty data was identified among the time series. This does not mean that the statuses provided by Company A are not reliable enough, rather than one of the detection rules presented here (i.e. flat value detection) may not be currently implemented inside their data warehouse.

4.3.2 Pumping Stations Data

There are a total of five pumping stations (PS) in City A system indicated with the following abbreviations: WPK, AVW, HLW, OSD and HLM (see Fig. 11).

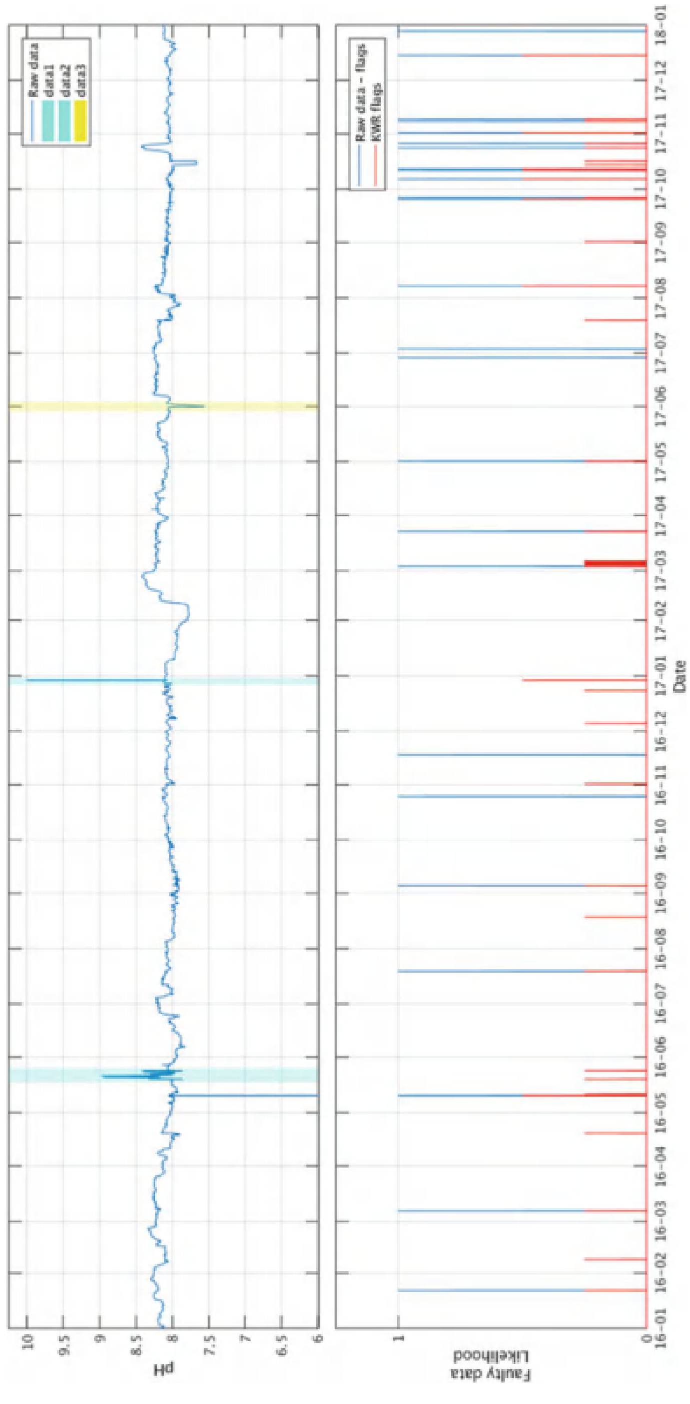


Fig. 7 Time series of PH IN WWTP I. Showing also system flags and proposed DQC. In Cyan, time windows in which the proposed data validation identified faulty data (no-yes), as May 2016 and January 2017. In Yellow, time windows in which there was no identification by any system (no-no) but visual inspection shows that there is faulty data as in June 2017

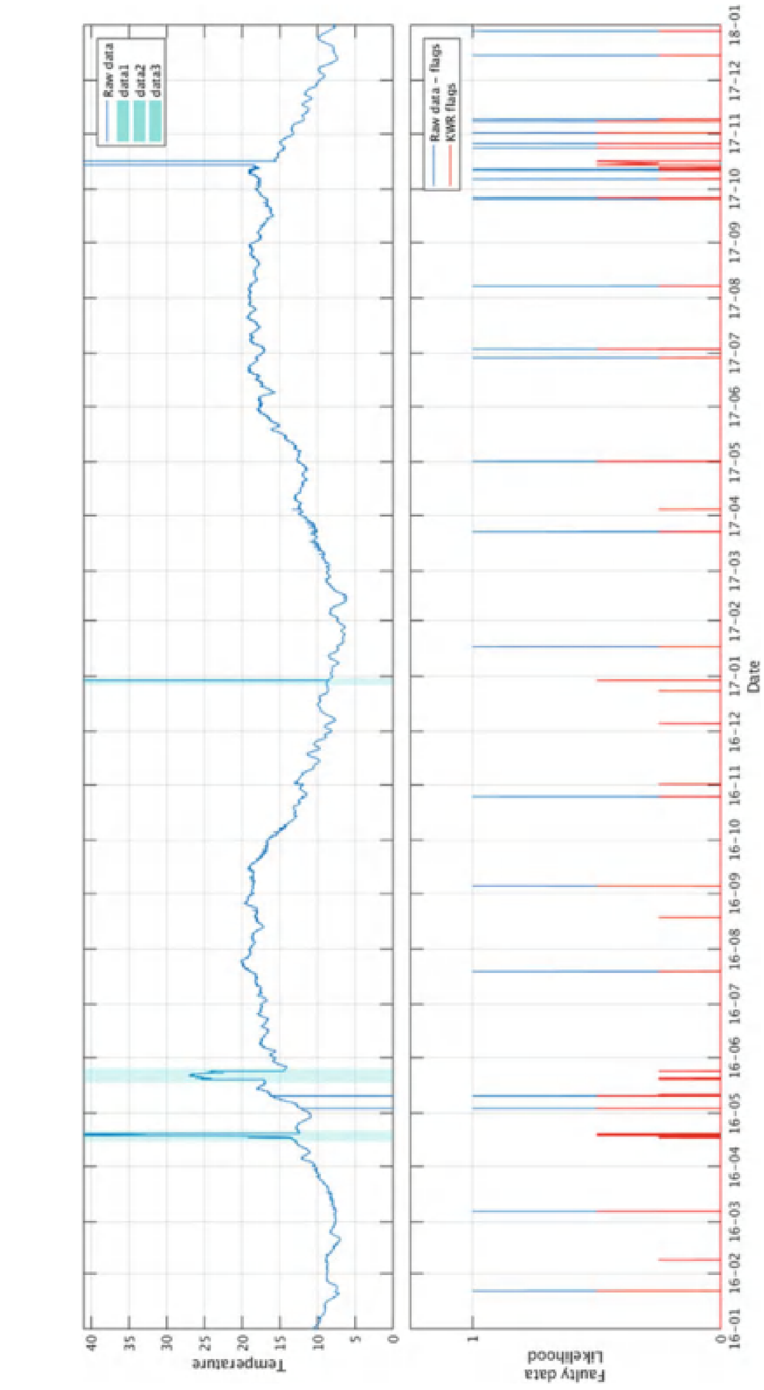


Fig. 8 Time series of temperature in WWTP I. Showing also system flags of company A and proposed DQC (KWR)

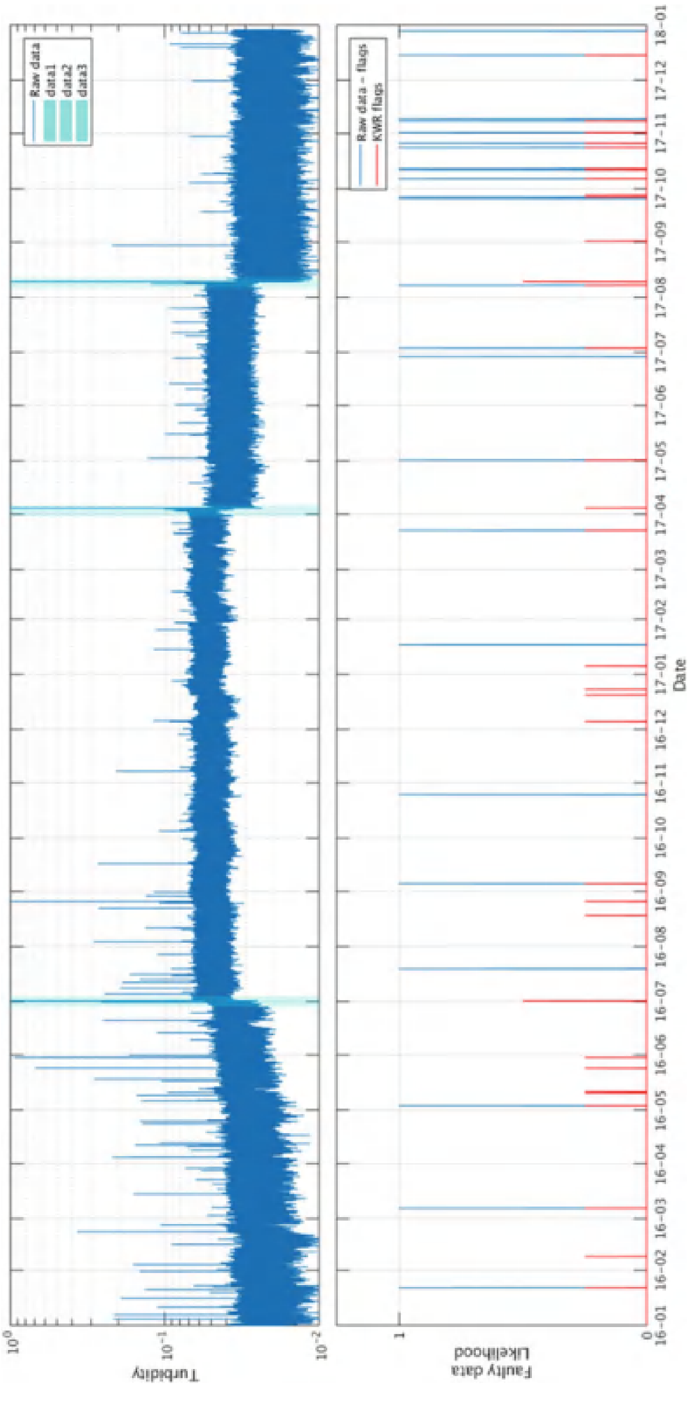


Fig. 9 Time series of turbidity in WWTP I. Showing flags obtained with current and proposed DQC (KWR)

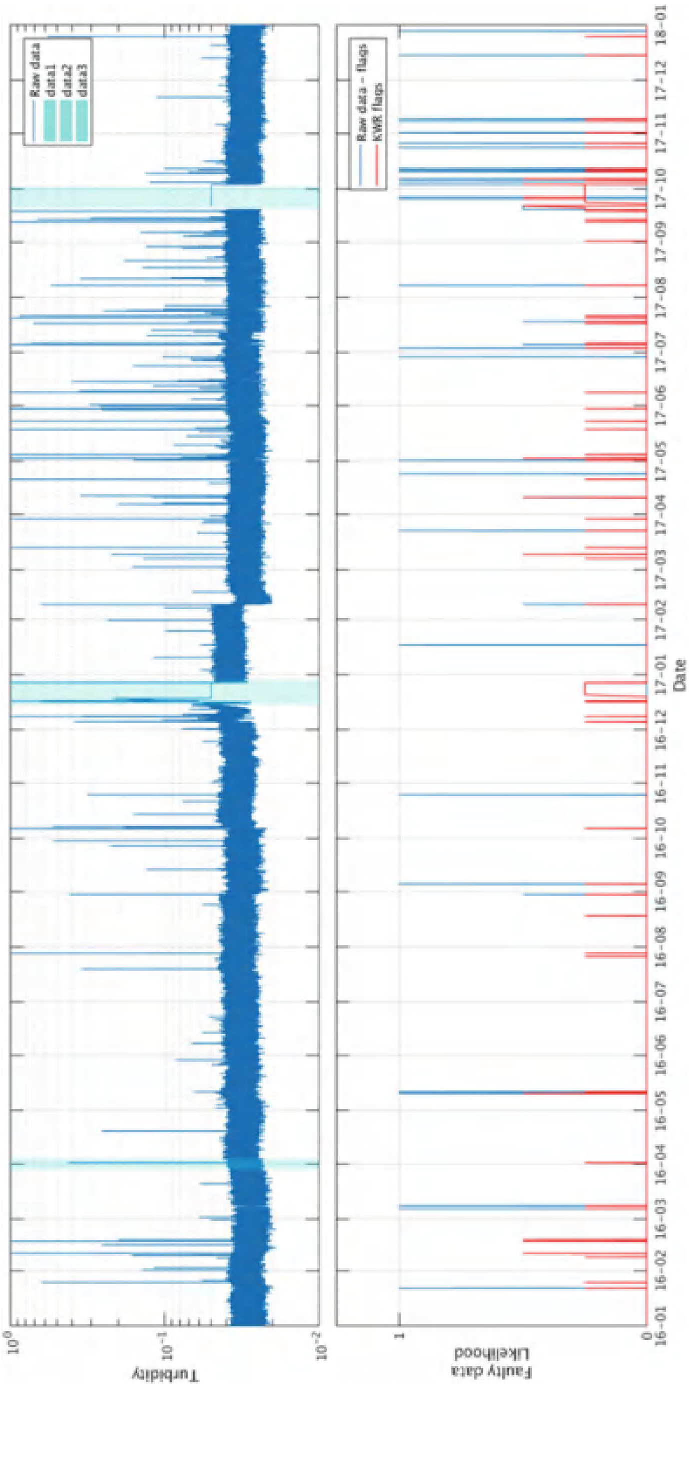


Fig. 10 Time series of turbidity in WWTP II. Showing also system flags and proposed DQC

Fig. 11 General location of pumping stations in city A (source: company A)

For each of them data of flow, pressure and energy use were available. Pressure is not presented here. Flows (m^3/h) have a time resolution of approximately 1 min, while energy use (kWh) contains data with a time resolution of approximately 15 min. The available data covers the period between 01 January 2016 and 31 December 2017. Data delivered by the utility contains the system′s flag for DQC. The flags are categorized as 0 and 1. A flag of 0 can be considered as Good data, while flags of 1 considered as Faulty, Dubious or out of range data.

The time series of the flows are presented in Fig. 12 and the energy use in Fig. 13 discriminated for each pumping station. In such figures, the flag status of the utility (Company A) is presented as red lines in the corresponding timestamp. If data is considered as valid or Good, the red line has a value of zero. The proposed data validation identified the timestamps at midnight as faulty data, but these have been removed.

For flow time series (see Fig. 12), the largest amount of flags is found at PS OSD (6,769 or 0.64% of the TS). The lowest number of flags is observed in the PS HLW (41 or 0.004%). In the case of PS HLM (see Fig. 12e), in the period between March 2016 and May 2016, there is no data of flows, and this drop is not identified by the system flags.

The total number of flags identified in the raw data for each pumping station and variable is presented in Table 5. It is of notice that no anomalies are identified by either the system of the water or the energy utility.

For energy time series, there are no flags identified by the system. The data is provided by a third party. Energy use at PS HLM (Fig. 13e) displays drops during extended period between February, March, April, and May 2016. After obtaining feedback from the utility, it was confirmed that this period corresponds to a maintenance of the pumping station. In addition for OSD PS (Fig. 13d), raw data contains a large number of pump switches (values = 0). This can indicate no registering of data or that indeed the pumps are shut off.

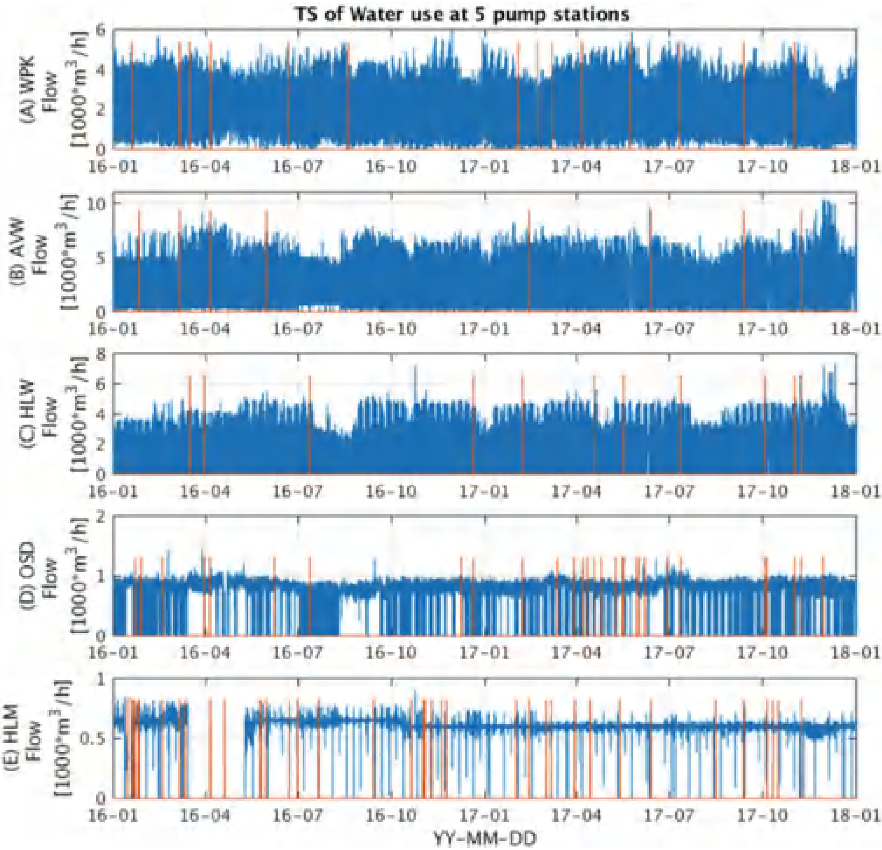


Fig. 12 Time series of flow at five pumping stations of company A. Faulty data reported by Company A displayed with red lines. Vertical axes are different to allow visibility of time series

4.4 Anomalies Within a Day

Subsequently, data from each time series has been processed to identify if there are particular periods, throughout a daily operation (in 24 h), when faulty data is more likely to occur. For this reason, data has been rearranged and categorized as hourly data, disregarding the dates. Such analysis is presented in Fig. 14 for flows and Fig. 15 for energy use. Due to these figures being a similar analysis of time series, red dots represent the same faulty data flags on current data validation by Company A (Table 5).

In the case of flows, most of the flags are present during the peak consumption hours. However, the faulty data detected does not correspond to high or low values either but to intermediate ones. In the case of PS WPK (see Fig. 14a), there are 3 values constantly picked by the system as faulty data near 4,000, 3,390 and 2,800 m^3/h. In the case of pumping station AVW (see Fig. 14b), there is a value

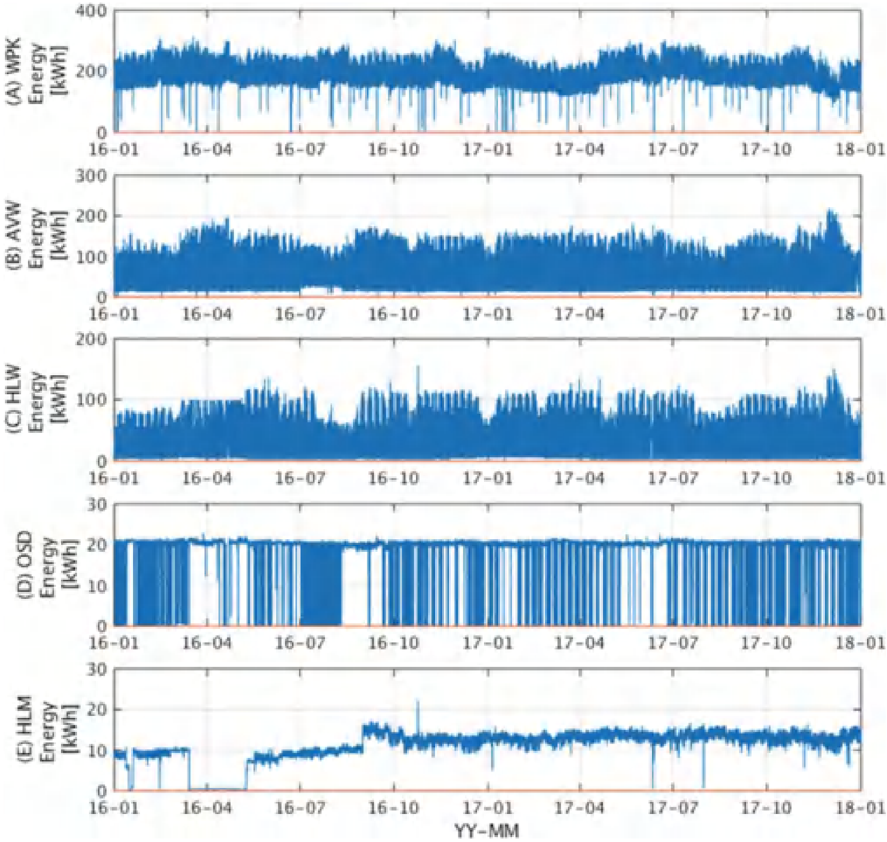


Fig. 13 Time series of energy use at five pumping stations of Company A

Table 5 Number of flags and percentage from total of timestamps from raw data provided by Company A

Pump station	Flow (m^3/h)	%	Energy* (kWh)	%
WPK	475	0.05	0	0.00
AVW	266	0.03	0	0.00
HLW	41	↓0.00	0	0.00
OSD	299	0.03	0	0.00
HLM	6,769	↑0.64	0	0.00

[a]Energy does not contain reported anomalies in the data. ↑ indicates highest percentage of anomalies. ↓ indicates lowest percentage of anomalies among pump stations

constantly picked by the system as faulty data near 4,510 m^3/h. In the case of pumping station OSD (see Fig. 14d), the predominant faulty detection value is 0 m^3/h between 9 am and 11 am. In the case of PS HLM (see Fig. 14e), there is a value constantly picked by the system as faulty data near 346 m^3/h (all day) and near 600 m^3/h (between 6:30 am and 13:30 am). Once again, this is consistent with a

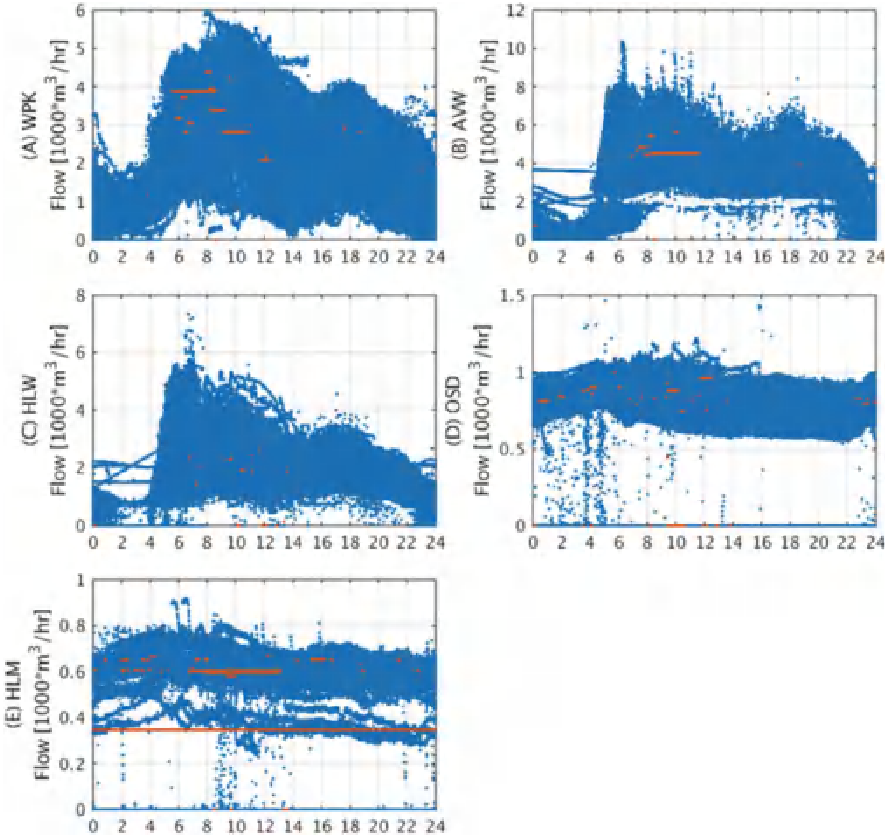


Fig. 14 Scatter of flows data at each pumping station during a day. Includes anomalies obtained by current data validation of Company A (red dots). Horizontal axis presents the hour of the day

disruption of the system. After a feedback session with the contact person from Company A, it was discovered that these specific values for which data is identified as faulty corresponds to the period when data is fetched to the data warehouse, and it is flagged as faulty by their system. The explanation provided is that because there is always a delay for the data transmission, triggering the flag in the system. This information became useful as it can be used to update their current DQC to take this into account.

The variation of energy consumption during the day shows that the three larger pumping stations (WPK, AVW and HLW) have a similar pattern to the one of flows. The highest range of variability of energy use during the day is present for WPK between 5:00 and 9:00 am (see Fig. 15a), midnight to 5:00 am for AVW and HLW (Fig. 15b, c). For OSD (Fig. 15d), there is a large variability of energy consumption from 11:00 pm and during the following 6 h of the day. In the case of HLM (Fig. 15e), the existence of gaps in the data (previously discussed) creates two

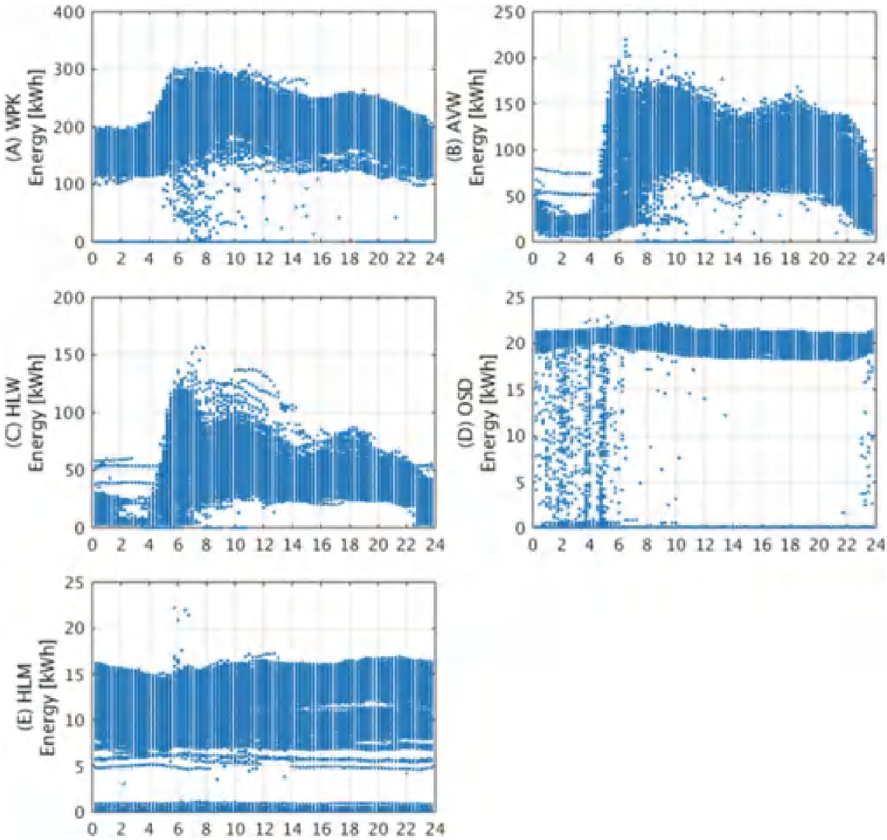


Fig. 15 Scatter of energy use at each pumping stations during a day. No anomalies reported by third party

different levels of energy consumption. A near-zero level (associated with maintenance periods) and what can be called a regular or central pattern.

4.4.1 Finding Faulty Data by Comparison. Water vs Energy

It is possible also to develop a comparative analysis of water production vs energy use. In general, one may expect to have a one-to-one relationship among both variables, and subsequently values which do not follow such behaviour can be associated as new faulty data.

For this, the data of each pumping station is aggregated for flows and energy to obtain the total. If a particular timestamp contains an anomaly on a pump station, this entire timestamp is flagged as dubious. Subsequently the data from total flows (every 1 min) has been aggregated to estimate the total flow delivered by Company A

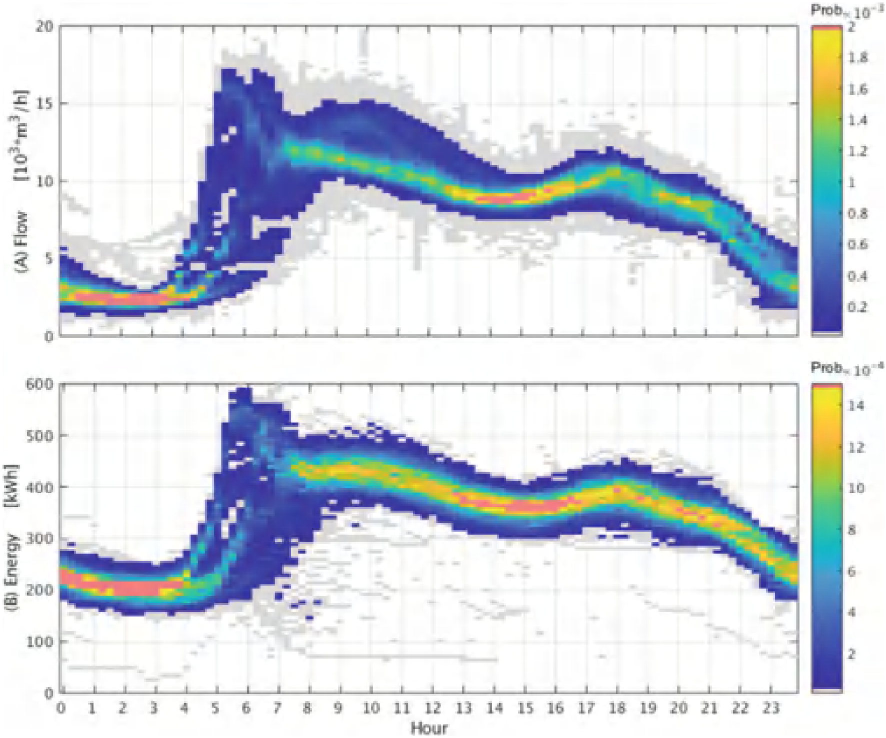


Fig. 16 Bivariate probability of (a) total demand consumption and (b) total energy use in city A. Anomalies have been removed. The colour represents the probability of total flow of City A at a certain 15 min interval during the day. Red values represent a higher probability, while grey values correspond to low probability

(every 15 min). The time series of accumulated flows itself will not provide additional information, as the anomalous timestamps are removed. For that reason, a daily frequency of the total flow of City A in time every 15 min (horizontal axis) and with a flow resolution of 250 m^3/h (vertical axis). The obtained result of the bivariate probability is presented in Fig. 16a. In fact, Fig. 16a is a representation of the flow pattern of the city for 2016.

Figure 16a shows a double peak consumption (as expected) during the early hours of the morning (05:00–09:00) and during the dinner time (17:00–19:00). On the other hand the Minimum Night Flow (MNF) occurs after midnight (01:00–04:00) with a high probability.

Another output which can be obtained from data validation perspective is the fact that some samples correspond to a large demand consumption around 10:00. Although their probability is low, it is evident that such events have occurred in City A during the last 2 years. In fact, the highest registered values for the entire system occur for this particular time of the day. In contrast, some samples show that there has been a lower demand consumption than the average trend at 11:30. There was no possible explanation identified by the utility for such atypical pattern variation.

Another possible use of this analysis for data validation is to perform a correlation analysis of flows with respect to the energy use obtained from the utility. In fact, as presented in Fig. 16b, the energy use pattern follows obviously a similar trend than that of the flows. As a matter of fact, this is not distant from the current operation of the system as a single District Metered Area (DMA), or fully interconnected WDN.

Similar to what was done in the case of flows of the five pumping stations. The obtained histogram of the bivariate probability of energy use in City A is presented in Fig. 16b. As it was the case of flows, there is a double peak in the energy consumption (this is expected) and a higher probability during the MNF.

From a data validation perspective, it is of particular interest that in this case a larger number of timestamps occur with low probability for energy consumptions under the average pattern (grey values), particularly during the hours of 06:00–10:00. However, given that there is no indication of status or flag data for any of these time series, it is not possible to conclude whether this is due to data anomalies or to a regular operation of the system. As a hypothesis, most of these low values below the trend of energy pattern can be attributed in part to the fact that there is a huge number of pump switches for all PS′ as it is presented in Fig. 15. Of notice is also that such pump switches identified in the energy consumption are not represented all the time in the flow data, mainly because pump switches occur at a 1 min resolution, while energy is a cumulative variable stored every 15 min.

From a feedback session with personnel from the utility, it was possible to determine that such behaviour of low energy registrations is possible. Sometimes during the mornings energy is self-produced by the utility from renewables, and the third party (energy company) is not aware of such energy influx. For that reason, there is a deviation between water and energy used for pumping which is not registered by the energy company. In this case, expert knowledge of the daily operations and workings of the utility became far more relevant; otherwise all such data would be flagged as faulty by a data validation system.

4.5 Best Practices and Issues in Data Validation Identified in the Case Studies

The following results are obtained from interviews (questionnaires, personal communications and feedback sessions). These are resented based on the analysis of four companies contributing data and not only based on the presented results of the previous sections. A summary of best practices and issues is identified (Table 6) and the issues identified during the pilot of this bird′s-eye view (Table 7).

A general issue that arises from the comparison of the four companies is the lack of standards. There are different types of registration protocols and standards used by

Table 6 Best practices identified during the interviews

Best practices	Company A	Company B	Company C	Company D
Data scientist works together with a domain expert				X
Clear responsibilities	x			X
Validation rules reported				X
Implementation of automatized routines which allows continuo validation of some datasets	x		x	X
Validation of aggregated data				X
Implementing pilot projects to learn from it		x		

Table 7 Issues found during this interviews

Issues	Company A	Company B	Company C	Company D
Data storage integration. Different databases or a single database. If multiple DB, then sometimes data is not linked one to one	x	x		
Lack of overview of metadata: Difficulties to track additional information, e.g. log books of maintenance, data is still in different databases stored	x	x	x	
Lack of priorities/time to (at least) tag the large number of signals and known events		x	x	
Problems related to the interfaces				X
Current approaches are often somewhat ad hoc		x	x	
A lot of techniques, but still data validation/correction is largely based on expert opinions	x			X
Own customize system, data models and tools (not compatible with other companies)	x		x	X
Only a small percentage of the data is validated	x	x	x	X
Vulnerability of failure of servers, data from third parties	x	x	x	
Black box in the built-in tools. Automatic filtering of suspicious data and not clear which rules they use to validate the data	x		x	

the companies. Additionally, there is a lack of knowledge about how data is validated by third parties, e.g. energy companies.

Among the utilities the purpose of validation is very diverse, for example, water flow, billing, determining the water balance, identifying leaks and changes in turbidity due to new filter installation. Depending on the objective the requirements of the validation change. The following issues have been identified that hinder the potential for implementation of more complex/advanced data validation techniques:

- Known deviations and operations are usually poorly logged by the utilities, or when this is done, it is very limited to some variables across utilities.
- Lack of specialized manpower to perform this task on a regular basis: a team consisting of both data scientists and hydraulic engineers is required.
- Specific techniques for data validation are still hard to be adopted because there is no overview regarding which data are needed for each of them.

Although only a small fraction of the data is validated, all of these data are potentially still used as input for different types of models, e.g. reliability analysis of predictions of systems performance under certain scenarios. However, this is dangerous as spurious data supplied to models will provide spurious results of model simulations.

Significant differences in the resolution of different parameters were also observed defined by the frequency with which they are stored in utility databases. For example, the year of installation of pipes usually has a resolution of 1 year. On the other hand, water quality data for turbidity can be stored with timestamps every 5 s in a database. However, each company has its own temporal resolutions for each variable.

There is still a lack of knowledge about how to first select the appropriate technique for validating a given dataset and secondly how to fine tune the parameters of each validation technique, to minimize the effort between identification of possible anomalies and extreme events in the system with high accuracy. All utilities face the same challenges when defining flags for data validation:

- Too many flags, (false-positives) operatives become reluctant to perform verification and data validation, and the trust on the data validation diminishes.
- Too few flags and robustness is lost, as water companies would not be able to differentiate between a regular event and extreme event and a real anomaly.
- Changes in the system is not always recorded or archived in the historic data. This is a challenge when the system configurations change dynamically (e.g. in the case of Company C).

5 Discussion and Recommendations for Future Work

5.1 Introduction

Data quality is a key consideration for the reliable functioning of drinking water systems, as data are used to monitor and operate systems, to bill customers, to report the performance of the company and to feed different types of models. Improving the quality of the data and making it more accessible will benefit every department of a company.

The following sections discuss a number of recommendations and future work in the field of DQC for drinking water utilities.

5.2 Recommendation Regarding Future Work

During the survey of the water companies, it has been evident that most utilities apply diverse methods of DQC. One of the features which is lacking across is a degree of standardization of DQC. A proposal for a Dutch initiative for standardization of DQC for drinking water utilities is suggested as a possible follow-up of this work. Such a task might require the development of a tool which could help operatives with the task of daily time series analysis, IVS, regression, interpolation, data smoothing, data aggregation and data correction.

5.2.1 Standardization

Defining when the quality of the data is good enough is case specific. Standardization can help to 1) understand common problems among utilities, 2) speak the same language so that similar issues can be addressed across utilities, 3) apply the same methods and 4) use the same tools. Challenges for which data standardization can provide a solution including:

- Allowing integration of data that come from different sources, origins and formats
- Automating data control and correction (support to automate processes) **–** less by hand and subjective handling of data
- Allowing faster and better analysis and understanding of processes (more objective, reproducible and comparable results)
- Improving and facilitating reporting and compliance
- Reducing cost (and time) by providing certainty of units, protocols, event types, etc.
- Allowing data sharing and implementation of hydroinformatics tools
- Facilitating interoperability of (IT) tools within a company and between companies
- Allowing comparison within departments or companies, e.g. benchmarking
- Enhancing transparency and clarity about what can and cannot be done with data

Water companies can use as a starting point, relevant experiences of other sectors. There are several standards which are relevant to the water sector, developed for instance within Internet of Things (IoT) initiatives (using smart appliances) [90, 91] or smart cities initiatives [92, 93]. Standardization within the water sector is a subject that the European Commission [14] is very interested to achieve and promote, and where different actions need to take place, but bottom-up action is also needed from utilities.

Data Model

It is recommended to adopt and adapt a framework where different dimensions and categories are clearly identified not only for the data content, data management, but also for diverse users considered. Currently, data from the utilities are highly variable in volume and resolution, format, metadata and shape. However, they all measure the same types of variables. It would be very helpful to have a standard data model for DQC control that can be agreed upon at a (inter)national level in due course. Given that the customers are relatively the same in the country, this implies that the water utilities can develop an intercompany data model for water accounting, with the possibility to extend it in the future as new variables are incorporated in their data warehouses.

Data Collection Redundancy

As presented in this chapter, the analysis of the data diet of large data collections may help utilities identify which sensors are more reliable and how much data storage is required. It is not yet clearly known by utilities which variables are correlated among different sensors. An effort should be made to develop an analysis of redundancy with the utilities using their data mining techniques discussed in this chapter.

Aggregation of Data

Aggregation of data is performed by all companies, for specific purposes. Additional data analysis which is of interest for all utilities can be introduced to obtain regular water balance calculations. All utilities have to account the water that is produced and billed. However, it was identified that the utilities have serious concerns about the limits of the anomalies in the water balance. Given that non-revenue water (NRW) is not a serious concern in the Netherlands, the main issue is to be able to identify when the water supply system presents a deviation from its regular pattern. As demonstrated in this chapter, for some utilities, the need to establish such boundaries for water balance is a current issue. Even with advanced tools for water accounting, there are deviations present in the data of water balance for all utilities. Therefore, it is recommended to implement advanced techniques such as model based validation to tackle this issue. For this, a pilot for a DMA configuration, with an optimal time step and additional info (e.g. pressure, flows) and expert knowledge (e.g. operators′ knowledge and logbooks of operations and maintenance), can be compared with a simulation model to determine anomalies.

Data Correction Techniques

Most methods applied for incomplete time series have been developed for surface and groundwater hydrology [39, 94, 95], and some applications have been made for WDN [61]. This chapter and its supporting research did not address the issue of data correction techniques. A proposal was made to continue the process of using hydroinformatics tools in the Dutch drinking water companies. The goal here would be identifying specific techniques for data correction which may be suitable for a subset of WDN variables. If for the case of data validation the spectrum of techniques was broad, a similar content is expected in the case of data correction techniques.

Data Reconciliation

The decision which must be made by utilities about which data to trust and which data not to trust is a constant struggle. Often, during the selection of data and case studies, a common concern was voiced by operatives of utilities about not trusting the data of a region or DMA. However, no quantification or metric was made available to express this in any of the cases and to our knowledge such metrics are not available in the Dutch context. This means that expert knowledge has been applied by utility′s experts, with prior experience in the management of their system. However, such expert knowledge is currently only encapsulated in the minds of experts. There are several methodologies available to transform such qualitative decision making into quantifiable rules for the determination of likelihood of data as being faulty. This can lead to improvements in data collection and reduce dependence of utilities on specific experts if they are not available.

5.2.2 Selection of Faulty Detection Techniques

Regarding DQC, there is no one technique fits all purposes. Depending on the monitored event/variable, different techniques with different parameters should be applied, also according the objective of the validation.

Software tools cannot validate data by themselves. Expert knowledge is always needed for a good determination of the parameters to perform the data validation, especially in complex systems, such as drinking water systems.

Given the variability of datasets, number of records, timestamps, time resolution and variables, only simple techniques for data validation have been applied for the water utilities data presented as case studies in this chapter. So it was recommended to the participating utilities that, in a future project, similar techniques to be specifically calibrated for subsets of data from the same variables. For example, in this chapter a short analysis of data validation for water balance is presented, but an extensive literature on the matter is also available. This means that the possibility to increase the identification of faulty data can be explored with many more techniques than the ones presented here on that subject. Once this is done, then a proper selection of best practice techniques for water balance can be established for the sector of Dutch drinking water companies.

5.2.3 Modelling for Anomaly Identification

In this manuscript, only a short portion of techniques for data validation was explored, and only simple tests were implemented. One of the biggest obstacles in applying data validation techniques corresponds to the use of modelling tools to identify anomalies (see again Fig. 5). Several methodologies are available which have not been implemented in this work, ranging from statistical, surrogate and physically based models with the aim to properly define when a sample of data corresponds to an anomaly or not. Having a calibrated model of a WDN presents the possibility to estimate deviations between measured data and simulated data and as such detect anomalies.

6 Conclusions

6.1 Specific Conclusions Based on the Cases Regarding Faulty Data Detection

Data validation applied to water quality demonstrated the validity of the utility′s flag system, as it was able to identify most anomalies. In general, data of Company A is of good quality with a low percentage of flags issued by the system as faulty data.

The implementation of a proposed data validation in this chapter using simple validation rules showed that the inclusion of a flat value test can improve the current data validation procedure of the drinking water companies for water quality data. Additionally for specific datasets such as turbidity time series, changes in the system can be identified by using jump detection (i.e. drifts or changes in average). Further development of the proposed data validation is the inclusion of more complex detection techniques.

From the validation of data for water balance, it can be concluded that the temporal data resolution is sufficient; however the process of data collection and aggregation is quite demanding. If a water balance needs to be performed at different intervals (e.g. 1 day, 1 week, 1 month, 1 year), the tasks of data validation would require extensive searches from diverse sources to confirm information from log-books, installation and maintenance records. This is highly time-consuming and it can be improved by automation by routines.

Expert knowledge from the utility helped to clarify the validity of large collections of data from the last 2 years, i.e. the pumping station maintenance in HLM. The integration of such expert knowledge in the system has yet to be implemented. This is evidenced in the fact that most queries of additional data validation were solved by internal communication ′via-via′ and not through a complete record of operations in any database.

6.2 General Conclusions Regarding Data Quality Control

Data quality control is a continuous process, instrumental in achieving large strategic objectives such as reliable and efficient drinking water system operations. Data validation for water companies is not an exclusive task of a data scientist. It is a collaboration between operators who understand the system and data analysts who must validate large proportions of data to improve models and, as a consequence, decisions. Currently only a small percentage of the data is validated, but we suggest that in the near future, there is a need to become even more active in data management at every level of a water company.

All involved water utilities implement individually data validation at different levels of complexity. Although water companies face similar issues, several customized tools/software are being developed per company. This is because each company has its own registration and database system, as well as different specifications regarding time steps, units, storage, etc. Working together on specific guidelines (standards) for the sector to define which datasets and methodologies are used for validation can facilitate and speed up implementation of DQC systems and be useful for potential future exchange of data, and, it may facilitate auditing operations as a nationwide goal.

From the analysed cases it was concluded that there is a need to exchange data and develop proper data models that consider not only the raw data but also metadata, formats and characteristics of the platform.

There are several techniques to validate the data. Regarding DQC, there is no such thing as one technique that fits all. Depending on the monitored event/variable, different techniques with different parameters should be applied also according the objective of the validation. Best practices and challenging issues have been identified (see again Sect. 4.3). To overcome the identified challenges a two-way implementation of DQC procedures is needed:

1. Strategic (top-down) by developing frameworks and standards for the water sector which are compatible with standards of other sectors
2. Operational (bottom-up) by implementing cases, evaluating case studies and sharing experiences across utilities.

   To progress both, it is envisioned that utilities can start with simple cases and techniques such as the ones presented in this chapter and steadily scale-up as the needs and goals of the utilities are met.

References

1. Bertrand-Krajewski J, Bardin J, Mourad M, Beranger Y (2003) Accounting for sensor calibration, data validation, measurement and sampling uncertainties in monitoring urban drainage systems. Water Sci Technol 47:95–102

2. Mourad M, Bertrand-Krajewski J (2002) A method for automatic validation of long time series of data in urban hydrology. Water Sci Tech 45:263–270
3. EC (2010) SeadataNet data quality control procedures
4. von Asmuth J (2012) Groundwater system identification through time series analysis
5. von Asmuth J (2015) Kwaliteitsborging grondwaterstands- en stijghoogtegegevens: protocol voor datakwaliteitscontrole (QC) KWR 2015.013. KWR Watecycle Research Insititute & TNO, Nieuwegein
6. von Asmuth J, van Geer F (2015) Kwaliteitsborging grondwaterstands- en stijghoogtegegevens: Systematiek en methodiek voor datakwaliteitscontrole (QC). KWR 2015.004. KWR Watercycle Research Institute, TNO, Nieuwegein, Utrecht
7. von Asmuth J, Maas K, Knotters M, Bierkens M, Bakker M, Olsthoorn T, Cirkel D, Leunk I, Schaars F, Asmuth D (2012) Menyanthes software for hydrogeologic time series analysis, interfacing data with physical insight. Environ Model Soft
8. Beuken R, Moerman A (2017) Uniforme storingsregistratie (USTORE). Praktijkcode voor het beheer van storingsregistratie van leidingnetten. KWR. PCD 9 2017, Nieuwegein
9. Paté-Cornell E (2012) On “black swans” and “perfect storms”: risk analysis and management when statistics are not enough. Risk Anal 32(11):1823–1833
10. Lynggaard-Jensen A, Hansen H and Bertrand-Krajewski J (2012) Real time integrated monitoring system supporting new data validation methods – Methodology guidelines and examples of application
11. Yoo C, Villez K, Lee I, Van Hulle S, Vanrolleghem P (2006) Sensor validation and reconciliation for a partial nitrification process. Water Sci Technol 53(4–5):513–521
12. Redman TC (1998) The impact of poor data quality on the typical enterprise. Commun ACM 41 (2):79–82
13. Alferes J, Tik S, Copp J, Vanrolleghem P (2013) Advanced monitoring of water systems using in situ measurement stations: data validation and fault detection. Water Sci Technol:1022–1030
14. Makropoulos C, van Thienen P, Agudelo-Vera C (2018) Towards a roadmap for hydroinformatics research BTO 2018.077
15. Tayi GK, Ballou DP (1998) Examining data quality. Commun ACM 41(2):54–57
16. Mutchek M, Williams E (2014) Moving towards sustainable and resilient smart water grids. Challenges 5(1):123–137
17. Mudumbe MJ, Abu-Mahfouz AM (2015) Smart water meter system for user-centric consumption measurement. In: 2015 IEEE 13th international conference on industrial informatics (INDIN)
18. Donhost MJ, Anfara VA (2010) Data-driven decision making. Middle Sch J 42(2):56–63
19. McAfee A, Brynjolfsson E, Davenport TH, Patil DJ, Barton D (2012) Big data: the management revolution. Harv Bus Rev 90(10):60–68
20. Kitchin R (2014) Big data, new epistemologies and paradigm shifts. Big Data Soc 1(1)
21. Ackoff RL (1989) From data to wisdom. J Appl Syst Anal 15:3–9
22. Deming WE (1986) Out of the crisis, vol 88. MIT Press, Cambridge
23. Hargesheimer E, Conio O, Popovicova J (2002) Online monitoring for drinking water utilities
24. Ishikawa K (1986) Guide to quality control. Asian Productivity Organization
25. Shewhart WA (1931) Economic control of quality of manufactured product. Martino Fine Books
26. English LP (2001) Information quality management: the next frontier. In ASQ world conference on quality and improvement proceedings. American Society for Quality
27. Stausberg J, Nonnemacher M, Weiland D, Antony G, Neuhäuser M (2006) Management of data quality – development of a computer-mediated guideline. Stud Health Technol Inform 124:477–482
28. Sobhani-Tehrani E, Khashayar K (2009) Fault detection and diagnosis. In: Fault diagnosis of nonlinear systems using a hybrid approach. Springer, Berlin

29. M. Di Zio, N. Fursova, T. Gelsema, S. Gießing, U. Guarnera, J. Petrauskienė, L. Quensel-von Kalben, M. Scanu, K. O. Bosch, M. Loo and K. Walsdorfer, "Methodology for data validation 1.0," 2016
30. Venkatasubramanian V, Rengaswamy R, Kavuri S, Yin K (2003) A review of process fault detection and diagnosis, part III: process history based methods. Comput Chem Eng 27:327–346
31. NITS (1996) Federal standard 1037C data stream. https://www.its.bldrdoc.gov/fs-1037/dir-010/_1451.htm. Accessed 20 Oct 2018
32. B. Gaag and J. Volz, "Real-time on-line monitoring of contaminants in water. Developing a research strategy from utility experiences and needs.," 2008
33. S. Sun, J. Bertrand-krajewski, A. Lynggaard-Jensen, J. Broeke, F. Edthofer, M. Céu Almeida, M. Silva Ribeiro and J. Menaia, "Literature review of data validation methods," 2011
34. EPA (2006) Data quality assessment, a reviewers guide. EPA QA/G-9R, US-Environmental Protection Agency, Washington
35. Venkatasubramanian V, Rengaswamy R, Yin K, Kavuri S (2003) A review of process fault detection and diagnosis, part I: quantitative methods. Comput Chem Eng 27:293–311
36. Branisqvljevic N, Kapelan Z, Prodanovic D (2011) Improved real-time data anomaly detection using context classification. Hydroinformatics
37. Waal T (2013) Selective editing: a quest for efficiency and data quality. J Off Stat 29:473–488
38. Wilson PW (1993) Detecting outliers in deterministic nonparametric frontier models with multiple outputs. J Bus Econ Stat 11:319–323
39. von Asmuth J (2011) Over de kwaliteit, frequentie en validatie van druksensorreeksen. KWR 2010.001. KWR Watercycle Research Institute, Nieuwegein
40. McKenna H, Klise K, Cruz V, Wilson M (2007) Event detection from water quality time series. In: Proceedings of world environmental and water resources congress, ASCE, Reston
41. Mounce S, Mounce R, Jackson T, Austin J, Boxall J (2014) Pattern matching and associative artificial neural networks for water distribution system time series data analysis. J Hydroinf 16:617–632
42. van Thienen P, Pieterse-Quirijnse I, Kater H, Duifhuizen J (2012) Nieuwe lekverliesbepalingsmethoden voor het drinkwaterdistributienet. H2O
43. Bakker M (2004) Optimised control and pipe burst detection by water demand forecastingOptimised control and pipe burst detection by water demand forecasting
44. Thienen PV, Vertommen I (2015) Automated feature recognition in CFPD analyses of DMA or supply area flow data. J Hydroinf 18(3):514–530
45. Clarke R (2013) Calculating uncertainty in regional estimates of trend in streamflow with both serial and spatial correlations. Water Resour Res 49:7120–7125
46. Hirsch RM, Moyer DL, Archfield SA (2010) Weighted regressions on time, discharge, and season (WRTDS), with an application to Chesapeake bay river inputs. J Am Water Resour Assoc 46:857–880
47. Furnival GM (1971) All possible regressions with less computation. Technometrics 13:403–408
48. Hocking R, Leslie N (1967) Selection of the best subset in regression analysis. Technometrics 9:531
49. Schatzoff M, Fienberg S, Tsao R (1968) Efficient calculations of all possible regressions. Technometrics 10:768
50. Lomb NR (1976) Least-squares frequency analysis of unequally spaced data. Astrophys Space Sci 39:447–462
51. Castelletti A, Galelli S, Ratto M, Soncini-Sessa R, Young PC (2012) A general framework for dynamic emulation modelling in environmental problems. Environ Model Soft 34:5–18
52. Galelli S, Humphrey G, Maier H, Castelletti A, Dandy G, Gibbs M (2014) An evaluation framework for input variable selection algorithms for environmental data-driven models. Environ Model Softw:33–51

53. Castro-Gama M, Popescu I, Li S, Mynett A, van Dam A (2014) Flood inference simulation using surrogate modelling for the Yellow River multiple reservoir system. Environ Model Softw:250–265
54. G. Box, G. Jenkins and G. Reinsel, Time series analysis: forecasting and control, 2008
55. Fatichi S, Barbosa S, Caporali E, Silva M (2009) Deterministic versus stochastic trends: detection and challenges. J Geophys Res Atmos 114
56. Beran J (2010) Long-range dependence. Wiley Interdisciplinary Rev Comput Stat 2:26–35
57. Beran J, Feng Y, Ghosh S, Kulik R (2013) Long-memory processes: probabilistic properties and statistical methods, pp 1–884
58. Ehsanzadeh E, Adamowski K (2010) Trends in timing of low stream flows in Canada: impact of autocorrelation and long-term persistence. Hydrol Process 24:970–980
59. Lennartz S, Bunde A (2009) Trend evaluation in records with long-term memory: application to global warming. Geophys Res Lett 36
60. Rea W, Oxley L, Reale M, Brown J (2009) Estimators for long range dependence: an empirical study. Electron J Stat
61. Oriani F, Borghi A, Straubhaar J, Mariethoz G, Renard P (2016) Missing data simulation inside flow rate time-series using multiple-point statistics. Environ Model Softw:264–276
62. Montanari A, Rosso R, Taqqu MS (2000) A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan. Water Resour Res 36:1249–1259
63. Montanari A, Taqqu MS, Teverovsky V (1999) Estimating long-range dependence in the presence of periodicity: an empirical study. Math Comput Model 29:217–228
64. Hirsch R, Alexander R, Smith R (1991) Selection of methods for the detection and estimation of trends in water quality. Water Resour Res 27:803–813
65. Darken PF, Zipper CE, Holtzman GI, Smith EP (2002) Serial correlation in water quality variables: estimation and implications for trend analysis. Water Resour Res 38:221–227
66. Palau C, Arregui F, Carlos M (2012) Burst detection in water networks using principal component analysis. J Water Resour Plan Manag 138(1):47–54
67. Vries D, van der Akker B, Vonk E, de Jong W, van Summeren J (2016) Application of machine learning techniques to predict anomalies in water supply networks. Water Sci Technol Water Suppl 16(6):1528–1535
68. Hill DJ, Minsker BS (2010) Anomaly detection in streaming environmental sensor data: a data-driven modeling approach. Environ Model Softw 25:1014–1022
69. P. Thienen, H.-J. Alphen, A. Brunner, Y. Fujita, B. Hillebrand, R. Sjerps, J. Summeren, A. Verschoor and B. Wullings, "Explorations in data mining for the water sector BTO 2018.085," 2018
70. Quinlan JR (1992) Learning with continuous classes. In: Proc. 5th Australian joint conf. on artificial intelligence, world scientific, Singapore
71. Allen M, Preis A, Iqbal M, Srirangarajan S, Lim H, Girod L, Whittle A (2011) The application of real-time in-network monitoring of the water distribution system to improve operational efficiency. J Am Water Works Assoc (JAWWA) 103:63–75
72. Mashford J, Silva DD, Marney D, Burn S (2009) An approach to leak detection in pipe networks using analysis of monitored pressure values by support vector machine. In: Proceedings of the 3rd international conference on network and system security, pp 1–6
73. Mounce S, Mounce R, Boxall J (2011) Novelty detection for time series data analysis in water distribution systems using support vector machines. J Hydroinf:672–686
74. Candelieri A, Soldi D, Archetti F (2014) Short-term forecasting of hourly water consumption by using automatic metering readers data. Proc Eng:844–853
75. Aksela K, Aksela M, Vahala R (2009) Leakage detection in a real distribution network using a self-organizing maps. Urban Water:279–289
76. Preis A, Whittle A, Ostfield A, Perelman L (2011) Efficient hydraulic state estimation technique using reduced models of urban water networks. J Water Resour Plan Manag 137:343–351
77. Perelman L, Ostfeld A (2011) Water distribution systems simplifications through clustering. J Water Resour Plan Manag 138(6):218–229

78. Mesman G, van Thienen P (2015) Lekzoeken met hydraulische modellen. BTO 2015.064. KWR Watercycle Research Institute, Nieuwegein
79. Martínez-Solano F, Iglesias-Rey P, Mora-Meliá D, Fuertes-Miquel V (2017) Exact skeletonization method in water distribution systems for hydraulic and quality models. Proc Eng 186:286–293
80. Ulanicki B, Zehnpfund A, Martinez F (1996) Simplification of water distribution network models. In: Proc., 2nd Int. Conf. on Hydroinformatics, Zurich, Switzerland
81. Anderson E, Al-Jamal K (1995) Hydraulic-network simplification. J Water Resour Plan Manag 121:235–240
82. Giustolisi O, Laucelli D, Berardi L, Savic D (2012) Computationally efficient modeling method for large water network analysis. J Hydraul Eng
83. Tsoukalas I, Kossieris P, Efstratiadis A, Makropoulos C (2016) Surrogate-enhanced evolutionary annealing simplex algorithm for effective and efficient optimization of water resources problems on a budget. Environ Model Softw 77:122–142
84. Bragalli C, Fortini M, Todini E (2016) Enhancing knowledge in water distribution networks via data assimilation. Water Resour Manag 186:3689–3706
85. Bragalli C, Fortini M, Todini E (2017) Data assimilation in water distribution systems. Proc Eng 186:506–513
86. Aisopou A, Stoianov I, Graham N (2012) In-pipe water quality monitoring in water supply systems under steady and unsteady state flow conditions: a quantitative assessment. Water Res 46:235–246
87. Vreeburg J, Vloerbergh I, Van Thienen P, De Bont R (2013) Shared failure data for strategic asset management. Water Sci Technol Water Suppl 13(4):1154–1160
88. Kwakkel M, Vloerbergh I, van Thienen P, Beuken R, Wols B, van Daal K (2015) 2015, uniform failure registration: from data to knowledge. Water Asset Manag Int 10(4):18–22
89. Moerman A, Beuken R, Wols B (2017) Review on the development of uniform failure registration (USTORE) in the netherlands. In: LESAM conference, Trondheim
90. EU (2019) AI & IoT standards. https://ec.europa.eu/knowledge4policy/foresight/topic/accelerating-technological-change-hyperconnectivity/online-resource/iot-standards_en. Geopend 31 Jan 2020
91. EU 2019 WiFi4EU portal. https://wifi4eu.ec.europa.eu/#/home. Geopend 31 Jan 2020
92. EU (2019). ec.europa.eu. https://ec.europa.eu/digital-single-market/en/internet-of-things. Geopend 30 Jan 2020
93. Makropoulos C, Savić D (2019) Urban hydroinformatics: past, present and future. Water 11 (10):1959
94. Veling E (2010) Approximations of impulse response curves based on the generalized moving Gaussian distribution function. Adv in Water Resour 33(5):546–561. https://doi.org/10.1016/j.advwatres.2010.02.009
95. von Asmuth J, van Geer F (2013) Kwaliteitsborging grondwaterstands- en stijghoogtegegevens: op weg naar een landelijke standaard. KWR 2013.027," KWR Watercycle Research Institute & TNO, Nieuwegein, Utrecht

Monitoring and Controlling a Smarter Wastewater Treatment System: A UK Perspective

Oliver Grievson

Contents

Abstract Monitoring wastewater has always been a challenge. In the wastewater collection network, historically, there has been a lack of permanent wastewater monitoring because of the propensity for fouling and the complications of monitoring both gravity and pressurised networks. In engineering and operational terms, the wastewater network has also been treated as a separate entity to the wastewater treatment works which is, in reality, part of the same system. The wastewater treatment works tends to be much better monitored depending upon the size of the works. However this monitoring has been very much based upon single system instrument-based control systems (e.g. a dissolved oxygen control system for an activated sludge plant) rather than a more holistic system approach balancing the different systems present on a typical treatment works.

O. Grievson (*)
Z-Tech Control Systems, Cambridge, UK
e-mail: oliver.grievson@z-tech.co.uk

Andrea Scozzari, Steve Mounce, Dawei Han, Francesco Soldovieri, and Dimitri Solomatine (eds.), ICT for Smart Water Systems: Measurements and Data Science, Hdb Env Chem (2021) 102: 111–140, https://doi.org/10.1007/698_2019_418,
© Springer Nature Switzerland AG 2019, Published online: 26 November 2019

The future of both the wastewater network and the wastewater treatment works is a much more holistic approach bringing the network and the treatment works together and treating it as a single system. In this way, rather than operating the wastewater treatment system for process control with the aim of protecting the water environment, it can also be operated for resource recovery and energy efficiency with a much wider environmental benefit.

Keywords Biosolid, Burst detection, Mogden formula, Multivariate process control, Respirometry, Septicity, Thermal hydrolysis, Urban drainage, Wastewater treatment

1 A Systematic Approach

Wastewater systems can vary in both size and complexity ranging from small and simple rural catchments to large and complex urban conurbations. In dry weather, the wastewater collection network receives both domestic wastewater from the customer and a variety of industrial inputs which can be very mixed in nature. The network, in the UK alone, is comprised of a variety of sizes and is hundreds of thousands of miles long, meaning that the residence time in the sewer can vary in time, depending on the size of the network, from a few hours to a few days. Longer residence times can create septic conditions and build-up of debris within the sewer (or within sewage pumping stations), depending upon whether the part of the sewer is under gravity or part of the pressurised sewer network.

In dry weather there is normally a single output to each wastewater system which is at the wastewater treatment works (WwTW), barring any problems within the network which create unplanned discharges. In wet weather this system changes and there are multiple outputs to the system (for combined systems such as those that operate in the UK). As flow inputs increase through rainfall runoff from road drains, the levels of flow in the sewer increase. Designed within the sewer are “relief valves” in the form of combined sewer (or storm) overflows (CSOs), which provide relief from the system when the flows are, in principle, dilute enough to have a minimal impact on the water environment. The only alternative, which can happen when the sewer is misused, is that the system floods up and into either external public areas or into the customer′s home.

At the sewage treatment works, the focus has always been to treat the received flows as efficiently as possible as they are received at works. This has meant designing works so that they can take flows of up to 300% (the flow to full treatment is often approximately three times the dry weather flow) and an increase of pollutant load of up to 40% more. Hence the works are often designed for just treating wastewater.

This philosophy is changing, and a more systematic approach is being taken. At the WTW this is designed based upon the philosophy of the resource factory and treating the outputs from the WwTW as a product. In Fig. 1 we see the wastewater treatment system in its entirety.

Looking at it from a systematic approach, and from a monitoring and control point of view, the inputs into the system are not currently monitored due to the difficulty in practically monitoring the inputs into the system. In the main, the wastewater discharged is a calculated factor based upon the proportion of water used if the customer is metered. If not, it is an estimated use depending upon the per capita consumption. As the use of universal smart metering proliferates, then this situation will improve (and is less of a problem in areas of the world with a much higher roll out of smart metering). Inputs into the system from industrial customers are monitored as they form the fundamental basis of the amount the industrial customer is charged. The cost is based upon both the flow and concentration that is being discharged utilising the Mogden formula. The last input into the system is rainfall and runoff, which by its very nature cannot be measured but can be monitored and predicted. This is where developments in both sensing technology and modelling approaches can be used to predict what the inputs into the system are [1].

There is one other “input” into the system that is even more difficult to monitor (i.e. almost impossible to directly monitor) and that is infiltration into the gravity system. As the wastewater collection network deteriorates, cracks or weaknesses around joins in the pipes or manholes develop, and these allow infiltration into the network. At its most serious, infiltration has been known to make up more flow into the system than all of the other inputs even within dry weather flow, as underlying soil conditions mean that, in wet weather, the soil fills and can take several months to release the stored water into the sewer environment.

All of this has to be taken into account when looking to monitor and control the wastewater system, and looking at it as a whole, it is important to look at the strategy of how the wastewater system needs to be monitored and controlled. There is underlying philosophy to a smart wastewater system and that is:

> To manage the system by monitoring and control where there is a defined need, and to have enough information to manage the system where there is not a defined need.

The ultimate aim for the wastewater system is for it to manage all of the flows that should enter the network whilst detecting flows that should not be entering and to convey them as optimally as possible to the WwTW in a balanced fashion. This allows stability within, what in essence will become, the effluent factory producing products such as water, nutrients (including biosolids) and energy.

We can achieve this by looking at the system holistically. In the rest of this chapter, we will look at the different elements of the system as a whole and look at the philosophy of operation that a smart system would put in place and the measurement and control needs.

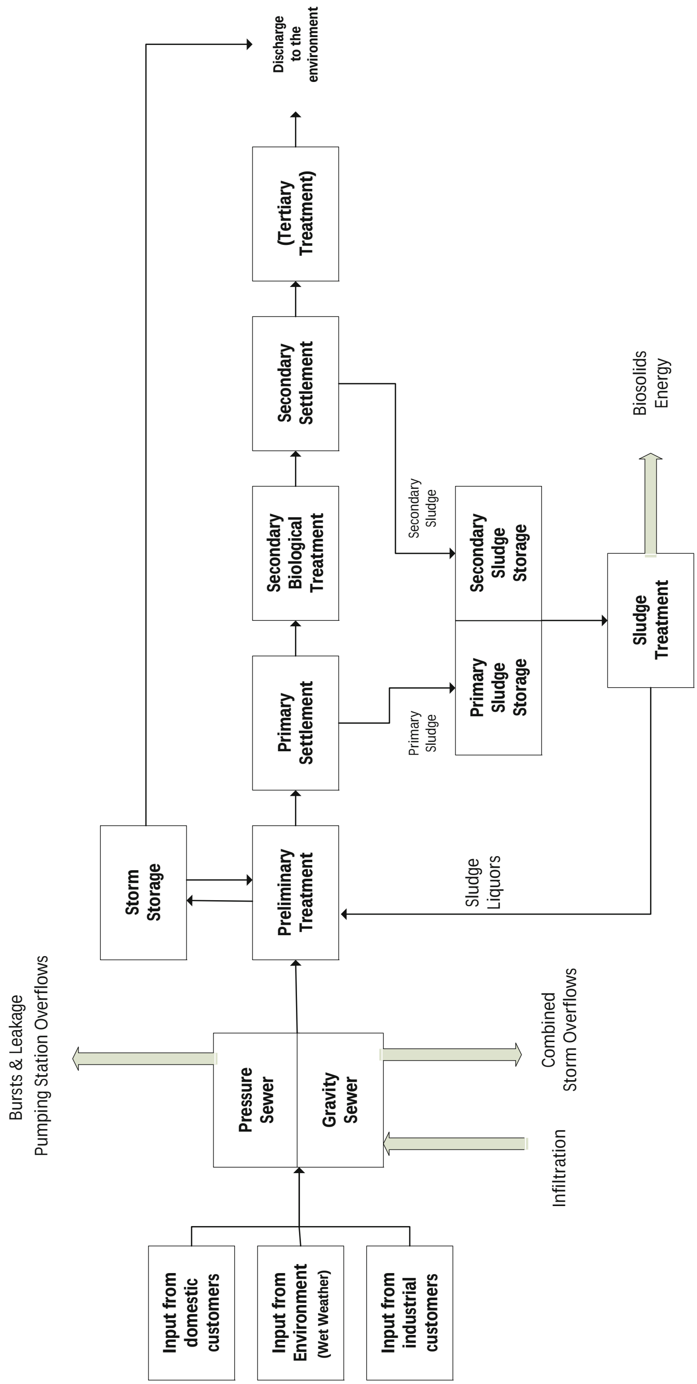


Fig. 1 Typical wastewater system

2 Smart Wastewater Network

The smart wastewater network has been around as a concept since the early 1970s when one was installed in Minnesota in the USA, and there have been a number of notable examples of smart wastewater networks in other countries [2]. In the Minnesota case study, the main aim of the project was to separate out the combined system. As this was very early days in both instrumentation and automation, the system was limited but worked well. There are numerous examples from Europe too, including the installation of a smart wastewater network in Barcelona for the 1992 Olympics due to flooding and pollution issues. The initial solution was a €150 million interceptor sewer under the centre of the city. The smart wastewater network solution, including three new storm water detention tanks, reduced the capital cost of the work needed to approximately 1/3 of the initial solution.

The aims of the wastewater network in its purest terms are to collect wastewater and to transfer it to the WTW. Where there is a combined system with combined storm overflows, there is a risk to the environment. As the overall environmental aim is to reduce the pollutant load, then controlling the wastewater network whilst maximising the throughput and minimising the losses (whilst also protecting the customer) is the aim of the smart network element of the wastewater system [2].

Several other smart wastewater network solutions have been built in cities around the world including five in Paris, some of which have been operating for over 25 years, and one in Tokyo in Japan. The first smart wastewater network to be built in the UK in the Eastney Catchment of Southern Water was largely based upon modelling of the network with Innovyze′s ICMLive modelling programme with various inputs from the network [3].

To summarise, what is the initial aim of the smart wastewater network, and what can we learn from the examples that have been installed around the world? The main aims are:

- To protect the customer from flooding from the sewer
- To detect blockages within the network from sewer misuse
- To prevent, where possible, the use of combined storm overflows especially in dry weather and funnel as much wastewater to the WwTW as possible
- To facilitate the efficient operation of the entire wastewater system (including both the wastewater collection system and the WwTW)

2.1 Philosophy of Operation

So, in order to achieve a smart wastewater network, what is the philosophy of operation? In the wastewater network, the philosophy of operation will actually change depending upon the underlying climatic condition. In dry weather the main aim, in the gravity sewer, is to manage the flows so that septicity is minimised by managing the throughput of the sewer and ensuring that the detention time in the sewer is minimised. However, periodic flushing of the sewer in dry weather is also beneficial in order to minimise the accumulation of debris within the sewer, whether this can be done in automated control or by manual cleaning of the sewer by jetting.

This is in the operation of the sewer in general, but, in dry weather, there is further work that can be done. This involves directly detecting debris accumulation or, by using analysis, detecting and locating areas that are either blocked or partially blocked that will potentially cause problems when flows increase and possibly causing overflowing of the network system. The last element of general operation is detecting where there is infiltration into the sewer. Infiltration simply adds to the basal flow of the sewer, taking up capacity within the network which will be required in storm conditions.

The gravity sewer provides the base flow which is transferred to the WTW in dry weather. Without installing gates within the sewer (which are popular in some European sewers), there is no method of controlling the gravity sewer. It is on the pumped sewer, where pumping stations can cause large variations in flow due to the size of the wet well and the impact that the pumping system can have on the treatment works. This is particularly the case where pump sizes have been increased to resolve local flooding issues that have, in turn, caused flooding issues on the treatment works. This is where the pumping system is larger than the capacity of the inlet works which as the pumps are oversized causes flooding of the works even in dry weather. Where there are combined gravity and pumped systems, the situation is further complicated. However, at their simplest, gravity systems provide the basal flow, and the pumped systems have the potential to cause problems with overflows to the environment.

It is when the sewer enters storm conditions that problems in operation can start. The flow from the gravity network cannot be controlled. The customer input, whether it is domestic or industrial, will remain the same. The runoff from the road network will increase, and so will any debris from the road surfaces thereby increasing the inorganic pollutant load in the sewer. The gravity network is largely unmonitored as it cannot be controlled. The pumped network is monitored to a certain extent, and it is here that smart systems can have the greatest potential and impact.

The philosophy of operation in the pumped wastewater network is to detect where the underlying flow condition within the sewer is going to increase and to take steps to increase the capacity of the network by pumping down the sewer to as empty as possible thus freeing up capacity to manage storm conditions. This is shown in Fig. 2.

Measuring level within the gravity sewer system, either measuring level in pumping stations or flow on the pressurised system, can enable a large element of control of the wastewater network. Where programmes, such as the Duration Monitoring (EDM) programme in the UK, which measures spills to the environment, exist, they can be used to warn of levels high enough so that wastewater spills to the environment. Thousands of CSO monitoring installations have been implemented with many more planned, involving data logging and level measurement to record spill events. Combined with rainfall radar data, online data analysis

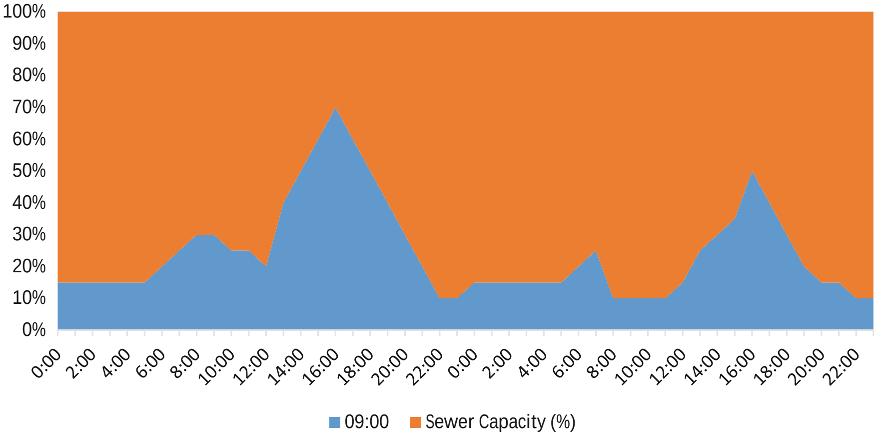


Fig. 2 Hypothetical control of the sewer levels storm without control (first 24 h) and with control (second 24 h)

systems have been proposed and implemented to perform new real-time event detection (e.g. [4]).

Add to this the potential of customer alarms at vulnerable customer spots on the network, then we can begin to consider the appearance of intelligence within the wastewater network through improved monitoring and analysis.

If we accept there is no control on the gravity system, it is by actively managing the pumped system that we can control the wastewater network by smoothing flows in periods of dry weather using the capacity of the wastewater collection network. In wet weather, especially if there is active monitoring of the meteorological conditions, then the sewer can be pumped down in advance to create capacity within the wastewater network. This is relatively simple within the pumped system, but complications occur when used within the gravity network. Active system control of the gravity can be used by installing gate systems to control the flow and use the gravity network as a flow detention system [5].

2.2 Monitoring the Inputs into the System

When you break the smart wastewater network down to its constituent parts, and look at the inputs into the system and how to measure them, there are some distinct challenges as well as technology that already exists.

Looking at the three inputs into the sewer as shown in Fig. 1, we have:

- Input from the domestic customer
- Input from the industrial customer
- Input from the environment

Looking at monitoring the domestic customer comes down to whether or not there is value in doing so or whether using an empirical (estimated) value which has been traditional in the water industry is sufficient for purpose. Traditionally, the amount of wastewater going into the wastewater collection network has been an estimate of an estimate, insofar as the amount of water supplied into the potable water network divided by an estimate of the number of customers (2.7 people per property on average) multiplied by 90% to take into account the amount of drinking water that never reaches the sewer (because of drinking water, water use on gardens, etc.). This has led to a per capita consumption (PCC) figure which, in the UK at least, has historically been 150 L per person per day.

This PCC figure has historically been used as the fundamental basis of design for the wastewater system. The supporting mechanical water meters were of little use however, because the meter readings were only taken every 6 months. With the advent of smart water meters, the situation has changed, and meter readings can now can be taken on a regular basis, e.g. hourly, thereby providing an appropriate balance.

This allows for a better visualisation of when water is consumed and, theoretically at least, could be used to imply impacts on the wastewater system. The amount of water metering across water companies vary greatly, and ranges from 40% metering in some places [6] to 100% (dumb) metering in others. This is set to increase drastically over the next few years, with most UK companies planning a significant proportion of smart water metering of customers.

Translating this into the volumes of wastewater that are produced is difficult in itself, and, typically, an industry standard of 90% of the potable water use has been used as a standard for charging the customer. The accuracy of this does depend upon the socio-economic category of the customer with more affluent people actually discharging less to the sewer because of potable water use elsewhere (e.g. watering gardens or topping up swimming pools) [7]. This shows that the estimations of customer discharges to the sewage collection network are poor at best.

There are domestic wastewater meters on the market; however their uptake is relatively low, not due to the cost of the meter itself, but more to do with the cost of installation which can vastly outweigh the purchase cost itself. In short, it is perceived that the cost of monitoring outweighs the benefit. This is a typical situation within the smart water industry where the cost-benefit is not truly known, and hence it is difficult to build business cases [8]. Technologies that are used for potable water metering are not appropriate as these technologies require a full pipe. The only technologies that are available would cost £200–300 installed (at minimum) which is approximately ten times the cost of a customer potable water meter. Over the life of a typical water meter which is normally 10 years, this would add £20–30 to a customer′s bill for seemingly very little benefit apart from control of the network.

This is not the case for an industrial customer. For trade effluent the charging structure, in the UK, is based upon the Mogden formula, which includes not only the volume but the strength of the wastewater that is discharged to the sewerage network [1]. As the cost is based upon the volume of the wastewater, there is an economic value to measuring the flow as accurately as possible, as any uncertainty in the measurement could mean that the cost to the customer is higher than necessary (or if the instrument is measuring low the cost could actually be reduced but normally instruments, if inaccurate, will read higher than actual).

All of this comes together in the normal definition of flows that pass to the WTW in dry weather or the DWF. This can defined as:

$$\text{DWF} = \mathsf{PG} + \mathsf{I} + \mathsf{E}$$

where:

DWF	Dry Weather Flow in m^3/day
P	Population served by the WTW (number of people)
G	the per capita consumption in m^3 day
I	Infiltration into the network in m^3 per day
E	trade effluent discharge into the network in m^3/day

In the past the dry weather flow was defined as:

> the average daily flow to the treatment works during seven consecutive days without rain (excluding a period which includes public holidays) following seven days in which the rainfall did not exceed 0.25mm on any one day

This was first defined in 1975 in the Institute of Water Pollution Control Glossary [9]. This was redefined in a 2006 study by UKWIR and Tynemarch. An alternative approach using the 20th percentile of the annual daily volumes of wastewater flow was proposed, which showed a good comparison over 4,447 work years of flow data and 3,123 work years of rainfall data [10]. This method became the Q_{80} method for dry weather flow measurement and, through manipulation of the DWF formula and knowledge of the different elements, can allow for estimation of the infiltration into the network (as long as the population, per capita consumption and trade effluent discharge volumes are known).

This is useful in looking at the different elements that make up the base flow going through the sewage collection network, before the environmental inputs are taken into account.

If measuring the customer is difficult, then measuring the environmental inputs is even more difficult and often relies on third parties which don′t necessarily measure in the right places for the sewerage collection network. For environmental inputs, in the main, there are two factors to consider:

- Inputs due to the underlying soil conditions and infiltration
- Inputs during rainfall/storm events

Infiltration into the wastewater network has been a perennial issue and depends upon the climatic conditions in a particular year and the underlying soil conditions. It is more of a long-term issue and can only be seen by looking at flow data on a day by day basis over a period of several years. The use of this is that infiltration into the sewer can contribute to the basal flow, especially in wet weather periods so that the sewer in a particular network can, for a period of up to several months, never return to a true dry weather flow. On a network monitoring and control basis, this takes up capacity in the network, but, by analysis, the very nature and behaviour of the flow, together with rainfall patterns, can give clues to the location of the break within the gravity sewer environment. This can allow for infiltration points to be localised with relatively small costs when compared with CCTV methodologies of infiltration identification.

The last input into the sewage collection network is, of course, rainfall. The main detection method is a combination of weather radar, rain gauges and artificial intelligence, enabling the prediction of the impact of a rainfall event on the sewer [11]. Although weather radar is well established, the current techniques in terms of the wastewater network have not allowed the resolution of data that is required, and there are very few techniques in place that will then translate this into impact on the sewerage environment using a model-based approach.

The main challenge is one of resolution. When the UK meteorological office looks at weather, they look at it on a national basis, and the resolution of the weather radar is for grids that are 13 km wide and 13 km long (169 m^2 grids). Allowing for the surface area of England and Wales, this splits the UK up into 895 grids. There are approximately 10,000 WTW in this area of which 3,742 works treat greater than 50 m^3/day and are numerically consented. Hence, in reality, the Meteorological office data has not currently got the resolution necessary for a smart wastewater network. In this case, the resolution of the weather radar data has to be much finer, and also work in conjunction with rainfall data to predict what the impact of a storm event will be on the network. This moves C-Band Weather radar (the technology that the Meteorological Office use) to X-Band Weather radar which measures on a much smaller scale, with the disadvantage that many more radar units are required [12]. Rico-Ramirez et al. [13] show the considerable uncertainty in radar rainfall estimates and how this raises implications for modelling. Also, in Schellart et al. [14], the difficulty of using rainfall nowcasts for predicting sewer flows is discussed.

From this it can be seen that the challenges of measuring the inputs into the wastewater network are large with:

- Domestic wastewater monitoring not established, as it is not financially viable.
- Industrial wastewater monitoring well established and input volumes known, although the accuracy of the installed flow meters are not always known, as the maintenance of the meters is sporadic.
- Inputs from infiltration can be estimated but are very changeable due to climatic changes and underlying soil conditions.
- Weather radar and the impact of rainfall on the network system are improving as technology improves, and there are some systems available at the current time that will allow this to happen.

2.3 Monitoring the Network and Outputs from the System

Within the wastewater network, the challenge is that they are three types of network, the more traditional gravity network that everyone considers the sewer to look like, the pumped network and those that are a mix between the two. Of course, the different types of network behave in different ways and require different types of instrumentation to monitor them, although the wastewater network is probably the largest area of the water industry infrastructure which is largely unmonitored.

The pumped network, in some ways, is very similar to its potable water cousin consisting of pumping stations and rising mains. The risks of this network are mainly centred around the risk of pipe bursts, due to blockages or pipe condition and overflowing of pumping stations caused by them being overwhelmed with flow, by power failures and by pump failures or simple blockage.

Like the potable water network, there are a few areas where a smart water industry can have large potential impacts around:

- Wastewater pumping station control
- Wastewater system control
- Wastewater network burst detection

Wastewater pumping stations have typically had simplified pumping station control, with either floats or level control to ensure that pumping stations pump when they need to pump in isolation from any other aspect of the wastewater network. They will normally have an overflow monitoring device that monitors flow through the emergency overflow to the environment, which is there in case of pump blockage or pump failure. As the size of the pumping station increases so does the complexity, but the principle remains the same.

As technology has developed, and the monitors and control systems can do more, then the intelligence of the sewage pumping station has increased. In addition to simple level sensing, other aspects of sensing have been incorporated including:

- Current monitoring for pump performance
- Flow monitoring for discharge rates
- Pressure monitoring for burst detection
- Implied flow meters from variable speed drives and virtual drop tests to imply inflow
- Control systems to reverse pumps that can attempt to free pump blockages automatically

Further developments in the terminal pumping stations can put control in place to ensure that the pumping stations coordinate with each other to smooth the flow entering the WTW, to ensure more efficient operation of the entire wastewater operational system.

Current monitoring strategies, as well as the use of Variable Speed Pump Drives used to imply flow from the power required to operate the pump, show the performance of the pump and can indicate when the pump is starting to wear. The installation of a flow meter in the rising main can provide a secondary verification of this, which is especially useful on large pumping stations. It can also be used to allow for visualisation of what is happening at a pumping station and inform how to react to alarms that are raised on particular asset. Imagine a scenario where it is raining, all of the pumps available are running, and the maximum output of the pumping station is being pumped; and yet there is a high-level alarm. In this situation everything possible is being done, but the sheer volume of flow is overwhelming the pumping station. The same situation without a flow monitor would mean it would have to be visited to assess the situation, whereas a flow meter would definitively confirm the actual situation.

More recently there has been work with pressure monitoring on the rising main aspect of the wastewater network. This uses high-resolution pressure monitoring to predict when the rising main is going to fail, which is working along the same lines as pressure transient monitoring in the potable water distribution network. The full understanding of the methodology of using pressure monitoring to predict pipe failure is not fully understood, but it is currently being studied.

On the pumped wastewater network, the ultimate aim is for it to work with the gravity network with the main control point being the inlet of the wastewater works. This provides flows to the treatment works that are, as far as possible, equalised across a 24 h period with the ability to react to a model-based approach to reduce the amount of wastewater discharged straight to the environment through storm overflows in wet weather.

The gravity network needs to work in conjunction with the pumped wastewater network although it behaves in a completely different way, and the monitoring of the gravity wastewater network has different challenges.

In the gravity network, the main question is what to measure, and where, within the network. There is a school of thought that believes that level monitoring of the sewerage system can be used in its entirety to monitor the situation in the network and allow the required element of situational awareness. Another perspective is that the use of flow monitoring in the wastewater network also adds another dimensional element. As there aren′t many measurement technologies that will measure part full pipes, flow measurement within the sewer is limited to Doppler style area-velocity devices that can either be submerged in the flow (and are at risk of damage due to fouling) or are remote from the flow (which have limitations over where they can be installed and need a minimum size of pipe). There are technologies that are currently being researched, looking into measuring the surface patterns of the flow inside the pipe to measure the flow rate [15].

What value do each of the measuring techniques bring to the wastewater network? Level within the wastewater network can be used to simply indicate how full the sewer is. This is a very powerful tool as it can indicate the current available capacity within the network and, together with a wastewater network model and meteorological detection methodologies, can be used to calculate whether there is sufficient capacity within the network to manage a potential storm event that is happening in the next few hours. This is a crucial element of a smart wastewater network. Level, when measured at several different points, can also indicate where there is potentially a blockage within the system as a raised level at one particular point, and a significantly lower downstream would indicate this. The problem of this is twofold:

1. How many sewer level monitors are needed to provide a resolution good enough to provide situational awareness and where within the network should they be located, as it is not financially viable to monitor everywhere?
2. How can sewer level monitors be installed at a reasonable cost?

Moving on from just using simple level measurements, the rate of change can also be used in a number of situations. The simplest of these methods is to measure the rate of change within a pumping station wet well. Some of the simpler control techniques use the rate of change to predict the flow rate within the wet well to imply flow by conducting a smaller version of a drop test. This requires an outflow pump rate (or flow measurement) to provide a measure of what is leaving the system and the rate of change within the wet well to measure the increase of volume over time (i.e. the flow rate).

Within the gravity network, any unusual change of level can be used to predict where there is a blockage (i.e. where a downstream level is lower than an upstream level there is a good indication that an unusual situation is occurring that needs further investigation).

When flow monitoring of the gravity network is brought into the equation, what benefit does it bring over level monitoring in isolation? Looking at flow monitoring, it is even more difficult to install than level monitoring and, unless it is used to control aspects of the network, or measure particular inputs into the network with an approach similar to the potable water networks DMA approach, then it could be used to indicate particular areas of risk of infiltration within the gravity network.

Wastewater networks require regular inspection in order to prioritise and perform effective maintenance. Currently wastewater networks are generally inspected using CCTV, taking one of two approaches. The first requires a camera, attached to a semi-rigid wire, to be pushed through the network. In doing so the collected footage is left to be analysed later by a trained engineer. Although quick to collect, the footage gathered is often of lower quality as the camera does not travel smoothly through the pipe. Alternatively a camera can be attached to a remote-controlled pig which is driven through the network. A skilled operator can often identify and record faults whilst operating the device. In doing so footage takes longer to collect, but does not require further analysis and is often of higher quality. CCTV (along with GPS) is the most commonly used technique for locating pipes and internal inspection [16]. The Water Research Centre (WRC) in UK devised the first condition grading scheme that provides protocols and guidelines for formally assessing current condition of individual pipes using CCTV inspection [17]. These can be somewhat subjective if performed by a human, and the latest research has been exploring the automation of CCTV footage.

With infiltration there are a number of passive technologies that have been developed to look into the state of the gravity wastewater network as well as, in some cases, the pressurised network. These range from looking at the thermal characteristics of flow entering the network, where infiltrated flows are warmer than the flows already within the network, to using acoustic waves technologies employed in a number of techniques that can be deployed externally or internally to the pipeline, to measure comparative pipe condition to see if the assets are deteriorating. Together with CCTV surveys and the utilisation of more advanced techniques such as pattern recognition, these methods can be used to identify possible areas of infiltration in gravity sewers, or potential for bursts in rising mains [18].

The most recent innovation for inspection is passive untethered platforms including free-swimming solutions (which advect with the ambient pipe velocity) that can complete long inspections in a single deployment. There is increased interest in the utilisation of robotics. A comprehensive journal review [19] of robots for pipeline inspection revealed that robots currently available are mainly laboratory prototypes designed for large diameter pipes, human-controlled, heavy (tens or hundreds of kg) single devices suitable for a single short-duration intervention. The last major area of monitoring within the wastewater network is at the combined sewer overflow (CSO), and this was driven in the UK by a strategic government decision under Ministerial Direction in July 2013, to monitor the vast majority of CSO′s within the water industry with event duration monitors (EDMs). This decision was made in order to protect the environment from discharges to water courses through CSOs and other discharges to the environment, including pumping station overflows and storm tank overflows, when they either shouldn′t be taking place or are happening more often than is reasonable.

A risk-based approach to the EDM programme was taken by the Environment Agency looking at CSO′s in areas where the amenity or economic value is high, and where an overflow is suspected to be a potential problem given the highest priority. This goes all the way to the other end of the spectrum where an overflow is of low amenity or economic value and is known not to overflow and hence is given the lowest priority. The philosophy of the programme was to discover where all the overflows were, and to monitor when and for how long they overflow to the environment. With the highest-risk areas requiring much more detailed monitoring, this effectively maps the exact area of impact that the sewerage network has on the wider aquatic environment. In the programme, only low amenity overflows with less than 20 spills per year were excluded, and, in practice, many of the water companies will have 100% coverage before 2025.

As with all of these initiatives, in practice, they have a much greater operational benefit than the project initially envisaged, and, as part of a “smart” wastewater network, the EDM programme has the potential to be used to measure network performance. It can also identify areas where the network is stressed and where investment is required in order to provide further capacity, either through reduction of the basal flow through infiltration reduction, increasing the capacity of the wastewater network, or potentially increasing the monitoring and control of the network to relieve periods when the wastewater network is having capacity issues and is under stress.

2.4 The Opportunities and Barriers for the Smart Wastewater Network

By putting intelligence into the wastewater network, there are a large amount of benefits. These include an improvement in environmental water quality and balancing network improvements to counteract the ever-tightening consents at the WTW. Consent conditions are reaching the point where the level of treatment required to comply with environmental permits are intensive enough to have both a large negative impact on the air environment through increased energy consumption and having an impact on resource issues through larger quantities of chemical consumption to treat to lower and more exacting levels. By providing better control of the impact, there is potential for providing a greater level of pollution control for a lesser overall environmental and financial impact.

However there are technological and financial barriers to the implementation of smart wastewater networks including:

- The development of wastewater network models for the purpose of operational control rather than engineering design, which is where most of the wastewater network models currently fit.
- Knowledge of the financial and environmental benefits of better monitoring and control in the wastewater network. There is a perceived benefit of getting a better environmental performance overall and being able to balance this with a potential loosening of environmental permits at the treatment works but as yet this is unproven.
- Integration and performance of meteorological artificial intelligence to feed into an operational model of a smart wastewater network. Enabling a measure of the impact of potential of storms on the network.
- The proliferation of measurement in the wastewater network, where the barrier at the current time is both technological and financial on the installation front, quite aside from the costs of maintenance of the instrumentation in terms of risk to operating staff and also the cost of conducting it.
- The methodology of the integration and interaction between the wastewater collection network and the WTW is largely unproven at the present time.

These issues need to be resolved before smart wastewater networks can proliferate throughout the wastewater treatment system as a whole, and there is a potential for them only to be installed on the larger, higher value networks where large populations are served, before the technology proliferates into the smaller networks systems.

3 Smart Wastewater Treatment

The key to the smart wastewater treatment plant is consistency. A WTW is a system that is naturally unstable. It is therefore through measurement and control that the operator can achieve some sort of stability within wastewater treatment in particular and the wastewater system in its entirety. One of the keys to this lies within the network, where possible smoothing out of peaks and troughs of instability in the flows and loads reaching the WTW could provide much more of a stable operating environment.

There are regulatory drivers (at least in England and Wales) that will push the industry into an environment that is much more rigid regarding the management of flows and, like the EDM programme in the wastewater network, will ensure that all flows that can possibly be treated, are.

The industry is moving towards a factory approach that was proposed by the Dutch Foundation for Applied Water Research (STOWA) in 2010 when they described the WTW of 2030 as water, energy and nutrient factories [20]. Their fundamental point is that the WTW adopts the factory approach, and, as with any factory, measurement and control is an essential part of managing the “factory” (system) as a whole.

3.1 Philosophy of Operation

What is the Aim of the WwTW?

The very fundamental aim is to treat wastewater to a standard where it does no harm to the environment that it is discharged to. This aim is incomplete though, as from an environmental perspective the aim can be modified to include an element of holistic environmental cost, insofar as doing no harm to the much wider environment will include an element of efficiency. It is relatively easy to treat wastewater to a standard so that virtually pure water leaves the treatment plant; however, the prohibitive energy and chemical costs to do this outweigh the overall environmental benefit.

This is where the aspect of the resource factory comes into play. Which products are available from a WwTW and how do we use the factory approach to produce them?

The products that can be produced from a WwTW are:

1. Water of varying qualities
2. Biosolids (rich in nutrients and used as a soil conditioner)
3. Phosphorus (as a fertiliser additive)
4. Energy and Biofuels (either directly as electricity or as a liquid gas)
5. Low-grade heating water
6. Cooling water (for use in data centres or other applications)
7. Plastics

Of these, at least the first five are actually already being used at the current time and are no longer in the research and development phase for potential use in WwTW of the future so, as we can see, the STOWA approach has become somewhat of a reality.

As well as producing resources, there is the drive towards energy and process efficiency, and there are a number of different tools enabling this, including instrumentation based control systems, as well as multivariate process control [21]. However, the overall most important thing, from an instrumentation and control point of view within the WTW, is having the information to hand to make informed operational decisions.

3.2 Measuring and Controlling the Wastewater Treatment Process

The WwTW can be split up into a number of different processes which are independent of the size of the works. However, in the main, the measurement and control processes are restricted to the medium and large treatment works, as the value of smart systems only becomes cost-efficient on the systems where economies of scale exist. This will govern not only where smart systems are installed on the WwTW but, in reality, the entire wastewater treatment system. This has been one of the stumbling blocks for the adoption of smart water systems across the wastewater industry.

Splitting the wastewater treatment system across its component parts, we can see the drivers of instrumentation and control systems across the works.

3.2.1 Measuring and Controlling Preliminary and Primary Treatment Processes

The purpose of the preliminary and primary treatment processes are threefold:

- Remove bulk debris from the incoming waste stream through the screening process.
- Remove grit.
- Manage flows by balancing and/or storm separation.

Looking at each of these processes in turn, there are potentials for increased efficiency of operation in some areas only.

Looking at the screening process, it is only there to remove solids, and the simple control system is designed to clean the screens when they become blocked (or blinded). The traditional measurement and control system look at differential levels up, and downstream of the screen, and when the differential increases to a preconfigured level, it initiates the screen cleaning process.

Grit removal typically does not have any measurement and control system and so from this aspect can be ignored.

The flow control and storm system within preliminary treatment is an area that has a vast potential for intelligence in operation. It is going to be an important area of development in England and Wales between 2020 and 2025 [22] and is all about controlling the flow to full treatment (FFT). There is a legal duty under a treatment work′s environmental permit for it to treat a certain flow before excess flows pass to storm storage tanks. If excess flows are seen at the works for greater than 2 h, they pass directly into the environment. Across the various treatment works in the industry, for various reasons, there are problems with this concept. FFT flow control is normally achieved by using either a static weir and flow control device (such as a flume or hydrobrake) or a modulating device complete with flow measurement (such as a modulating penstock). The modulating methodology can, on occasion, have problems due to poor engineering practices.

As a result of this, there are moves within the English and Welsh wastewater industry to install flow measurement and sensing on storm splits as well as on the effluent point of the WwTW storm storage tanks. The retrofitting of flow to full treatment flow measurement for flow control is disproportionately expensive. However, it does have further benefits for advanced control of the wastewater treatment system.

Measurement of the FFT at a treatment works allows for the performance of the wastewater collection system along with the potential for control of any terminal pumping station. At the most basic, any large changes in the amount of flow over a defined period of time can indicate whether there is a blockage within the wastewater system. At its simplest, if a WTW doesn′t see flow over a period of time, then this is a good indication that a blockage within the network is causing the flow that is normally seen to escape elsewhere (normally through a CSO). It will vary from system to system depending upon how many pumped flows and how many gravity flows feed the treatment works, but a simple algorithm of a defined low flow rate over a defined time will allow for anomaly detection.

In a more complicated system, there is potential to automate the pumped feeds into the treatment works by using the FFT flow meter and automated control on the terminal pumping stations. The latter is achieved using a priority based-system related to the capacity within each wet well, and possibly artificial intelligence to manage the contents of the wet wells, to smooth flows entering the treatment works. This looks at the capacity of the network and the performance of the inlet works to see what flows (and loads) can be received at the WTW whilst also minimising the risk to both the environment and the customer. The current growing “big data” context provides abundant opportunities for the application of artificial intelligence to such problems as optimisation of sewerage networks′ operation, predicting urban flooding [23], CSO monitoring and analysis [24, 25] and automatic control of sewer pumping stations utilising fuzzy logic [26]. Latest approaches have proposed more distributed real-time control of urban drainage systems to locally manage flooding and overflow [27, 28]. Artificial neural networks (ANNs) have become an increasingly popular data-driven approach for water industry applications. ANNs have been widely applied in different fields of engineering and are a modelling approach based on how biological neural systems are believed to work. Li et al. review the applicability of ANNs to urban hydraulics and hydrology. An area that could be usefully explored includes using ANN models to predict flows at treatment works.

With the advent of event duration monitors and performance targets on up-time on the weirs to and from storm tanks, there is a push within the industry to manage flows at the front end of the treatment works. This would both protect and reduce risk to the environment (in using storm tank capacity) a lot more stringently.

The system as a whole can be seen in Fig. 3. It is slightly overcomplicated with the addition of flow monitoring on the feed pumping stations and on the storm tank return and yet could be considered slightly simplified as large works could potentially have many more than just two feed pumping stations. However, the principal is there insofar as there is no control over the gravity feed to the works, but there is control over the two feed pumping stations that could hold back feed flows. However, this would be dependent upon the capacity in the network upstream of the pumping stations, and in the pumping station wet wells themselves.

This is the point where the inlet works of the WTW becomes an integral part of the wastewater collection network and vice versa.

After flows pass forward to treatment, the first stage that they will typically pass through is primary settlement. Although not always present as a treatment stage on the larger works, it is normal practice. This is because the separation of gross solids in the sewage is a valuable energy source that, if not removed, is simply load, and therefore an energy loss, to be treated in the next stage of the process. As it is a simple physical settlement stage, there is very little to control, but measurement of the wastage of primary solids is something that should be considered in an energy/resource factory, as it is a major source of energy. Primary settlement tanks can be used as a thickening stage using type 3 (hindered) settlement and type 4 (compression) settlement. In this way there is the potential to use a combination of both sludge blanket level and flow/solids measurement on the primary de-sludge line from the settlement stage to manage the sludge inventory. In reality, flows and loads coming into a sewage treatment works are ultimately predictable, and with predictability the primary sludge inventory can be managed with the right measurement. This also needs to feed forward to sites with anaerobic sludge digestion to create an almost on demand availability of sludge allowing for the correct blend of import and secondary sludges to maximise the energy output of the wastewater treatment facility.

3.2.2 Measuring and Controlling Secondary Treatment

The biological (or secondary) stage of wastewater treatment very much depends upon the size and the type of the plant and what the treatment goals are. Biological filters are still the predominant type of wastewater treatment, especially on small and medium treatment works, and, apart from simple recirculation flow control to keep biological filters wet, there is very little measurement and control needed in this type of secondary treatment. As the process is so simple, there is also very little that can

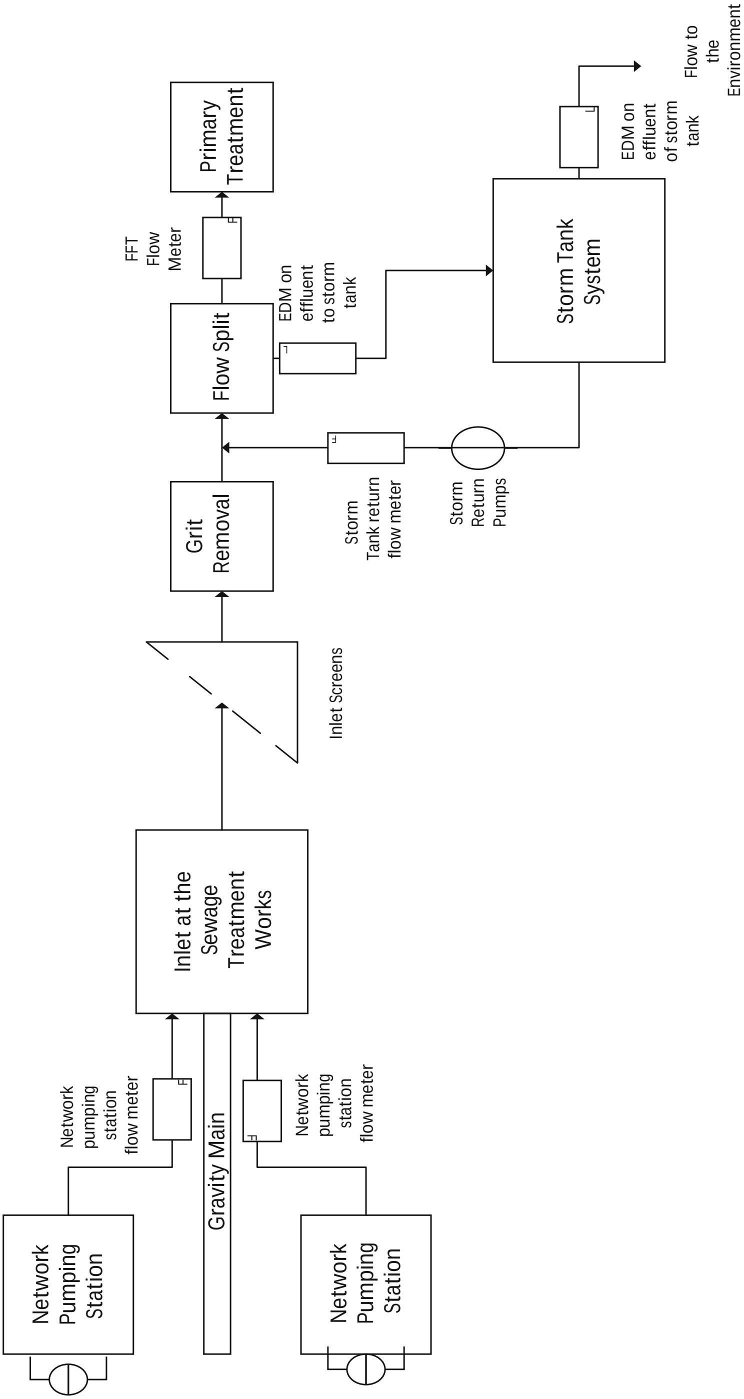


Fig. 3 Simplified process flow diagram of a potential smart wastewater simplified network and treatment works inlet

be done to make efficiency savings and very few “resources” that can be recovered from biological filter treatment streams.

With the development of tighter standards, and certainly on large facilities, activated sludge plants are much more common as a biological treatment stage along with numerous variants that are designed for nutrient control (Biological Nutrient Removal plants and Enhanced Biological Phosphorus Removal Plants). The overall measurement and control aim is to maintain the biological mass at the right levels to achieve the desired treatment goals.

Activated sludge plants tend to be very energy heavy as, the tighter the consent, the more intensive the treatment process is. This is especially true of aeration control. This is also applicable to managing the biological balance within the plant, through the control and return of the bacterial mass, and waste of biological solids.

The simplest of these control systems is direct control of the air within the activated sludge plant, by using the measurement of dissolved oxygen and ramping blowers up and down in line with the demand for air. This is not a very common technique in practice, as the variation in oxygen being measured will cause constant ramping of the blower system. Instead, a much more common approach in a simple aeration control system is to pressurise an aeration header and maintain the pressure using blower ramping and dissolved oxygen measurement to control the position of modulating valves. How these valves are controlled is where the efficiencies in aeration control lie, including feedback control in-line with measured concentrations of target parameters, or feed forward control using Activated Sludge Models. Controlling the whole balance of the system, including the sludge solids, is actually the way to manage the efficiency of the whole activated sludge system for maximum benefit. Moving to the resource factory approach, this can be tailored depending upon the treatment outcomes. Figure 4 shows a conventional activated sludge plant system including its control systems.

At the most basic, the dissolved oxygen concentration is used to adjust the aeration lane valves (open or closed). The level of the dissolved oxygen measured in the lane creates the pressure to drop in the aeration manifold when the valve is opened, and this pressure drop starts the aeration blowers to maintain a set point pressure. In the slightly more advanced systems, there is a feedback loop from the ammonia concentration which sets a dynamic oxygen set point which the plant attempts to maintain.

Of course, on top of this is the control of the actual biomass within the activated sludge system. The importance of this is reflected in the oxygen required to treat the load to the biological process.

Oxygen requirements in an activated sludge plant can be split into four main areas:

- Oxygen required to treat the carbonaceous load (BOD)
- Oxygen required to treat the nitrogenous load (ammonia)
- Oxygen required by the biomass to breathe
- A credit caused by the conversion of nitrate to nitrogen gas in the anoxic zone

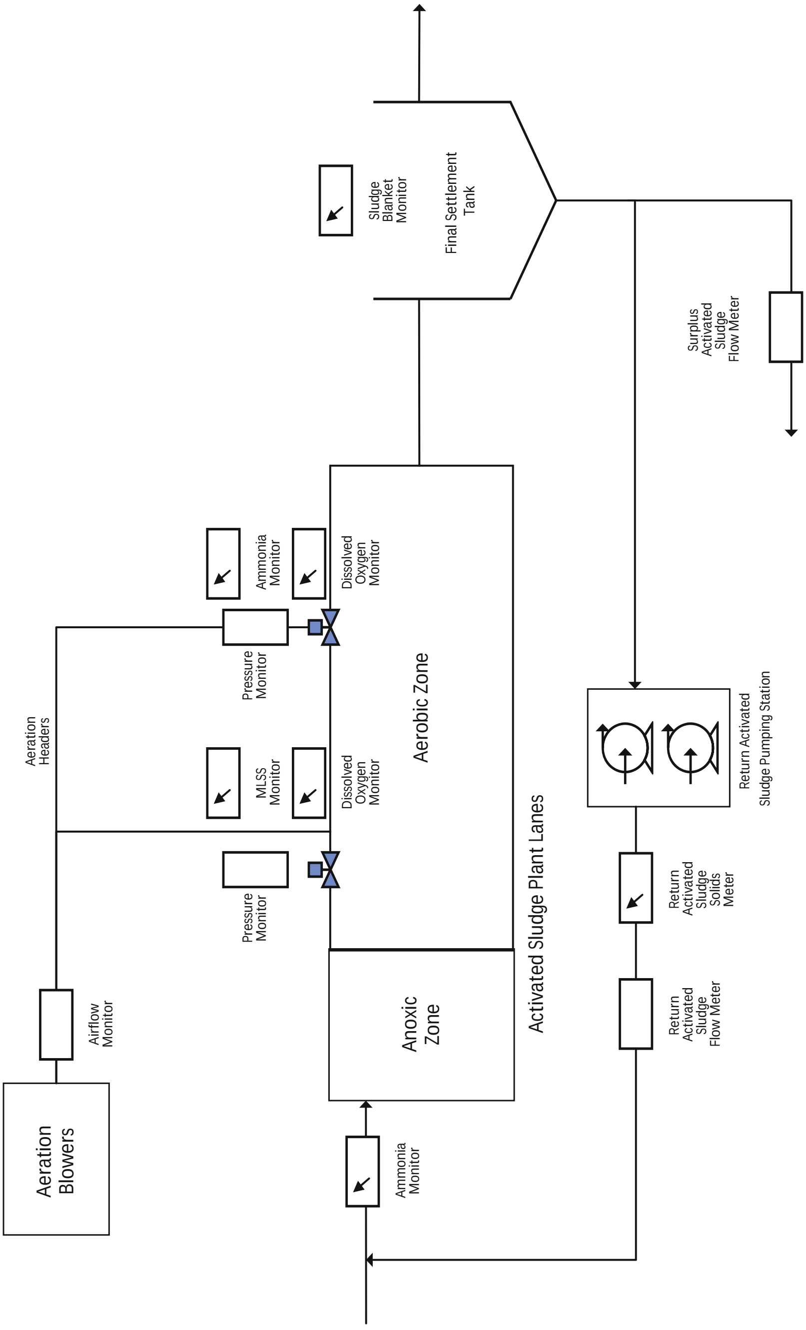


Fig. 4 A typical conventional activated sludge plant control system

All of this is added together (or taken away for the credit) and is used to calculate the aeration system requirements and normally includes a peaking factor to allow for peaks in load of up to 40%. This all creates inefficiencies in the activated sludge plant design.

There are four areas where advanced activated sludge plant control can create efficiencies in operation:

- The first, and most obvious, is to limit the biomass oxygen demand by limiting the physical quantity of biomass required to successfully treat the wastewater.
- The second is to control the oxygen required to treat the nitrogenous load to just more than is needed (to allow for a small safety factor) by using feedback ammonia control, feed forward ammonia control or potentially a combination of both.
- To control and stabilise the flows (and load) coming through the works and, where possible, eliminate the peaking factor as much as possible.
- Maximise the denitrification credit by increasing the size of the anoxic zone by utilising the concept of swing zones, thereby allowing a greater size anoxic zone when the temperature allows.

There are numerous ways to control the activated sludge plant and its many variants to maximise the treatment objectives. These include using many more sensors than the basic methods including:

- Respirometry that will look at the bacterial health and calculate the specific oxygen utilisation rate
- Sludge Volume Index that will look at speed of the settleability of the sewage sludge and waste sludge, based upon the capacity of the final settlement tanks to hold the mass of sludge
- Nitrous oxide sensors that will look at the performance of the denitrification process, the stress levels of the bacterial population and control the efficiency of the wastewater treatment process

Practically, there are many types of secondary treatment processes, and the aim of the process needs to be understood thoroughly. Is it a case of removing all pollutants within the biological treatment stage? Or is it using the process for resource recovery? This is an area that the industry is heading towards, potentially setting things up for a tertiary stage where products are removed. This is all going to affect the monitoring and control strategy of the secondary (and potentially tertiary) treatment process.

3.2.3 Measuring and Controlling Sludge and Resource Recovery Processes

It is when the wastewater treatment works is a sludge treatment centre that the real value of wastewater comes into force, and this is an area of the business that should resemble “the factory approach” much more than is currently the case. There are, however, complications in the measurement techniques that make this a difficult proposition, especially as the sludge from the primary and secondary treatment processes moves from a Newtonian to a non-Newtonian fluid (usually around the 5–8% dry solids mark).

The processes within sludge treatment often see sludges that start at 2–3% dry solids and are thickened to around 8% dry solids. If this is on a sludge treatment centre, it is more about processing the sludges for use in subsequent treatment steps, but often this is done on satellite sites to ensure that as small amount of water as possible is transported via tankers.

Once thickened sludges are produced it becomes very difficult, although not impossible, to measure what is actually happening. This is especially the case when sludges are de-watered in a centrifuge process up to dry solids between 23–28%. At this level they resemble a powder. This depends upon what is being done to the sludges in terms of treatment steps.

The most common treatment step in the UK is mesophilic anaerobic digestion (MAD) although quite often nowadays there is an Advanced Thermal Digestion step before MAD to speed up the throughput of the process [29]. In MAD, sludge is kept in an anaerobic environment normally for upwards of 12 days at $37 \pm 2^{\circ}$C. This converts a proportion of the sewage sludge to methane gas as it goes through the anaerobic digestion process. The temperature of the process is very critical, as is the mixing and the complete absence of oxygen. Thus, temperature monitoring is an absolutely critical part of the process and is normally carried out at three points within the digester. However, this is not the only critical factor as the heating loops for the digester heating have to be maintained and monitored precisely to ensure the maximum output of the gas quality. Measurement of the gas quality is becoming more and more common to measure the performance of the overall system.

What is essential for the digestion process is consistency and, in an ideal world, if the same volume of sludge is fed to the digesters consistently throughout the day ad infinitum the process would run without error. However, this consistency is not typical due to problems within treatment processes and general operational challenges that are the normal daily routine within the water industry. Thus checking the anaerobic environment becomes a priority. For example, when a digester is overfed, the amount of acids that are produced as part of the digestion process can build, and a digester will turn acidic. It is in these circumstances where monitoring the pH as well as the output gas quality can become important.

Taking it back a stage, the whole process can be controlled by a greater level of monitoring and control at the sludge thickening (or dewatering stage if using thermal hydrolysis), to get a consistent output in the sludge quality feeding the digesters and basically processing on an on-demand basis. This level of consistency is where an automated measurement and control system comes into its own. Increasingly, the industry is seeing the use of advanced process control on sludge thickening and dewatering systems using sludge flow measurement into the thickening stage, as well as the dry solids content. This approach allows a consistency of application of the polymers used in the treatment process. In advanced systems, the flocculation stage, where the polymer and sludge are mixed, is optimised. Also, changing the angle of the unit to speed up or slow down the process enables control of the thickening and dewatering stages in drum thickening and centrifugation, allowing for thinner or thicker product which is then transferred to the digestion stage.

If consistency is achieved, then it spreads throughout the entire process, achieving the ultimate aim. However, in reality, consistency is not usually achieved throughout the whole process, and the challenge is to bring balance from imbalance. This is where the measurement and control systems come into play.

3.3 Holistic Control: Model-Based and Multivariate Process Control

The WTW as a whole can be controlled manually by analysing and responding to the data produced by the instrumentation, but this is a labour-intensive process which requires continuous adjustments. On the larger scale treatment works, there is normally some sort of automation within the process to realise process-based efficiencies over standard manual operation. This can be as simple as a single instrument-based control loop, or as complicated as a multivariate process control system.

Within the past 5–8 years, the principles of individual process control on the activated sludge plant have been gathered into integrated advanced process control systems in the UK by two commercial organizations and one or two of the water utilities. The water companies themselves have tried to do this with PID loop controls and, in some cases with cascade loop control systems, some of which have had some success in Europe.

Hach Lange developed the Water Treatment Optimization system (WTOS), which is very much an advanced process control system based upon instrumentation. The WTOS system utilized Hach Lange instrumentation and a process model based upon the ASM1 activated sludge model [21].

One of the first full-scale installations controlled with the WTOS system was a 250,000 population equivalent four-stage Bardenpho plant with methanol addition in the second anoxic zone. The system and controller that was developed for this treatment plant looked to monitor and automate the whole process, including the nitrification and methanol dosing. This first installation conducted a trial over a 10-week period and managed to achieve a 20% reduction in the amount of aeration, control of the amount of ammonia that was discharged, and a 50% reduction in the amount of methanol that was consumed.

Since its first implementation in 2008, this technology has developed even further with other control modules including a nitrification module (which includes sludge age control) specifically designed for the activated sludge plant, as well as modules that are designed for other plant processes.

The second approach to advanced process control has again been based on model-based controllers but is less reliant on instrumentation and more reliant on the intelligence of the system as a whole; any failings in the implementation of APC have been due to poor data quality from the instruments. This approach put more intelligence into the control system to identify when an instrument becomes unreliable and, for the system as a whole, to replace the unreliable data with an inferred value based upon the readings being received from other instruments within the system.

For example, the control model “knows” what each DO sensor should measure at any given time, given the influent flow, blower load, valve positions, manifold pressures and treated water quality. If any probes report values that are significantly different from those that are expected, an alarm is raised, and the inferred value is used to exercise control of the process. Optimized control can be maintained even when real-time measurements become unreliable.

The multivariate process approach has advantages of being a system based upon the control element and is much more widespread within the plant, taking into account the whole treatment facility rather than just the activated sludge plant on its own. Case studies of this approach in three UK water and sewage plants realized savings between 20 and 35% of the aeration costs whilst also reducing the risk of compliance failure as the treatment plant operates more efficiently under automated control [30].

All of these systems are designed on the basis of the International Water Association Activated Sludge Models (ASM), and it is these models that are providing the fundamental basis for the control systems of the secondary treatment plant stages. The challenge to the industry is to use these models at their core and stretch them further into not only the treatment works level but across the whole wastewater system. Although there are individual models for aspects of control within the wastewater system, it is likely that the real complication in controlling the entire system will be melding together the different models that are available. These include hydraulic collection network models, multivariate process control models for individual aspects of the treatment system and SCADA and control systems. This does not take into account sporadic inputs into the wastewater system such as from customers and of course weather.

4 A Smarter Wastewater Industry

Ubiquitous sensing will create many opportunities and threats for urban water management and calls for a digital transformation [31]. Increasing amounts of data is only of real business value if this valuable resource is ultimately used to inform and support decision-making, i.e. data to information to insight to action. Smart water network technologies have the potential to deliver an improved service to customers and cost-effective performance improvements for the water industry. On the wastewater side of the water industry, the “smart water” approach that we have seen applied to applications in potable water (such as water leakage) is much more difficult. The first barrier is the value of wastewater, as it is very much seen as something to be treated as efficiently as possible but actually something with a negative value as it is considered as a waste. In the past few years, due to a growing awareness of environmental issues and a sustainable circular economy, this has very much changed towards elements of wastewater being considered a resource. An example of this is energy and biosolids from wastewater sludges, but there are many more resources to be gained including water, phosphorous and many other nutrients within the matrix being reused as resources. It is a balance however because, if phosphorus is stripped from the biosolids, then it is not available in the biosolid when it is applied to agricultural land. So, when thinking of wastewater and biosolids as a resource, the industry has to consider what the market is, and what product is being made available for that market. As the WwTW becomes a production factory, the processes within that factory must be measured and controlled to ensure efficiency of production and the quality of the products.

There are also efficiencies to be gained in how production processes are controlled. The value of this is not to think about the WwTW and the wastewater collection network as separate entities, as in the past, but to think of it as part of one dynamic system. Controlling the flow of wastewater through the collection network and funnelling it towards the WwTW not only has energy benefits by smoothing at the flow profile but has the potential benefit of protecting both the customer and the environment.

By measuring the state of the system and utilising operational models, there are large benefits to be gained. The minimum benefit comes from simple operational visualisation that allows for potential pollution detection, and the maximum benefits are realised by using advanced control of the entire wastewater system. By utilising automated control of the system and operational models within the wider wastewater system, efficiencies in energy consumption and customer and environmental protection can be achieved. These include balancing flows through the network, monitoring areas that are at risk of flooding, preventing sewer overflows where possible and allowing the maximum flow through the treatment system. Some aspects of this smarter approach are currently implemented, but there is not yet one system consistently using all these tools at this time.

References

1. UKWIR (2012) A review of the effectiveness of Mogden Formula charging when meeting modern sewage treatment works consents. UKWIR, London
2. Kellagher R, Osbourne M (2013) Factors limiting the use of ASC on sewerage systems in the UK project report. UKWIR, London
3. Kellagher R, Osbourne M (2013) The use of active system control when designing sewerage systems: a guide. UKWIR, London
4. Mounce SR, Shepherd W, Sailor G, Saul A, Boxall JB (2014) Application of artificial neural networks to assess CSO performance. In: 13th international conference on urban drainage, Sarawak, Malaysia, 7–12 Sept 2014

5. Ostojin S, Shepherd W, Williams K, Simões N, Steinhardt J (2017) CENTAUR: smart utilisation of wastewater storage capacity to prevent flooding. In: Proceedings of CIWEM Urban Drainage Group autumn conference and exhibition 2017. CIWEM, London
6. OFWAT (2015) Water meters – your questions answered. OFWAT, Birmingham
7. Harding E, Homewood S, Kowalski M, Palfrey R, Poole K, Snowdon C (2012) The relationship between per capita consumption and wastewater flows project report. UKWIR, London
8. UKWIR (2013) Cost benefit analysis of ubiquitous data collection in water distribution – CBA scenarios. UKWIR, London. 13/DW/12/2. ISBN: 1 84057 692 8
9. Institute of Water Pollution Control (1975) Glossary of terms used in water pollution control. Institute of Water Pollution Control, Kent
10. Heywood G, Starr M (2005) Alternative measurements of sewage treatment works dry weather flow. UKWIR, London
11. Cockroft J, Body R (2015) Portsmouth flood alleviation: early warning system managing storm flows using the an early warning system moving from ′just in case′ to ′just in time′. UK Water Projects 2015
12. Heinonen M, Jokelainen M, Fred T, Koistinen J, Hohti H (2013) Improved wet weather wastewater influent modelling at Viikinmäki WWTP by on-line weather radar information. Water Sci Technol 68(3):499–505
13. Rico-Ramirez MA, Liguori S, Schellart AN (2015) Quantifying radar-rainfall uncertainties in urban drainage flow modelling. J Hydrol 528:17–28
14. Schellart AN, Liguori S, Kraemer S, Saul A, Rico-Ramirez M (2014) Comparing quantitative precipitation forecast methods for prediction of sewer flows in a small urban area. Hydrol Sci J 59(7):1418–1436
15. Nichols A, Tait S, Horoshenkov K, Shepherd S, Zhang Y (2013) Low cost on-line non-invasive sewer flow monitoring. Water Pract Technol 8(3–4):461–468
16. Baik HS, Jeong HS, Abraham DM (2006) Estimating transition probabilities in Markov chain-based deterioration models for management of wastewater systems. J Water Resour Plan Manag ASCE 132(1):15–24
17. WRc (ed) (2013) Manual of sewer condition classification.5th edn. WRc Publications, Swindon
18. Myrans J, Everson R, Kapelan Z (2018) Automated detection of faults in sewers using CCTV image sequences. Autom Constr 95:64–71
19. Mills GH, Jackson AE, Richardson RC (2017) Advances in the inspection of unpiggable pipelines. Robotics 6:36. https://doi.org/10.3390/robotics6040036
20. Roelveld P, Roorda J, Schaafsma M (2010) The Dutch roadmap for the WWTP of 2030. Dutch Foundation for Applied Water Research (STOWA), Amersfoort
21. Thornton A, Sunner N, Haeck M (2010) Real time control for reduced aeration and chemical consumption – a full scale study. Water Sci Technol 69:2169–2175
22. Grievson O (2018) Wastewater flow control. Inst Water 200:42–43
23. Duncan AP, Chen AS, Keedwell EC, Djordjević S, Savić DA (2013) RAPIDS: early warning system for urban flooding and water quality hazards. In: MaLWaS Sympsium, AISB-IACA conference, University of Exeter, Apr 2013. ISBN: 978-1-908187-33-8
24. Mounce SR, Shepherd W, Sailor G, Shucksmith J, Saul AJ (2014) Predicting CSO chamber depth using artificial neural networks with rainfall radar data. IWA Water Sci Technol 69 (6):1326–1333
25. Zhang D, Lindholm G, Ratnaweera H (2018) Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring. J Hydrol 556(2018):409–418
26. Ostojin S, Mounce SR, Boxall JB (2011) An artificial intelligence approach for optimising pumping in sewer systems. J Hydroinf 13(3):295–306
27. Garofalo G, Giordano A, Piro P, Spezzano G, Vinci A (2017) A distributed real-time approach for mitigating CSO and flooding in urban drainage systems. J Netw Comput Appl 78:30–42
28. Shepherd W, Mounce SR, Ostojin S, Abdel-Aal M, Schellart A, Skipworth PJ, Tait S (2017) Optimising a fuzzy logic real-time control system for sewer flooding reduction using a genetic algorithm. In: Proceedings of computing and control for the water industry (CCWI2017), Sheffield, UK, 5–7 Sept 2017

29. Kepp U, Machenbach I, Weisz N, Solheim O (2000) Enhanced stabilisation of sewage sludge through thermal hydrolysis – three years of experience with full scale plant. Water Sci Technol 42(9):89–96
30. Wall A, Hart M, Mack J, O′Brien M (2009) The use of advanced process monitoring and control optimize energy use on an activated sludge plant. Proc Water Environ Fed 13:3788–3797
31. Blumensaat F, Leitão JP, Ort C, Rieckermann J, Scheidegger S, Vanrolleghem PA, Villez K (2019) How urban storm- and wastewater management prepares for emerging opportunities and threats: digital transformation, ubiquitous sensing, new data sources, and beyond – a horizon scan. Environ Sci Technol 53(15):3488–8498

Using Radial Basis Function for Water Quality Events Detection

Eyal Brill

Contents

Abstract The current chapter demonstrates utilization of radial basis function (RBF) as a tool for detection and classification of abnormal events in water quality. The methodology is based on calibration of a RBF based on historical true events classified by human experts. The aim of the process is selection of parameters that ensure zero false negative events. The chapter describes the main method of using RBF and then compares four different kernel functions which are used for implementing the RBF. The case study part of the chapter illustrates actual analysis of real-world data as well as an illustrative example. The chapter concludes with some practical advice on how kernel functions should be selected for this task.

Keywords Abnormality, Radial basis function, Water quality events

E. Brill (*)
Faculty of Technology Management (MOT), Holon Institute of Technology (HIT), Holon, Israel
e-mail: eyalb@hit.ac.il

Andrea Scozzari, Steve Mounce, Dawei Han, Francesco Soldovieri, and Dimitri Solomatine (eds.), ICT for Smart Water Systems: Measurements and Data Science, Hdb Env Chem (2021) 102: 141–166, https://doi.org/10.1007/698_2019_424,
© Springer Nature Switzerland AG 2019, Published online: 9 May 2020

1 Introduction: The Problem of Water Quality Events Classification

According to ISO24522[1] (ISO standard for water quality events detection in drinking water and wastewater systems), a water quality event (henceforth WQE) is defined as a situation in which "measurements of water quality are not according to what they are expected to be." This does not necessarily mean that these measurements violate regulation rules.

Online monitoring uses several common measurements of water qualities in order to ensure the safety of using it for drinking and sanitation. The most common combination is free chlorine, turbidity, and pH as can be seen in several common systems (http://ec.europa.eu/environment/water/water-drink/index_en.html). However, several other parameters were listed as can be seen in (https://www.epa.gov/wqs-tech/water-quality-standards-handbook).

Given that, in some countries, the regulatory limit for measurement of turbidity is 1.0 NTU[2], any water with measurement above this value is not recommended for drinking. However, in a site in which the average measurement of turbidity is around 0.1 NTU with standard deviation of 0.05 NTU, a measurement of turbidity of 0.75 NTU may be considered abnormal and should be investigated even if it does not violate regulation limits. If it is known that somewhere in the upstream of the sampling point, a pipe repair or maintenance work was performed prior to the event in which such turbidity was measured, this may explain this result and may cause this event to be "expected." The problem of identifying WQE becomes more complicated when measurements include several parameters. In this case additionally to examine each measurement separately, a multiparameter approach should be used. When using such measurements, it is advisable to use the likelihood of the combination additionally to values of the individual sensors. See Amit and Brill [1] and Mounce et al. [2].

Several general methods have been suggested in the past for identifying and classifying multiparameter abnormal events in general cases. These general methods include supervised methods such as regression or regression trees and methods that make use of unsupervised learning such as clustering.

Examples for identifying abnormalities using distance-based or density-based method have been demonstrated in the past in many areas. Knorr and Ng [3] demonstrated distance-based methods. Further improvements were demonstrated by Knorr and Ng [4], for distance-based clustering such as the kMean algorithm. More recent work in this field is presented in Angiulli and Pizzuti [5] and in Ramaswamy et al. [6]. The methods proposed in these works are also based on the kNN algorithm (where kNN stands for multi-k-nearest neighbors). Bay and

[1]ISO24522 is under publication procedures and will be available during winter 2019.

[2]NTU are the standard units for measuring turbidity.

Schwabacher [7] addressed pruning for distance-based outlier algorithms applied for large real-world data sets introduced.

Another clustering philosophy is based on density of points. Examples of such a philosophy are given by Breunig et al. [8, 9], in relation to relative density. Jin et al. [10] showed how some of the calculations of the density function can be skipped or estimated based on distribution type. Tang et al. [11] further improved clustering by adding the idea of a connectivity-based outlier factor, which refers to the number of connections between points.

With regard to WQE specifically, a major work has been done under the EPA (US Environment Protection Agency) framework and including the development of the Canary software. Canary is a free software tool developed by the EPA which aims to detect abnormality in water network [12, 13]. An extensive comparison work (https://www.epa.gov/sites/production/files/2015-07/documents/water_quality_event_detection_system_challenge_methodology_and_findings.pdf) done by the EPA compared the Canary algorithm to other tools. It was reported based on comparison between actual and expected value of water quality measurements that this algorithm does not have general superiority over other machine learning algorithms. That is, using linear regression to predict water quality does not do better than other methods. In this specific report, the Canary was ranked third in many tests based on the amount of false alarm reported. Another conclusion which was pointed by this report is the fact that the Canary has a narrow time window from which its algorithm learns and hence can′t provide conclusions based on large data set. Abnormality detection has two main methods as was detailed by Clark and Hakim [14]: supervised and unsupervised. Examples for WQE detection have also been provided by Skadsen [15], Story et al. [16], Yang et al. [17], Chang et al. [18], and Helbling and VanBriesen [19]. A set of more commercial reports with many useful pieces of information can be found at the following EPA website: http://www.epa.gov/nhsrc/pubs.html.

The current chapter examines the efficiency of radial basis function (RBF) as an identification and classification tool for WQE. RBF is known in the literature as a tool for abnormality detection in other areas. Examples are given by Mellisa et al. [20] for mammograms classification; Padmapriya et al. [21] for brain tumor detection; Rajab and Salleh [22] for classification of diabetes; Mansourkhaki et al. [23] for traffic prediction; and Chun-Cheng Lin and Weichih Hu [24] for detection of abnormal intra-QRS pulses.

The current chapter demonstrates the implementation of RBF in the WQE detection domain. The structure of the chapter is as follows: the second section of the chapter gives a description of RBF and RBF networks. The third section gives a description of parameter values selection. Section 4 describes a lab-based example. Section 5 is the analysis of the data set with the traditional form of RBF. Section 6 shows a similar analysis with several other forms of RBF kernel function. Section 7 concludes the chapter.

2 RBF: Structure and Basic Description

A simple RBF can be mathematically described by Eq. (1). The function refers to a single dimension function which has two parameters. The first is the distance of the value of x from a constant value depicted by μ. And the other one is a scale factor depicted by γ. Figure 1 refers to a specific function with value of $\mu = 5$ and the value of $\gamma = 1$. The figure shows the resulting value of $h(x)$ over a range of x from 0 to 10.

Equation 1: Basic RBF

$$h(x) = e^{-\gamma(x-\mu)^2} \tag{1}$$

The value of μ sets the location of the $h()$ peak on the horizontal axis. The value of γ sets the height of this peak.

The last parameter has an additional influence as can be seen in Fig. 2. As γ absolute value becomes smaller, the function becomes wider and vice versa.

Define parameter μ as the "centroid." In case of function (1), it is a single point on a one-dimensional axis. However, in case of a multidimensional problem, μ becomes a vector. It describes a set of values (one for each dimension), i.e., it is a point in a multidimensional space.

Figure 3 illustrates the calculation of RBF for a single point (the red point) in a two-dimensional space with three centroids (i.e., $\mu = 5$).

As it can be seen, the red point has three distances on the horizontal axis (axis X1) one to each centroid. These are numbered as 1, 2, and 3 in Fig. 3. There are three

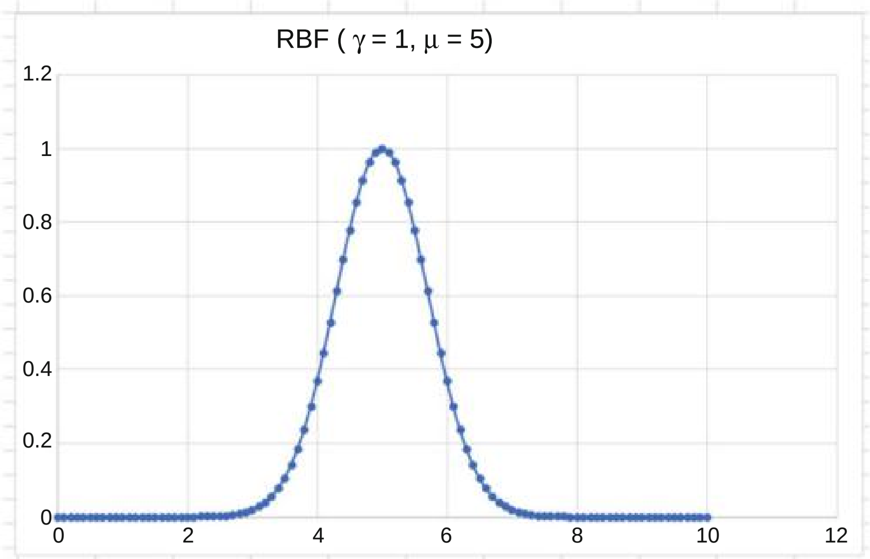


Fig. 1 Chart of function (1) in the range 0 to 10 with $\gamma = 1$ and $\mu = 5$

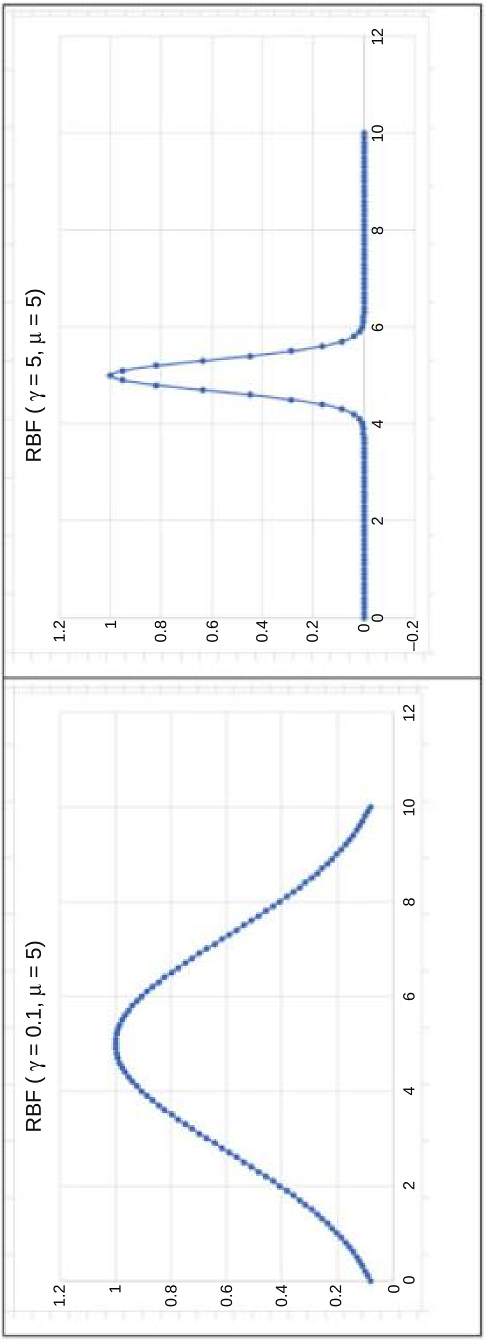


Fig. 2 Influence of γ on the shape of h(x)

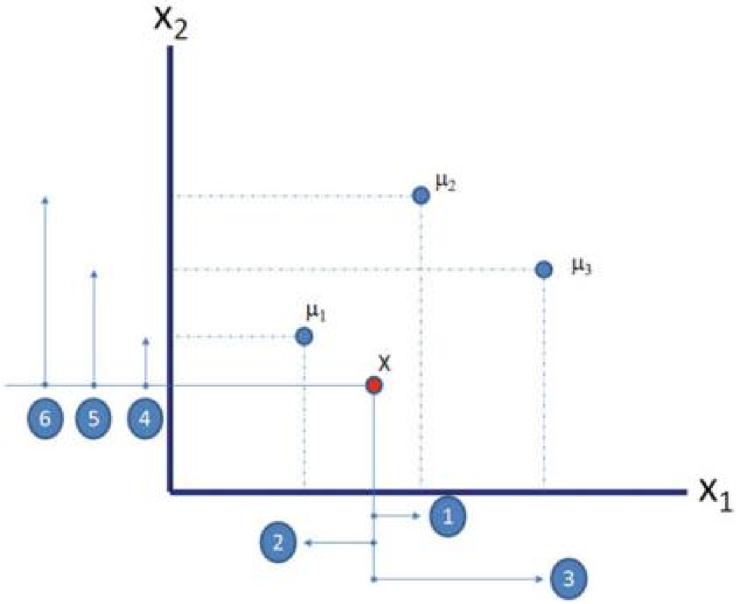
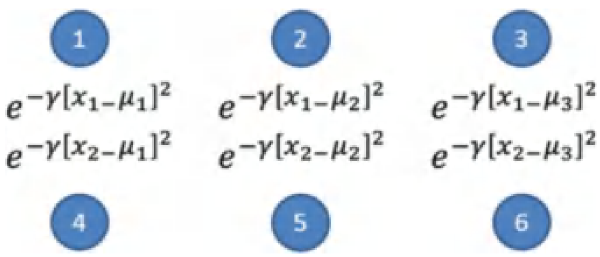


Fig. 3 Illustration of RBF in two dimensions

Fig. 4 RBF in a matrix form



additional distances on the vertical axis (axis X2). These are numbered 4, 5, and 6 in Fig. 3. Figure 4 demonstrates this calculation in a matrix form. As it can be seen, each distance numbered in Fig. 3 populates a cell in the matrix with form as shown in Fig. 4.

In order to complete the implementation of RBF for events classification, two components are needed. First, each axis should be given a different weight. This weight reflects the importance of this dimension/measurement over the final value of the RBF. A second component is a classification policy. The classification policy defines how the result of the weighted summation of all distances over all dimensions is translated into one of two values about each point in the space. These values are true or false. This value specifies that a single point in the multidimensional space, when investigated and given its neighboring centroids, should be considered as a true WQE or false WQE.

It should be noted that some systems give also probability for the classification. However, this probability is based on distance and is affected by the variables normalization and thus may be biased. Hence, the current chapter displays a method which deals only with the binary classification, assuming that a true WQE will require additional human investigation.

It should be noted that the classification process must take into consideration the number of centroids, for example, by normalizing the summation of the distance by the number of centroids (see step 2 in what follows). Examples for the classification process as used in this paper include:

- Step1: Calculate the RBF for a given point.
- Step 2: Divide the result of step 1 by the number of centroids used for this calculation.
- Step 3: If the result of step 2 is above a threshold (henceforth high RBF level (HRL)), classify as true and classify others as false.

The numerical example given in the data set analysis section of this chapter will illustrate this process.

Equation (2) gives an implicit form of this function. The term DET refers to detection process as described above.

Equation 2: RBF with classification

$$h(x) = \mathrm{Det}\left\{\sum_{m=1}^{N} w_m e^{\left(-\gamma[x_m-\mu_m]^2\right)}\right\} \tag{2}$$

Once again, using a matrix the general notation yields Eq. 3.

Equation 3: General form of a RBF where the squared matrix is denoted by Φ

$$\begin{bmatrix} e^{-\gamma[x_1-\mu_1]^2} & e^{-\gamma[x_1-\mu_2]^2} & \cdots & e^{-\gamma[x_1-\mu_N]^2} \\ e^{-\gamma[x_2-\mu_1]^2} & e^{-\gamma[x_2-\mu_2]^2} & \cdots & e^{-\gamma[x_2-\mu_N]^2} \\ \vdots & \vdots & \cdots & \vdots \\ e^{-\gamma[x_K-\mu_1]^2} & e^{-\gamma[x_K-\mu_2]^2} & \cdots & e^{-\gamma[x_K-\mu_N]^2} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} \tag{3}$$

The equation set above can be written in short as $\Phi w = y$ where Φ is the term in the first squared parenthesis and w is the correspondent weights vector.

3 RBF Parameters Selection

As it was set out earlier, the values of W(s) and HRL should be defined or initialized in some manner. We turn now into the last stage of the methodology. This is the stage that involves setting the values of these parameters. Let us assume that the set of known centroids (which will be called the training set) is based on known and

Table 1 Results of detection

	Model classification		
Actual classification		True	False
	True	True positive (TP)	False negative (FN)
	False	False positive (FP)	True negative (TN)

classified events. A human expert has manually classified each one of these events as true. Additionally, we have a set of other K points with the same structure. This set will be named as the calibration set. For this set we have for each point a human-defined classification. However, this set (unlike the training set) includes both points that were classified as true and points that were classified as false. For each point of the calibration set (called a selected point), one can calculate using the RBF algorithm, the value of h() based on the set of neighboring points from the training set. As it was explained previously, points which are located near the selected point will have high influence on the value of h(), while points located far from the selected point will have low influence on the h() value of the selected point, or even zero influence. Based on the calculated result of the h() and the classification policy of HRL, the selected point will be classified by the algorithm as true or false. This means that each point in the test set will have two classifications (reference human classification and algorithmic). The human classification can also be called the actual ("correct") classification since the human expert bases it on actual observation of reality. And the model classification is the classification obtained by the algorithm for each point in the test set.

After this process is repeated for all points from the calibration set, the results can be assembled as a confusion matrix as shown in Table 1. Such a matrix has been introduced, for example, by Cohen′s kappa (1960).

- True Positive – a point that was classified as true by both the algorithm and the human expert
- False Positive – a point that was classified as true by the algorithm and as false by the human expert
- False Negative – a point that was classified as false by the algorithm and as true by the human expert
- True Negative – a point that was classified as false by both the algorithm and the human expert

Each of the TP, FP, FN, and TN values is a count of the number of events satisfied in the correlated condition as shown in Table 1 (and in the four points above). Using these numbers, it is also possible to calculate the sensitivity and specificity of the results. Sensitivity (also called the TP rate, the recall, or the probability of detection) measures the proportion of actual positives that are correctly identified out of total number of true cases – i.e., TP/(TP + FN). Specificity (also called the TN rate) measures the proportion of actual negatives that are correctly identified out of the total number of false cases – i.e., TN/(TN + FP). In some cases where the cost of FP or the cost of FN is extremely high, these ratios are very important.

A special note should be made about how the counting of events is performed. Since many monitoring systems produce a reading every minute, if each reading is counted separately, a biased result would be achieved due to the existence of long events and short events. For example, if (as an extreme case) a situation with a wrong measurement (due to instrument dropouts or calibration drift, etc.) produces a FP condition which lasts for several days (until the measuring device is fixed), a large amount of records classified as FP will be counted and will adversely affect the result of the model accuracy. In order to avoid such a situation, counting of events should be referenced to events and not to single records. Hence, an event needs to be defined.

An event is a situation in which the value of RBF crossed the HRL to the high side for a substantial amount of time (this amount of time should be set by the users) after being located below the HRL for a substantial amount of time (also this amount of time should be set by the user).

The above definitions imply that two time lags should be defined. The first one is how long the algorithm should wait, when the process crosses the HRL to the upper side, to declare an event. This one will be called a Delay On time. The second one is how long the algorithm should wait when a process crosses the HRL to the lower side before an end of event should be declared. This one will be called a Delay Off time. Based on the two delay time definitions, the algorithm classification method should be updated with the following additional rules:

Rule 1 In case of a long period of no event, the process crosses the HRL to the upper side and then goes back to the lower side of the HRL, and a TN event will be declared. See Fig. 5 as an example of Delay On.

Figure 5 is a representation of a TN since in this case the system (the algorithm) did not generate an alarm even if a violation of the HRL occurred. This is due to the Delay On mechanism.

Rule 2 In the case that, after an event has started, the process oscillates above and below the HRL without achieving either the Delay On or Delay Off amount of time, the counting of the events will not be incremented. See Fig. 6 for an example.

The values of the Delay On time and the Delay Off time should also be added to the process of algorithm parameter selection. Thus, the target of the algorithm tuning is a selection of the values of W(s), HRL, Delay On, and Delay Off in order to optimize the results described in Table 1. Once again it should be noted that optimizing the values in Table 1 means optimizing the Cohen′s kappa, the sensitivity and specificity.

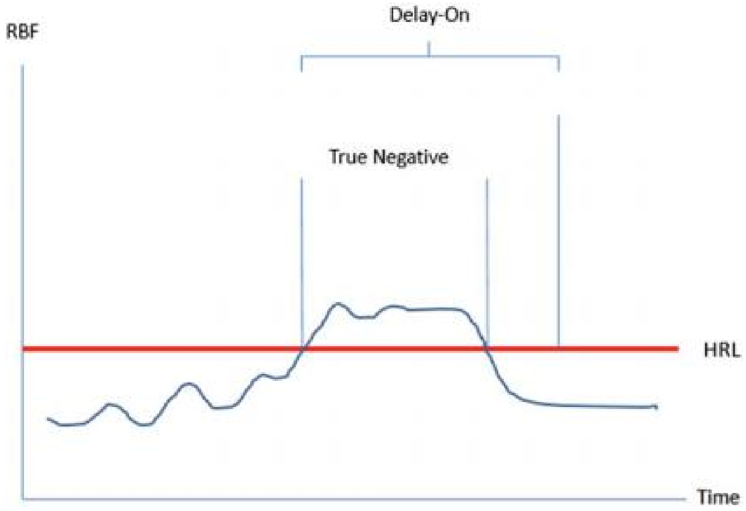


Fig. 5 Example of a true negative

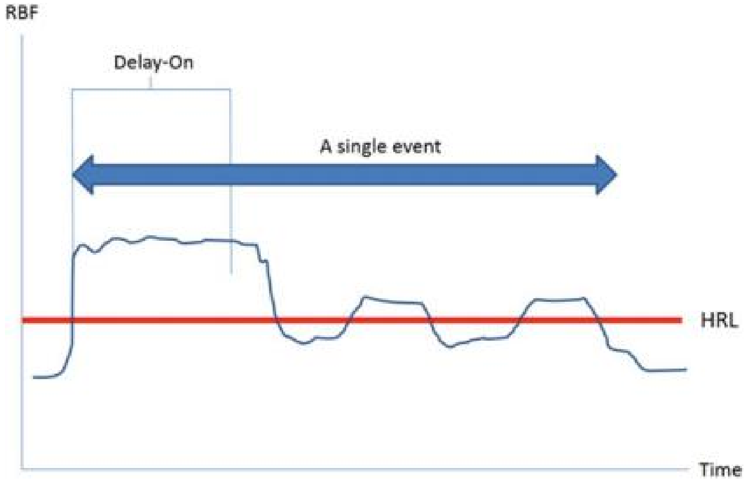


Fig. 6 RBF fluctuations as a single event

For the purpose of this chapter, it is defined that the optimal parameter selection is such that all true events are detected with zero FN events and minimal FP events. This policy is aligned with the idea that a contamination event may be hazardous for humans. Hence, the target is to avoid FN events completely.

4 An Illustrative Case Study

This section describes a numerical example, which implements the above framework, and utilizes laboratory rig generated results. The purpose of this example is to demonstrate the implementation of the algorithm to a level in which the reader can follow the calculation results. The data set in this example contains 1-day samples with several artificial events than have been injected into the data in order to illustrate the detection of these events using the RBF.

Figure 7 shows a 1-day chart of water quality with three measurements: turbidity, free chlorine, and pH.

Data is recorded with 5-min intervals between records. This gives $24*12 = 288$ records per day. The pH measurement is displayed with blue color. Its actual range varies between 7.0 and 7.5 with two local peaks at a level near 8.0 around record 190 and 270. The free chlorine and turbidity are displayed with green and red colors, respectively. Turbidity normal level is around 0.37, and free chlorine normal level oscillates around 0.4 with two additional noise effects. First, during midday (around record number 140) when the temperature is high, the level of free chlorine is lower due to extensive evaporation. Second, the level of free chlorine is affected by the dosing system which works in "open loop control." These two effects create the smooth U shape with hourly fluctuation of the free chlorine curve.

Five abnormal events can be noticed in Fig. 7. They are numbered 1–5. Events 1, 2, 3, and 5 were classified as false. Only event 4 was classified as true by an expert. The task now is to calibrate parameter values of the RBF algorithm in order to achieve the same classification. The steps for the manual calculation to obtain the optimal values are explained below.

The five events shown in Fig. 7 form the calibration set. The training set for this problem includes two points. These points are listed in Table 2.

Given the values of the centroids in Table 2 for each point in Fig. 7, a value of RBF was calculated based on the inner part of Eq. 2. For example, the value for the first point of chart 7 has the values of pH = 7.077, free Cl = 0.407, and turbidity = 0.322. Hence, its RBF value is calculated using:

$$
\begin{aligned}
&\exp^{-(7{:}077-7{:}5)^2} + \exp^{-(4{:}07-0{:}18)^2} + \exp^{-(0{:}322-0{:}70)^2} + \exp^{-(7{:}077-7{:}2)^2}\\
&\quad + \exp^{-(4{:}07-0{:}28)^2} + \exp^{-(0{:}322-0{:}55)^2}\\
&\quad = 5{:}571
\end{aligned}
$$

Performing the calculation for each of the points in Fig. 7 yields Fig. 8, which shows the RBF curve for the same group of records.

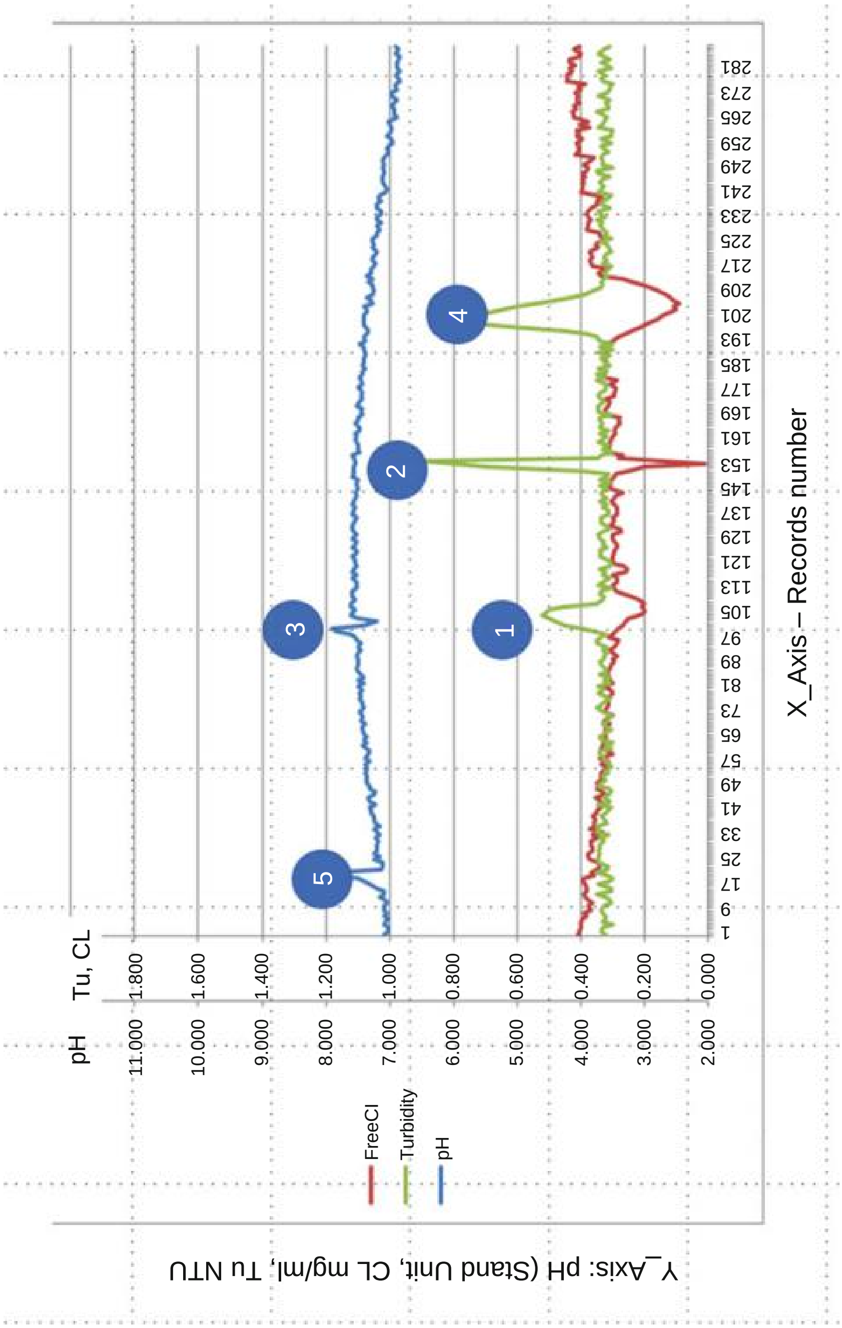


Fig. 7 One day of quality samples

Table 2 Centroids for the case study

	pH	Free chlorine	Turbidity
Weight	1.0	1.0	1.0
Point 1	7.5	0.18	0.7
Point 2	7.2	0.28	0.55

As it can be seen from Fig. 8, the most common values of RBF are between 5.7 and 5.8. Since the lower is the value of the RBF, the further it is from centroids, and since centroids represent in our case true events, values below 5.7 have no interest. Figure 8 has three abnormal spikes labelled as A, B, and C. These spikes correspond to points 1, 2, and 4 in Fig. 4. Setting the HRL to a value between 5.87 and 5.90 and setting the Delay On time to a value of more than 3 min will enable the algorithm to label point C as true and points A and B as false. This will make point C in Fig. 8 (which is point 4 in Fig. 7) a TP event and point B in Fig. 7 (which is point 2 in Fig. 7) and point A in Fig. 8 (which is point 1 in Fig. 7) TN events. This setting will achieve the target of detecting all TP events (according to the manual classification) with no FN events. This section shows a very simplified example for RBF with few injected events. In reality, however, training and calibration sets may contain a large amount of records with many events. The number of dimensions may be 4, 5, or 6 and the length of an event may very between a few minutes to a few hours. The next section shows the results of the implementation of the RBF algorithm to real-world data.

The RBF algorithm has additional parameters such as variables weight (the w values) and function height normalizer (γ). These parameters may give an additional degree of freedom for the calibration process. However, for simplicity these parameters have been kept fixed in the current analysis with a value of 1.

5 Real-World Data Analysis

The above algorithm has been implemented on a “real-world” data set taken from the monitoring station located in a large city with more than half a million residences. The data set includes 2 years of data with sampling intervals of 1 min. The data set includes data from January 1, 2017, to September 30, 2018. The first year (2017) was used as the training set. This period included 41 events that were tagged manually. Out of this list, seven events were classified as false events, and the rest were classified as true events. As explained, only the true events were used. The second part of the data which was used as calibration set included 25 events. From which 4 events were classified as true events and the remaining 21 as false events. The difference between the ratio of true and false events with respect to the training and calibration set may seem strange. However, this is a real-world data set that was manually classified.

Based on the above calibration set, an RBF value was calculated for each record. The result of the RBF is shown in Fig. 9.

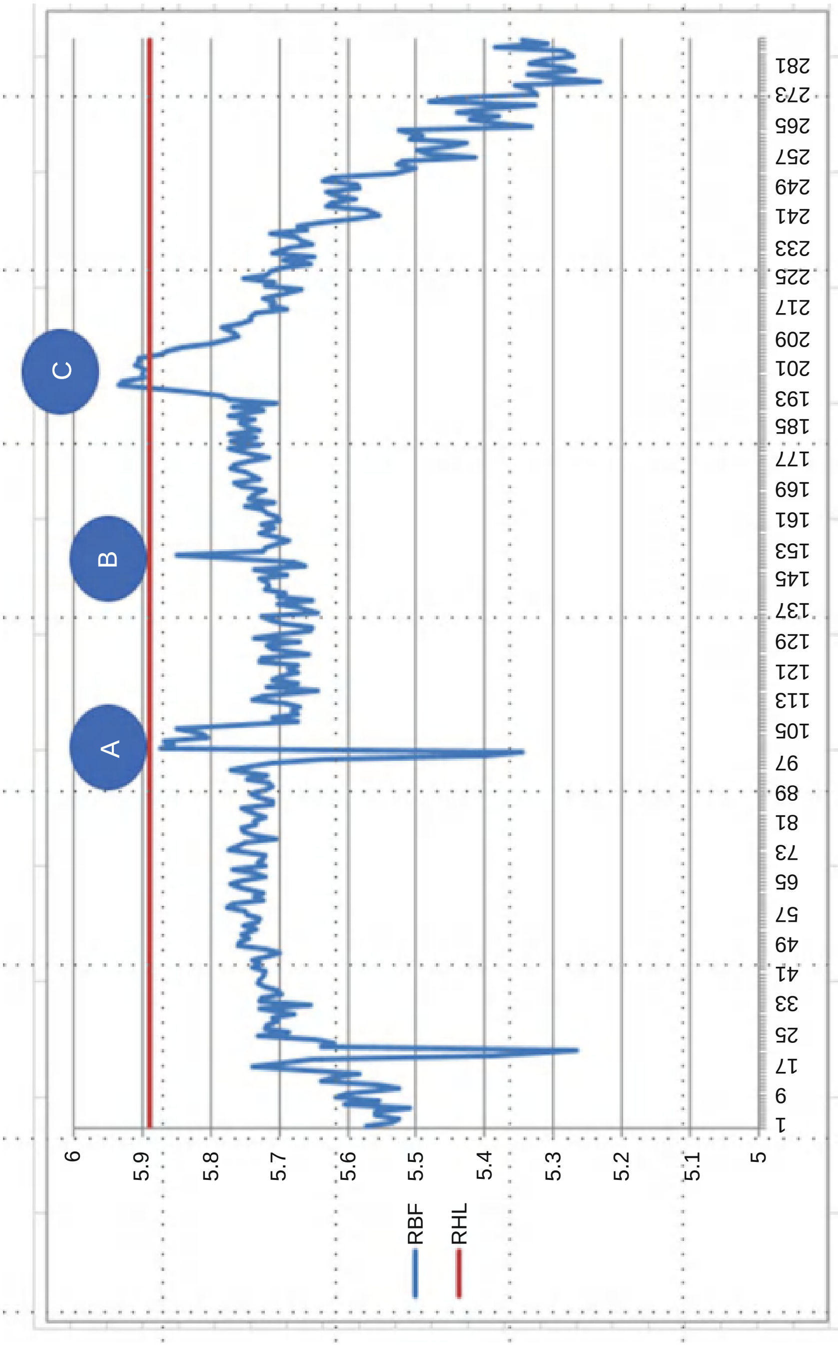


Fig. 8 RBF values for 1-day sample

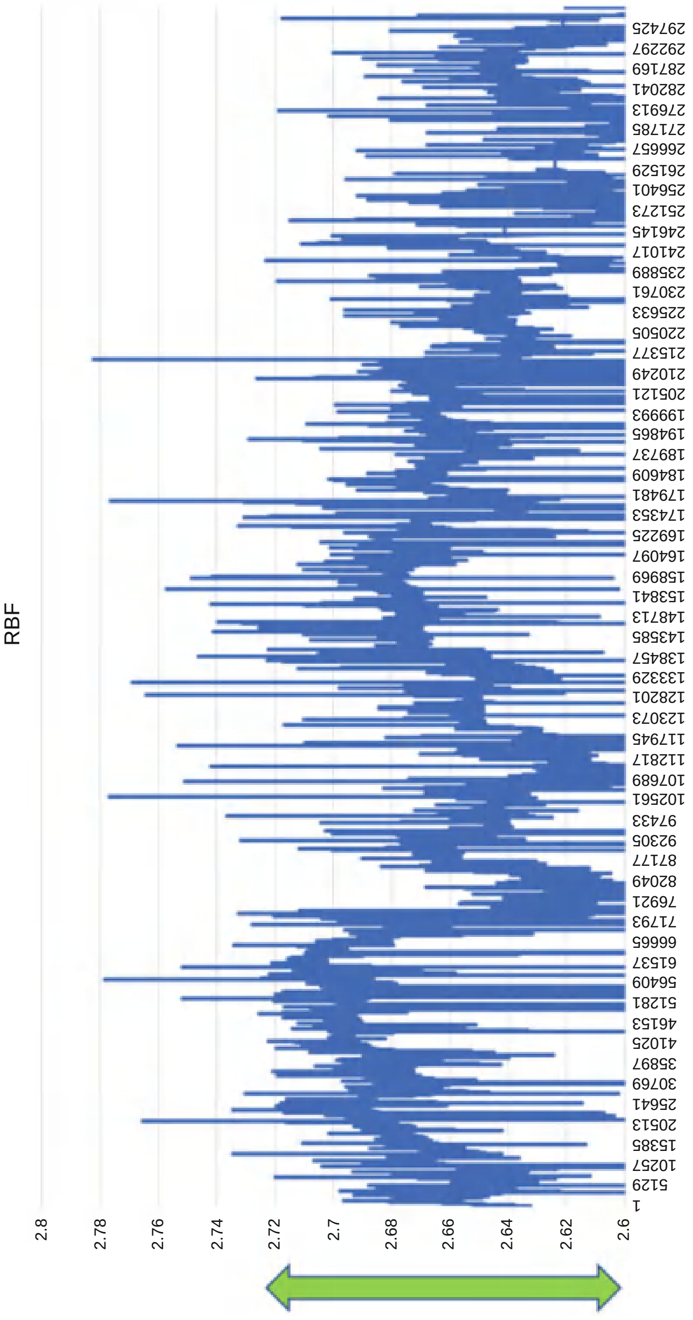


Fig. 9 RBF chart of the calibration set

Table 3 Summary results of Run 1

HRL	Delay	TP	TN	FP	FN	Sen	Spe
2.61	10	4	2	19	0	1.00	0.10
2.62	10	3	3	18	1	0.75	0.14
2.61	15	4	2	19	0	1.00	0.10
2.62	15	3	3	18	1	0.75	0.14
2.61	30	4	5	16	0	1.00	0.24
2.62	30	3	5	16	1	0.75	0.24
2.6	60	4	6	15	0	1.00	0.29
2.61	60	3	7	14	1	0.75	0.33
2.6	90	4	8	13	0	1.00	0.38
2.61	90	2	10	11	2	0.50	0.48
2.6	120	4	9	12	0	1.00	0.43
2.61	120	2	11	10	2	0.50	0.52

The RBF algorithm was used to detect abnormalities in the data. The full detailed results are given in the table in Appendix. According to what it is shown there, the HRL range was between 2.58 and 2.72, and the delay time range was between 10 and 120 min. A summary of the results (lines highlighted in yellow in the table in Appendix) is shown in Table 3 below. The rows in Table 3 were selected with a specific policy. This policy was aimed to demonstrate for each Delay On time (10, 15, 30, etc.), in which HRL the lowest result of non-zero FN is obtained. The reason for this policy is that the water utility would like first of all to reduce the FN to minimum, and for that it has to choose the correct HRL. For example, for a delay time of 10 min, an HRL value of 2.61 results in zero FN events, while a HRL value of 2.62 results in 1 FN event.

As it can be seen from Table 3 (and from the Appendix), when 10 min delay is used, the optimal value for HRL that detects all TP events is a value of 2.61. If a value of 2.62 is used, the system will “miss” one TP event in favor of one FN event. Assuming that zero FN events are the ultimate target in a water quality system, this should therefore be the setup. However, as it can be seen from the first row of Table 3, detecting all TP events with such a setup will come with an “organizational cost” of 19 FP events. This is labelled as an “organizational cost” because every FP event requires allocation of resources for verification, for example, sending a sampling team to the field.

One may ask whether other setups might enable the same level of detection with a lower level of false alarm. The answer is yes. As it can be seen from Table 1, using a longer delay time will enable a similar level of detection (all four true events are detected). This can be seen from the last two lines of Table 3 where the HRL of 2.6 ends with four TP events, zero FN events, and ten FP events. However, these results should be examined bearing in mind the following. The difference between a delay of 10 min and a delay of 120 min in a big water system, in case of real contamination, may be the difference between affecting several thousands of citizens and affecting several hundreds of thousands of citizens. This is due to the additional network sections that will be contaminated during the additional time. The trade-off between

10 and 19 FP events needs examining. The meaning of ten false events over 9 months is on average one false alarm every month. The meaning of 19 false alarms over 9 months is on average a false alarm every second week. In both cases, a water utility will need a sampling team that will be ready to go out per call and take manual samples to approve or disapprove the indication of the automatic system. Choosing between the two options is strictly a managerial decision. However, it will be logical to assume that in most cases an experienced manager will not let an automatic system shut down water supply without a second manual examination of samples from the "contaminated" area.

Obviously, if the level of false alarms in the case of a short delay of 10 min would result in hundreds of alarms per month, the correct policy would be to set a delay of 2 h or more. In this case it seems that a selection of RHL $= 2.61$ and delay time of 10 min seem logical.

6 Using Other Kernel Functions for RBF

As it was explained earlier in this chapter, other forms of functions may be used as a kernel function for the RBF. In what follows three different forms of kernel functions are examined and compared to the first functional form, which has been used until now. These functions are:

- Run 2: Inverse quadratic – $\varnothing(r) = \frac{1}{r}$
- Run 3: Inverse quadratic – $\varnothing(r) = \frac{1}{1+(r)^2}$
- Run 4: Inverse multi-quadratic – $\varnothing(r) = \frac{1}{\sqrt{1+(r)^2}}$

The results and RBF chart of Run 2 are displayed in Fig. 10 and Table 4, respectively.

Once again, the analysis of the function performance is focused on the point where it loses the "first TP event." This happens when the number of FP events is 18 with a delay of 10 min (see second row in Table 4). In case of a delay of 120 min, this happens at a level of 16 FP events. As it can be seen from Fig. 10, the noise generated by the second RBF is substantially large relative to the range of the output of the RBF. The two red arrows in Fig. 10 show that the ratio between the noise and the range of the RBF is in some periods 50% to 75% of the range. This is a major disadvantage for this function.

This means that relative to the performance of Run 1 (see Table 3), the Run 2 function performed poorly.

The results and RBF chart of Run 3 are displayed in Fig. 11 and Table 5, respectively.

Once again, the results are worse than Run 1. Losing the first TP event reduces only four FP events and increases the delay time from 10 to 120 min. However, as it

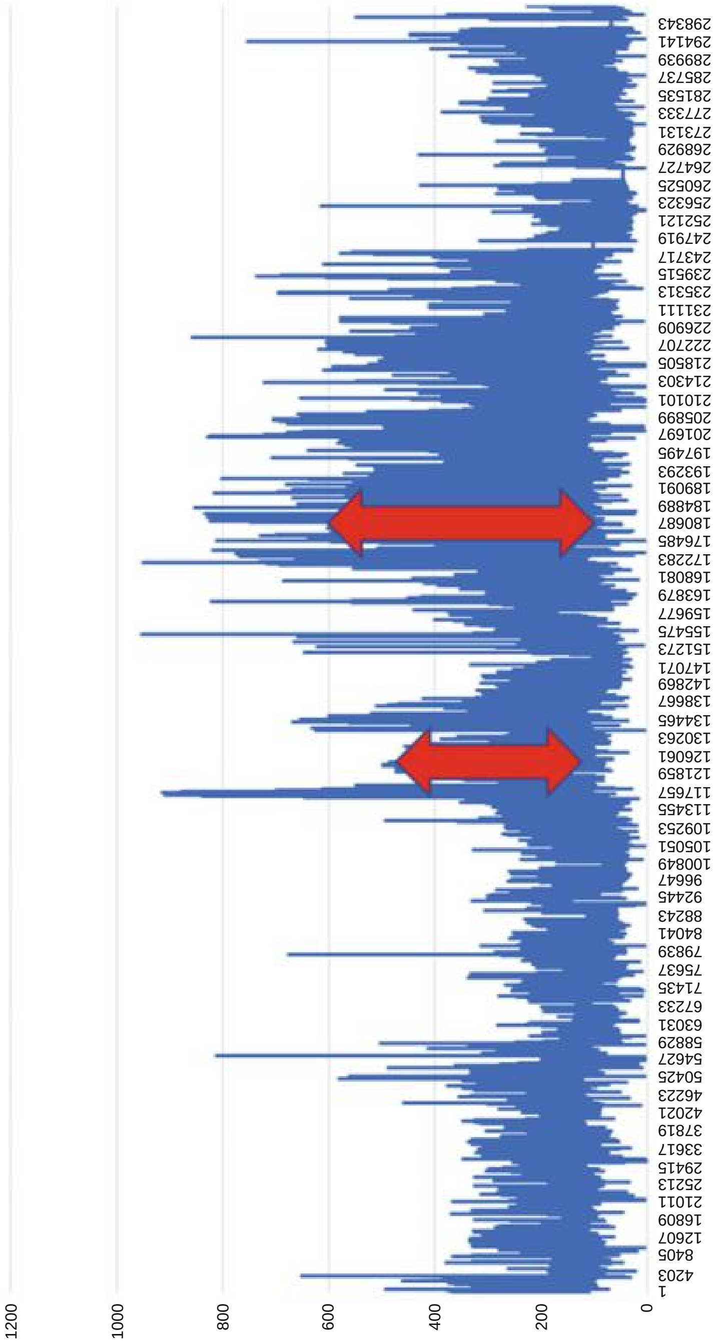


Fig. 10 RBF of Run 2

Table 4 Results of Run 2

HRL	Delay	TP	TN	FP	FN	Sen	Spe
75	10	4	1	20	0	1.00	0.05
100	10	1	3	18	3	0.25	0.14
25	120	4	5	16	0	1.00	0.24
50	120	3	5	16	1	0.75	0.24

can be seen from Fig. 11, the noise of this function is relatively small. This can be seen from the red arrows in Fig. 11.

The last results of Run 4 are shown in Fig. 12 and Table 6.

This function performs better than the other runs. As it can be seen from Table 6, by choosing HRL between 2.82 and 2.83, it is possible to reduce the number of FP events from 20 to 13 with a cost of a single FN event without increasing the delay time.

One may ask at this point, what differs the four functions one from the other, and is it possible just to be looking at the RBF chart to "predict" which RBF will perform well and which will not? A closer look at Figs. 3, 4, 5, and 6 may reveal the "patterns" of good functions and "bad" functions. Function 2, which had the worst performance, has a relatively wide "blue band." This band is marked with red arrows in Figs. 4 and 5. And as it can be seen, the band of function 2 is bigger. A bigger band makes it difficult to observe when a sample leaves the normal zone for the abnormal zone. A second characteristic, which differs good from bad functions, is the size of the range where fluctuation of the RBF is located. Comparing Figs. 3, 4, 5, and 6, one may observe that in Fig. 9 the majority of RBF values are located between 2.6 and 2.72 (see green arrow in Fig. 9). However, in Fig. 12 the majority is between values 2.8 and 2.86 (see green arrow as Fig. 12). Hence the range of fluctuations in Fig. 9 is twice as big as the range in Fig. 12. This gives the second criteria for different RBFs in the vertical axis.

The smaller the range of fluctuations, the easier it is to detect abnormality.

7 Concluding Remarks

This chapter has demonstrated one approach of how RBF can be used to classify abnormal events in water supply systems directly from manually tagged sensor data. The methodology is based on prior (manual) classification of events into true and false events. Once enough true events exist, the set of these events may be used as a training set. From this stage a second set of events (which contains true and false) are used for the calibration of the RBF algorithm. The parameters of the algorithm include the value of the RBF, which is used as a threshold, and the delay time before an alert is declared. As it was explained early, for the sake of simplicity, the additional parameters of the RBF have been kept fixed in the current analysis with a value of 1.

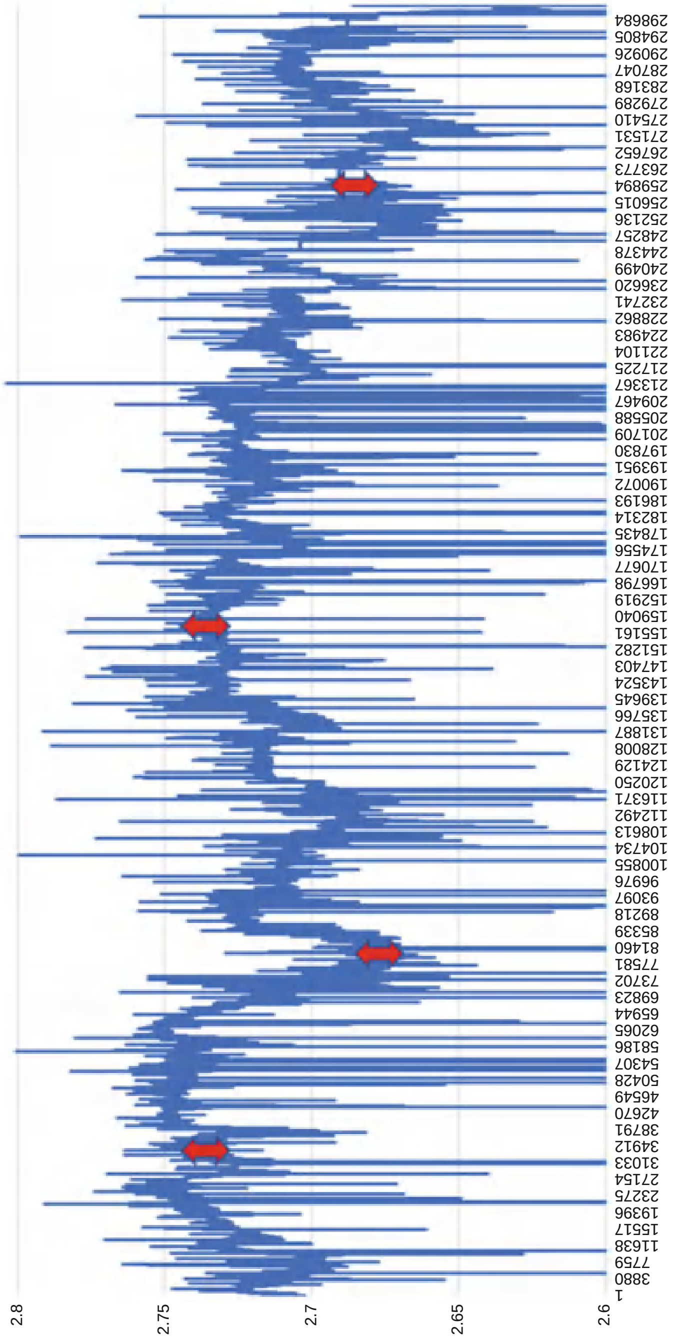


Fig. 11 RFB of Run 3

Table 5 Results of Run 3

HRL	Delay	TP	TN	FP	FN	Sen	Spe
2.68	10	4	1	20	0	1.00	0.05
2.69	10	3	2	19	1	0.75	0.10
2.66	120	4	5	16	0	1.00	0.24
2.67	120	3	5	16	1	0.75	0.24

The performance analysis of each kernel function focused on the main target of not losing true events, i.e., zero FN events. Hence the simulation was looking for those parameters, which differ between 100% TP events with minimal FP events. Based on real-world data and for a specific data set, it was demonstrated that it is possible to achieve such a target with a level of FP events between one and two FP events per month. The analysis also clarified the implication of reducing the level of FP events in relation to the delay time. As it can be seen from the results, going from two FP events per month to one FP event (in relation to the demonstration data) implies extending the delay time from 10 to 120 min. This change has implications for the amount of water users affected by a contamination event.

Finally, the analysis focused on the possibility to “predict” when RBF may perform well or not. As it was explained earlier, this includes two characteristics. The function should have a “thin” band and small range of fluctuations. Although this is not an exact quantitative characteristic (and somewhat subjective), it gives an efficient tool to select and rank RBFs visually with respect to one another.

The above analysis has both pros and cons. On the negative side, it is limited to a single data set, and selection of algorithm parameters was done manually, and not all possible degrees of freedom were used; hence optimal configuration is not guaranteed. On the positive side, it is a simple approach that can be implemented relatively quickly with small computational effort. Future research may include additional data sets with automatic methods for parameter selection.

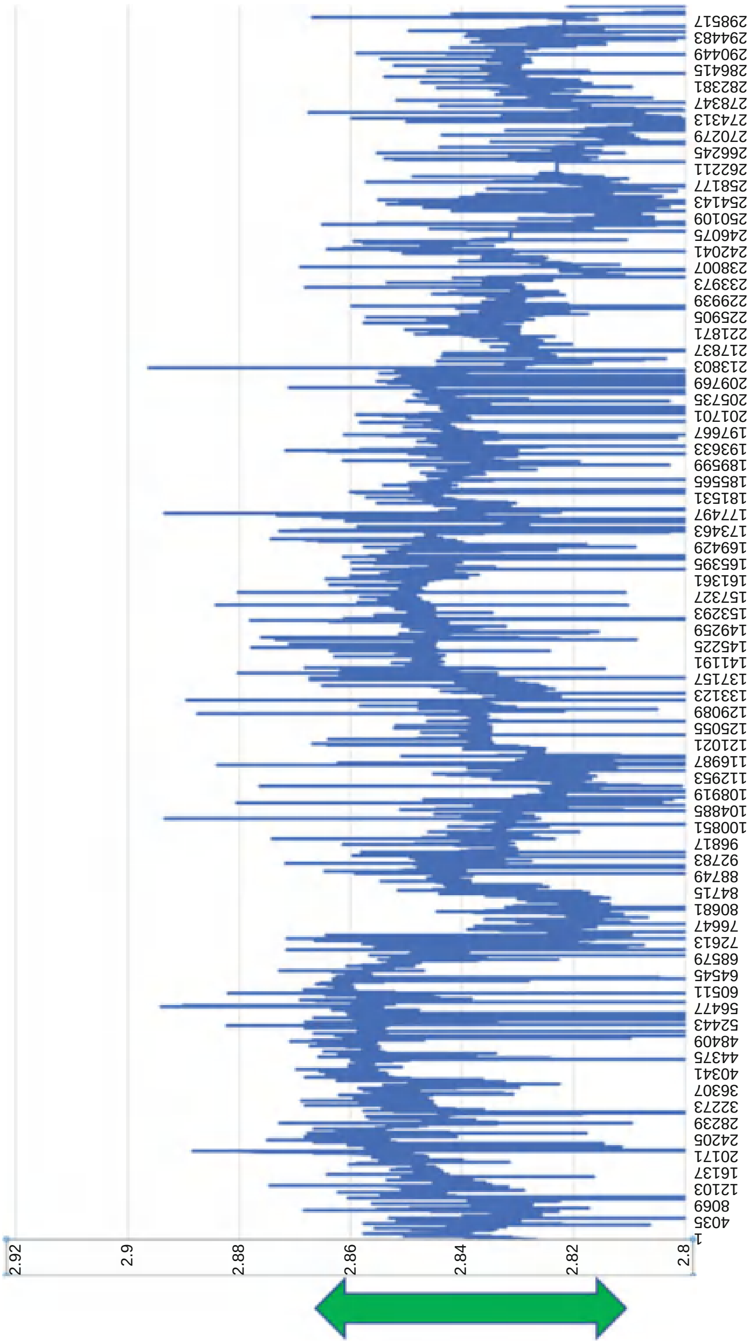


Fig. 12 RFB of Run 4

Table 6 Results of Run 4

HRL	Delay	TP	TN	FP	FN	Sen	Spe
2.82	10	4	1	20	0	1.00	0.05
2.83	10	3	8	13	1	0.75	0.38
2.81	120	4	5	16	0	1.00	0.24
2.82	120	3	5	16	1	0.75	0.24

Appendix: Result of Run 1

HRL	Delay	TP	TN	FP	FN
2.58	10	4	0	21	0
2.59	10	4	0	21	0
2.6	10	4	1	20	0
2.61	10	4	2	19	0
2.62	10	3	3	18	1
2.63	10	3	6	15	1
2.64	10	3	9	12	1
2.65	10	3	9	12	1
2.66	10	3	9	12	1
2.67	10	1	10	11	3
2.68	10	1	10	11	3
2.69	10	1	11	10	3
2.7	10	0	21	0	4
2.71	10	0	21	0	4
2.72	10	0	21	0	4
2.58	15	4	1	20	0
2.59	15	4	1	20	0
2.6	15	4	2	19	0
2.61	15	4	2	19	0
2.62	15	3	3	18	1
2.63	15	3	6	15	1
2.64	15	3	9	12	1
2.65	15	3	10	11	1
2.66	15	2	10	11	2
2.67	15	1	10	11	3
2.68	15	1	10	11	3
2.69	15	1	12	9	3
2.7	15	0	21	0	4
2.71	15	0	21	0	4
2.72	15	0	21	0	4
2.58	30	4	3	18	0
2.59	30	4	3	18	0
2.6	30	4	4	17	0
2.61	30	4	5	16	0

(continued)

HRL	Delay	TP	TN	FP	FN
2.62	30	3	5	16	1
2.63	30	3	7	14	1
2.64	30	2	10	11	2
2.65	30	2	10	11	2
2.66	30	2	10	11	2
2.67	30	1	10	11	3
2.68	30	1	10	11	3
2.69	30	1	16	5	3
2.7	30	0	21	0	4
2.71	30	0	21	0	4
2.72	30	0	21	0	4
2.58	60	4	5	16	0
2.59	60	4	5	16	0
2.6	60	4	6	15	0
2.61	60	3	7	14	1
2.62	60	3	8	13	1
2.63	60	3	9	12	1
2.64	60	1	10	11	3
2.65	60	1	10	11	3
2.66	60	1	10	11	3
2.67	60	0	10	11	4
2.68	60	0	10	11	4
2.69	60	0	17	4	4
2.7	60	0	21	0	4
2.71	60	0	21	0	4
2.72	60	0	21	0	4
2.58	90	4	7	14	0
2.59	90	4	7	14	0
2.6	90	4	8	13	0
2.61	90	2	10	11	2
2.62	90	2	10	11	2
2.63	90	2	12	9	2
2.64	90	1	12	9	3
2.65	90	1	12	9	3
2.66	90	1	12	9	3
2.67	90	0	12	9	4
2.68	90	0	12	9	4
2.69	90	0	18	3	4
2.7	90	0	21	0	4
2.71	90	0	21	0	4
2.72	90	0	21	0	4
2.58	120	4	8	13	0
2.59	120	4	8	13	0

(continued)

HRL	Delay	TP	TN	FP	FN
2.6	120	4	9	12	0
2.61	120	2	11	10	2
2.62	120	2	11	10	2
2.63	120	2	13	8	2
2.64	120	1	13	8	3
2.65	120	1	13	8	3
2.66	120	1	13	8	3
2.67	120	0	13	8	4
2.68	120	0	13	8	4
2.69	120	0	18	3	4
2.7	120	0	21	0	4
2.71	120	0	21	0	4
2.72	120	0	21	0	4

References

1. Amit B, Brill E (2018) New approach for estimation of detention time and prediction of quality in water networks. Water Qual Res J 53(2):72–85
2. Mounce S, Machell J, Boxall J (2012) Water quality event detection and customer complaint clustering analysis in distribution systems. Water Supply 12(5):580–587
3. Knorr EM, Ng RT (1997) A unified approach for mining outliers. In: Proceedings conference of the Centre for Advanced Studies on Collaborative Research (CASCON), Toronto, Canada
4. Knorr EM, Ng RT (1998) Algorithms for mining distance-based outliers in large datasets. In: Proceedings of the international conference on very large data bases (VLDB), New York, NY
5. Angiulli F, Pizzuti C (2002) Fast outlier detection in high dimensional spaces. In: Proceedings of the European conference on principles of knowledge discovery and data mining, Helsinki, Finland
6. Ramaswamy S, Rastogi R, Shim K (2000) Efficient algorithms for mining outliers from large data sets. In: Proceedings of the ACM SIGMOD international conference on management of data (SIGMOD), Dallas, TX
7. Bay SD, Schwabacher M (2003) Mining distance-based outliers in near linear time with randomization and a simple pruning rule. In: Proceedings of the international conference on knowledge discovery and data mining (KDD), Washington, DC
8. Breunig MM, Kriegel H-P, Ng RT, Sander J (1999) OPTICS-OF: identifying local outliers. In: Proceedings of the European conference on principles of data mining and knowledge discovery (PKDD), Prague, Czech Republic
9. Breunig MM, Kriegel H-P, Ng RT, Sander J (2000) LOF: identifying density-based local outliers. In: Proceedings of the ACM SIGMOD international conference on management of data (SIGMOD), Dallas, TX
10. Jin W, Tung A, Han J (2001) Mining top-n local outliers in large databases. In: Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining (SIGKDD), San Francisco, CA
11. Jie T, Hang L, Yunbo C, Zhaohui T (2005) Email data cleaning. KDD′05, August 21-24, 2005, Chicago, IL, USA. Copyright 2005 ACM 1-59593-135-X/05/0008

12. EPA (2010) Water quality event detection systems for drinking water contamination warning systems: development, testing and application of CANARY. EPA/600/R-10/036. US Environmental Protection Agency, Washington, DC. http://www.epa.gov/ord
13. Murray R, Haxton T, McKenna SA, Hart DB, Klise K, Koch M, Vugrin ED, Martin S, Wilson M, Cruz V, Cutler L (2010) Water quality event detection systems for drinking water contamination warning systems development, testing, and application of CANARY. EPA/600/R-10/036, Cincinnati, OH
14. Brill E (2014) Implementing machine learning algorithms for water quality event detection: theory and practice. In: Clark RM, Hakim S (eds) Securing water and wastewater systems series: protecting critical infrastructure, vol 2. Springer, Cham, pp 107–122
15. Skadsen J (2008) Distribution system on-line monitoring for detecting contamination and water quality changes. J AWWA 100:7
16. Story MV, Van der Gaag B, Burns B (2011) Advances in on-line drinking water quality monitoring and early warning systems. Water Res 45:741–747
17. Jeffrey Yang Y, Haught RC, Goodrich JA (2009) Real-time contaminant detection and classification in a drinking water pipe using conventional water quality sensors: techniques and experimental results. J Environ Manage 90(8):2494–2506
18. Chang N-B, Pongsanone NP, Ernest A (2012) A rule-based decision support system for sensor deployment in small drinking water networks. J Clean Prod 29–30:28–37
19. Helbling DE, VanBriesen JM (2008) Continuous monitoring of residual chlorine concentrations in response to controlled microbial intrusions in a laboratory-scale distribution system. Water Res 42(12):3162–3172
20. Mellisa P, Alexandera JH, Sakka N (2015) Mammograms classification using gray-level co-occurrence matrix and radial basis function neural network. International conference on computer science and computational intelligence (ICCSCI 2015). Procedia Comput Sci 59:83–91
21. Padmapriya P, Manikandan K, Jeyanthi K, Renuga V, Sivaraman J (2016) Detection and classification of brain tumor using radial basis function. Indian J Sci Technol 9(1). https://doi.org/10.17485/ijst/2016/v9i1/85758
22. Rajab T, Salleh R (2017) Classification of diabetes disease using backpropagation and radial basis function network. UTM computing proceedings innovations in computing technology and applications, vol 2
23. Mansourkhaki A, Berangi M, Haghiri M (2018) Comparative application of radial basis function and multilayer perceptron neural networks to predict traffic noise pollution in Tehran roads. J Ecol Eng 19(1):113–121
24. Chun-Cheng L, Weichih H (2011) A radial basis function neural network for the detection of abnormal intra-QRS potentials. Computing in cardiology conference 2011, Hangzhou, China

Promoting Smart Water Systems in Developing Countries Through Innovation Partnerships: Evidence from VIA Water-Supported Projects in Africa

Silas Mvulirwenande and Uta Wehn

Contents

Abstract Over the past decades, the potential of Information and Communication Technologies (ICTs) to improve water management has been demonstrated.

S. Mvulirwenande (*) and U. Wehn
IHE Delft Institute for Water Education, Delft, The Netherlands
e-mail: nzanyi@yahoo.fr

Andrea Scozzari, Steve Mounce, Dawei Han, Francesco Soldovieri, and Dimitri Solomatine (eds.), ICT for Smart Water Systems: Measurements and Data Science, Hdb Env Chem (2021) 102: 167–208, https://doi.org/10.1007/698_2019_422,
© Springer Nature Switzerland AG 2019, Published online: 6 December 2019

However, the development and/or adoption of ICT-focused water innovations in developing countries does not seem to occur at the expected pace, which calls for suitable innovation approaches. This chapter investigates how these innovations can be fostered through partnerships. The explorative analysis of 24 ICT-focused water innovation partnerships (ICT-WIPs) implemented in Africa leads to two important findings. First, it appears that these partnerships enable effective exchange of complementary tangible and intangible resources and co-creation of ICT-focused water solutions in a cost-effective and timely manner but also pose collaboration challenges due to the heterogeneity of innovation partners. Second, the analysis demonstrates the importance of concurrent use of relational (trust-based) and structural (legally binding control-based) partnership governance mechanisms to mitigate these challenges, notably by reducing opportunistic behaviours and increasing clarity of partners′ commitments and rights. We conclude that well-designed and -implemented ICT-WIPs can contribute to enhancing the capabilities of developing countries to implement the smart water systems agenda.

Keywords ICTs, Innovation partnerships, Smart water systems, VIA water

1 Introduction

Access to water is indispensable for achieving sustainable development and has been acknowledged as a fundamental human right [1]. However, due to global challenges such as climate change, rapid urbanisation and population growth, water systems around the world are increasingly getting stressed, which threatens water security, particularly in developing countries [2, 3]. As argued by the World Economic Forum [4], water crises will be among the greatest social risks for the coming years. Already, 10 years ago the UN-Water and FAO [5] claimed that the world was facing a crisis of water security as globally roughly 1.2 billion people experienced water scarcity problems. Current predictions also show that by 2050, at least one in four people are likely to live in a country with a shortage of freshwater [6].

In view of these global challenges, sustainable water management is more needed today than ever. Over the past years, "smart water systems" have emerged and got promoted as a potential approach to ensure wise management of water resources, particularly in developing countries [7]. Also referred to as "ICT-enabled systems", the purpose of smart water systems is to increase efficiency in water management by deploying solutions provided by Information and Communication Technologies (ICTs). This is achieved notably by enabling constant generation and transmission – to relevant stakeholders – of real-time data and information related to all aspects of the water cycle, which enables knowledge-based decisions and real-time solutions to the challenges faced by water systems [8, 9]. The areas in which ICTs can bring improvements in water management include, but are not limited to, mapping of water resources and weather forecasting, mapping and monitoring of physical water infrastructure, early warning systems, monitoring of water quality, consumer service delivery, water governance and operations of water sector organisations [10, 11].

Despite the benefits associated with the implementation of ICTs in water management, the development and adoption of these innovations by water sector institutions seem not to occur at the expected speed, and there are great variations in the way ICTs are adopted [12]. On the one hand, this situation can be attributed to the specific nature of the water sector. For example, Tutusaus et al. [13] describe how the idiosyncrasies of the water services sector – e.g. monopolistic nature, social and economic importance and commercialisation of public water utilities – influence the adoption of ICTs. On the other hand, the innovation strategies used to foster ICTs matter. Particularly, it has been argued that the promotion of ICTs in the water sector can only be achieved through a partnership approach, based on open innovation; only then can appropriate ICT-based solutions be developed and integrated in water systems [7, 10, 14]. However, what the ICT-focused water innovation partnerships (ICT-WIPs) approach entails, the challenges associated with it and how they can best be prevented or dealt with are issues that are rarely addressed in the ICT and water literature. This chapter aims to bridge this gap.

The research reported in this chapter is explorative, and the analysis is based on 24 ICT-WIPs. The innovations are funded by VIA Water – a Dutch programme to foster water innovation in African cities [15]. Drawing on insights from three theoretical perspectives – innovation systems, open innovation and governance perspectives – this chapter sheds light on the complexity of ICT-WIPs and on the need to devise appropriate partnership governance mechanisms. The study finds that the partnerships analysed make a clear distinction between at least two categories of partners: implementing partners and non-implementing partners. Partnering organisations exchange resources (material and immaterial) that enable cost-effective and timely generation and/or adoption of ICT-focused water innovations. The study further establishes that the partnerships simultaneously utilise relational governance (trust) and structural governance mechanisms to ensure effective resource exchange. However, we observe limited use of legally enforceable instruments (such as contractual agreements) in the ICT-WIPs analysed in this study, which partly explains the problems reported in some of the partnerships. We conclude that successful ICT-WIPs rely on careful selection of partners, design of appropriate resources exchange mechanisms, genuine engagement of non-implementing partners and complementarity between legally binding control and trust-based governance mechanisms.

The remainder of the chapter is structured as follows. Section 2 briefly describes the potential of ICTs for effective and efficient management of water systems, particularly in developing countries. Section 3 presents the theoretical context of the study and defines the focus of the analysis. Section 4 discusses the methodology followed in this chapter. Section 5 presents and analyses the results. Section 6 discusses the findings using the theoretical insights presented in Section 3. The final section concludes the chapter, with a reflection on the potential of innovation partnerships as an approach to foster the smart water systems agenda in the context of developing countries, the practical implications (of the major insights arising from the study) for effective design and execution of innovation partnerships for smart water systems in developing countries, as well as recommendations for future research and for VIA Water.

2 ICTs and Water Management in Developing Countries

Fresh and adequate water resources are crucial for achieving sustainable development. However, water systems around the world are expected to be stressed by societal challenges such as climate change and its effects (e.g. flooding and droughts), rapid human population growth and urbanisation [16, 17]. These phenomena will have negative effects on the quality and quantity of water resources, the capacity of water infrastructure and the cost of water services, therefore threatening water security, particularly in developing countries [2, 3]. Water security problems in developing countries are further exacerbated by current inefficient and unsustainable water management practices, as evidenced by high water loss rates (in terms of non-revenue water [NRW])[1] observed in most water service providers and situations of over-exploitation of water resources which cause physical water scarcity in many countries.

Ineffective management of water systems stems also from the unreliability of conventional approaches, due to the increasing complexity of water problems. For example, extreme weather variability due to climate change has rendered reliance on historical hydrologic weather patterns for predicting future variables impracticable; traditional water data and information gathering and transmission mechanisms are no longer (cost) effective; and with increasing concerns about good governance in the water sector around the world – particularly in developing countries – traditional top-down management approaches are no longer preferred. Thus, efficient water management is more needed today than ever, and there is growing consensus on the need to rethink the overall management of water resources.

Over the past decades, the potential of ICTs to improve water management in developing countries has been demonstrated, notably through implementation of the smart water systems concept. Smart water systems use ICT applications at all levels of the water cycle to maximise efficiency of water management, thus promoting water security. Table 1 displays the main areas of water management where ICTs can bring about improvements in developing countries, with examples of ICT applications that are relevant for the water sector and associated benefits.

[1]NRW from water distribution systems worldwide is estimated at 48 billion cubic meters per year of which 55% occurs in developing countries [18]. An important proportion of NRW (physical losses) is generally attributed to the ageing and subsequent deterioration of water infrastructure, because of poor operation and maintenance activities as well as slow the replacement process.

Table 1 Key areas of ICT in water management and ICT tools

Areas for ICTs in water management	Examples of ICT tools	Benefits for water management
1. Weather forecasting	Remote sensing satellite systems; in situ terrestrial sensing systems; wireless sensor networks; geographical information systems (GIS)	High-quality and standardised observations of the atmosphere and ocean surface; real-time exchange of meteorological data and information
2. Mapping of water resources	GIS; satellite mapping; water portal systems; supervisory control and data acquisition (SCADA)	Improved understanding of the water resource base; improved knowledge of current levels of water abstractions and use; improved prediction of water resources supply and demand
3. Asset management	GIS, buried asset identification and electronic tagging; smart pipes, hand pumps and meters; supervisory control and data acquisition (SCADA)	Improved management of distribution networks; reduced water losses; reduced network damage And deterioration; reduced risk of infection in the water system; shortened response time, reduced maintenance costs
4. Early warning systems	GIS; sensor networks; early warning websites; mobile phone applications; digital delta	Improved reservoir management; flood mapping; improved data management (quick acquisition, processing, analysis and dissemination to warn the public)
5. Water demand forecasting	GIS, ground penetrating radars; optical and pressure sensors; cloud computing; SCADA	Rain/storm water harvesting; managed aquifer recharge; improvements in water resource management
6. Service delivery	e-payment systems; GIS; call centres	Improved service delivery: timely access to water information, operational efficiency of water sector institutions – shortened response time, improved financial management, increased revenue collection
7. Governance and visualisation	Smart mobile phone applications; websites	Improved public participation, transparency and accountability; improved customer relations

Based on Mauree [19] and Ndaw [10]

3 Theoretical Context and Focus of the Study

This analysis conducted in this chapter draws on three theoretical perspectives. The innovation systems and open innovation perspectives help to explain the rationale behind innovation partnerships in fostering smart water systems in developing countries (and beyond) and what these partnerships entail for ICT-focused water innovation processes. The governance perspective provides insights into the mechanisms through which these innovation partnerships can best be managed to maximise their benefits and mitigate possible problems associated with their implementation.

Before discussing these theoretical perspectives and describing the focus of the ICT-WIPs analysis, we define the concept of innovation partnerships as used in this study.

3.1 Innovation Partnerships

The literature on inter-organisational cooperation emphasises that the internal capabilities even of large organisations are often limited and these organisations are forced to compensate their weaknesses by teaming up with others [20]. Innovation partnerships are one of the possible strategies that organisations can use to strengthen their innovation capabilities. Other mechanisms include, for example, acquisitions and mergers [21]. A variety of terms are used in the literature that are closely related to the concept of innovation partnerships used in this study. For example, Hagedoon and Schakenraad [22] describe *strategic technology partnering* as the "establishment of cooperative agreements aimed at joint innovative efforts or technology transfer that can have a lasting effect on the product-market positioning of participating companies". In the same vein, DeMan and Duysters [23] define *strategic technology alliances* as "cooperative agreements in which two or more separate organizations team up in order to share reciprocal inputs while maintaining their own corporate identities". These terms describe the same reality: a type of relationship in which organisations agree to jointly conduct innovative activities. However, the terms appear to be essentially biased in the sense that they seem to suggest that cooperation agreements between innovating organisations must always be for strategic reasons. The word "technology" included in these terms also seems to suggest that these agreements basically concern technological innovations, involving research and development in high-tech industries. These biases may be due to the fact that these authors draw on corporate sector experiences.

In this chapter, we propose the term "innovation partnership" because we deem it to be more neutral and inclusive. Departing from the above definitions, we refer to innovation partnerships in this study as *arrangements in which innovating organisations and other relevant stakeholders cooperate with the objective to innovate together while maintaining their own identities. This objective is achieved through ongoing exchange of resources that are necessary to successfully conduct innovation activities.* Defined in this way, the concept of innovation partnerships accommodates cooperation agreements for both technological and non-technological innovation ventures and those entered into for reasons other than "strategic". In the corporate sector, strategic reasons for which companies enter into innovation partnerships often refer to reasons such as cost and risk reduction, exploration of new markets and market niches, etc. However, partnerships can also be crafted to pursue innovations that are not market/business-oriented and which can involve a variety of other actors in the operating environment. This is mostly the case in public sector innovations where relevant stakeholders are engaged to share their knowledge, information and experiences and to generate innovations that are relevant to them.

Finally, we define innovation partnerships as arrangements that can be short or long term, depending on the focal issues being considered. As described in the sections below, innovation partnerships as an approach is rooted in the broad theories of open innovation and innovation systems.

3.2 Open Innovation and Innovation System Theories

Traditionally, innovation has been conceived as a closed activity – being driven and controlled internally by innovating firms in the private sector and knowledge institutions (e.g. universities, research centres) in the case of science-driven innovation. Following this approach, innovating organisations are expected to generate their own ideas and transform them into business opportunities on their own (and using their own resources). Over the past decades, "open" innovation has emerged as a new perspective on innovation, based on the assumption that innovating organisations can and should use internal as well as external ideas (and other resources) and internal and external paths to market. This means that internal ideas can be taken to market through external channels, but ideas can also start outside the firm′s own labs and move inside [24]. Thus, in today′s globalised world, organisations no longer develop innovations in isolation; they partner with other organisations to develop innovations which they would hardly realise without the supplement of resources of a network of actors [25]. It should be emphasised that while innovation in the private sector is generally driven by the desire to remain competitive in the market and increase profits, the main driver for public sector innovations is to create greater public value or improvements in the public sphere – e.g. by introducing new working practices and approaches (such as citizen participation in government projects, devolution of decision-making powers).

The open innovation literature acknowledges partnerships as an excellent way to innovate cost-effectively and time efficiently [26]. In line with the resource-based view (RBV) (more specifically the knowledge-based view) of the firm [27, 28], "open innovation" as an approach acknowledges that companies in an industry (such as the water sector) are heterogeneous regarding the resources they possess and that this heterogeneity is partially preserved by the difficult mobility of these resources. A firm′s resources fall into two categories, material and immaterial, and they span from all assets to capabilities, organisational processes and knowledge that it uses strategically to gain competitive advantage [27]. Organisational knowledge (e.g. embodied in its staff and systems) is considered to be the most strategically important resource and enabler of innovation, particularly tacit knowledge which is generally difficult to imitate by competitors [28]. Under these circumstances, it is argued that innovation partnerships enable partnering firms to overcome the resource immobility problem. Partnerships are essentially crafted and executed to allow resource flows between organisations and, as such, create new entities with strengthened innovation capabilities [29, 30]. Thus, in selecting innovation partners, companies ought to carefully examine the extent to which their resources will be complemented by those of the partners. According to Nooteboom [31], knowledge transfer between partners is the core of every innovation partnership, and human resource exchange is recommended as an effective way to ensure effective transfer of both tacit and explicit knowledge [32]. However, innovation partnerships have other motives such as allowing partnership members to share innovation costs as well as risks [33].

Open innovation is consistent with the systems approach to innovation. The latter emphasises that innovation does not take place in vacuum, but rather it is developed and implemented in a collaborative process whereby innovating organisations interact, rely and learn from other entities in their operating environment. The systems approach became popular in the 1980s through seminal works by researchers such as Freeman [34], Lundvall [35] and Nelson [36] and as an alternative to the linear model of innovation. The linear perspective assumes that innovation starts with basic research which leads to development and then development leads to production and production to marketing and diffusion [37]. In contrast, the innovation systems perspective pulls away from the view that innovation is necessarily and primarily related to research activities and acknowledges the role of other players in the innovation process. Innovation is seen as part of a larger system of actors and institutions and thus a complex and interactive process [37, 38]. Viewed as systems, innovation partnerships can involve loose and or contractual arrangements between companies; they can also involve weak and strong ties among partnership members [39].

Innovation partnerships are associated with many challenges and risks – not just benefits. In addition to the challenges associated with the transfer on knowledge itself (e.g. lack of capabilities to value and tap into the knowledge possessed by other partners), innovation partnerships raise questions such as how to prevent exchanged knowledge from being used opportunistically by some partners and how to deal with issues such as conflicting goals of innovation partners, partnership coordination costs and the problem of appropriation of innovation outputs [28, 40–42]. As described below, good governance mechanisms are required to overcome such partnership challenges and risks.

3.3 Structural and Relational Perspectives on Partnerships

The literature suggests a variety of mechanisms that enable the creation and execution of stable and effective innovation partnerships. These mechanisms generally draw on two major theoretical perspectives: structural and relational [43, 44]. The structural perspective, which is based on transaction costs economics [45], considers innovation partners as rational, calculating and self-interested actors, which may induce in fact opportunistic behaviour. Thus, this perspective posits that partnerships should be formalised through the use of official administrative coordination and control mechanisms. Structural governance involves frameworks that specify the obligations and rights of innovation partners and are codified into written documents [46, 47]. These frameworks may include court enforceable contracts and noncontractual agreements that are internally enforceable documents – such as job descriptions, division of tasks, rules and procedures [48]. Critics of structural governance mechanisms argue that they are too expensive (due to the transaction costs involved) and too restrictive concerning creativity and flexibility [49–51].

The relational perspective builds on social exchange theory according to which human beings are social, capable to trust and be trusted [52–54]. As described by Gil [55], relational governance is one in which the partners′ personal relations (e.g. social norms such as trust, cooperation and solidarity) are deeply intertwined with their economic exchange. This perspective suggests that opportunistic attitude and behaviour of partners can be overcome as they get to know each other. Trust-based relationships are therefore believed to be the potential asset that needs to be developed over time for the partnership to yield high returns [56]. With trust, partners expect each other′s behaviour and willingness to adhere to commonly accepted principles or individual commitments [57, 58]. Governance mechanisms associated with this perspective emphasise safeguards that are self-enforcing through social interaction and control. They further rely on informal norms and rules (not written in any documents) to indicate how responsibilities and rights are distributed among partners [59, 60]. Some literature often describes relational governance and structural governance as being mutually exclusive [61]. However this view has been challenged by evidence suggesting that the two perspectives are complementary. Structural governance is increasingly perceived as providing a solid basis for creating trust, especially in partnerships where members do not know each other or have not collaborated before [44, 46, 58, 62].

In this discussion, the influence of national culture on the extent to which partnering organisations use and/or prefer either structural or relational perspectives must be emphasised. In line with Hofstede′s [63] widely used cultural dimensions (uncertainty avoidance, individualism versus collectivism, masculinity versus feminity and power distance),[2] it is expected that organisations from a country dominated by a certain culture will prefer a particular governance perspective over another. For example, organisations coming from individualistic cultures would rely more on structural mechanisms, while those from collectivist cultures would prefer

---

[2] Power distance is defined as the extent to which the less powerful members of institutions and organisations within a country expect and accept that power is distributed unequally (p. 28). Individualism pertains to societies in which the ties between individuals are loose: everyone is expected to look after himself or herself and his or her immediate family. Collectivism as its opposite pertains to societies in which people from birth onward are integrated into strong and cohesive in-groups, which throughout people′s lifetimes continue to protect them in exchange for unquestioning loyalty (p. 51). Masculinity pertains to societies in which social gender roles are clearly distinct (i.e. men are supposed to be assertive, tough and focused on material success, whereas women are supposed to be more modest, tender and concerned with the quality of life. Femininity pertains to societies in which social gender roles overlap (i.e. both men and women are supposed to be modest, tender and concerned with the quality of life) (pp. 82–83). Uncertainty avoidance is defined as the extent to which the members of a culture feel threatened by uncertain or unknown situations (p. 167).

relational mechanisms. The reason being that in individualistic societies, people and organisations tend to be more rational in their dealings and comfortable with performance and incentive systems that are contract-based, for instance. Conversely, members of collectivist societies are more strongly driven by social factors and values such as seeking long-term relationships, trust and harmony, all of which are compatible with relational governance. In a similar vein, in cultures where people and organisations are more comfortable with ambiguity and uncertainty, innovation partners are likely to favour relational governance mechanisms, whereas in uncertainty avoidance cultures, characterised by a desire for discipline and clear rules, organisations would feel comfortable with structural governance mechanisms [64, 65].

3.4 Focus of the Study

Drawing on the concepts and theories discussed in this section (and in Sect. 2), we propose to focus the analysis of ICT-WIPs on the following three sets of variables: (1) characteristics of ICT-WIPs, (2) resources exchange during the execution of partnerships, and (3) ICT-WIPs governance mechanisms. These variables represent three important areas that are critical for understanding the dynamics of innovation partnerships that aim at fostering ICT-enabled water systems in developing countries. This understanding is a prerequisite for successfully setting up and managing these partnerships. Figure 1 schematically represents the focus areas of our analysis.

4 Methodology

4.1 Selection of Cases

The research reported in this chapter focused on ICT-WIPs implemented in Africa within the framework of VIA Water, a programme supporting innovative projects for water problems facing cities in Benin, Ghana, Kenya, Mali, Mozambique, Rwanda, South Sudan, Senegal and Ethiopia. Funded by the Dutch Ministry of Foreign Affairs, VIA Water was hosted (until December 2017) by the IHE Delft Institute for Water Education and its fund managed by Aqua for All – both based in the Netherlands. From January 2018 both VIA Water secretariat and fund are managed by Aqua for All. In addition to seed capital investment, VIA Water provides other support services (such as coaching, training and networking) to the innovation partnership members. The programme focuses on 12 strategic innovation areas, the so-called pressing water needs: drinking water, sanitation, water in urban agriculture, water harvesting, groundwater, water quality, data, institutional strengthening, water allocation, financial arrangements, urban planning and floods and droughts.

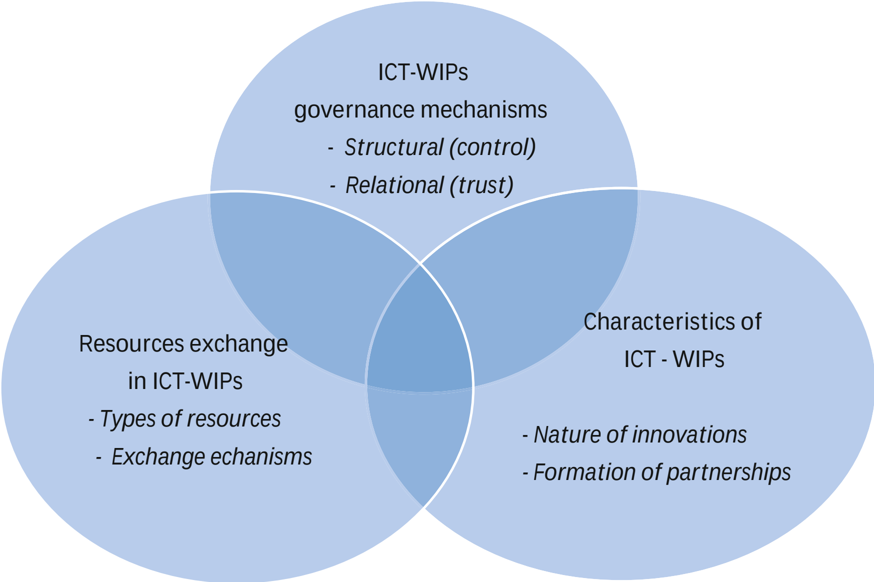


Fig. 1 Focus areas for analysing ICT-WIPs

Further details on this programme can be found in Nagel et al. [15] and Mvulirwenande et al. [66]. The reasons for selecting cases from this programme are twofold. On the one hand, most VIA Water projects are implemented through partnerships. The programme promotes this approach and encourages lead applicants to partner with other organisations to ensure successful innovation processes – both at pilot- and large-scale implementation stages. As of 2018, the programme had supported 64 water innovation projects of which 25 are ICT-focused (they are either entirely ICT or have an important ICT component) (see description in Table 2). The analysis in this chapter is focused on the 24 ICT-focused projects that are implemented through partnerships (only 1 is not); they cover a variety of areas in which ICTs can make a contribution in promoting smart management of water resources and sanitation. This is a reasonable number of cases on the basis of which insights can be generated on how to successfully design and manage innovation partnerships for smart water systems. On the other hand, the fact that all 24 projects were developed and supported through one support programme implies that we are able to have easy access to data and information related to these innovations – notably through the VIA Water database. Of note is that most of the ICT-WIPs analysed in this study are not yet completed, and, for some, little information is available about their performance. This implies that it is still too early to evaluate their success rates and success factors.

Table 2 Description of VIA Water-supported ICT projects

Name of project	Short description	Lead innovator
Rwanda		
1. Scaling mobile IT for safe water enterprises	Builds an open-source IT platform for safe water enterprises (SWEs) to be able to monitor who purchases water and where they live, how and how well it′s delivered and how well it is produced (water quality, volumes, etc.)	dloHaiti
2. Pay your relatives′ water bills	Offers an online platform on which Rwandan diaspora can pay various services to benefit their relatives back in Rwanda, making sure the money sent by Rwandan diaspora will definitely pay water only. The payment technology is integrated with the public water company and the water kiosks systems	Markets Merger Ltd
3. Storm forecasts for Musanze	Helps city residents and farmers get weather predictions with alerts on floods and lightning on their phones. The innovation works with innovative low-cost lightning data detectors to track lightning strikes in an area and produce alerts	Delft University of Technology
Ghana		
4. Self-billing and payment of water bills	Deploys a mobile phone-based billing system that will allow customers of Ghana Water Company Limited (GWCL) to self-bill and pay water tariffs. Project further provides a platform through a social control check that will help prevent non-revenue water losses	SkyFox Ltd
5. Knowledge building through Water Technology Academy	Introduces so-called online Water Technology Academy, allowing local operators and technicians have 24/7 access to all the knowledge they need to successfully fulfil their tasks. Users can also use the academy to share their knowledge mutually and easily create a knowledge and experience database	Empower People
6. GARV smart public toilets	Introduces toilets that are regularly maintained through self-cleaning mechanisms and real-time monitoring and equipped with a biodigester tank	SnapEX Overseas
7. Flash flood forecasting app	Provides early warning for flooding via state-of-the-art modelling	Royal HaskoningDHV

(continued)

Table 2 (continued)

Name of project	Short description	Lead innovator
	techniques. Satellite data is transformed into highly detailed rainfall data. An innovative algorithm provides rainfall forecasts. This is input for a flood model that provides flood maps on street level	
8. Flood risk assessment methods	Collects urban drainage data through crowd sourcing; uses a multidisciplinary approach with social, technical and institutional aspects and a communication strategy (i.e. by social media) to create awareness and commitment among citizens	HKV Consultants
Kenya		
9. TAP21 purified water distribution in the twenty-first century	Introduces a franchise business model, whereby vendors dispense purified water via 24/7 prepaid water ATMs. Payments and performance are monitored online	Maji Milele Ltd
10. Maji Mkononi – helping communities in Kibera	Enables community members to use their mobile phone to acquire information about location and real-time availability of water at water points while providing water providers with essential information and water level data from their water points	MobiTech Water Solutions
11. Reducing water loss by improved data systems	Seeks to minimise water losses by developing wireless sensors to collect data on water flow, pressure, levels – readings serve as early warning signals for burst and leakages or water theft. Sensors are used under the free radio frequency spectrum to transmit data into a consolidated dashboard (alongside SIM card-based data loggers) to decision-makers	Upande Ltd
12. Exploration of sales channels for organic fertiliser	Uses online methods (social media) and a mobile platform to disseminate information about their products – biosolids-based organic fertiliser and insect-based animal feed all made from human waste	Sanergy
13. eSOS: efficient and intelligent toilets	eSOS (emergency Sanitation Operation System) makes maximum use of novel ICTs combined with the state-of-the-art toilet and novel technology for faecal and septic sludge treatment. The eSOS concept provides a unique	IHE Delft Foundation

(continued)

Table 2 (continued)

Name of project	Short description	Lead innovator
	set of features that allows integrated smart and real-time monitoring, operation and maintenance of sanitation system components	
14. Chezo serious gaming	Gives gaming companies an ICT-based perspective into creating games other than purely fun games, which could add revenue streams	Upande Ltd
Mozambique		
15. Smart water metre reading, Mozambique	Allows water utilities bill their customers more efficiently and transparently using advanced smartphone technology that registers water consumption by taking a picture and automatically communicates validated usage values to the billing system of operators	Mobile Water Management
16. Water quality monitoring with a DNA-based field device	Introduces a DNA-based detection technology that can be used in the field to improve water quality monitoring and generate more digital data on water quality as well as making data faster, more affordable and more accurate. Detects DNA of human pathogens in water samples using an innovative field device and by integrating the digital data generated directly into web-based monitoring systems	Orvion B.V.
17. App service for faecal sludge management (PULA)	Develops a cloud-based application called PULA, which provides real-time data to both businesses and municipalities/city authorities for improved faecal sludge management FSM	BoP Innovation Center and Water & Sanitation for the Urban Poor (WSUP)
18. Virtual latrines for improved sanitation	Introduces a sanitation marketing tool which eases the communication of complex messages and needs little literacy. Equips activation agents with a smartphone and a lightweight (and low-cost) VR headset, as they go door by door. By offering a VR experience viewers are presented with a virtual bathroom (either entirely based on 3D models or sourced from 360 video footage), which gives them a better understanding of what an investment in	BoP Innovation Center

(continued)

Table 2 (continued)

Name of project	Short description	Lead innovator
	sanitation could bring to their immediate environment	
Mali		
19. Map action Bamako, Mali	Introduces an online and interactive map of Bamako and a mobile application through which community members are able to report WASH problems and propose solutions recommendations. A report is created and transmitted to the companies subscribed to the platform that are responsible for taking action	Kaicedra-Consulting
20. Geolocating water quality	Mobile application called Akvo Caddisfly and pocket-size hardware attachments are used for the first time to test water quality. Akvo Caddisfly is a simple, low-cost, open-source, smartphone-based water quality testing system connected to an online data platform. Real-time GPS-based water quality data is shared via Akvo′s data collection platform	World Waternet
21. Innovative sanitation services for Bamako	Develops a call centre and an innovative approach of sanitation marketing to optimise the encounter between offer (emptiers) and demand (households). It also creates a smartphone application as an instrument to monitor the emptiers′ technical and financial performances	Protos
22. Water quality data of the Niger River	Uses ICTs for data collection on the water quality of the Niger River, storage and analysis (smartphones and underwater drone, central repository) and sharing it in appropriate formats for various target groups	Agence du Bassin du Fleuve Niger (ABFN) Niger River Basin Agency
23. Mapping and monitoring critical developments in Niger Delta	Develops a platform with data on the Niger Delta. The analysed data is aggregated from (a combination) of satellite data, online (social) media, remote sensing and user-generated data sources and will lead to information on the longer-term drivers and effects of floods and droughts	FloodTags
Benin		

(continued)

Table 2 (continued)

Name of project	Short description	Lead innovator
24. Warning system for water shortages (ALERTE)	Establishes a warning system for water shortages in the Société Nationale des Eaux du Bénin (SONEB) water company network, by using ICT. An Internet- and SMS-based alert system informs clients about planned and unplanned cuts in the water supply and possible options they have to store, treat or get water	Benin Country Water Partnership (PNE-Bénin)
Senegal		
25. The Greening Plastic project in Senegal	Introduces portable sampling devices – passive samplers – in the monitoring of soil and water quality. Once located in the soil or water, passive samplers collect lab information on chemical concentrations of substances, which can then be analysed in a remote lab	Deltares

[a]Note that there are no ICT-focused water projects supported by VIA Water in South Soudan and Ethiopia
[b]This project is not implemented through a partnership. So the analysis in this chapter focuses of 24 projects

4.2 Data Collection and Analysis

Our analysis draws on secondary as well as primary data. We reviewed a variety of relevant documents related to the ICT-WIPs available in the VIA Water database. The documents include project proposals submitted to VIA Water by the innovating organisations, project contracts signed between the VIA Water Fund Manager and the lead partner, as well as interim and final project reports. We also consulted VIA Water corporate documents, reports (including those produced by consultants hired for specific assignments) and website. In the VIA Water online learning platform, project owners post useful information related to the day-to-day management of their innovations. Our analysis further draws on interviews with representatives of 10 ICT-WIPs in Ghana (July and November 2017), Kenya (September 2017) and Mozambique (February 2018), and discussions held with VIA Water managers – all of which resulted in primary data and information. Relevant information was also collected via informal discussions held with representatives of the ICT-WIPs at VIA Water events such as the sharing skills seminar in Ghana (November, 2017) and the "VIA GO" event organised at the Amsterdam Water Week Conference (October, 2017). The analysis conducted in this research is purely qualitative. For the analysis, we used the key themes selected based on the theoretical framework used in this chapter. With these themes, matrices were first created in which the data and information collected was systematically captured; then a meta-analysis was conducted.

5 Results and Analysis

This section presents a meta-analysis of the 24 VIA Water ICT-WIPs, using the 3 sets of variables presented in Sect. 3.4 as a structuring framework. The variables are (1) characteristics of ICT-WIPs (nature of innovations, types of partners, formation of partnerships), (2) resources exchange during the execution of partnerships (types and exchange mechanisms) and (3) partnership governance mechanisms. This section further analyses some of the challenges reported by the VIA Water ICT-WIPs.

5.1 Characteristics of VIA Water ICT-WIPs

5.1.1 Nature of Innovations

As indicated previously, the VIA Water programme supports innovations that address water challenges in urban environments – so are the ICT-focused water innovations analysed in this study. Figure 2 illustrates the relationship between the components of a (simplified) urban water system and different ICT-focused water innovations supported by VIA Water as described in Table 2. The numbers indicated in Fig. 1 correspond to the numbers assigned to the innovation projects in Table 2. It should be noted that some innovation projects relate to several components (of the water cycle) simultaneously, which makes it difficult to assign them to one particular component. For example, the (online) interactive map and associated mobile application proposed in the Map action Bamako project in Mali (# 19) allow community members to report any WASH (water, sanitation and hygiene)-related problems, and these could be identified at any of the components of the urban water cycle. Such ICT-focused water innovation projects were placed in the centre of Fig. 2 (see the quad arrow pointing in multiple directions).

The ICT-focused water innovations supported by VIA Water are diverse, but they all aim at improving urban water management by enabling real-time monitoring of urban water systems and knowledge-based decision-making (with regard to different components of the urban water cycle). The water innovations that serve the objective of “real-time monitoring” relate to the monitoring of both water quantity and quality. They involve ICTs that allow, among other things, real-time leak detection and real-time networks monitoring (e.g. Upande′s Non-Revenue Water reduction project in Kenya, # 11) and real-time water quality management (e.g. World Waternet′s Geolocating water quality project in Mali, #20, Orvion BV′s Water quality

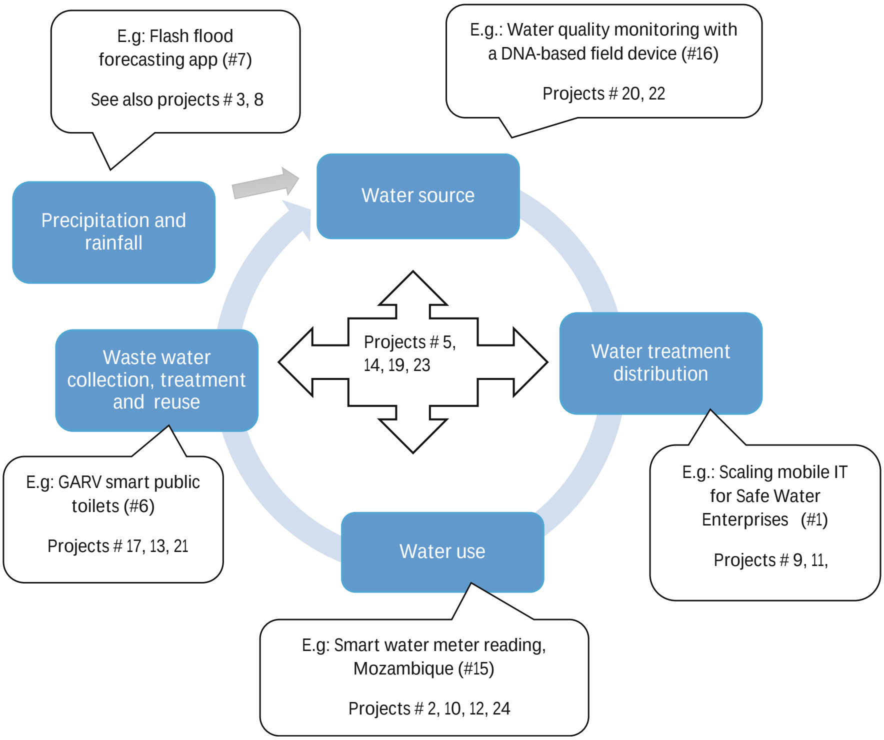


Fig. 2 VIA Water ICT-focused innovations in the urban water cycle

monitoring with a DNA-based field device in Mozambique, # 16). Other innovation projects serve the objective of "visualisation and decision support" for different types of relevant stakeholders. For example, the Maji Mkononi project in Kenya (#10) and the Warning system for water shortages (ALERTE) in Benin (#24) use mobile phone applications to provide water consumers with better information about water supplies (availability, location, planned and unplanned cuts). Based on this information, people can decide when and where to get water. In the similar vein, Storm forecasts for Musanze in Rwanda (#3) and Flash flood forecasting app in Ghana (#7) set up early warning systems that can be used by a variety of actors (citizens, municipalities, etc.) to mitigate the risks associated with events such as floods and lightning.

It is also observed that the ICT-focused water innovations analysed in this study make use of a variety of ICT tools including mobile phones, GPS mapping and websites. However, mobile phones are used in more than 17 projects (out of 24). This is not surprising though; in fact, given that mobile technology is a fast growing market in Africa, innovators in many sectors (such as health, agriculture, energy, water) are leveraging this increasingly available resource on the continent. According to GSMA Intelligence, sub-Saharan Africa alone accounts for nearly a tenth of the global mobile subscribers. In 2016, the penetration rate in this region was estimated to be 43%, with 420 million unique mobile subscribers of which smartphone connections were nearly 200 million [67].

Finally, the ICT-focused water innovations analysed in this study seem to fit with the realities of African countries – they arguably tend to be low-cost and affordable to potential customers. For example, in the Reducing water loss by improved data systems project in Kenya, Upande develops and tests low-cost wireless water flow meters and low-cost pressure and level sensors that can be sold to water utilities at affordable prices. Within the Storm forecasts for Musanze project, the Trans-African Hydro-Meteorological Observatory (TAHMO) is introducing its inexpensive[3] but robust weather stations in Rwanda. Many of the mobile phone applications being supported by the VIA Water programme enable citizens to easily take measurements on water levels (e.g. of rivers), evaluate them and send information and images by phone to relevant authorities. Using these applications seems to be far more cost-effective, reliable and timely than traditional data collection and transmission infrastructure. Although it is still early to conclude that these technological innovations really provide robust solutions to the problems facing water systems in African countries, it is interesting to see that they all were developed with the same spirit of "being easy to use, cheap and durable". Our analysis suggests that this was triggered by the fact that the VIA Water selection criteria emphasise both technical and social sustainability aspects of the innovations. The applicants had therefore to ensure that the proposed innovations are appropriate to the local circumstances and affordable for local users.

5.1.2 Formation of ICT-WIPs

The analysis conducted in this study suggests that the VIA Water programme played an important role in the formation of investigated ICT-WIPs, by creating conditions that encouraged innovators to team up. As a matter of fact, the VIA Water programme actively promotes innovation partnerships, notably between African and foreign organisations. This is clearly indicated in the programme′s entry criteria [69]. Applicants outside Africa need to have at least one African partner. Thus, as most of the ICT-WIPs analysed in this study have a Western lead innovator, the VIA Water conditions required them to have an African partner. In the case of African applicants, VIA Water encouraged these to team up with possible Western partners, but this was not a "hard" condition. Our empirical findings suggest that, in some cases, lead innovators selected their working partners based on earlier experience working with them, perceived ability of the partners to complement their resource gaps and/or their strategic positions in the water sector of the country of project

[3]The cost of a station is evaluated at only $500: this cheap price is achieved notably by leveraging on already existing low-cost sensors (as found in objects ranging from washing machines to cars and smart phones) and using them as weather or water sensors. For example, the simple piezo buzzer (costing $1), which is used in fire alarms, is used to measure rainfall intensity [68].

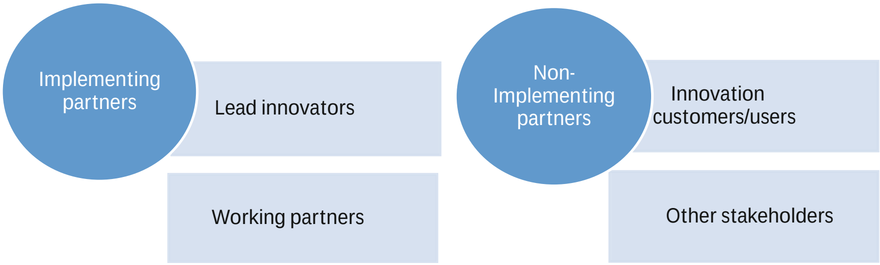


Fig. 3 Types of partners in ICT-WIPs

implementation. The study further found that VIA Water has played an instrumental role in facilitating the actual formation of many of the partnerships, thus acting as an innovation broker [70]. The partnership formation support was provided through matchmaking mechanisms. During the project selection phase, VIA Water managers would scan their network and propose (where relevant) possible cooperation partners (e.g. universities, companies in the Netherlands) to the applicants. Examples include the GARV toilet project in Ghana in which VIA Water facilitated linkages between SnapEX Overseas – the lead partner – and Advocates and Trainers for Women′s Welfare Advancement and Rights (ACTWAR). By establishing the online learning community, VIA Water further anticipated – among other things – that some organisations could use the platform to link up with potential partners [66].

The VIA Water-supported ICT-WIPs generally make a distinction between two major categories of partners (Fig. 3). The "implementing partners" category consists to begin with of the lead innovator (s) – that is the organisation (s) who, usually, comes up with an ICT-based water innovation idea and identifies prospective partners with whom to implement the idea. In the context of VIA Water, lead innovators submit the so-called teaser – a two page document in which they describe their innovative idea, interact (on behalf of other partners) with VIA Water managers and are responsible for the coordination of day-to-day management of the innovation process. Another important partner are the working partners – these work closely with the lead innovator and actively participate in the day-to-day implementation of the innovation process. Participation is either through joint activities with the lead innovator or by executing specialised tasks based on the distinct capabilities that they possess. Working partners in the VIA Water ICT-WIPs include mainly providers of technologies, customers and contributors of specialised knowledge that is not available at the lead innovator.

Overall, implementing partners are directly involved in the development and implementation of the innovation project and contribute knowledge and other types of resources to ensure successful innovation. When ICT-WIPs are profit-oriented, implementing partners are also entitled to share the profits of the innovation process and liable for the debts of the partnership.

The "non-implementing partners" category is made up of partners who are indirectly involved in or are affected by (and can affect) the ICT-focused water innovation project. In project/programme management literature, these partners are generally referred to as project stakeholders. In the context of ICTs and smart water management, ITU [71] describes a stakeholder as "...any individual, group, or institution that has a vested interest in smart water management by being potentially directly or indirectly affected by its projects, activities, policies, and/or has the ability to influence smart water management′s outcomes". In the ICT-WIPs analysed in this study, the non-implementing partners consist of a wide variety of actors spanning from customers and users of innovations[4] to government organisations (e.g. municipalities), water service providers, water regulatory agencies, ministries, funders (including donors) and so on.

It has been observed that in developing countries, the promotion of ICTs in the water sector is still heavily reliant on external (donor) funding mechanisms [10]. In that regard, the ICT-focused water innovations analysed in this chapter are not an exception: in each project, the main funding mechanism is through the VIA Water seed capital fund. Thus, the donor organisation is generally described as a "funding partner". However, the VIA Water programme is not just a funding mechanism: it functions as an incubator and is thus involved at different stages of the innovation process, not in the implementation of the projects per se but through a variety of support services (e.g. coaching, training and networking) provided to the partnerships. A detailed analysis of the innovation support services provided by VIA Water can be found in Mvulirwenande et al. [66].

The analysis of the VIA Water ICT-WIPs shows that lead innovators tend to be mainly enterprises, most of which are of small and medium size (or SMEs).[5] These are followed by non-governmental organisations (NGOs), consulting companies and universities and research institutes. Figure 4 displays the relative importance of the different lead innovators in the partnerships analysed in this chapter. Conventional innovators – notably universities, research institutions and large companies – are not well-represented in the VIA Water-supported partnerships. Our analysis further shows that lead innovators are mostly foreign organisations or local organisations established and owned by foreign entrepreneurs. For example, the three academic and research institutions are all Dutch: Delft University of Technology, IHE Delft Institute for Water Education and Deltares. Many of the enterprises (SMEs) and NGOs also either come from the Netherlands (e.g. Royal HaskoningDHV, Mobile Water Management, Orvion B.V, BoP Innovation Center) or are locally registered

[4]In this chapter, we distinguish between "end-users" (or simply users) and "customers" of an innovation. The former term refers to a person or entity that uses an innovation (e.g. product), and the latter refers to a person or entity that purchases it. Note that in some cases, the end-user and customer of an innovation are the same (e.g. a water utility purchasing low-cost sensors and using them to detect leakages in its distribution network), while they are different in other cases (e.g. an NGO purchasing a mobile application for a selected number of citizens who then use the application to share data and information about water issues in their community).

[5]Small and medium enterprises (SME) – the European Union defines these as enterprises employing fewer than 250 persons and have either an annual turnover not exceeding EUR 50 million or an annual balance sheet total not exceeding EUR 43 million.

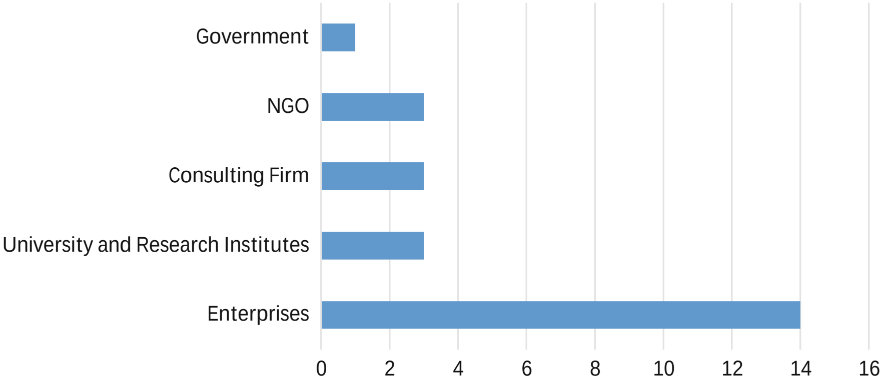


Fig. 4 Importance of ICT innovators leading the VIA Water-supported partnerships (n ¼ 24) (Source: authors)

companies that were established by Dutch entrepreneurs (e.g. Maji Milele Ltd, Upande Ltd). The few lead innovating organisations that are African-owned were initiated by tech entrepreneurs from different backgrounds, spanning from African Diaspora returnees (e.g. Markets Mergers LTD) to locally trained techies (e.g. Mobile Tech solutions) and consulting firms providing innovative technical and social solutions to challenges facing many sectors, including water (e.g. Caicedra consulting Mali, SkyFox Limited). Annex provides an overview of lead innovators, their type of organisations and their country of origin.

In the context of VIA Water projects, the lead contract partners are not necessarily the initiator of the innovation (idea); in some cases, they are chosen to lead the partnership for pragmatic reasons. For instance, in the case of App service for faecal sludge management (PULA) project in Mozambique, BoP Inc. is listed as lead partner (along with WSUP who is actually the initiator) notably because of its connection to the Netherlands.

These findings are not surprising though they are a reflection of the general situation of innovation and entrepreneurship in Africa. According to some authors, the most dynamic SMEs on this continent are mainly in the hands of non-Africans, implying that the level of indigenous innovation and entrepreneurship (including in the ICT domain) in general is still low [72, 73]. This is particularly true for the "tech" industry in Africa which is still in its infancy.

5.2 Resources Exchange in VIA Water ICT-WIPs

As elaborated in the theoretical section of this chapter, (innovation) partnerships between two or more organisations are in essence established to tap into the distinct resources possessed by other organisations [29, 30]. In the VIA Water partnerships, lead innovators generally decided to partner with organisations who they thought would bring resources that complement their own. We discuss below the immaterial and material resources exchanged in the examined partnerships as well as the mechanisms used.

5.2.1 Exchange of Immaterial Resources

The immaterial resources mostly exchanged in the ICT-WIPs analysed in this study appear to be knowledge. In the first place, those who contribute knowledge resources are the implementing partners. For example, many of the working partners have joined the partnerships either as suppliers of ICT knowledge (embedded in specific technologies) or as providers of expert knowledge that is not possessed by the lead innovator. To illustrate, in the Flash flood forecasting application project in Ghana, Royal HaskoningDHV – the Dutch lead innovator – provided meteorological and hydrological experts to develop the application (including staff from its local Ghana office). Whereas Infoplaza – one of the working partners – provided meteorological data (rainfall data) for the validation of the system and set up an App service (i.e. sms/smartphone) that integrates meteorological and hydrological forecasts and provides effective warnings to the local people. Nelen and Schuurmans – another working partner – brought to the partnership the expertise to setup the 3Di hydrological model that is able to predict flash floods based upon the input of rainfall forecasts. In the Water quality monitoring with a DNA-based field device project in Mozambique, Orvion B.V – the lead innovator – brings to the partnership the expertise to characterise and quantify micro-organisms in water using innovative DNA-techniques, while WE Consult – the working partner – contributes its expertise in water quality measuring and sampling and knowledge of the local context.

Different mechanisms are used by implementing partners to exchange knowledge. These mechanisms include the creation of structures such as joint project implementation teams through which regular meetings (e.g. weekly teleconference meetings, kick off meetings) are organised, serving as channels for sharing both tacit and explicit knowledge. Online collaborative systems are also used through which participants from partnering organisations share information relevant to the innovation they are developing and or promoting. Through these structural arrangements, participants further continuously cocreate knowledge and use it during joint partnership activities. Most of this co-creation takes place through combination of the participants′ respective personal/tacit knowledge [74]. As we explain in the next section, the level of knowledge exchange is dependent on the extent to which members of these teams relate to and trust each other. Task division among partnering organisations is another important mechanism to exchange knowledge resources in the partnerships. This is particularly the case when some partners are best qualified (and or specialised) in producing some components of a particular innovation and where cooperation through joint team activities would make little sense. As argued by Batterink [75], task division improves the performance of innovation partnerships, by allowing a more efficient use of the resources possessed by partners.

Knowledge that is used in the production of ICT-focused water innovations also comes from non-implementing partners. The reports consulted and the interviews conducted in this study showed that the knowledge possessed by these partners was accessed through different mechanisms and at different degrees. Some of the innovators obtained knowledge from potential customers and end-users of their innovations by actively engaging them in some stages of the innovation process. For example, in the App service for faecal sludge management (PULA) project in Mozambique, the innovators applied a "user-centred design" approach. This iterative design approach helped to understand the customers'[6] needs and requirements and to develop a prototype that considered these realities. The customers were involved in the design process through a mixture of methods, including surveys and interviews as well as brainstorming workshops. Through these consultation processes, the innovators collected data and information about several aspects: features needed in the application, level of complexity desired, how much potential customers currently pay for similar services, their willingness to pay for and interest in the innovation and so on.

Other strategies used in the ICT-WIPs to access the knowledge possessed by non-implementing partners include involvement (of these partners) as members of project steering committees, ad hoc consultations to seek their expert advice on some aspects of the innovation and invitations to attend project events and activities. For instance, in the Self-Billing and payment of water bills project in Ghana, a steering committee was formed under the auspices of the Water Directorate of the Ministry of Water Resources, Works and Housing and headed by the Director for Water. The committee is responsible for strategic oversight and provides recommendations at each stage of the project. Such a platform offers an opportunity for committee members to contribute knowledge that can be used in the innovation process. In the case of Chezo serious gaming project in Kenya, it was reported that a few stakeholders and water experts were invited to the project kick-off event which aimed, among other things, to communicate the innovation idea to them and have an exchange on how they could best make use of (computer-based) serious games. Some of the stakeholders also attended the training on serious gaming.

All in all, by engaging with relevant stakeholders, the innovators were able, among other things, to acquire necessary knowledge for their innovation processes and to easily make a case (and secure strategic support) for their ICT innovations in water management.

[6]The potential customers of this application are of two categories: (1) vacuum tankers and small-scale service providers providing services to low income consumers, and (2) municipalities, utilities, regulators and financing institutions who need faecal sludge management-related data to improve regulation of services.

5.2.2 Exchange of Material Resources

The study results show that there is exchange of material resources in the VIA Water-supported partnerships. Material resources exchanged in the ICT-WIPs by partnering organisations include the provision of ICTs themselves, licence fees and other costs charged. For example, in the partnership between Delft University of Technology, Severe Weather Consult and TAHMO[7] to implement the Storm forecasts for Musanze in Rwanda – TAHMO covers the costs of weather stations as resource contribution to the project. In other partnerships, companies contributed office space (and associated equipment and services) to host the projects, access to already existing technologies developed by one partnering organisations and so on. For instance, in the TAP21 Purified Water distribution in the twenty-first century in Kenya, Akvo (one of the partners) availed its FLOW[8] technology to the partnership – FLOW is supposed to be integrated with the "Susteq" electronic payment system to collect real-time data. In many cases, the implementing partners contributed financial resources – although a big portion of the required budget to implement the projects is provided by VIA Water.

It is important to emphasise the use of networks (of partner organisations) as a mechanism to exchange material and immaterial resources in the partnerships analysed in this study [58, 62]. Our study results show that, in many of the cases, partners′ networks enabled access to information and other resources through personal referrals, for instance; lead innovators decided to partner with specific organisations because they wanted to make use of their strong network of relationships in the local operating environment. This is notably the case in the partnerships where the lead innovator is a foreign organisation with limited local footing. In the App service for faecal sludge management (PULA) project in Mozambique, the BoP Innovation Center teamed up with Water & Sanitation for the Urban Poor (WSUP) partly because the latter could offer its strong relationships with public and private actors responsible for faecal sludge management in targeted cities in Mozambique, while the former could offer its connections in the Netherlands. In a similar vein, in the GARV smart public toilets in Ghana, SnapEX Overseas partnered with M4 Group (a Ghanaian company) to benefit from its strong network. The M4 Group supposedly provides sales and distribution support for scaling up of GARV Toilets – notably by helping SnapEX Overseas to acquire required government approvals, land for toilets construction and access to corporate partnerships. SnapEX Overseas also partners in this project with IRC Ghana who provides support pro bono through its network of partners.

Our research found that provision of access to networks (e.g. potential business partners and customers) is a core activity of the VIA Water programme and that this occurs through a variety of mechanisms. The programme has established an

[7]TAHMO (Trans-African Hydro-Meteorological Observatory)

[8]FLOW is an open-source mapping software used for data collection and monitoring of the functionality of water access points.

"online learning community" which aims at connecting not only VIA Water project owners among themselves but also different other target groups with an interest in urban water issues in Africa. These groups include (young) water professionals, experts working with NGOs, universities, knowledge institutes, the business sector and governments. The representatives of the ICT-WIPs interviewed in this study acknowledged that they have made useful connections with or through community members. The exchange of networks and connections also occurs during the physical events organised by VIA Water in the target countries (e.g. skills sharing seminars organised every year, learning tours, knowledge cafés). These face-to-face encounters increase trust among participants and allow them to eventually exchange useful connections. Finally, the programme supports project owners to pitch their innovations at national and international events (e.g. international conferences or bilateral meetings between the Netherlands and the target countries), which provides ICT-WIPs with opportunities to connect with potential collaborators.

5.3 Governance Mechanisms of VIA Water WIPs

5.3.1 Structural Governance Mechanisms

This study identified different structural elements that are used to effectively manage the VIA Water ICT-WIPs. To start with, each lead innovator ought to sign a contract with VIA Water in which funding conditions are described. Besides, in the project proposals submitted to VIA Water by lead innovators, there is always mention of partnering organisations – with their tasks and responsibilities described. In many cases, the submitting lead innovator must provide further proofs that the partners mentioned in the proposal have actually agreed to be involved in the project. The proofs include here copies of memoranda of understanding (MoU) and support letters from partnering organisations. All these documents spell out what is to be provided by each partner (and by which means) and sometimes describe how innovation benefits will be shared – they serve as structural governance mechanisms. Contracts between the lead innovator and working partners are also used but only in a few partnerships. For example, in the case of eSOS: efficient and intelligent toilets project in Kenya, a clear division of tasks is made – based on the specialisation area of the partners. Manufacturing of specific electronic parts unique to this toilet and software development are carried out by SYSTECH.ba, design responsibility was given to Flex Design, whereas the implementation and testing are led by IHE Delft with support of Sanergy. IHE Delft – the lead innovator – signed a separate contract with each of the three partners. The use of joint project steering committees was observed in 7 ICT-WIPs analysed in this study. These committees are expected to meet periodically and are responsible for guiding and overseeing the execution of the partnership. They are also responsible for advising the respective project teams in all matters of the innovation project and give comments on the project outline, progress and outputs where appropriate.

5.3.2 Relational Governance Mechanisms

The findings of our study suggest that – in addition to structural governance mechanisms – the VIA Water ICT-WIPs rely on trust in several regards. Some of the reasons described in the partnerships′ documents (notably project proposals) on why involved organisations decided to team up illustrate the importance of trust as a governance mechanism in these partnerships. The reasons include previous cooperation in other contexts, belief in the positive intentions of the partners about solving targeted water pressing needs, shared beliefs in the potential innovations being developed, clarity of incentives to partners to implement the projects. For example, the fact that some of the VIA Water ICT-WIPs involved organisations who had worked together in the past implies that they got the opportunity to learn about each other′s honesty and ability to comply with the promises and agreements they make. The partnerships were based on already existing relational trust. This could explain to some extent why some partners in such partnerships did not even bother about signing binding cooperation agreements – they believed that their counterparts would deliver on their duties as agreed informally. However, as described in Sect. 5.4, not every partner came up to expectations. The fact that many ICT-WIPs give importance to relational governance could also be associated with the national culture of the countries where partnership members come from. For example, our analysis showed that many of the lead innovators come from the Netherlands (Annex); and, with its low scores on Hofstede′s [63] "power distance" and "uncertainty avoidance" cultural dimensions,[9] the Netherlands is known to be a high trust and less bureaucratic country. These cultural characteristics imply that Dutch organisations involved in partnerships would generally prefer relational governance mechanisms or less constraining structural mechanisms (e.g. simple MoUs instead of legally binding contracts).

However, there are many other ICT-WIPs that were created just to respond to the VIA Water funding opportunity, i.e. partnerships in which members had not worked together previously. In these cases, the level of trust needed to effectively execute the partnerships needed to be cultivated throughout the implementation of the partnerships. The mechanisms to establish and maintain trust that were observed in the ICT-WIPs include recognition of the importance of other partners′ contributions (e.g. by assigning specialised tasks and consulting them on matters that are subject to their expertise), openness and transparent communication (e.g. use of drop box and other online information systems shared by all team members throughout the project) and collaborative decision-making (e.g. through joint committees).

[9] On these two dimensions, the Netherlands has the scores (out of 100) of 38 and 53, respectively.

5.4 Some Challenges for the VIA Water ICT-WIPs

Our study found that in spite of the structural and relational governance mechanisms discussed above, some of the VIA Water ICT-WIPs still experience challenges. It should be indicated that for some of the partnerships, not much information is available yet to judge their performance, be it in terms of achieving partnership objectives or the functioning of the partnership itself. The challenges discussed in this section were identified in those partnerships that are far advanced in the implementation of the pilot stage of their innovation processes or have just completed it. The challenges fall under the following four categories:

- Low levels of partner commitment: in some of the analysed partnerships (e.g. in Kenya), this issue resulted in problems of poor delivery of expected contributions from partnering organisations – in at least one partnership, this led to early termination of the cooperation. The reasons cited for poor delivery include long delays in providing expected resources, lack of prioritisation of partnership activities by some partners and weak communication between partners. These findings suggest that the partners who failed to deliver did not give adequate importance to the partnership. Yet, research has shown that the strategic importance of a partnership for a company to achieve its goals determines what that company can invest in the partnership (e.g. number of staff a company assign to the partnership, amount of resources dedicated to the partnership) [76].
- Difficult transfer of tacit knowledge: challenges related to knowledge exchange (and transfer) were reported in one partnership in Ghana. In this partnership, translation of the knowledge of experts into a digital manual that could be used by the operators with no experience at all proved to be difficult. Not surprising though, because a big portion of the knowledge possessed by water sector experts is tacit in nature (i.e. personal, context-dependant and based on practice and experience in nature) and thus very difficult to formalise and communicate. The best mechanism to transfer tacit knowledge to others is through sharing mutual experiences and through active participation in real-time and face-to-face interactions [74, 77].
- Asymmetries in resource exchange: in one partnership in Ghana, the lead innovator faced the challenge to access the data and information (e.g. financial reports) possessed by a working partner, despite of an existing MoU. This could suggest that there were concerns about confidentiality and the need for an agreement along these lines. However, in other cases, the failure to access a partner′s data/knowledge base could result from issues to do with the partner organisation′s management style (e.g. a hierarchical structure, rigid procedures about sharing company data and information), which then needs to be understood before starting the partnership. In this regard, the lead partner in this partnership further reported that at the initial stages of the project, they focused more on involving top managers of the working partner. However, it proved difficult to work with them because they were always busy with other priorities and often out

of the country. To correct this, the lead innovator had to broaden his focus to include other key staff members of the working partner who made the processes smoother and quicker than before. This experience suggests that lead innovators should target not just top managers of their potential partners. Instead, they should aim to create a multilayered network in their partner organisations so as to lower the risks that the partnership activities will be disrupted should a key contact be reassigned or leave the organisation. In other partnerships, the interviewees reported cases of partners who did not contribute resources in expected quantities and qualities – e.g. supply of poorly performing technology, secondment of low calibre staff to joint innovation project teams.

– Ambiguity of partnership objectives: this challenge was reported in one partnership in Ghana. In an interview with one partner, it was disclosed that lack of enough consultation among partners led to an ICT application with a lot of problems (as it looks now). The final project report states that the project was supposed to develop a tool that enables citizens to become more flood resilient, which the project succeeded to do. But this report further suggests that there were clarity problems among the partners, not only of partnership objectives but also of expected contributions (lack of understanding about who was expected to bring what to the partnership), which led to poor cooperation.

6 Discussion

In the context of globalised and competitive world, organisations increasingly innovate in partnership with others (and with relevant stakeholders such as customers/users) in order to bridge their resource gaps and thus produce cost-effective and timely innovations. Drawing on evidence from the VIA Water-supported ICT-focused water innovation projects, the explorative research reported in this chapter sought to generate insights into how to best foster smart water systems in developing countries through innovation partnerships. We did this by using the lenses of innovation theories, notably open innovation [24] and innovation systems [35] theories, and governance perspectives [43, 44]. The analysis was based on 24 ICT-focused water innovation partnerships (ICT-WIPs) developed and implemented in seven African countries (Kenya, Ghana, Rwanda, Mozambique, Benin, Senegal and Mali). The analysis generated several insights that could inform and guide efforts aimed to form and execute successful innovation partnerships for smart water management. We discuss these below.

The study results show that of the 25 ICT-focused water innovation projects supported by VIA Water, 24 involve at least 1 partner. These were the focus of this study. The partnerships analysed generally distinguish between two categories of partners – implementing partners and non-implementing partners (or stakeholders). The major motivation for the partnering organisations appeared to be the search for complementary resources between them. This finding is consistent with the assumptions made in the open innovation (and resource-based view) literature about why organisations are forced to team up with others in their innovation endeavours [24, 78]. The resources exchanged spanned from technologies and other forms of material resources, expert knowledge in specific areas and information related to the targeted markets. The importance of cooperation with suppliers of technologies (e.g. in the cases of the eSOS: efficient and intelligent toilets project in Kenya and the Storm forecasts for Musanze project in Rwanda) and with clients of innovations (such as in the case of the Reducing water loss by improved data systems project in Kenya in which Upande Ltd partners with water utilities) has been studied in other contexts [79, 80]. The exchange of resources in the ICT-WIPs takes place through channels such as joint project teams (i.e. bringing together staff from partnering organisations) and division of labour among partnership members. In line with knowledge management theories, the meetings and other project events organised by such joint teams allow partnership members to share explicit knowledge (e.g. through information exchange) as well as tacit knowledge (e.g. through human resource exchange) [74, 77].

While cooperation with suppliers generally aims at complementing research and development (R&D)-related resources, partnering with customers and/or end-users provides access to relevant information and knowledge about market aspects such as customer preferences and prices they can afford and the market size – all of which reduces market uncertainties [81, 82]. The results in this study showed that conventional innovators such as universities and research institutions are not well represented in the partnerships. This finding is surprising though! Because the VIA Water programme supports water innovations that have just come out of the research phase and require a piloting period before scale up [15], one would expect knowledge institutes to be involved in such early stage innovations as they embody the related knowledge. Yet we also know that, traditionally, alliances with the aforementioned institutions are crafted when innovating companies need sophisticated and intensive R&D infrastructure and knowledge [83].

The innovation partnerships analysed in this study appear to also be motivated by the objective to reduce risks or costs associated with the development and implementation of their innovations. Although the ICT-focused water innovations supported by VIA Water are relatively small projects, the fact that most of them are essentially at the pilot stage implies that they involve huge risks and uncertainties which can be reduced through partnerships. The resources required to implement the pilots are not that huge, and they are in big part covered through the VIA Water seed capital fund; thus, the cost burden at the moment is relatively not heavy as the partnerships are not obliged to return the seed money in case of failure (or lack of innovation uptake). However, cost and risk reduction concerns will definitely increase when the partnerships start large-scale implementation of their innovation projects. This is the innovation stage that usually requires colossal amounts of money (generally acquired through bank loans or Venture capitalists) and exposes innovators to serious risks.

The fact that the partnerships acknowledge the importance of (and involve) stakeholders in their water innovation processes is in line with the literature on stakeholder engagement in the water sector. Wehn et al. [84] argue that when conceived as a social learning process, stakeholder engagement can lead to shared understanding and concerted actions to improve water governance. The value of stakeholder engagement in fostering smart water systems is further emphasised in the International Telecommunication Union′s report on ICTs and smart water management [71]. By involving stakeholders (the non-implementing partners), the VIA Water-supported partnerships are expected to be able to not only secure acceptance and ownership of their ICT-focused water innovations but also to have access to resources such as data and information possessed by these stakeholders which could only be acquired through cooperation with them. Partnering with relevant stakeholders in ICT-WIPs is finally consistent with insights from the innovation systems literature: notably that innovation is (and should be conceived as) resulting from complex interactions among various actors and institutions who both affect and are affected by the innovation process [35].

The two types of governance mechanisms – structural and relational – were found to be used concurrently in the partnerships. This finding does not support many of the previous studies which emphasised that relational governance is often preferred in innovation alliances and or networks due to its assumed advantages (i.e. flexibility) as compared to structural governance which is criticised for being costly and rigid [52–54]. In contrast, the finding here is consistent with some studies which underscore the importance of combining the two categories of governance mechanisms [46, 85]. In a study on 18 innovation networks in the Dutch agri-food sector, Tepic et al. [44] demonstrated the complementarity of both governance mechanisms. They conclude, among other things, that structural governance increases clarity of partnerships, while relational governance helps prevent attrition in highly uncertain conditions, especially in newly established innovation networks with limited previous cooperation. Similarly, Garbade et al. [48] in their analysis of 94 innovation alliances (in the Netherlands, Belgium, Germany and Austria) establish that structural agreements play a key role in setting up a platform for trust on which relational governance can strive.

Finally, this study has established that some of the VIA Water ICT-WIPs still experienced problems in spite of the governance mechanisms deployed. This suggests that there is still room for improvement on the mechanisms selected by the partnerships. We observed, for example, that the structural governance mechanisms that are mostly used include agreements such as MoUs and reference letters/endorsement letters. Although such documents are useful to make partnership obligations and rights clearer and, as such, to lower uncertainties to some extent, it is important to highlight that they are not legally enforceable agreements. They are just used by partnering organisations to express their willingness to participate, indicating an intended common line of action, and they are only settled internally. Agreements that imply a legal commitment are however extremely important to handle some of the key conflicts that might arise from the ICT-based water innovation partnerships. Appropriability of innovation outputs and distribution of innovation revenues is one of the areas where such agreements are needed [86].

For many of the partnerships analysed in this study, these issues do not seem to be a priority yet. In our discussions with some of these partnerships, they argued that their innovations are not “profit-based”, implying that once established they would be in the public domain. This was particularly the case when innovation partners considered themselves to be social enterprises. However, in other partnerships, lead innovators (e.g. Upande Ltd in Kenya) were thinking of securing intellectual property rights (IPR) for their innovation – although many were actually ignorant of how the acquisition of IPR works and how they should proceed. Thus, they had not anticipated any contractual arrangement to answer the key question often raised in open innovations projects, namely, who of the partnering organisations owns the property right for the innovation being developed [87]. Legally binding agreements are equally important for managing potential risks. We argue that when the VIA Water ICT-WIPs will start upscaling their innovations, the risks incurred will be extremely high – which justifies the design of risk-related agreements. Such agreements are useful notably to handle situations as when one partner goes bankrupt and is no longer able to deliver what he promised or decides to terminate the cooperation early in the execution of the partnership. Upfront agreements about all these issues will be a must for ICT-WIPs that are “for profit”.

7 Conclusions

In this section, we consider the major lessons that can be extracted from the explorative analysis conducted in this study. On the one hand, we reflect on the potential of innovation partnerships as an approach to boost the smart water systems agenda in the context of developing countries. On the other, we discuss the practical implications of the study results and formulate some recommendations.

7.1 Innovation Partnerships and Smart Water Systems in Developing Countries

The results of this study reinforce the argument made in the literature that smart water systems can be effectively promoted through a partnership approach, based on open innovation [7, 10, 14]. This is particularly true in the developing countries context where the capabilities to innovate are far less developed. As exemplified by the 24 ICT-focused water innovation projects analysed in this study, innovation partnerships can play an important role in enhancing capabilities of developing countries to implement the smart water systems agenda. To start with, in many of these countries, the adoption of smart water systems approach is slowed down by a lack of awareness and consciousness about the positive impact these systems can have if implemented on a large scale. This study has shown that ICT-WIPs allow a variety of relevant stakeholders in the water sector (such as municipalities, ministries, large utilities and regulatory agencies) to work together and get sensitised about the potential of smart water systems. Thus, these processes can foster the emergence of champions and advocates of smart water systems and, eventually, the creation of a critical mass of people and institutions who could push the smart water systems agenda beyond successful but isolated pilots. This push could eventually result in the establishment of national and sectoral institutional frameworks that promote and regulate the use of ICTs in the management of water systems.

The ICT-WIPs further represent a potentially transformational framework for promoting smart water systems in developing countries. The results of this study suggest that genuine partnerships can trigger a departure from the old-fashioned hierarchical model of North-South cooperation in which resources (financial, technology, knowledge, etc.) flow from the North to the South. Although most of the lead innovators in the analysed partnerships come from Western countries, our interviews suggest that the partners have a general feeling of being mutually and equally accountable for the innovation being developed and/or promoted. The innovation partnership approach indeed fosters the mindset that any partner – irrespective of their country affiliation – can be in the driver′s seat (e.g. being the lead partner) in the promotion of smart water systems.

In the same way, the innovation partnership approach promotes the culture of mutual learning, thus allowing partners to strengthen each other′s innovation competences relating to smart water systems in developing countries. This is particularly true when – as in the case of the ICT-WIPs analysed in this study – some partners come from foreign countries and thus need the knowledge and experience of local partners (e.g. on problems facing water systems, existing solutions and their weaknesses, possible local risks). The ICT-WIPs analysed in this study involved various types of mechanisms that foster a culture of mutual learning. Most of the partnerships adopted a model of decision-making that was collaborative and supported by open dialogue. Mechanisms such as regular joint planning and review events (either physically of via Skype calls) and joint committees enabled information and knowledge sharing among partnership members. In particular, through such mechanisms domestic organisations operating in local environments could contribute their local knowledge which is critical for developing relevant ICT solutions for water management.

Finally, the nature of the ICT-focused water innovations analysed in this study leads to the insight that fostering smart water systems in developing countries requires a “rethink” of not only the technologies themselves but also the business models around them. In other words, what is needed is not just ICTs but the development of ICTs that fit the peculiar problems (technological, financial, social, etc.) of developing countries. For instance, it is well known that weak ICT infrastructure (e.g. low Internet bandwidth) is one of the key barriers for implementing smart water systems in developing countries. There are also issues regarding affordability of mainstream ICT solutions for water management, which explains why the implementation of smart water systems in many developing countries often relies on external (and unsustainable) donor funding [10]. This implies that the ICTs developed and used in developed countries are not necessarily appropriate for developing countries. Thus, in line with frugal innovation literature [88, 89], we argue that ICT-focused water innovations needed in developing countries are those that are easy to use, low-cost and robust enough to solve the complex problems facing water systems in developing countries. One way to achieve the parallel objectives of offering quality ICTs (for water) at an attractive cost in these countries seems to be the use of innovation partnerships. As exemplified by the VIA Water innovations analysed in this study, these ICTs can be codeveloped and/or co-promoted by either domestic or foreign organisations (e.g. Upande′s low-cost sensors, TAHMO low-cost weather stations). They could also be produced through “polycentric innovation” [90] in which ICT multinational enterprises (such as IBM and Google) and local organisations collaborate.

7.2 Practical Implications

This study has the following practical implications. First, the findings show that ICT-focused water innovation partnerships are aimed at exchanging knowledge resources and that exchanging tacit knowledge can be a difficult task. A practical implication is that such partnerships should pay heed to the design of appropriate mechanisms to exchange tacit knowledge. This can be done notably by setting up strategies that enable an open flow of information among partnering organisations and continuous co-creation and use of knowledge by the partners throughout the entire life time of the partnership. Some of the strategies identified in the ICT-WIPs analysed in this study include the establishment of joint innovation teams, division of labour, human resource exchange and online collaborative systems.

Second, the study results showed that fruitful ICT-WIPs do not naively rely on mutual trust, as evidenced by the partnership challenges described in Sect. 5.4 (e.g. low levels of partner commitment, asymmetries in resource exchange and ambiguity of partnership objectives). This calls for a careful combination of structural and relational partnership governance mechanisms. Organisations partnering to implement ICT-focused water innovation projects should acknowledge that formal control and trust-based governance mechanisms reinforce each other in safeguarding the exchange of resources in the partnerships. Thus, even when they have all good reasons to trust in the ability of their partners to deliver on their promises, they should strive to set up formalised agreements right from the start of the partnership. Structural agreements should be envisioned in a variety of areas, spanning from dispute resolution mechanisms to knowledge protection, decision-making procedures, incentive systems, performance monitoring and evaluation, intellectual property rights, appropriability of innovation results and sharing of innovation revenues [62].

Third, the study confirmed the increasing recognition that ICT-focused water innovations cannot be implemented in a vacuum. Designing and executing effective partnerships for smart water systems requires, therefore, that partners implementing

ICT-focused water innovation projects fully recognise the importance of genuine engagement of non-implementing partners. Only then can they learn from (and with) them about the nature of the problems faced in water management in particular contexts (or countries) and thus provide appropriate solutions to those problems.

Fourth, the results of this study suggest that effective ICT-WIPs rely on the calibre of partnering organisations, i.e. their capability to contribute what they promised (in expected quantities and qualities). This implies that partner selection is a critical stage in the partnership formation process; it should be carefully conducted in order to determine and choose partners who can really add value to the partnership. With regard to partner selection criteria, Geringer [91] proposes two categories: task-related and partner-related criteria. The former concerns typical and distinct resources that a prospective partner would bring to the partnership – to bridge the resource gaps of other partners – given the requirements of the innovation project. The latter category refers to criteria pertaining to the ability of potential partners to efficiently and effectively work together in a partnership. They include prior experience of the partners with partnerships, trust in each other′s capability to deliver, market familiarity and so on. One could also include strategic criteria such as the partners′ convergent expectations for starting the innovation project, the possibility for diversification and future business and the likelihood to share investments costs [91–93].

Finally, this study showed the instrumental role of VIA Water in facilitating the creation of many of the ICT-WIPs analysed. A practical implication from this finding is that innovating organisations who intend to team up with others should make use of innovation brokers or like-minded entities. Innovation brokers help to identify reliable partners and to fully understand local risks and how to mitigate them, among other things. They can also help understand and interpret local culture. Another key implication linked to the issue of “culture interpretation” is that organisations who want to partner with others for innovation purposes should acknowledge that some partnership governance mechanisms might be more relevant and effective in specific cultural contexts than in others. For example, trust-based mechanisms cannot be relied upon to the same extent across all cultural settings. Thus, in order to effectively manage partnerships, innovating organisations must fully consider the societal norms and value systems of the countries where they select partners.

7.3 Recommendations

7.3.1 Recommendations for Future Research

This study focused on ICT-WIPs that are still in early stages of implementation. Subsequent research should investigate the performance of these partnerships at later stages of the innovation process, notably during the scale-up stage. Future research should also go beyond the sample limitation of our research, i.e. the fact that we focused on only the ICT-WIPs supported by the VIA Water programme in Africa. We assume that the collection of data on ICT-WIPs crafted and executed in other conditions (e.g. not funded by the same partner, other continents) might produce additional insights. Finally, it would be highly relevant to undertake comparative studies to find out how ICT-focused water innovations developed and implemented through partnerships compare (in terms of success) with those that are not generated through partnerships.

7.3.2 Recommendations for VIA Water

This study results showed that the analysed ICT-WIPs faced some challenges, which can be partly explained by the limited use of legally enforceable partnership governance instruments (contractual agreements). It also appeared from this study that VIA Water facilitated linkages between partners in the ICT-WIPs and requested lead innovators to provide proof of commitment (to collaborate) by selected working partners. In most cases, such proof consisted of MoUs and reference letters. It is recommended that, for future partnerships, VIA Water – in its capacity as innovation broker – takes this a step further and advises partners to formalise their partnerships. Notably, in the same way that VIA Water itself signs a contract with lead innovators, these should be encouraged (or perhaps requested) to sign formal contracts too with their working partners (e.g. a partnership agreement). Arguably, this would increase the level of commitment of the different partners to the partnership and the successful generation of the envisaged innovation.

Acknowledgements This research was financially supported by the Dutch Ministry of Foreign Affairs to generate insights on water innovation processes in Africa. We are grateful for the constructive comments by Dick Bouman, Senior Fund Manager at Aqua for All, on a draft version of this chapter.

Annex: Overview of Lead Innovators, Type of Organisations They Are and Their Country of Origin

Lead innovator	Type of organisation	Country of origin
1. dloHaiti	Enterprise (company)	USA
2. Markets Merger Ltd	Enterprise (start-up)	Rwanda
3. Delft University of Technology	University	Dutch
4. SkyFox Ltd	Enterprise (start-up)	Ghana
5. Empower People	Enterprise (start-up)	Dutch
6. SnapEX Overseas	Enterprise (company)	Indian
7. Royal HaskoningDHV	Consulting firm	Dutch
8. HKV Consultants	Consulting firm	Dutch

(continued)

Lead innovator	Type of organisation	Country of origin
9. Maji Milele Ltd	Enterprise (start-up)	Kenyan (owned by a Dutch entrepreneur)
10. MobiTech Water Solutions	Enterprise (start-up)	Kenyan
11. Sanergy Kenya Ltd	Enterprise (start-up)	Kenyan (USA founded)
12. IHE Delft Foundation	University	Dutch
13. Upande Ltd	Enterprise (start-up)	Kenyan (owned by a "Dutch" entrepreneur)
14. Mobile Water Management	Enterprise (start-up)	Dutch
15. Orvion B.V.	Enterprise (company)	Dutch
16. BoP Innovation Center	Consultancy	Dutch
17. Kaicedra-Consulting	Enterprise (start-up)	Mali
18. World Waternet	Utility Branch with NGO status	Dutch
19. Protos	NGO	Belgian
20. Niger River Basin Agency	Intergovernmental	Regional
21. Flood Tags	Enterprise (company)	Dutch
22. Benin Country Water Partnership	NGO	Benin
23. Water & Sanitation for the Urban Poor	Not-for-profit company	United Kingdom
24. Deltares	Research Institute	Dutch

References

1. United Nations General Assembly (2010) The human right to water and sanitation. Resolution 64/292
2. ADB (Asian Development Bank) (2013) Asian Development Outlook: measuring water security in Asia and the Pacific. Asian Development Bank, Mandaluyong
3. UNESCO (2012) Managing water under uncertainty and risk, United Nations World Water Development. Report 4, vol 2 (Knowledge Base). UNESCO, Paris
4. World Economic Forum (2012) Global risks 2012. World Economic Forum, Geneva
5. UN-Water & FAO (2007) Coping with water scarcity. Challenge of the twenty-first century. http://www.fao.org/nr/water/docs/escarcity.pdf
6. United Nations General Assembly (2015) Transforming our world: the 2030 Agenda for sustainable development. United Nations, New York
7. ITU (2010) ICT as an enabler for smart water management. ITU-T Technology. Watch Report October 2010
8. Kappor T (2006) Role of information and communication technology in adaptive integrated water resource management. World Environmental and Water Resource Congress, pp 1–10. Retrieved from https://doi.org/10.1061/40856%28200%291
9. Meena M, Singh K (2012) Information and communication technology for sustainable natural resource management. Retrieved from http://mpra.ub.uni-muenchen.de/45818/1/MPRA_paper_45818.pdf
10. Ndaw MF (2015) Unlocking the potential of information communications technology to improve water and sanitation services. Water and Sanitation Programme, The World Bank

11. De Souza PF, Mackintosh GS, Suites C, Town C, Management E, Africa S (2014) Lessons learnt from the implementation of ICT applications to support the water sector, pp 1–16. Retrieved from http://greenseesaw.files.wordpress.com/2014/05/wisa2014-081-schaub-jones-et-al-b.pdf
12. Wehn U, Montalvo C (2018) Exploring the dynamics of water innovation: foundations for water innovation studies. J Clean Prod 171(S):1–19
13. Tutusaus M, Schwartz K, Smit S (2017) The ambiguity of innovation drivers: the adoption of information and communication technologies by public water utilities. J Clean Prod 171:79–85
14. Hope R, Foster T, Money A, Rouse M, Money N, Thomas M (2011) Smart water systems. Project report to UK DFID, April 2011. Oxford University, Oxford
15. Nagel W, Wouters T, van der Weerd K (2018) VIA Water: a programme to support co-created water innovations in African cities. J Clean Prod 171:140–143
16. Ntelekos AA, Oppenheimer M, Smith JA, Miller AJ (2010) Urbanization, climate change and flood policy in the United States. Clim Change 103:597
17. Sun G, McNulty SG, Myers JAM, Cohen EC (2008) Impacts of climate change, population growth, land use change, and groundwater availability on water supply and demand across the conterminous U.S. AWRA Watershed Update 6(2):1
18. Kindom B, Liemberger R, Marin P (2006) The challenge of reducing non-revenue water in developing countries. The World Bank, Washington
19. Mauree V (2010) ICTs as an enabler for smart water management. ITU-Technology Watch Report
20. Vossen RW (1998) Combining small and large firm advantages in innovation: theory and practice, Graduate School/Research Institute Systems, Organisation and, Groningen
21. Hagedoorn J, Duysters G (2002) The effect of mergers and acquisitions on the technological performance of companies in a high-tech environment. Tech Anal Strat Manag 14(1):68–85
22. Hagedoon J, Schakenraad J (1994) The effect of strategic technology alliances on company performance. Strateg Manag J 15:291–309
23. DeMan AP, Duysters G (2005) Collaboration and innovation: a review of the effects of mergers, acquisitions and alliances on innovation. Technovation 25:1377–1387
24. Chesbrough H (2003) Open innovation: the new imperative for creating and profiting from technology. Harvard Business School, Boston
25. De Man A-P, Duysters G (2005) Collaboration and innovation: a review of the effects of mergers, acquisitions and alliances on innovation. Technovation 25(12):1377–1387
26. Chesbrough HW, Vanhaverbeke W, West J (eds) (2006) Open innovation: researching a new paradigm. Oxford University Press, Oxford
27. Barney JB (1991) Firm resources and sustained competitive advantage. J Manag 17:99–120
28. Grant RM (1996) Toward a knowledge-based theory of the firm. Strateg Manag J 17:109–122
29. Harrison JS, Hitt MA, Hoskisson RE, Ireland RD (2001a) Resource complementarity in business combinations: extending the logic to organizational alliances. J Manag 27:679–690
30. Harrison JS, Hitt MA, Hoskisson RE, Ireland RD (2001b) Resource complementarity in business combinations: extending the logic to organizational alliances. J Manag 27(6):679–690
31. Nooteboom B (2000) Learning by interaction: absorptive capacity, cognitive distance and governance. J Manag Gov 4(1–2):69–92
32. Nonaka I, von Krogh G (2009) Perspective-tacit knowledge and knowledge conversion: Controversy and advancement in organizational knowledge creation theory. Org Sci 20 (3):635–652
33. Eisenhardt KM, Schoonhoven CB (1996) Resource-based view of strategic alliance formation: strategic and social effects in entrepreneurial firms. Organ Sci 7:136–150
34. Freeman C (2002) Continental, national and sub-national innovation systems complementarity and economic growth. Res Policy 31:191–211
35. Lundvall B-A (ed) (1992) National systems of innovation. Anthem Press, London
36. Nelson RR (ed) (1993) National innovation systems. Oxford University Press, New York

37. Godin B (2007) National Innovation System: The system′s approach in historical perspective. Working Paper No. 36, Project on the History and Sociology of STI Statistics (Montreal, Canada)
38. Edquist C (ed) (1997) Systems of innovation: technologies, institutions, and organizations. Pinter, London
39. Enzing C (2009) Product innovation in the Dutch food and beverage industry: a study on the impact of the innovation process, strategy and network on the product′s short-and long-term market performance PhD thesis, Wageningen University
40. Deeds DL, Hill CW (1999) An examination of opportunistic action within research alliances: Evidence from the biotechnology industry. J Bus Ventur 14:141–163
41. Farr CM, Fischer WA (1992) Managing international high technology cooperative projects. R&D Manag 22(1):55–68
42. Hamel G (1991) Competition for competence and inter-partner learning within international strategic alliances. Strateg Manag J 12(S1):83–103
43. Faems D, Janssens M, Madhok A, Van Looy B (2008) Toward an integrative perspective on alliance governance: connecting contract design, trust dynamics and contract application. Acad Manage J 51:1053–1078
44. Tepic M, Omta OS, Trienekens JH, Fortuin FT (2011) The role of structural and relational governance in creating stable innovation networks: insights from sustainability-oriented Dutch innovation networks. J Chain Netw Sci 11(3):197–211
45. Williamson OE (1989) Transaction cost economics. In: Schmalensee R, Willig R (eds) Handbook of industrial organization. North-Holland, New York
46. Poppo L, Zenger T (2002) Do formal contracts and relational governance function as substitutes or complements? Strateg Manag J 23:707–725
47. Vlaar PWL, Bosch FAJ, Van den Volberda HW (2007) On evolution of trust, distrust and formal coordination and control in inter-organisational relationships: toward an integrative framework. Group Org Manag 32:407–429
48. Garbade PJP, Omta SWF, Fortuin FTJM (2015) The interplay of structural and relational governance in innovation alliances. J Chain Netw Sci 16(2):117–134
49. Nooteboom B (1999) Inter-firm alliances: analysis and design. Routledge, London
50. Uzzi B (1997) Social structure and competition in interfirm networks: the paradox of embeddedness. Adm Sci Q 42(1):35–67
51. Volberda HW (1998) Building the flexible firm: how to remain competitive. Oxford University Press, Oxford
52. Dyer JH, Singh H (1998) The relational view: cooperative strategy and sources of inter-organizational competitive advantage. Acad Manage Rev 23(4):660–679
53. Gulati R (1995) Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances. Acad Manage J 38:85–112
54. Larson A (1992) Network dyads in entrepreneurial settings: a study of the governance of exchange relationships. Adm Sci Q 37:76–104
55. Gil N (2009) Developing cooperative project client-supplier relationships: how much to expect from relational contracts? Calif Manage Rev 51(2):144–169
56. Madhok A (1995) Opportunism and trust in joint venture relationships: an exploratory study and a model. Scand J Manag 11:57–74
57. Das TK, Teng BS (1998) Between trust and control: developing confidence in partner cooperation in alliances. Acad Manage Rev 23:491–512
58. Jones GR, George JM (1998) The experience and evolution of trust: implications for cooperation and teamwork. Acad Manag Rev 23:531–546
59. Zenger T, Lazzarini S, Poppo L (2002) Informal and formal organization in new institutional economics. Adv Strateg Manage 19:277–305
60. Dekker HC (2004) Control of inter-organizational relationships: evidence on appropriation concerns and coordination requirements. Acc Organ Soc 29:27–49

61. Hartmann A, Caerteling J (2010) Subcontractor procurement in construction: the interplay of price and trust. Supply Chain Manage 15(5):354–362
62. Gulati R (2007) Managing network resources: alliances, affiliations and other relational assets. Oxford University Press, Oxford
63. Hofstede G (1997) Cultures and organizations: software of the mind. McGraw-Hill, London
64. Cannon JP, Doney PM, Mullen MR, Petersen KJ (2010) Building long-term orientation in buyer–supplier relationships: the moderating role of culture. J Oper Manag 28(6):506–521
65. Handley SM, Angst CM (2015) The impact of culture on the relationship between governance and opportunism in outsourcing relationships. Strateg Manag J 36:1412–1434
66. Mvulirwenande S Wehn U, Leliveld A (2017) Analysing frugal innovation incubation programmes: a case study from the water sector. Conference paper 7–8 November 2017. Centre for Frugal Innovation, Leiden, The Netherlands
67. GSMA Intelligence (2017) The mobile economy. Sub-Saharan Africa. www.gsmaintelligence.com
68. Snow JT, Bonizella B, Benchwick G, Georgie G, Hoedjes J, Miller A, Usher J (2016) A new vision for weather and climate services in Africa. UNDP, New York
69. VIA Water (2018) Programme database. VIA Water, Delft
70. Winch GM, Courtney R (2007) The organization of innovation brokers: an international review. Tech Anal Strat Manag 19:747–763
71. ITU (2014) Partnering for solutions. ICTs in Smart Water Management. ITU, Geneva
72. Biggs T, Shah MK (2006) African small and medium enterprises, networks, and manufacturing performance. World Bank Policy Research Working Paper 3855
73. Tshikuku K (2001) Culture, entrepreneurship and development in Africa. http://citeseerx.ist.psu.edu/viewdoc/download?doi¼10.1.1.202.3380&rep¼rep1&type¼pdf
74. Nonaka I, Takeuchi H (1995) The knowledge-creating company. How Japanese companies create the dynamics of innovation. Oxford University Press, New York
75. Batterink M (2009) Profiting from external knowledge: how firms use different knowledge acquisition strategies to improve their innovation performance. Wageningen Academic Publishers, Wageningen, 376 p
76. Garbade PJP, Omta O, Fortuin FTJM (2015) The interplay of structural and relational governance in innovation alliances. J Chain Netw Sci 16(2):1–18
77. Polanyi M (1966) The tacit dimension. Routledge and Kegan Paul, London
78. Grant R, Bade-Fuller CA (2004) A knowledge accessing theory of strategic alliances. J Manag Stud 41(1):61–84
79. Bidault F, Despres C, Butler C (1998) The drivers of cooperation between buyers and suppliers for product innovation. Res Policy 24:97–114
80. Tether BS (2002) Who co-operates for innovation, and why. An empirical analysis. Res Policy 31:947–967
81. Kotabe M, Swan KS (1995) The role of strategic alliances in high-technology new product development. Strateg Manag J 16:621–636
82. Shaw B (1994) User–supplier links and innovation. In: Dodgson M, Rothwell R (eds) The handbook of industrial innovation. Edward Elgar, Cheltenham
83. Archibugi D, Coco A (2004) International partnerships for knowledge in business academia: a comparison between Europe and the USA. Technovation 24(7):517–528
84. Wehn U, Collins K, Anema K, Basco-Carrera L, Lerebours A (2018) Stakeholder engagement in water governance as social learning: lessons from practice. Water Int 43(1):34–59
85. Faems D, Janssens M, Van Looy B (2007) The initiation and evolution of interfirm knowledge transfer in R&D relationships. Organ Stud 28(11):1699–1728
86. Teece DJ (1986) Profiting from technological innovation: implications for integration, collaboration, licensing and public policy. Res Policy 15:285–305
87. Shan W (1990) An empirical analysis of organizational strategies by entrepreneurial high-technology firms. Strateg Manag J 11:129–139
88. Bhatti Y (2012) What is frugal, what is innovation? Towards a theory of frugal innovation. Working Paper Said Business School, Oxford University, Oxford

89. Radjou N, Prabhu J (2015) Frugal innovation: how to do more with less. The Economist Books, London
90. van Beers C, Knorringa P, Leliveld A (2014) Frugal innovation in Africa. Towards a research agenda. Position paper
91. Geringer JM (1991) Strategic determinants of partner selection criteria in international joint ventures. J Int Bus Stud (1st Qtr) 22(1):41–62
92. Contractor FJ, Lorange P (1988) Why should firms cooperate? The strategy and economics basis for cooperative ventures. In: Contractor FJ, Lorange P (eds) Cooperative strategies in international business. Lexington Books, Lexington, pp 3–28
93. Harrigan KR (1985) Strategies for joint ventures. Lexington Books, Lexington

Website

https://www.viawater.nl

Exploring Assimilation of Crowdsourcing Observations into Flood Models

M. Mazzoleni, Leonardo Alfonso, and D. P. Solomatine

Contents

Abstract This chapter aims to describe the latest innovative approaches for integrating heterogeneous observations from static social sensors within hydrological and hydrodynamic modelling to improve flood prediction. The distinctive characteristic of such sensors, with respect to the traditional ones, is their varying lifespan and space-time coverage as well as their spatial distribution. The main part of the chapter is dedicated to the optimal assimilation of heterogeneous intermittent data within hydrological and hydraulic models. These approaches are designed to

M. Mazzoleni (*)
Department of Earth Sciences, Uppsala University, Uppsala, Sweden

Centre of Natural Hazards and Disaster Science (CNDS), Uppsala, Sweden
e-mail: maurizio.mazzoleni@geo.uu.se

L. Alfonso
IHE Delft Institute for Water Education, Delft, The Netherlands

D. P. Solomatine
IHE Delft Institute for Water Education, Delft, The Netherlands

Delft University of Technology, Delft, The Netherlands

Andrea Scozzari, Steve Mounce, Dawei Han, Francesco Soldovieri, and Dimitri Solomatine (eds.), ICT for Smart Water Systems: Measurements and Data Science, Hdb Env Chem (2021) 102: 209–234, https://doi.org/10.1007/698_2019_403,
© Springer Nature Switzerland AG 2019, Published online: 13 September 2019

account for the intrinsic uncertainty contained into hydrological observations and model structure, states and parameters. Two case studies, the Brue and Bacchiglione catchments, are considered. Finally, the evaluation of the developed methods is provided. This study demonstrates that networks of low-cost static and dynamic social sensors can complement traditional networks of static physical sensors, for the purpose of improving flood forecasting accuracy. This can be a potential application of recent efforts to build citizen observatories of water, in which citizens not only can play an active role in information capturing, evaluation and communication but also can help improve models and increase flood resilience.

Keywords Crowdsourced observations, Data assimilation, Flood forecasting, Hydraulic modelling, Hydrological modelling

1 Introduction

The impact of natural hazards on societies and economies has drastically increased in the last years due to many natural and anthropogenic factors, including climate change [1, 2]. For this reason, the demands for non-structural measures able to accurately and timely forecast in real-time river water level to allow decision-makers to take the most effective and timely decisions for reducing harm or loss have significantly increased [3–5]. Among different types of water system models, hydrological and hydrodynamic models are the most utilised ones in flood early warning systems in river basins.

Unfortunately, deterministic predictions contain an intrinsic uncertainty due to many sources of error that propagate through the model and therefore affect its output [6]. In fact, uncertainty can be due to either the inherent stochastic nature and variability of hydrological processes, i.e. aleatory uncertainty [7, 8], or to our imperfect state of knowledge of the hydrological system and our limitedness to model it, i.e. epistemic uncertainty [9–12]. Three main sources of uncertainty can be identified [13] in hydrological and hydrodynamic modelling: (a) observation uncertainty, which is the approximation in the observed hydrological variables used as input or calibration data (e.g. rainfall, temperature and river discharge); (b) parameter uncertainty, which is induced by imperfect model calibration; and (c) model structural uncertainty, which is a result of the inability of models to perfectly schematize the physical processes involved. Epistemic uncertainty can be associated with the latest two sources of uncertainty previously mentioned due to limited knowledge about the physical behaviour of the system.

A reliable characterisation and reduction of the uncertainties affecting hydrological and hydrodynamic processes is an important scientific and operational challenge [14–17]. Different approaches like the first-order reliability method [18], probabilistic Monte Carlo (MC) and fuzzy rule-based methods [19–21] can be used to assess model uncertainty.

Several research activities aimed to reduce such uncertainty in the flood estimation, predictive uncertainty, have been carried out due to its importance to the decision of issuing a flood warning [5, 22, 23]. Methods like the UNcertainty Estimation based on Local Errors and Clustering (UNEEC, [24–26]), Generalised Likelihood Uncertainty Estimation (GLUE, [27, 28]) and Machine Learning in parameter Uncertainty Estimation (MLUE, [29, 30]) can be employed to assess uncertainty in water system models and estimate predictive uncertainty (see, e.g. [31, 32]). However, such tools are often not used in operational forecasting by environmental agencies and river basin authorities, perhaps because of their belief that uncertainty analysis cannot be incorporated into the decision-making process and because uncertainty analysis is too subjective, among others [5, 11, 33].

In the last decades, model updating techniques for reducing predictive uncertainty approaches have been increasingly studied and implemented in water-related applications. These approaches allow for changing model input, states, parameters or output in response of new observations coming into the model in order to improve the prediction accuracy and quantifying uncertainty [3, 14, 34]. In most of the cases, model updating occurs only in form of data assimilation using information of streamflow, soil moisture, etc. coming from static physical stations. Model updating techniques are rarely implemented in operational forecasting due to the lack of approaches to quantify the uncertainty in real-time observations from multiple sources across a range of spatiotemporal scales and methods to integrate these new information in an appropriate and transparent way. In this respect, in operational practice it is preferred to correct the model inputs (in most of the cases), states, initial conditions and parameters in an empirical and subjective way rather than apply advanced (optimal) data assimilation techniques for improving hydrologic forecast [35]. Welles et al. [36] and Liu et al. [34] pointed out how the need for implementing reliable data assimilation methods in operational forecast is increasing in order to fill the mentioned gap with the scientific world.

Traditionally, static physical sensors, such as pressure sensors, water level sensors, and pluviometers, are commonly used by water authorities to calibrate, validate and (in some cases) update physical models in real time. However, the main problem of physical sensors is the proper maintenance which can be very expensive in case of a vast network as well as the limited data that existing sparse monitoring networks can provide to this end.

The continued technological advances have stimulated the spread of low-cost sensors that has triggered crowdsourcing as a way to obtain observations of hydrological variables in a more distributed way than the classic static physical sensors [37]. The main advantage of using these types of sensors is that they can be used not only by technicians, as is the case of traditional physical sensors, but also by regular citizens. Recently, citizen science activities have been widely promoted in order to allow citizens to participate in different aspects of environmental planning and management. One of the most common activities to achieve such goal includes involving citizens in data collection, or crowdsourcing (CS). In particular, observations of hydrological variables can generate additional knowledge, in relation to the water cycle, and use such knowledge in decision-making [38, 39]. However, because of their relatively limited reliability, and random accuracy in time and space, crowdsourced observations have not been widely integrated in hydrological and/or hydraulic models for flood forecasting applications. Instead, they have generally been used to validate model results against observations, in post-event analyses. Different studies addressed the issue of assimilation of distributed observations in distributed and semi-distributed hydrological models (e.g. [40–43]). Neither of the previous studies considers the dynamic nature of data from heterogeneous sensors which provide an intermittent signal in time and space. In fact, the information coming from a specific sensor might be sent just once, occasionally or in time steps that are non-consecutive, i.e. with intermittent observations having different lifespans.

A number of studies have developed methods for using crowdsourced citizens-based observations in water-related models [44–56]. In particular, crowdsourced information are used for directly creating deterministic or probabilistic flood maps [48], derive stream discharges and flow velocities fields [57] and flood extent [52]. In alternative, crowdsourced data have been used for validating flood models [44, 56]. A detailed review on the use of citizen observations for flood modelling applications is provided in Assumpção et al. [58]. However, none of the previous studies assessed the usefulness of citizen observations for improving flood predictions [39, 59]. The first attempts to study the effects of assimilating crowdsourced citizen observations in hydrological and hydraulic models for improving flood prediction in real-time applications are reported in Mazzoleni et al. [60–62] and Mazzoleni [63]. Just recently, Mazzoleni et al. [64] proposed two innovative approaches to assimilated qualitative flow data within hydrologic routing models.

In this chapter, we describe the proposed innovative methods to assimilate heterogeneous intermittent observations, coming from social sensors, within hydrological and hydrodynamic modelling to improve flood prediction. This research was carried out under the framework of the European project WeSenseIt (https://www.wesenseit.com/) [65].

2 Crowdsourced Observations

In this chapter, we consider two different types of sensors to measure hydrological variables such as water level: static physical (StPh) and static social (StSc) sensors (see Fig. 1). In addition, also dynamic social sensors may be used but are not included in this chapter. An example of a static social sensor is a staff gauge located in a strategic point of the river used by citizens to estimate water depth values using a mobile phone app to send CS observations using the QR code as geographical reference point. An example of dynamic sensor is a mobile app allowing any citizen to send the information related to the distance between the water profile and the river bank using a mobile app at random locations along the river. It might be in fact difficult to estimate the water depth value without having any indication about river

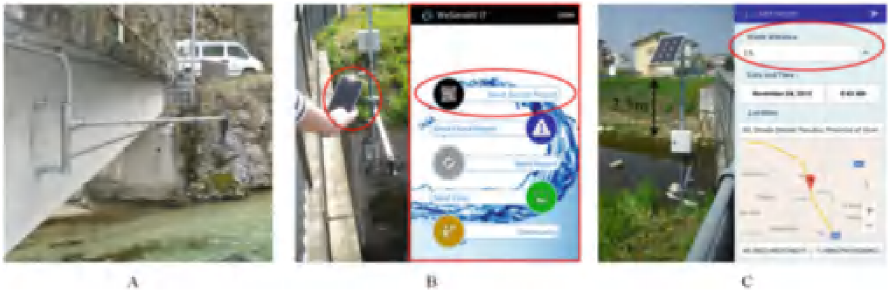


Fig. 1 Proposed sensors classification with (a) static physical sensors (StPh), (b) static social sensors (StSc), and (c) dynamic social sensors (DySc)

depth. In this case, the CS observations have higher degree of uncertainty due to the indirect method used to estimate water depth value.

According to the nature of the sensor, uncertainty can be defined either as a probability distribution (quantitative observation) or a fuzzy set (qualitative or semi-qualitative observations).

During the last decades, probability theory has been applied in order to represent epistemic or observational uncertainty in mathematical models. In particular, quantitative observations of physical variables can be expressed as a stochastic variable with a given probability distribution which represents the likelihood of that variable value to take on a given value. In most of the cases, stochastic variables are represented using a normal distribution with assigned mean and standard deviation. The higher is the standard deviation, the higher the uncertainty of that variable is.

Examples of qualitative information can be found in verbal or text messages coming from social networks (Twitter, Facebook, etc.). Fuzzy logic emerged as a more general form of logic that can handle the concept of possibilistic values or partial truth. This approach has been used recently [64] as a qualitative modelling methodology since it allows for an easier transition between human and computers for decision-making (transition from fuzzy to numerical data), and it is able to handle imprecise and uncertain information [66]. From a statistical point of view, a physical variable can be associated to a deterministic value plus a given degree of uncertainty, expressed as a pdf, or the second or third order moment. In fuzzy logic-based approach, a physical variable value (e.g. precipitation) would belong to a specific fuzzy set having given characteristic (e.g. low, medium, high precipitation).

3 Case Studies and Water-Related Models

Two different case studies having different hydrometeorological characteristics are analysed in this book chapter. The case studies are the Brue catchment (UK) and the Bacchiglione catchment (Italy). Different hydrological and hydraulic models are implemented within each case study. In particular, a semi-distributed version of a continuous Kalinin-Milyukov-Nash (KMN) cascade hydrological model is applied on the Brue catchment, while a semi-distributed hydrological and hydraulic model developed by the Alto Adriatico Water Authority is implemented in the Bacchiglione catchment. In this study, synthetic flow observations derived from observed and simulated quantitative streamflow are used. Synthetic data are used to evaluate the potential of the proposed approaches as real qualitative observations may be affected by different unpredictable errors.

3.1 Brue Catchment (UK)

The Brue catchment is located in Somerset, South West England, with a drainage area of about 135 km^2 and a time of concentration of 10 h at the catchment outlet, Lovington. Hourly precipitation data are supplied by the British Atmospheric Data Centre from the NERC Hydrological Radar Experiment Dataset (HYREX) project [67, 68] and available at 49 automatic rain stations; average annual rainfall of 867 mm is measured in the period between 1961 and 1990. Discharge is measured at the catchment outlet by one station at a 15 min time step resolution, having an average value of 1.92 m^3/s. For both precipitation and discharge data, a 3-year complete data set, between 1994 and 1996, is available.

A semi-distributed hydrological model is used to assess the flood hydrograph at the outlet section of the Brue catchment and to represent the spatial variability of the CS flow observations. The Brue catchment is divided into 68 sub-catchments having a small drainage area (on average around 2 km^2) so that any observation at a random location in a given sub-catchment would provide the same information content that an observation at the outlet of same sub-catchment [60]. For each sub-catchment, a conceptual lumped hydrological model, continuous Kalinin-Milyukov-Nash (KMN) cascade, is implemented to estimate the outflow discharge [69]. The KMN model considers a cascade of storage elements (or reservoirs), assuming that the relation between stage, discharge and stored water volume is linear and that the water storage x_t is only a function of the outflow of the reach Q_t [60]. Subsequently, the KMN is represented as a dynamic state-space system to apply data assimilation techniques as explained in the previous section. In the case of the linear systems, the discrete state-space system can be represented as follows [69]:

$$x_t = \boldsymbol{\Phi} x_{t-1} + \boldsymbol{\Gamma} I_t + w_t \tag{1}$$

$$z_t = H x_t + v_t \tag{2}$$

where t is the time step, x is vector of the model states (stored water volume in m^3), $\boldsymbol{\Phi}$ is the state-transition matrix (function of the model parameters n and k), $\boldsymbol{\Gamma}$ is the input-transition matrix, H is the output matrix and I and z are the input (forcing) and model output, while w and v are the system and measurements errors.

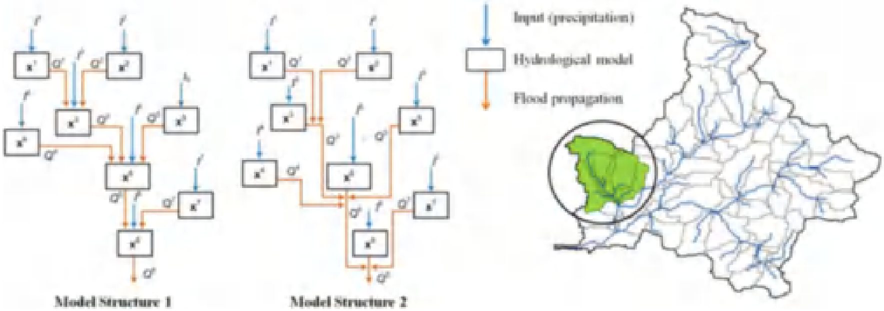


Fig. 2 Considered model structures (MS) for the semi-distributed hydrological model

Muskingum channel routing method [70] is used for flow propagation between sub-catchments; for details see Mazzoleni et al. [60]. The semi-distributed model is structured in such a way that the sub-catchments are sequentially connected and the output of the upstream sub-catchments is used as input in the downstream ones (see Fig. 2). More details about the model calibration are reported in Mazzoleni et al. [60].

3.2 Bacchiglione Catchment (Italy)

The Bacchiglione River catchment is located in the north-east of Italy and tributary of the River Brenta which flows into the Adriatic Sea at the south of the Venetian Lagoon and at the north of the River Po delta. The considered area is the upstream part of the Bacchiglione River, which has an overall area of about 400 km^2, river length of about 50 km, river width of 40 m and river slope of about 0.5% [71]. The main urban area is Vicenza, located in the downstream part of the study area, where recent floods were registered during the springs of 2010 and 2013. Within the activities of the WeSenseIt project [72], one StPh sensor and ten StSc sensors (staff gauges complemented by a QR code, as represented in Fig. 1) were installed in the Bacchiglione River to measure water level (see Fig. 3). Hourly information related to rainfall, temperature, wind direction and intensity, humidity, snow, solar radiation and water level are available for the last 12 years.

In order to represent the distributed hydrological response of this catchment, a semi-distributed model, in which the output of the hydrological model is used as boundary conditions in the hydraulic model, has been implemented.

The hydrological response of the catchment is estimated using the hydrological model developed by the Alto Adriatico Water Authority (AAWA) that considers the routines for runoff generation, having precipitation as model forcing, and a simple routing procedure. The processes related to runoff generation are modelled mathematically by applying the water balance to a control volume, of soil depth, representative of the active soil at the sub-catchment scale. The water content is estimated

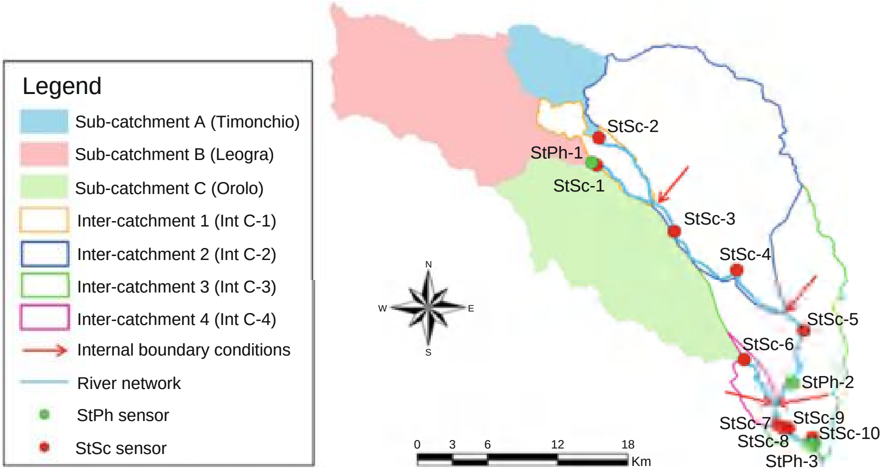


Fig. 3 Structure of the semi-distributed model for the Bacchiglione catchment and location of the static physical (StPh) and social (StSc) sensors

as function of the precipitation, evapotranspiration, surface runoff, sub-surface runoff and deep percolation. The propagation process in the river channel is represent using a distributed Muskingum-Cunge model discretized each 1,000 m. More details about these models can be found in Mazzoleni et al. [62].

The calibration of the hydrological model parameters was performed by AAWA using an adaptation of the “SCE-UA” algorithm [73], considering the time series of precipitation from 2000 to 2010, in order to minimise the root mean square error between observed and simulated values of water level at PA (Vicenza) gauged station. For the Muskingum-Cunge model, the only parameter that is calibrated in this chapter is the Manning coefficient n, used to estimate the water level along the river. The semi-distributed hydrological-hydraulic model in the Bacchiglione catchment is then validated considering the flood events that occurred in May 2013, November 2014 and February 2016.

In order to apply data assimilation, both the hydrological and Muskingum models are represented using the stochastic state-space form reported in the previous section. In particular, for the Muskingum model, the approach proposed by Georgakakos et al. [74] is used.

4 Model Updating Techniques

Operational forecast can be seen as combination of water models (e.g. hydrological and hydrodynamic) and an updating module. In fact, in the last decades model updating techniques have been intensively used within water system models [75, 76], in order to reduce predictive uncertainty. The hydrological and hydrodynamic models utilise input variables, which are either measured or estimated (e.g. areal precipitation, air temperature, potential evapotranspiration), into a set of equations that contain state variables and parameters. Typically, the parameters remain constant, while the state variables vary in time, even if there are different examples of parameter updating approaches such as Moradkhani et al. [77, 78], Salomon and Feyen [79] and Lü et al. [80]. The feedback process of assimilating the new available information into the forecasting procedure is referred to as updating [75] or DA [76].

The assimilation methods can be divided according to the variables modified during the updating process. In the frequently cited WMO report [76], updating is understood in a wide sense, and input, parameters, states and output updating techniques are distinguished. Recently, Liu et al. [34] provided a detailed review of the status, progresses, challenges and opportunities in advancing DA in operational hydrological forecasting. There are many data assimilation techniques that can be used to integrate hydrological observations within water-related models. In this chapter we will focus mainly on Kalman filter and ensemble Kalman filter.

4.1 Kalman Filter

Kalman filter (KF, [81]) is an approach which allows to optimally estimate the state of a dynamic uncertain model as response of real-time (noisy) observations [3, 14, 77, 82–85]. KF update model states considering only the last available observation allowing for a faster computation. However, KF is optimal only in the case of linear dynamic systems. Kalman filter procedure can be divided in two steps: time update equations, namely, forecast (background) equations, Eqs. (3) and (4),

$$x_t^- = \mathbf{\Phi} x_{t-1}^+ + \mathbf{\Gamma} I_t + w_t \tag{3}$$

$$P_t^- = \mathbf{\Phi} P_{t-1}^+ \mathbf{\Phi}^T + S_t \tag{4}$$

and update (or analysis) Eqs. (5), (6) and (7):

$$K_t = \frac{P_t^- H^T}{H P_t^- H^T + R_t} \tag{5}$$

$$x_t^+ = x_t^- + K_t \cdot \left(z_t^o - H x_t^-\right) \tag{6}$$

$$P_t^+ = (I - K_t H) P_t^- \tag{7}$$

where x is the $n_{state} \times 1$ state matrix at time t and t−1, K_t is the $n_{states} \times n_{obs}$ Kalman gain matrix, P is the $n_{states} \times n_{states}$ error covariance matrix and z^0 is the new observation. The superscripts + and – indicate, respectively, the updated and background state values, and $\mathbf{\Phi}$ and $\mathbf{\Gamma}$ represent the state-transition and input-transition matrices, which change according to the model type and structure. The system and measurement error w_t is assumed to be normally distributed with zero mean and covariance R. In the application considered in this chapter, the matrix R is time dependent as the error in the measurement is assumed variable because of the varying behaviour in time and space of the crowdsourcing observations.

A key issue in the implementation of the Kalman filter is the determination of model errors. In fact, an overestimation of model errors can reduce the confidence in the model bringing the KF closer to the observations and vice versa [86]. In this study, the modified version of KF, which accounts for the intermittency of crowdsourced observations in between two model time steps, proposed in Mazzoleni et al. [62] is used.

4.2 Ensemble Kalman Filter

Ensemble Kalman filter [87–90] is a widely used data assimilation method for non-linear dynamic model. The main idea of the EnKF is to represent the forecasted pdf estimate with a set of random samples and estimate the updated probability density function (pdf) of the model states as a combination between data likelihood and forecasted pdf of model states by means of a Bayesian update. In this way, the evaluation of the model error covariance matrix is performed as proposed by Evensen [87]:

$$P_t^- = \frac{1}{N_{ens} - 1} EE^T \tag{8}$$

where N_{ens} is the number of ensemble members and E is the ensemble anomaly [40] for each ensemble member:

$$E_t = \left(x_{t,1}^- - \bar{x}, x_{t,2}^- - \bar{x}, \cdots, x_{t,i}^- - \bar{x}, \cdots, x_{t,N_{ens}}^- - \bar{x}\right) \tag{9}$$

where $\bar{x}$ is the ensemble mean. The update states and Kalman gain are calculated using Eqs. (5) and (6). Because the EnKF performance is influenced by the spread of the ensemble [91–93], it is important to properly perturb the system in a way to obtain a reliable spread of the ensemble within a meaningful range [94]. For this reason, in this study we used the approach proposed by Anderson [91] to perturb the system and to evaluate the quality of the ensemble spread. More details are provided in Mazzoleni [63].

In order to implement EnKF, an ensemble of model realisations is generated perturbing the forcing data and the model parameters using a uniform distribution. The observation error is assessed using the approach described in the section below.

4.3 Synthetic Flow Observations

Synthetic flow observations are used because of the lack of distributed crowdsourced observations at the time of this study within the considered case study [62]. Such synthetic observations are generated by two different approaches for the two catchments. On the one hand, for the Brue catchment, the approach used to generate the synthetic values of river flow is very similar to the one used by Weerts and El Serafy [90], in which the model forcing is perturbed by means of a time series normally distributed with zero mean and given standard deviation.

On the other hand, for the Bacchiglione catchment, the observed time series of precipitation are used as input for the hydrological models of the sub-catchments and inter-catchments to generate synthetic discharges and then propagate them with the hydraulic model down to the outlet point of the catchment. In this way, the synthetic WL values at the outlet of the sub-catchments or inter-catchments and at each spatial discretization of the six reaches of the Bacchiglione River are estimated and assumed as observed variables in the assimilation process.

4.4 Estimation of the Observational Error

The correct estimation of the model and observational error is crucial for implementing data assimilation methods. Few studies in the past have addressed this issue (e.g. [95]), but further research is needed. For this reason, we adopted a simplified approach to quantify observational errors. Here, the covariance matrix R is assessed using the approach described in Weerts and El Serafy [90], Rakovec et al. [43] and Mazzoleni [63]:

$$R_t = \left(\alpha_t \cdot Q_t^{synth}\right)^2 \tag{10}$$

where α is a variable related to the accuracy level (i.e. degree to which the measurement is correct overall) of the flow measurement and Q^{synth} is the synthetic flow observation. In the case of CS observations, accuracy levels vary temporally and spatially.

Table 1 summarises the distribution of the coefficient α of the observational error of Eq. (10). The distribution of the coefficient α does not pretend to be exhaustive in

Table 1 Assumed observational errors for the different types of sensors

Sensor type	Assumed accuracy level	Coefficient α	Temporal and spatial variability
Static physical (StPh)	High	$\alpha = 0.1$	Fixed location Constant in time
Static social (StSc)	Medium	$\alpha = U(0.1, 0.3)$	Fixed location Intermittent arrival

accounting for different inaccuracies of observations coming from physical and social sensors and is subject for further research.

For static social sensors, α values are higher than for static physical sensors and are considered to be a random stochastic variable uniformly distributed in time and space. More details can be found in Mazzoleni et al. [61].

5 Assimilation of Flow Observations from Static Heterogeneous Sensors

This section aims to explore the benefits of assimilating flow observations from a network of static heterogeneous sensors in the case of synchronous (Sect. 5.1) or asynchronous (Sect. 5.2) social observations, depending on the predictability of the arrival time of the observations. In particular, we assume that social sensors provide intermittent observations that can lie either in a specific model time step (synchronous) or in between two model time steps (asynchronous). In addition, social sensors may be distributed within the catchment or be located in a specific point.

5.1 Assimilation of Synchronous Observations

Here, we show the model performance after the assimilation of intermittent synchronous observations, i.e. their arrival time matches the model time step, within the semi-distributed hydrological models of the Brue catchment. We can divide this section in two parts: first, the flow observations are assimilated from different social sensors located within the catchment; second, social sensors are integrated with a network of physical sensors to evaluate the added value of crowdsourced sensors in the assimilation process. A straightforward and pragmatic method (based on EnKF) is used to assimilate the intermittent observations into the hydrological model updating the model states matrix only when observations are available, while when there are no observations, it is assumed that the state covariance error does not change at that time step [60, 96].

5.1.1 Assimilation of Flow Observations Only from Social Sensors

In the first part of this section, we considered MS1 and three different spatial configurations (SC) of static social sensors within the catchments (called scenarios SC1, SC2 and SC3). In particular, SC1 refers to social sensors located along the main river channel, SC2 to sensors located on the upstream part of the main river channel, while SC3 to sensors located close to the catchment outlet. Model results obtaining assimilating flow observations from physical sensors are considered as

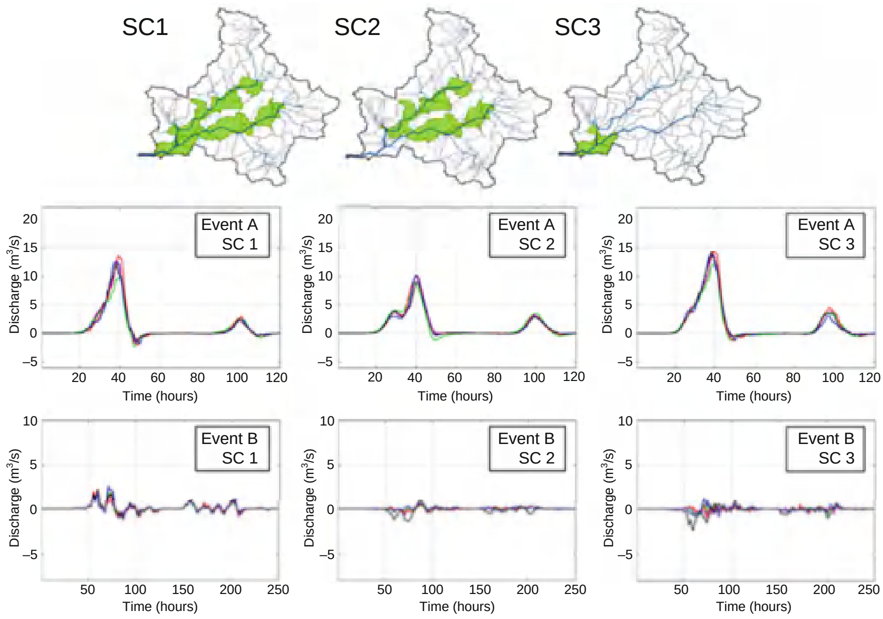


Fig. 4 Differences between assimilation of synthetic physical and social flow data in terms of outflow hydrographs under different intermittency configurations of the social sensors (different colour lines) for MS1 (source [60])

benchmark in order to compare the assimilation performances using social sensors. A main assumption of this study is that flow observations from social sensors are accepted to be less accurate, with random observation error both in time and space, than the ones from physical sensors.

The difference between the outflow hydrograph estimated assimilating physical and social data (in the same location) is represented in Fig. 4. The different colours of the hydrographs represent the different intermittency configurations of the social sensors, i.e. the unpredictable arrival time of the social observation. The smaller the value of difference, the smaller the sensitivity of the model to assimilation of observations from social sensors. Two different flood events are analysed.

As expected, the assimilation performances change with the different locations of social sensor within the catchment. Considering the flood event A, it can be seen that model outputs are affected by changing from physical to social flow data mainly for SC3. Physically, this can be due to the particular structure of the hydrological model. In particular, the discharge differences in flood event B are smaller than in flood event A due to the different performances of the model without assimilation. In fact, for flood event B, additional real-time observations of discharge slightly improved the model results since the model tends to better estimate the observed value of discharge even without assimilation. It is worth noting that results do not seem to be very sensitive to the intermittency scenarios (different colours of Fig. 4).

Table 2 NSE index values obtained assimilating streamflow observations from different spatial configuration of physical and social sensors for MS1

Spatial configuration	NoDA	1	2	3
Physical sensor	0.46	0.77	0.69	0.75
Social sensors	0.46	0.58	0.51	0.47

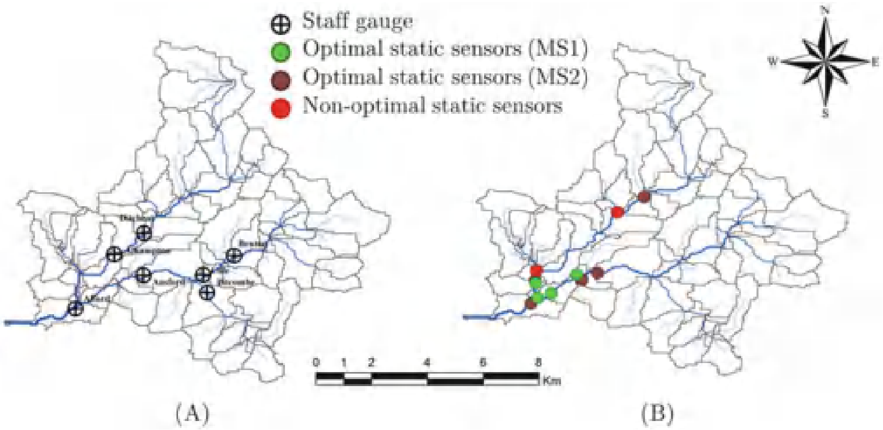


Fig. 5 Representation of distribution of static physical and social sensors along the Brue Basin for MS1 and MS2, respectively [63]

The results reported in Table 2 show a large difference in the NSE between assimilations from physical and social sensors. Table 2 underlines that the best model performances are not obtained when the assimilation of flow data is performed using sensors located at the outlet section of the catchment but when sensors are located along the main river channel, i.e. SC1 [60].

5.1.2 Assimilation of Flow Observations from Both Physical and Social Sensors

As a matter of fact, the location of the social sensors should typically follow some rules and be subjected to specific constraints. For example, existence of multiple sensors in remote areas of the catchment is quite unlikely due to economical and management reasons. For this reason, in the second part of this section, we assume a realistic configuration of the social sensors closer to the main urbanised area within the catchment (see Fig. 5). The network of social static sensors is integrated with the optimal network of static physical sensors (α equal to 0.1) for MS1 and MS2, respectively.

Different scenarios are introduced based on assumption on the intermittency and availability of CS data and on the possible integration between uncertain CS data and optimal/nonoptimal network of static physical sensors (see Table 3).

Table 3 Description of the different settings

Setting	Social sensors			Physical sensors	
	Intermittent	Daily timing	Daily and peak timing	Optimal	Nonoptimal
1	–	X	–	–	–
2	X	X	–	–	–
3	–	–	X	–	–
4	X	–	X	–	–
5	–	–	–	X	–
6	–	X	–	X	–
7	X	X	–	X	–
8	–	–	X	X	–
9	X	–	X	X	–
10	–	–	–	–	X
11	–	X	–	–	X
12	X	X	–	–	X
13	–	–	X	–	X
14	X	–	X	–	X

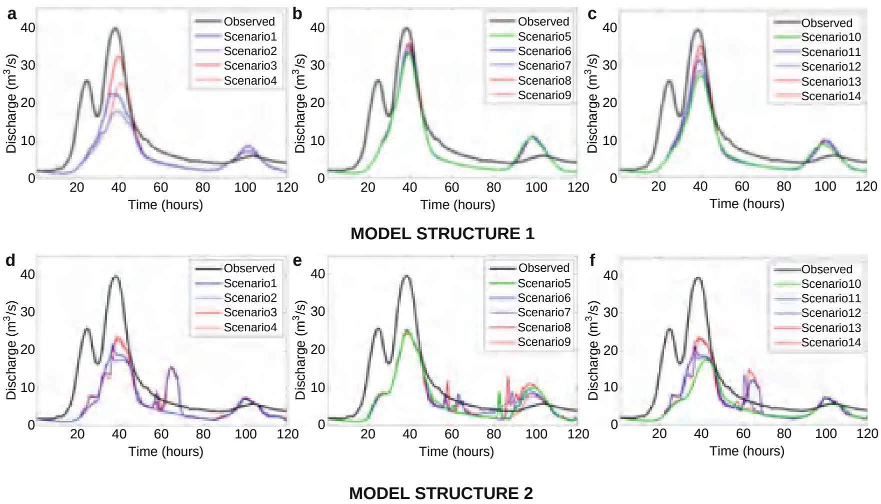


Fig. 6 Outflow hydrographs resulting from the assimilation of physical, social and intermittent observations in the case of realistic scenarios (from a to f) of spatial and temporal distribution of static sensors [63] for MS1 (first row) and MS2 (second row)

We demonstrate that assimilation of uncertain discharge observations measured at seven staff gauges by social sensors could improve the model results, however, still with the underestimation of the peak flow for scenarios 1 and 2 (see Fig. 6). Assimilation of observations coming from trained volunteers in the time of the peak flow (scenarios 3 and 4) showed a satisfactory improvement of the discharge hydrograph (higher for the model structure 1 than for the structure 2). Intermittent observations do not improve the model results in the same way that social observations coming continuously in time do.

Figure 6b confirms that similar improvements for the scenario 3 are achieved assimilating observations coming from the optimally located static sensors running continuously in time (scenario 5). In addition, a combined assimilation of intermittent observations (during daylight time) and static observations from optimal and nonoptimal network of static sensors tends to slightly improve the model output.

Figure 6 demonstrates that considering this type of hydrological model in this particular basin, in the case of an inappropriate distribution of static physical sensors within the basin (scenario 10), the model performances can be improved. However, there is an evident limitation of the model in providing biased hydrographs (especially for MS2), underestimated when compared to the observed one. Biased models can affect the DA results [34].

5.2 Assimilation of Asynchronous Observations

In the previous analysis, social data are provided at the same time of the model time step. However, in case of CS observations, the arrival moment might have lower frequency than the model time step (asynchronous observations), as reported in Mazzoleni et al. [62]. Various experimental scenarios representing different configurations of arrival frequency, number and accuracy of the flow observations are reported in Fig. 7. In order to remove the random behaviour related to the irregular arrival frequency and observation accuracy, different model runs (100 in this case) are carried out, assuming different random values of arrival and accuracy (coefficient α in Eq.10) during each model run, for a given number of observations and lead time. The NSE value is estimated for each model run, so μ(NSE) represents the mean of the different values of NSE.

5.2.1 Assimilation of Flow Observations Only from Social Sensors

A lumped hydrological model based on the KMN model is applied to the Brue catchment in order to assimilate synthetic asynchronous observations using the modified version of KF reported in Mazzoleni et al. [62]. Two flood events and experimental scenarios from 1 to 9 (see Fig. 7) are considered in this section.

As it can be seen from Fig. 8, increasing the number of social observations within the observation window results in the improvement of the NSE, but it becomes negligible for more than ten observations. This means that the additional social observations do not add information useful for improving the model performance.

From Fig. 8 it can be seen that, overall, assimilation of crowdsourced observations improves model performances in all the considered flood events. In the case of scenarios 2 and 3 (represented using warm, red and orange, colours in Fig. 8, for lead

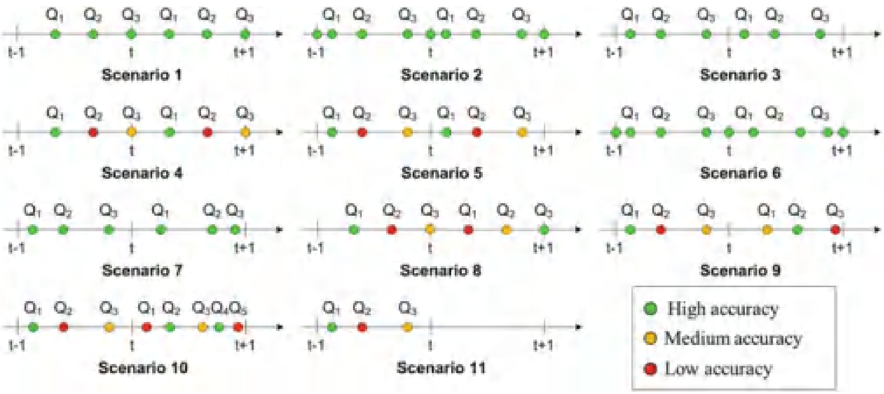


Fig. 7 The experimental scenarios representing different configurations of arrival frequency, number and accuracy of the streamflow observations [62]

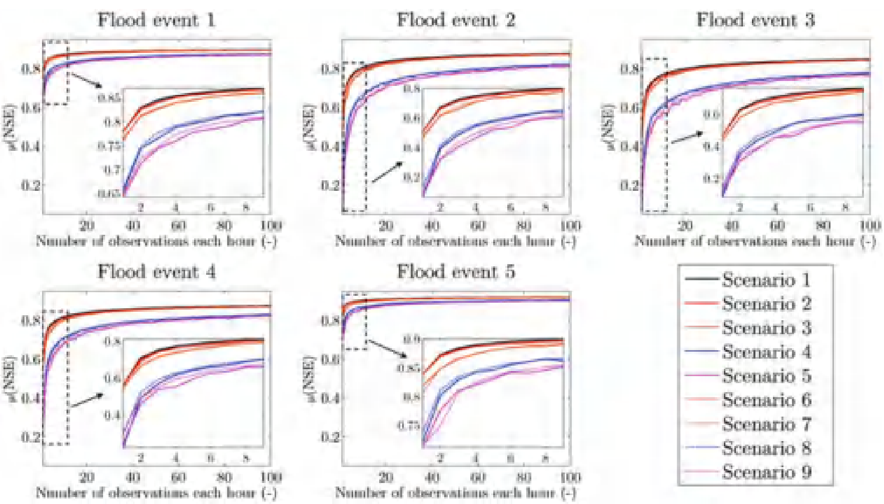


Fig. 8 μ(NSE) values estimated for varying number of assimilated flow observations, for the intermittency scenarios for the different flood events [63]

time equal to 24 h), i.e. random arrival frequency with fixed/controlled accuracy, the average values of NSE, μ(NSE), are smaller but comparable with the ones obtained in case of scenario 1 for all the considered flood events. In particular, scenario 3 has lower μ(NSE) than scenario 2. This can be related to the fact that both scenarios have random arrival frequency; however, in scenario 3 observations are not provided at the model time step, as opposed to scenario 2. In scenario 4, represented using cold blue colour, observations are considered coming at regular time steps but having random accuracy. Figure 8 shows that μ(NSE) values are lower in case of scenario 4 rather than scenarios 2 and 3. This can be related to the higher influence of observation accuracy if compared to arrival frequency. The combined effects of random arrival frequency and observation accuracy are represented in scenario 5 using a magenta colour (i.e. the combination of warm and cold colours) in Fig. 8. As expected, this scenario is the one with the lower values of μ(NSE) if compared to the previous ones. The remaining scenarios, from 6 to 9, are equivalent to the ones from 2 to 5 with the only difference that they are non-periodic in time. For this reason, in Fig. 8, scenarios from 6 to 9 have the same colour of scenarios 2–5 but indicated with dashed line in order to underline their non-periodic behaviour. Overall it can be observed that non-periodic scenarios have similar μ(NSE) values to their corresponding periodic scenario. However, their smoother μ(NSE) trends are due to lower variability of NSE values which means that model performances are less dependent to the non-periodic nature of the crowdsourced observations than their periodic behaviour. Overall, σ(NSE) tends to decrease for the high number of observations.

5.2.2 Assimilation of Flow Observations from Both Physical and Social Sensors

In the following, the contribution of assimilating synthetic flow data from a heterogeneous network of physical and social sensors on the semi-distributed model implemented in the Bacchiglione catchment is analysed. Streamflow observations from physical sensors are assumed to be synchronous with hourly frequency, while social observations are considered asynchronous with higher and irregular frequency. Five different experimental settings are introduced and represented in Fig. 9, corresponding to different types of sensors used.

The physical and social observations are assimilated in order to improve the poor model prediction at the catchment outlet (city of Vicenza) affected by an underestimation of the 3-day rainfall forecast used as normal input in flood forecasting practice in this area. Scenarios 10 and 11, described in Fig. 7, are used in this experiment in order to represent an irregular and random behaviour of the social observations.

Figure 10 shows the results obtained from the experiment settings represented in case of observations from distributed physical and social sensors. One of the main outcomes of these analyses is that the replacement of a physical sensor for a social sensor at only one location (settings B) does not improve the model performance in terms of NSE for different lead time values. Distributed locations of social sensors (setting C) can provide higher value of NSE than a single physical sensor, even for low number of observations in both regular and intermittent social observations. It is interesting to note that in case of integration between physical and social sensors (setting D), the NSE is higher than in case of setting C for low number of observations. However, with the higher number of observations, setting C is the one providing the best model improvement for low lead time values. Best model improvement is achieved in case of setting E. In case of intermittent observations (d, e and f), it can be noticed that the setting D provides higher improvement than

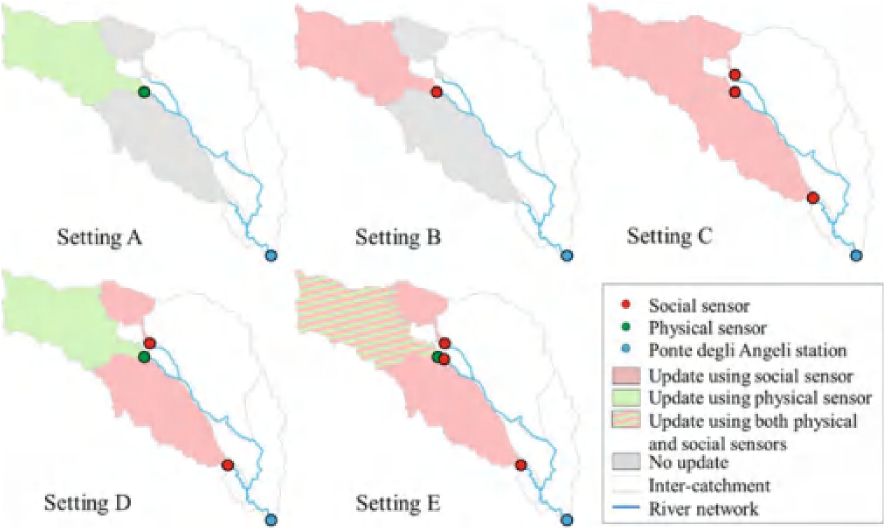


Fig. 9 Different experimental settings implemented within the Bacchiglione catchment (based on [62])

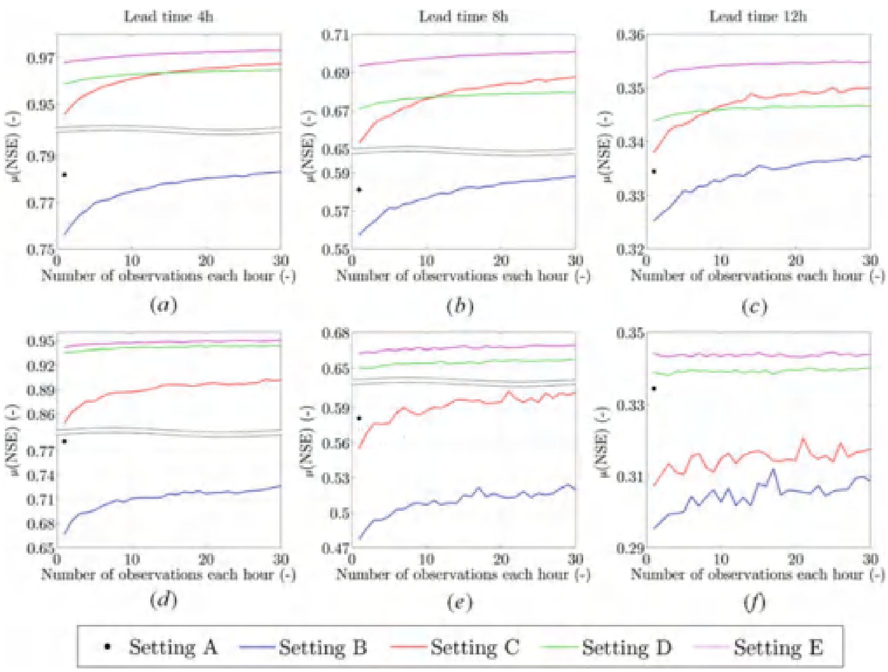


Fig. 10 Model performance expressed as μ(NSE) – assimilating different number of crowdsourced observations, for the three lead time values, having characteristic of scenario 10 (first row) and 11 (second row) (based on [63])

setting C. In case of high lead time value (12 h), results of setting C tend to be similar to the ones obtained with setting B. As in case of scenario 10, also in case of scenario 11, the best results are achieved in case of setting E.

6 Conclusions

This chapter describes the novel methods mainly developed within the EU-FP7 WeSenseIt project, aimed to optimally assimilate heterogeneous intermittent observations, coming from static social sensors, to improve hydrological and hydrodynamic models for flood prediction. The proposed methods used to assimilate crowdsourced observations are applied to the Brue and Bacchiglione catchments, in which different hydrological and hydraulic models are implemented. A Kalman filter and ensemble Kalman filter are used to assimilate flow observations in linear and non-linear models, respectively. Observational error is assumed uniformly distributed with multiplying factors of 0.1 and 0.3 as minimum and maximum values for the static social sensors, respectively. It is worth noting that because real crowdsourced observations from citizen were not available at the time of this study, model-based synthetic realistic flow observations are used instead.

This study demonstrated that crowdsourced citizen-based observations can significantly improve flood prediction if integrated into hydrological and hydraulic models. In addition, networks of low-cost static and dynamic social sensors can actually complement traditional networks of static physical sensors, for the purpose of improving flood forecasting accuracy. This can be one of the potential applications of increasing efforts to build citizen observatories of water. On the one hand, citizens can play an active role in information capturing, evaluation and communication, and on the other hand, they can also help in improving models and increasing flood resilience.

In particular, assimilation of streamflow observations from static social sensors provides improvements in model performance which depends on the location of such observations and the structure of the considered hydrological model. Flood forecasts are influenced by the total number of social sensors and their locations in the case of semi-distributed model with sub-catchments connected in parallel, while results achieved with sub-catchment connected in series are more sensitive to the locations of the static physical sensors but not to their number.

This research proved that assimilation of asynchronous observations results in a significant improvement of NSE for different lead time values. Increasing the number of assimilated crowdsourced asynchronous observations within two model time steps induces an improvement in the NSE. However, after a threshold number of crowdsourced observations, NSE asymptotically approaches a certain value meaning that no improvement is achieved with additional observations.

Besides these important results, this work has still certain limitations which should be mentioned. Additional analyses on different case studies and hydrological/hydraulic model have to be carried out to draw more general conclusions about assimilation of the crowdsourced observations and their additional value in different types of catchments. In addition, the adopted simple hydrologic and flow propagation models neglect some of the physical processes in complex floodplains (e.g. lamination/reservoir effects). The internal states of the hydrologic model where crowdsourced observations are supposed to be observed should be calibrated, since unbiased models are necessary to optimise data assimilation frameworks [34]. Moreover, real-life crowdsourced observations provided by citizens using static social and dynamic social sensors have to be used to further validate the results obtained in this research.

Overall, with this research we demonstrated that the choice of the proper mathematical model and updating technique to be used for flood forecasting may vary according to the data availability, location of the sensors, type of forecast, etc.

Acknowledgements This research was funded in the framework of the European FP7 Project WeSenseIt: Citizen Observatory of Water, grant agreement No. 308429.

References

1. Hinkel J, Lincke D, Vafeidis AT, Perrette M, Nicholls RJ, Tol RSJ, Marzeion B, Fettweis X, Ionescu C, Levermann A (2014) Coastal flood damage and adaptation costs under 21st century sea-level rise. Proc Natl Acad Sci 111(9):3292–3297. https://doi.org/10.1073/pnas.1222469111
2. Jongman B, Hochrainer-Stigler S, Feyen L, Aerts JCJH, Mechler R, Botzen WJW, Bouwer LM, Pflug G, Rojas R, Ward PJ (2014) Increasing stress on disaster-risk finance due to large floods. Nat Clim Chang 4(4):264–268. https://doi.org/10.1038/nclimate2124
3. McLaughlin D (2002) An integrated approach to hydrologic data assimilation: interpolation, smoothing, and filtering. Adv Water Resour 25(8–12):1275–1286. https://doi.org/10.1016/S0309-1708(02)00055-6
4. Solomatine DP, Wagener T (2011) Hydrological modelling. In: Wilderer P (ed) Treatise on water science. Elsevier, Amsterdam, pp 435–457
5. Todini E, Alberoni P, Butts M, Collier C, Khatibi R, Samuels P, Weerts A (2005) ACTIF best practice paper–understanding and reducing uncertainty in flood forecasting. In: Balabanis P, Lumbroso D, Samuels P (eds) International conference on innovation, advances and implementation of flood forecasting technology, Troms, Norway
6. Pappenberger F, Matgen P, Beven KJ, Henry J-B, Pfister L, de Fraipont P (2006) Influence of uncertain boundary conditions and model structure on flood inundation predictions. Adv Water Resour 29(10):1430–1449. https://doi.org/10.1016/j.advwatres.2005.11.012
7. Koutsoyiannis D (2010) HESS opinions “a random walk on water”. Hydrol Earth Syst Sci 14(3):585–601. https://doi.org/10.5194/hess-14-585-2010
8. Montanari A, Koutsoyiannis D (2012) A blueprint for process-based modeling of uncertain hydrological systems. Water Resour Res 48(9):W09555. https://doi.org/10.1029/2011WR011412
9. Alfonso L, Tefferi M (2015) Effects of uncertain control in transport of water in a river-wetland system of the Low Magdalena River, Colombia. Transport of water versus transport over water. Springer, Cham, pp 131–144
10. Domeneghetti A, Vorogushyn S, Castellarin A, Merz B, Brath A (2013) Probabilistic flood hazard mapping: effects of uncertain boundary conditions. Hydrol Earth Syst Sci 17(8):3127–3140. https://doi.org/10.5194/hess-17-3127-2013

11. Hall J, Solomatine D (2008) A framework for uncertainty analysis in flood risk management decisions. Int J River Basin Manag 6(2):85–98. https://doi.org/10.1080/15715124.2008.9635339
12. Merz B, Thieken AH (2005) Separating natural and epistemic uncertainty in flood frequency analysis. J Hydrol 309(1–4):114–132. https://doi.org/10.1016/j.jhydrol.2004.11.015
13. Goetzinger J, Bardossy A (2008) Generic error model for calibration and uncertainty estimation of hydrological models. Water Resour Res 44:W00B07. https://doi.org/10.1029/2007WR006691
14. Liu Y, Gupta HV (2007) Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. Water Resour Res 43(7):1–18. https://doi.org/10.1029/2006WR005756
15. Quinonero-Candela J, Rasmussen CE, Sinz F, Bousquet O, Schölkopf B (2006) Evaluating predictive uncertainty challenge. Machine learning challenges. Evaluating predictive uncertainty, visual object classification, and recognising tectual entailment. Springer, New York. http://link.springer.com/10.1007%2F11736790_1. Accessed 2 Mar 2016, pp 1–27
16. Renard B, Kavetski D, Kuczera G, Thyer M, Franks SW (2010) Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors. Water Resour Res 46(5):W05521. https://doi.org/10.1029/2009WR008328
17. Wagener T, Gupta HV (2005) Model identification for hydrological forecasting under uncertainty. Stoch Environ Res Risk Assess 19(6):378–387. https://doi.org/10.1007/s00477-005-0006-5
18. Melchers RE (1999) Structural reliability analysis and prediction, 2nd edn. Wiley, New York
19. Abebe AJ, Solomatine DP, Venneker RGW (2000) Application of adaptive fuzzy rule-based models for reconstruction of missing precipitation events. Hydrol Sci J 45(3):425–436
20. Bárdossy A, Bronstert A, Merz B (1995) 1-, 2- and 3-dimensional modeling of water movement in the unsaturated soil matrix using a fuzzy approach. Adv Water Resour 18(4):237–251
21. Hundecha Y, Bardossy A, Theisen HW (2001) Development of a fuzzy logic-based rainfall-runoff model. Hydrol Sci J 46(3):363–376
22. Plate E, Shahzad K (2015) Uncertainty analysis of multi-model flood forecasts. Water 7(12):6788–6809. https://doi.org/10.3390/w7126654
23. Xuan Y, Cluckie ID, Wang Y (2009) Uncertainty analysis of hydrological ensemble forecasts in a distributed model utilising short-range rainfall prediction. Hydrol Earth Syst Sci 13(3):293–303
24. Dogulu N, López López P, Solomatine DP, Weerts AH, Shrestha DL (2015) Estimation of predictive hydrologic uncertainty using the quantile regression and UNEEC methods and their comparison on contrasting catchments. Hydrol Earth Syst Sci 19:3181–3201. https://doi.org/10.5194/hess-19-3181-2015
25. Shrestha DL, Solomatine DP (2006) Machine learning approaches for estimation of prediction interval for the model output. Neural Netw 19:225–235. https://doi.org/10.1016/j.neunet.2006.01.012
26. Shrestha DL, Rodriguez J, Price RK, Solomatine DP (2006) Assessing model prediction limits using fuzzy clustering and machine learning. Proceedings of the 7th international conference on hydroinformatics, 4–8 September, Nice, France
27. Beven K, Binley A (2014) GLUE: 20 years on. Hydrol Process 28:5897–5918. https://doi.org/10.1002/hyp.10082
28. Beven K, Binley A (1992) The future of distributed models: model calibration and uncertainty prediction. Hydrol Process 6:279–298. https://doi.org/10.1002/hyp.3360060305
29. Shrestha DL, Kayastha N, Solomatine D (2009) A novel approach to parameter uncertainty analysis of hydrological models using neural networks. Hydrol Earth Syst Sci 13:1235–1248
30. Solomatine D, Shrestha DL (2009) A novel method to estimate total model uncertainty using machine learning techniques. Water Resour Res 45:W00B11. https://doi.org/10.1029/2008WR006839

31. Beven K, Freer J (2001) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J Hydrol 249:11–29
32. Mantovan P, Todini E (2006) Hydrological forecasting uncertainty assessment: incoherence of the GLUE methodology. J Hydrol 330(1–2):368–381. https://doi.org/10.1016/j.jhydrol.2006.04.046
33. Pappenberger F, Beven KJ (2006) Ignorance is bliss: or seven reasons not to use uncertainty analysis. Water Resour Res 42(5):1–8. https://doi.org/10.1029/2005WR004820
34. Liu Y, Weerts AH, Clark M, Hendricks Franssen HJ, Kumar S, Moradkhani H, Seo DJ, Schwanenberg D, Smith P, Van Dijk AIJM, Van Velzen N, He M, Lee H, Noh SJ, Rakovec O, Restrepo P (2012) Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. Hydrol Earth Syst Sci 16 (10):3863–3887. https://doi.org/10.5194/hess-16-3863-2012
35. Seo DJ, Cajina L, Corby R, Howieson T (2009) Automatic state updating for operational streamflow forecasting via variational data assimilation. J Hydrol 367(3–4):255–275. https://doi.org/10.1016/j.jhydrol.2009.01.019
36. Welles E, Sorooshian S, Carter G, Olsen B (2007) Hydrologic verification: a call for action and collaboration. Bull Am Meteorol Soc 88:503–511
37. Yarvis M, Kushalnagar N, Singh H, Rangarajan A, Liu Y, Singh S (2005) Exploiting heterogeneity in sensor networks. Proceedings IEEE INFOCOM 2005. 24th annual joint conference of the IEEE computer and communications societies, vol 2, pp 878–890
38. Bonney R, Shirk JL, Phillips TB, Wiggins A, Ballard HL, Miller-Rushing AJ, Parrish JK (2014) Next steps for citizen science. Science 343(6178):1436–1437. https://doi.org/10.1126/science.1251554
39. Buytaert W, Zulkafli Z, Grainger S, Acosta L, Alemie TC, Bastiaensen J, De BiÃvre B, Bhusal J, Clark J, Dewulf A, Foggin M, Hannah DM, Hergarten C, Isaeva A, Karpouzoglou T, Pandeya B, Paudel D, Sharma K, Steenhuis T, Tilahun S, Van Hecken G, Zhumanova M (2014) Citizen science in hydrology and water resources: opportunities for knowledge generation, ecosystem service management, and sustainable development. Front Earth Sci 2:1–21. https://doi.org/10.3389/feart.2014.00026
40. Clark MP, Rupp DE, Woods RA, Zheng X, Ibbitt RP, Slater AG, Schmidt J, Uddstrom MJ (2008) Hydrological data assimilation with the ensemble Kalman filter: use of streamflow observations to update states in a distributed hydrological model. Adv Water Resour 31 (10):1309–1324. https://doi.org/10.1016/j.advwatres.2008.06.005
41. Mazzoleni M, Alfonso L, Solomatine D (2016) Influence of spatial distribution of sensors and observation accuracy on the assimilation of distributed streamflow data in hydrological modelling. Hydrol Sci J 62(3):389–407. https://doi.org/10.1080/02626667.2016.1247211
42. Mazzoleni M, Noh SJ, Lee H, Liu Y, Seo DJ, Amaranto A, Alfonso L, Solomatine DP (2018) Real-time assimilation of streamflow observations into a hydrological routing model: effects of model structures and updating methods. Hydrol Sci J 63:386–407
43. Rakovec O, Weerts AH, Hazenberg P, Torfs PJJF, Uijlenhoet R (2012) State updating of a distributed hydrological model with ensemble Kalman filtering: effects of updating frequency and observation network density on forecast accuracy. Hydrol Earth Syst Sci 16(9):3435–3449. https://doi.org/10.5194/hess-16-3435-2012
44. Alfonso L, Lobbrecht A, Price R (2010) Using mobile phones to validate models of extreme events. 9th international conference on hydroinformatics, Tianjin, China, pp 1447–1454
45. de Vos L, Leijnse H, Overeem A, Uijlenhoet R (2017) The potential of urban rainfall monitoring with crowdsourced automatic weather stations in Amsterdam. Hydrol Earth Syst Sci 21:765–777. https://doi.org/10.5194/hess-21-765-2017
46. Etter S, Strobl B, Seibert J, van Meerveld I (2018) Value of uncertain streamflow observations for hydrological modelling. Hydrol Earth Syst Sci Discuss 22(10):5243–5257. https://doi.org/10.5194/hess-2018-355

47. Fava C, Santana G, Bressiani DA, Rosa A, Horita FEA, Souza VCB, Mendiondo EM (2014) Integration of information technology systems for flood forecasting with hybrid data sources. International conference of flood management, Sao Paolo, Brazil
48. Fohringer J, Dransch D, Kreibich H, Schröter K (2015) Social media as an information source for rapid flood inundation mapping. Nat Hazards Earth Syst Sci 15:2725–2738. https://doi.org/10.5194/nhess-15-2725-2015
49. Gaitan S, van de Giesen NC, ten Veldhuis JAE (2016) Can urban pluvial flooding be predicted by open spatial data and weather data? Environ Model Softw 85:156–171. https://doi.org/10.1016/j.envsoft.2016.08.007
50. Giuliani M, Castelletti A, Fedorov R, Fraternali P (2016) Using crowdsourced web content for informing water systems operations in snow-dominated catchments. Hydrol Earth Syst Sci 20:5049–5062. https://doi.org/10.5194/hess-20-5049-2016
51. Rollason E, Bracken LJ, Hardy RJ, Large ARG (2018) The importance of volunteered geographic information for the validation of flood inundation models. J Hydrol 562:267–280
52. Rosser JF, Leibovici DG, Jackson MJ (2017) Rapid flood inundation mapping using social media, remote sensing and topographic data. Nat Hazards 87:103–120
53. Schneider P, Castell N, Vogt M, Dauge FR, Lahoz W, Bartonova A (2017) Mapping urban air quality in near real-time using observations from lowcost sensors and model information. Environ Int 106:234–247
54. Smith L, Liang Q, James P, Lin W (2015) Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. J Flood Risk Manag 10:370–380. https://doi.org/10.1111/jfr3.12154
55. Starkey E, Parkin G, Birkinshaw S, Large A, Quinn P, Gibson C (2017) Demonstrating the value of community-based (“citizen science”) observations for catchment modelling and characterisation. J Hydrol 548:801–817. https://doi.org/10.1016/j.jhydrol.2017.03.019
56. Yu D, Yin J, Liu M (2016) Validating city-scale surface water flood modelling using crowdsourced data. Environ Res Lett 11:124011. https://doi.org/10.1088/1748-9326/11/12/124011
57. Le Coz J, Patalano A, Collins D, Guillén NF, García CM, Smart GM, Bind J, Chiaverinica A, Le Boursicauda R, Dramaisa G, Braud I (2016) Crowdsourced data for flood hydrology: feedback from recent citizen science projects in Argentina, France and New Zealand. J Hydrol 541:766–777
58. Assumpção TH, Popescu I, Jonoski A, Solomatine DP (2018) Citizen observations contributing to flood modelling: opportunities and challenges. Hydrol Earth Syst Sci 22:1473–1489. https://doi.org/10.5194/hess-22-1473-2018
59. Shanley L, Burns R, Bastian Z, Robson E (2013) Tweeting up a storm: the promise and perils of crisis mapping, available SSRN 2464599. https://ssrn.com/abstract=2464599. Accessed 20 Mar 2016
60. Mazzoleni M, Alfonso L, Chacon-Hurtado J, Solomatine D (2015) Assimilating uncertain, dynamic and intermittent streamflow observations in hydrological models. Adv Water Resour 83:323–339
61. Mazzoleni M, Cortes Arevalo VJ, Wehn U, Alfonso L, Norbiato D, Monego M, Ferri M, Solomatine DP (2018) Exploring the influence of citizen involvement on the assimilation of crowdsourced observations: a modelling study based on the 2013 flood event in the Bacchiglione catchment (Italy). Hydrol Earth Syst Sci 22:391–416. https://doi.org/10.5194/hess-22-391-2018
62. Mazzoleni M, Verlaan M, Alfonso L, Monego M, Norbiato D, Ferri M, Solomatine DP (2017) Can assimilation of crowdsourced data in hydrological modelling improve flood prediction? Hydrol Earth Syst Sci 21:839–861. https://doi.org/10.5194/hess-21-839-2017
63. Mazzoleni M (2017) Improving flood prediction assimilating uncertain crowdsourced data into hydrologic and hydraulic models. UNESCO-IHE PhD thesis series, CRC Press/Balkema, Leiden

64. Mazzoleni M, Amaranto A, Solomatine DP (2019) Integrating qualitative flow observations in a lumped hydrologic routing model. Water Resour Res 55. https://doi.org/10.1029/2018WR023768
65. WeSenseIt (2016) WeSenseIt: citizen water observatories. http://wesenseit.eu/. Accessed 19 Feb 2016
66. Sugeno M, Yasukawa T (1993) A fuzzy-logic-based approach to qualitative modelling. IEEE Trans Fuzzy Syst 1:7–31
67. Moore RJ, Jones DA, Cox DR, Isham VS (2000) Design of the HYREX raingauge network. Hydrol Earth Syst Sci 4(4):521–530. https://doi.org/10.5194/hess-4-521-2000
68. Wood SJ, Jones DA, Moore RJ (2000) Accuracy of rainfall measurement for scales of hydrological interest. Hydrol Earth Syst Sci Discuss 4(4):531–543
69. Szilagyi J, Szollosi-Nagy A (2010) Recursive streamflow forecasting: a state space approach. CRC Press, Leiden
70. Cunge JA (1969) On the subject of a flood propagation computation method (Muskingum method). J Hydraul Res 7(2):205–230
71. Ferri M, Monego M, Norbiato D, Baruffi F, Toffolon C, Casarin R (2012) La piattaforma previsionale per i bacini idrografici del Nord Est Adriatico (I). Proceedings XXXIII conference of hydraulics and hydraulic engineering, Brescia, p 10
72. Huwald H, Barrenetxea G, de Jong S, Ferri M, Carvalho R, Lanfranchi V, McCarthy S, Glorioso G, Prior S, Solà E, Gil-Roldàn E, Alfonso L, Wehn de Montalvo U, Onencan A, Solomatine D, Lobbrecht A (2013) D1.11 sensor technology requirement analysis. Confidential deliverable, the WeSenseIt project (FP7/2007-2013 grant agreement no 308429)
73. Duan Q, Ajami NK, Gao X, Sorooshian S (2007) Multi-model ensemble hydrologic prediction using Bayesian model averaging. Adv Water Resour 30(5):1371–1386. https://doi.org/10.1016/j.advwatres.2006.11.014
74. Georgakakos AP, Georgakakos KP, Baltas EA (1990) A state-space model for hydrologic river routing. Water Resour Res 26:827–838
75. Refsgaard JC (1997) Validation and intercomparison of different updating procedures for real-time forecasting. Nord Hydrol 28(2):65–84. https://doi.org/10.2166/nh.1997.005
76. WMO (1992) Simulated real-time intercomparison of hydrological models. World Meteorological Organization, Geneva
77. Moradkhani H, Hsu KL, Gupta H, Sorooshian S (2005) Uncertainty assessment of hydrologic model states and parameters: sequential data assimilation using the particle filter. Water Resour Res 41(5):W05012. https://doi.org/10.1029/2004WR003604
78. Moradkhani H, Sorooshian S, Gupta HV, Houser PR (2005) Dual state–parameter estimation of hydrological models using ensemble Kalman filter. Adv Water Resour 28(2):135–147
79. Salamon P, Feyen L (2009) Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. J Hydrol 376(3-4):428–442
80. Lü H, Yu Z, Zhu Y, Drake S, Hao Z, Sudicky EA (2011) Dual state-parameter estimation of root zone soil moisture by optimal parameter estimation and extended Kalman filter data assimilation. Adv Water Resour 34(3):395–406
81. Kalman RE (1960) A new approach to linear filtering and prediction problems. J Basic Eng 82(1):35–45. https://doi.org/10.1115/1.3662552
82. Heemink AW, Segers AJ (2002) Modeling and prediction of environmental data in space and time using Kalman filtering. Stoch Environ Res Risk Assess 16(3):225–240. https://doi.org/10.1007/s00477-002-0097-1
83. Reichle RH, Crow WT, Keppenne CL (2008) An adaptive ensemble Kalman filter for soil moisture data assimilation. Water Resour Res 44(3):W03423. https://doi.org/10.1029/2007WR006357
84. Robinson AR, Lermusiaux PFJ, Sloan III NQ (1998) Data assimilation. Sea 10:541–594
85. Walker JP, Houser PR (2005) Hydrologic data assimilation. Adv Water Sci Methodol 41:233. https://doi.org/10.5772/1112

86. Sun L, Seidou O, Nistor I, Liu K (2015) Review of the Kalman type hydrological data assimilation. Hydrol Sci J 61(13):2348–2366. https://doi.org/10.1080/02626667.2015.1127376
87. Evensen G (2003) The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn 53:343–367
88. Heemink AW, Verlaan M, Segers AJ (2001) Variance reduced ensemble Kalman filtering. Mon Weather Rev 129(7):1718–1728
89. Reichle R, McLaughlin DB, Entekhabi D (2002) Hydrologic data assimilation with the ensemble Kalman filter. Am Meteorol Soc 130(1):103–114. https://doi.org/10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2
90. Weerts AH, El Serafy GYH (2006) Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. Water Resour Res 42(9):1–17. https://doi.org/10.1029/2005WR004093
91. Anderson JL (2001) An ensemble adjustment kalman filter for data assimilation. Mon Weather Rev 129:2884–2903
92. Murphy JM (1988) The impact of ensemble forecasts on predictability. Q J Roy Meteorol Soc 114(480):463–493. https://doi.org/10.1002/qj.49711448010
93. Pauwels VRN, De Lannoy GJM (2009) Ensemble-based assimilation of discharge into rainfall-runoff models: a comparison of approaches to mapping observational information to state space. Water Resour Res 45(8):W08428. https://doi.org/10.1029/2008WR007590
94. De Lannoy GJM, Reichle RH, Houser PR, Pauwels VRN, Verhoest NEC (2007) Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. Water Resour Res 43 (9):W09410. https://doi.org/10.1029/2006WR005449
95. Brouwer T, Eilander D, Van Loenen A, Booij MJ, Wijnberg KM, Verkade JS, Wagemaker J (2017) Probabilistic flood extent estimates from social media flood observations. Nat Hazards Earth Syst Sci 17(5):735
96. Cipra T, Romera R (1997) Kalman filter with outliers and missing observations. TEST 6 (2):379–395. https://doi.org/10.1007/BF02564705

Precipitation Measurement with Weather Radars

Nergui Nanding and Miguel Angel Rico-Ramirez

Contents

Abstract Weather radar is a remote sensing instrument that has been increasingly used to estimate precipitation for a variety of hydrological and meteorological applications, including real-time flood forecasting, severe weather monitoring and warning, and short-term precipitation forecasting. Weather radar provides unique observations of precipitating systems at fine spatial and temporal resolutions, which are difficult to obtain through conventional raingauge networks. The potential benefit of using radar rainfall in hydrology is huge, but practical hydrological applications of radar have been limited by the inherent uncertainties and errors in radar rainfall estimates. Uncertainties in radar rainfall estimates can lead to large errors in flood forecasting applications, so radar rainfall measurements must be

N. Nanding (*)
Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, School of Atmospheric Sciences, Sun Yat-Sen University, Guangzhou, China
e-mail: nanding@mail.sysu.edu.cn

M. A. Rico-Ramirez
Department of Civil Engineering, University of Bristol, Bristol, UK

Andrea Scozzari, Steve Mounce, Dawei Han, Francesco Soldovieri, and Dimitri Solomatine (eds.), ICT for Smart Water Systems: Measurements and Data Science, Hdb Env Chem (2021) 102: 235–258, https://doi.org/10.1007/698_2019_404,
© Springer Nature Switzerland AG 2019, Published online: 23 November 2019

corrected before the data are used quantitatively. This chapter discusses some of the latest advances in the measurement and forecasting of precipitation with weather radar and some of the techniques proposed in the literature to correct and adjust radar rainfall estimates.

Keywords Bias correction, Flood forecasting, Precipitation forecasting, Rainfall estimation, Urban hydrology, Weather radar

1 Introduction

Water and environmental management increasingly require rainfall products with good spatial and temporal resolutions over the region of interest for planning and risk assessment. Weather radars are instruments capable to provide rainfall measurements with suitable spatial and temporal resolutions. The radar (an acronym for RAdio Detection And Ranging) was developed during the Second World War to detect enemy aircraft at sufficiently long distances to react to the threat. However, military users of radar often found radar echoes cluttered with precipitation targets. They realized that radar systems were sensitive enough to be able to detect precipitation. One of the first observations of precipitation made by radar was in 1941 [1]. Originally, weather radars were used for tracking balloons to determine upper winds and for detection of precipitating cloud systems [2]. Since then, huge progress has been made, both in terms of hardware and algorithm development. Nowadays, weather radars can be used to estimate precipitation over large regions for hydrological and meteorological purposes such hydrological modelling, short-term precipitation forecasting, real-time flood forecasting, improving the initial conditions of numerical weather models through data assimilation, cloud research, etc.

The operational principle of weather radar is that radar transmits short pulses (with a wavelength between 2 and 10 cm) of electromagnetic radiation to precipitation particles. The returned signal from precipitation particles has information related to their physical characteristics (e.g. target range, echo strength and velocity) within the illuminated volume by the radar beam. The returned signal (or power reflected) (P_r) from precipitation particles is given by [3]

$$P_r = \frac{C|K|^2 Z}{r^2} \tag{1}$$

where C is a constant depending on the radar characteristics (e.g. transmitted power, antenna gain, beamwidth, pulse length, wavelength), $|K|^2$ is the dielectric constant of the precipitation particles (e.g. 0.93 for liquid water and 0.176 for ice), r is the range between the radar and target and Z is the radar reflectivity factor. The power reflected from precipitation particles must be converted into meteorologically meaningful products (e.g. rainfall rate). Because P_r is measured by the radar, therefore Eq. (1)

can be used to compute Z. Radar reflectivity factor Z is a measure of the distribution of particles present within the radar sampling volume. If the precipitation particles are much smaller than the radar wavelength (Rayleigh scatterers), Z can be represented as the sixth moment of the drop size distribution, that is

$$Z = \int_0^\infty D^6 N(D) dD \tag{2}$$

where $N(D)$ is the raindrop size distribution (DSD) and represents the number of raindrops of diameter D per unit volume. Z in Eq. (1) is given in linear units ($mm^6\ m^{-3}$) and it can range from very small values (e.g. 0.1 $mm^6\ m^{-3}$) in drizzle to very large values (e.g. $10^6\ mm^6\ m^{-3}$) in very heavy precipitation or hail thunderstorms. Therefore, it is convenient to express the reflectivity factor in logarithmic units (dBZ):

$$dBZ = 10 \log_{10}(Z) \tag{3}$$

It is worth to mention that if the precipitation particles do not behave as Rayleigh scatterers (e.g. large snowflakes or large ice particles), then the radar reflectivity factor is known as the equivalent reflectivity factor (Z_e) or just as reflectivity. Z is equivalent to Z_e if the precipitation particles are Rayleigh scatterers and are made of liquid water. The rainfall rate (in mm/h) can be expressed as

$$R = 0.0006\pi \int_0^\infty \nu(D) D^3 N(D) dD \tag{4}$$

where $\nu(D)$ is the terminal velocity ($m\ s^{-1}$) of raindrops with a diameter D in mm. The terminal velocity can be approximated as a function of particle diameter, which is given by $\nu(D) = 3.78D^{0.67}$ [4] in the absence of vertical air motions. If we use this terminal velocity, it can be seen that the rainfall rate R represents the 3.67th moment of the DSD, while radar reflectivity factor Z represents the sixth moment of DSD. This indicates that Z is largely affected by the larger drops, even if there is a large fraction of smaller raindrops. This produces a source of uncertainty because both, Z and R depend to different extend on the DSD, which can continuously change during a rainfall event and the DSD is known to vary with rainfall intensity and type of precipitation. Thus, a good knowledge of the DSD is crucial to provide radar rainfall estimates with good accuracy. The measured reflectivity Z can be transformed to an estimate of precipitation R by using a $Z-R$ relationship. There are many $Z-R$ relationships in the literature and the most commonly used power-law relationship has the form $Z = aR^b$, where a and b are parameters that depend on the DSD [5]. For instance, the Marshall–Palmer $Z-R$ relationship $Z = 200R^{1.6}$ [6] is the most widely used equation in stratiform precipitation, but there are often many different $Z-R$ relationships quoted in the literature as summarized by Battan [7]. The choice of the $Z-R$ equation depends on the type of precipitation expected in that region and the parameters of this equation can be calibrated either using radar–raingauge measurements or disdrometer (instruments that measure DSDs) observations.

Operational weather radars can be classified into Single-Polarization (SP) and Dual-Polarization (DP) weather radars. DP radars are sensitive to size, shape, orientation and thermodynamic phase of the precipitation particles [8]. Operational DP radars alternately or simultaneously transmit vertically and horizontally polarized electromagnetic waves and receive polarized backscattered signals, whereas SP radars transmit and receive electromagnetic waves using single polarization only (either horizontal or vertical). SP radars can measure the reflectivity (Z) only and if the radar has Doppler capability, they also measure the radial velocities of precipitation particles. DP radars can measure additional variables such as the horizontal and vertical reflectivities (Z_h and Z_v), the differential reflectivity (Z_{dr}), the linear depolarization ratio (LDR), the correlation coefficient (ρ_{hv}) and the differential phase (Φ_{dp}). These additional measurements from DP radars have shown to provide significant improvements in terms of data quality and rainfall estimation compared with SP radars.

2 Sources of Uncertainty in the Estimation of Precipitation with Radar

Although recent advances in weather radar technology has helped to improve our understanding of the microphysics of precipitation as well as better rainfall estimates, there are still many challenges to improve the estimation of precipitation at ground level [9–19]. Rainfall estimation using weather radars can be subject to different sources of errors such as radar calibration, variations of the DSD, radar signal attenuation, echoes due to non-meteorological origin, variation of the vertical profile of reflectivity, radar beam blocking, etc. The following sections describe some of the work carried out to mitigate some of these errors.

2.1 Radar Calibration

Accurate precipitation estimates using weather radar rely on stable hardware components (e.g. transmitter and receiver) with an accurate calibration. Inaccurate determination of the radar constant C (hereafter referred as radar calibration bias) can cause a significant error source to the radar precipitation estimations [15]. This error can cause significant differences in radar rainfall and therefore C must be carefully monitored. By using up-to-date hardware, radar calibration bias can be limited to within 2 dB or 36% error in precipitation rate [13]. Many techniques have been developed to monitor and adjust the radar calibration bias. For instance, Whiton et al. [20] proposed a calibration technique using solar interference, and it has been widely applied on operational weather radars for monitoring the sensitivity of the radar receiver and antenna pointing accuracy [21–23]. Wolff et al. [24] also demonstrated the calibration technique using statistical analysis of the echo power returns from fixed targets (e.g. high ground). For a radar network, the radar calibration bias can be monitored by joint observations from two or more radars [25]. This technique ensures the stability of radar calibration by comparing the radar reflectivity values of two or more radars in the same area. By using dual-polarization radar, Gorgucci et al. [26] developed a procedure for radar calibration, based on the self-consistency between the radar reflectivity at horizontal polarization (Z_H), differential reflectivity (Z_{dr}) and specific differential phase shift (K_{dp}). However, the self-consistency technique is also sensitive to the variations of the drop size distribution and raindrop shape [27]. Furthermore, the stability of radar calibration can also be monitored by comparing with raingauge accumulations [28]. However, these comparisons are more suitable for long-term adjustment, due to the differences in spatial-temporal samplings of the two sensors and the high variability of the $Z-R$ relationship [29]. Recently, a new methodology [30] has been developed to match the precipitation observations from ground-based and space-borne radars for the determination of calibration biases in ground-based radar systems. It has been shown that the radar calibration bias can be less than 1.5 dB in well-calibrated ground-based radars. This technique can be a useful tool for the systematic monitoring of the radar calibration bias.

2.2 Echoes Due to Non-meteorological Origin

The radar usually scans at low elevation angles to obtain measurements close to the ground surface. Echoes from mountains or buildings can be misinterpreted as heavy precipitation, which are known as ground clutter. Such echoes are often permanent under standard beam propagation conditions, and thus techniques using a map of ground clutter locations are often successful in removing them [31]. However, echoes from targets under atmospheric super-refraction conditions are unpredictable in terms of location. This is known as anomalous propagation (AP), where the radar beam is bent toward the Earth′s surface due to changes in the atmospheric temperature and humidity distributions [7]. AP is an important source of error in radar rainfall measurements. For instance, the presence of AP echoes may produce reflectivities reaching 60 dBZ, which is comparable with echoes observed during severe thunderstorms [32].

Several methods to identify and suppress clutter echoes have been developed. For Doppler radar systems, filtering of the radial velocity signal can discriminate clutter and AP echoes from meteorological echoes. The assumption is that ground clutter echoes can be characterized as having zero-velocity and narrow spectral widths compared to weather echoes [33]. However, precipitation echoes may also have near-zero radial velocity and low spectral widths, which is commonly observed in

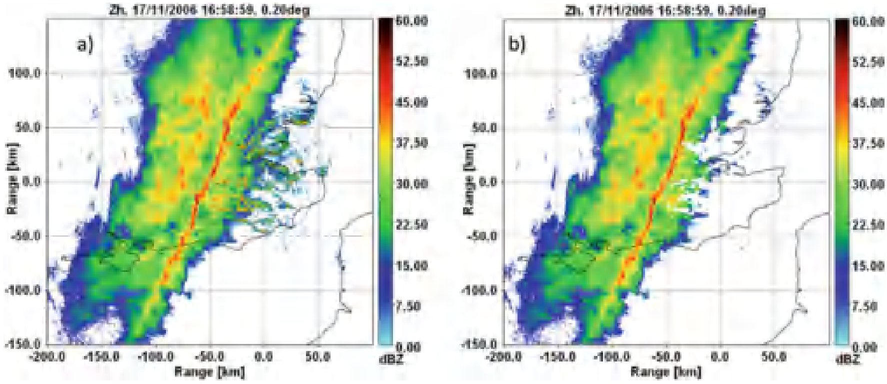


Fig. 1 Raw reflectivity scan (left) and reflectivity scan with clutter being removed (right)

widespread stratiform rain, or when the precipitation system is moving perpendicular to the direction of the radar beam (i.e. the radial velocity component is zero). Moreover, the notch filtering of near-zero velocity echoes is ineffective for AP over the sea as waves have true measurable velocities.

Other techniques developed more specifically to tackle AP echoes are mainly on analyzing quantities derived from the spatial and temporal information of the reflectivity field. Spatial information is usually presented in the form of gradients in the reflectivity field between adjacent range gates in either the horizontal or vertical dimensions [34]. The common descriptions of the gradient of the reflectivity field are texture, the reflectivity fluctuations and the statistical features (e.g. mean, median, mode and standard deviation). These reflectivity fields usually have different probability distribution functions (PDFs) for echoes from clutter, AP or precipitation. Parameters derived from the reflectivity gradient fields have been used in probabilistic classification algorithms, such as Bayesian [35, 36], fuzzy logic [37–39] and neural networks [40, 41] classification algorithms. Recently, a number of classification methods based on dual-polarization radar measurements were also introduced [42–44]. The advantage of multiparameter weather radars is their ability to obtain measurements of hydrometeor characteristics such as the size, shape, spatial orientation, phase state and fall behaviour [8]. The use of DP radar measurements has enabled more accurate classifications of non-meteorological echoes [36, 42]. Figure 1 shows an example of squall line moving eastwards; the figure on the left shows the raw reflectivity data, whereas the figure on the right shows the same scan with the clutter echoes being removed using the textures of the DP radar measurements. The use of the textures of the DP radar measurements has enabled a more accurate classification of non-meteorological echoes. This has also been demonstrated for the classification of sea clutter [36], echoes due to wind farms [45] and biological targets (e.g. birds and insects) [46] using fuzzy logic-based classifiers.

2.3 Attenuation

Attenuation is the gradual loss of power resulting from absorption and scattering as the radar signal travels through precipitation. The amount of attenuation depends on the precipitation particles present along the path of the radar beam and the radar frequency. The main absorbing substances that cause attenuation of microwaves in the atmosphere are water vapour and precipitation. Attenuation caused by precipitation increases steadily with radar frequency. At frequencies below 3 GHz (wavelengths greater than 10 cm), attenuation is relatively small. However, attenuation of radar signals by precipitation is a significant problem and one that becomes increasingly severe at wavelengths shorter than 10 cm. For instance, for a uniform rain rate of 20 mm/h on a 10-km path, the path integrated attenuation (PIA) is around 50 times greater at X-band frequencies than at S-band frequencies. However, the amount of PIA depends upon the rainfall rate and the length of the path. The attenuation effects remain relatively moderate at the C-band radar (5.4-cm wavelength) with a factor of less than four compared to S-band [33]. For a given wavelength, the amount of attenuation grows proportionally with rainfall intensity, but its effects are cumulative with range. In practice, heavy rainfall may lead to a complete loss of radar signal at X-band frequencies, severely limiting the maximum detectable range, whereas at C-band frequencies, the radar signals can still penetrate through even the most intense precipitation. Furthermore, a thin film of water forms on the radome surface in rain causing additional attenuation, particularly at shorter wavelengths. As a result, rain attenuation and radome attenuation are important error sources that affect the radar rainfall estimates. However, modern radomes have water-repellant coatings (e.g. hydrophobic coating) that might help to reduce radome attenuation. It is therefore important that steps are taken to mitigate attenuation effects if reliable radar rainfall estimates are required. Technically, the choice of the radar system with a longer wavelength (e.g. S-band) is a practical solution to mitigate this specific issue. However, this comes at a high cost due to the larger antenna of S-band radar and the higher transmitted power to retain a reasonable resolution and sensitivity. Shorter wavelength radars (e.g. X-band) have their own advantages, including smaller-sized antenna and higher sensitivity of the differential phase shift, which is immune to attenuation and can be used to estimate rain rates in heavy precipitation. However, several X-band radars are often required to measure precipitation over a particular region in order to mitigate potential problems of radar signal loss due to rain attenuation at these frequencies.

Different techniques have been developed to mitigate attenuation effects on radar systems at shorter wavelengths (e.g. X-band or C-band) [47–50]. Attenuation correction algorithms that use reflectivity measurements only are known to be unstable [48]. Early attenuation correction approaches were iterative, correcting the range gate from the first resolution volume (where attenuation is considered negligible) and moving to continuous range gates along the beam as it penetrates the precipitation cells. However, such gate-to-gate algorithms are inherently unstable and certain constraints must be imposed on the maximum amount of attenuation

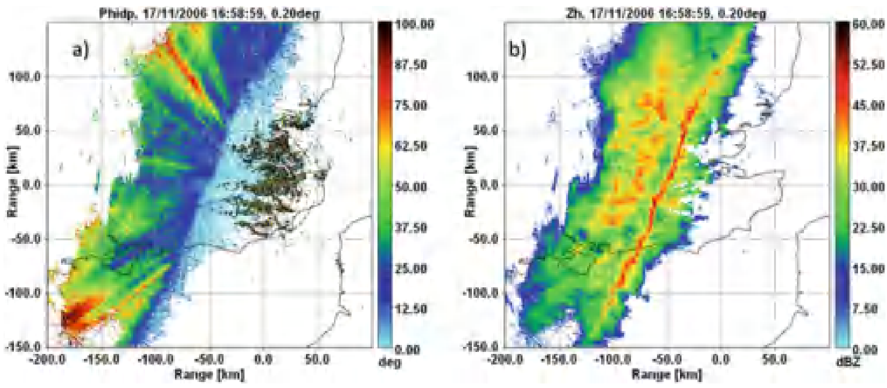


Fig. 2 Differential phase measurements (left) and attenuation-corrected reflectivity (right)

correction [48]. An overestimation on the closer range gates can worsen the attenuation correction for range gates further away. Attenuation correction procedures can be greatly improved if the total path-integrated attenuation is available as a constraint, for example, using dual-frequency radars or dual-polarization radars.

Advances in polarimetric weather radar technology can provide additional phase measurements that can be used to correct reflectivity measurements for rain attenuation at C-band and X-band frequencies [51]. The attenuation can be estimated by calculating the total differential propagation phase shift between the vertical and horizontal orthogonal signals (Φ_{dp}). The total differential propagation phase shift across a rain cell can be used as a constraint to estimate the PIA due to the fact that a linear relation exists between the two at typical radar frequencies (3–10 GHz). Differential phase measurements can be used to correct for attenuation in the reflectivity using algorithms of the form $\mathsf{A} = \alpha \mathsf{K}_{dp}^{\beta}$ [8], where K_{dp} (specific differential phase) is the derivative of Φ_{dp} along the range. However, the parameter α is temperature dependent, but a technique has been developed to estimate this parameter in real-time and taking into account the total PIA as a constraint [52].

Figure 1 shows a squall line that produced strong attenuation on the west side of the radar scan. Figure 2a shows the differential phase measurements. Φ_{dp} shows large differential phase shifts on the west side of the squall line indicative of strong attenuation in the reflectivity. Figure 2b shows the attenuation-corrected reflectivity using the algorithm proposed by Bringi et al. [52]. By comparing the attenuation-corrected reflectivity (Fig. 2) with the original reflectivity scan shown in Fig. 1, it becomes clear that there are some regions on the west side of the squall line that showed a strong attenuation of around 10 dB, which can produce a large source of uncertainty when the reflectivity is transformed to an estimate of rainfall rate if no correction for attenuation is performed.

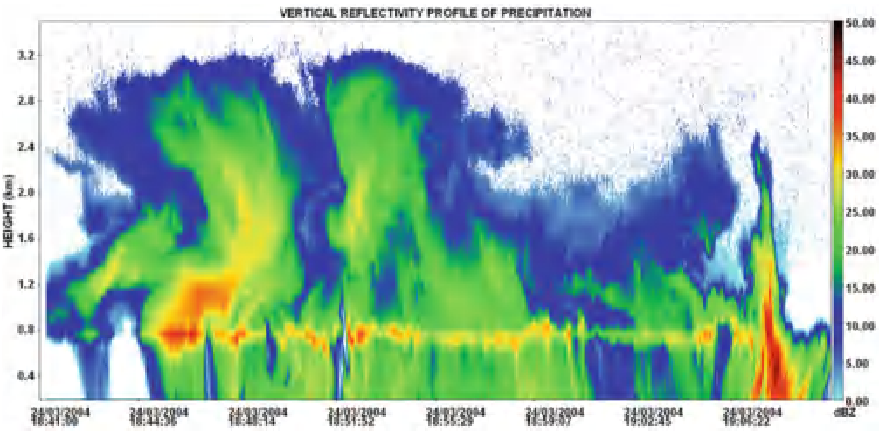


Fig. 3 VPR measured with a vertically pointing radar in the UK. The BB is shown at a height of around 800 m

2.4 Variations in the Vertical Profile of Reflectivity

Microphysics determines the hydrometeors present in the atmosphere by affecting their growth or evaporation and thermodynamic phase, which in turn shapes the structure of the reflectivity with height [1]. This is known as the vertical profile of reflectivity (VPR). Significant variability in the VPR occurs as a result of precipitation growth, evaporation, melting of ice particles and snow flakes and wind effects. Such variations indicate that there are large differences between the radar precipitation estimation at certain altitude and that falling at the ground surface (see Fig. 3). At low radar elevation angles, the height of the radar beam increases with distance. As a result, the precipitation particles intercepted by the radar sampling volume might be due to rain, melting snow, snow, ice, etc. or a combination of different precipitation particles. This variability affects reflectivity measurements and the estimation of precipitation may not be representative of the rainfall rate at the ground. Variations in the VPR are particularly pronounced where melting occurs. Snowflakes are generally low-density aggregates and when they start to melt they look like big raindrops to the radar, resulting in larger values of reflectivities compared to the expected reflectivity below the melting layer [7]. The enhanced reflectivity in the melting layer is known as the bright band (BB) and it can cause significant overestimates of precipitation. To overcome ground clutter and partial beam blockage due to high ground, a weather radar scans at several elevation angles. The height of the radar beam will increase with distance from the radar site due to both scan elevation angle and curvature of the earth. The radar beam is likely to overshoot the shallow precipitation at longer ranges, resulting in underestimation of the precipitation rate or complete failure to detect the shallow precipitation. On the other hand, the radar beam can intercept the melting layer at long ranges leading to overestimation of precipitation. This can produce range-dependent biases in the radar rainfall estimates [11, 13, 15, 53].

A number of techniques have been proposed to deal with the VPR and BB effects on radar measurements, including climatological corrections depending on seasons [54, 55] or rain types [56], characterization of VPR by estimating the altitude and peak of the BB [57], and retrieval of the VPR by filtering the beam-sampling effects from the comparison of radar reflectivity at different distances and altitudes [58, 59]. Fabry and Zawadzki [60] studied the structure of the BB using long-term observations from a vertically pointing radar with high spatial (15 m) and temporal (2 s) resolutions. Their results highlighted the importance of the shape, density and fall speed of the ice particles in the existence of a BB. Andrieu and Creutin [58] proposed an inverse method for retrieving VPR from a two-elevation scanning radar based on reflectivity ratio (reflectivity at high elevation divided by reflectivity at low elevation) function. However, this method assumes VPR homogeneity. Further improvements in the VPR retrieval methods have been developed by using volume radar scans (scans taken a different elevation angles) [61–65]. A full-volume scan radar provides an estimate of the VPR, which can be smoothed by the characteristics of the radar beam. Bellon et al. [66, 67] highlighted the influence of the spatial heterogeneity of VPRs on the resulting corrections associated with volumetric sampling strategy. Kitchen et al. [57] developed a method to correct the BB by using an idealized reflectivity profile convolved with the radar beam power profile. The current UK operational correction method is based on Kitchen et al.′s [57] algorithm which relies on forecasts of freezing level heights in addition to a fixed BB thickness of 700 m. However, this approach does not allow for spatial and thickness irregularities in the BB which can occur due to atmospheric variability. Smyth and Illingworth [68] have emphasized that it is important to use a correction procedure which uses different VPRs for different precipitation types (i.e. stratiform and convective). Moreover, the advance of polarimetric radar enabled new techniques to classify hydrometeors for BB correction, such as the decision tree method, classic statistical decision theory, neural network techniques and fuzzy logic [43, 69]. Rico-Ramirez et al. [70] developed a fuzzy logic classifier based on S-band DP radar measurements to identify the BB and showed that the combination of this classifier with Kitchen et al.′s [57] algorithm can be used to identify and remove the BB. This algorithm has also been implemented at operational C-band frequencies and has shown some skill in identifying and removing the BB [71].

2.5 Variations of the DSD and Radar Rainfall Estimation

Additional errors and uncertainties could be introduced when converting the radar reflectivity Z into an estimate of precipitation intensity R at ground level [13, 19, 72, 73]. The general form of the $\mathsf{Z}-\mathsf{R}$ relationship is a power-law given by $\mathsf{Z}=\mathsf{a}\mathsf{R}^{\mathsf{b}}$, where a and b are the parameters that depend on the DSD. The DSD parameters are obtained empirically by establishing a climatological $\mathsf{Z}-\mathsf{R}$ relationship or by

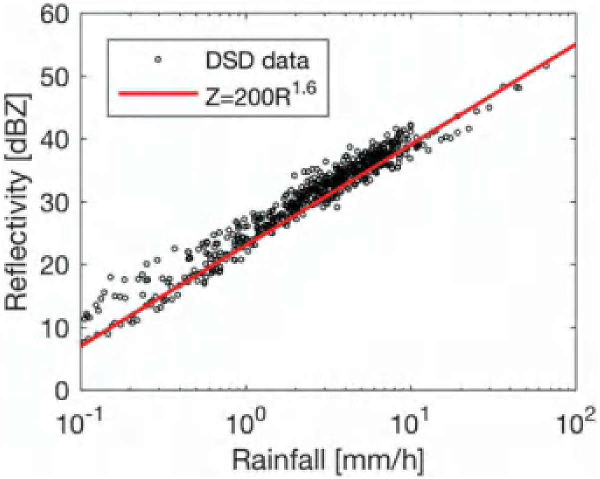


Fig. 4 Z – R measurements computed from DSD data measured by a disdrometer in the UK. The red line shows the climatological Z – R equation used operationally in the UK to estimate radar rainfall rates

simulating Z and R over a wide range of DSDs (see Fig. 4). The relationship between radar reflectivity Z and rainfall intensity R relies on the actual DSD, which varies between different types of storms [7, 13, 72] and within storms [74–76]. Therefore, changes in the DSD introduce a time-varying bias in radar rainfall estimates, because of the use of climatological Z – R relationships. Updrafts and downdrafts can also cause the Z – R relationship to differ greatly from the one obtained in still air [7].

Some approaches address DSD variability with conventional radars such as identifying different precipitation types and applying different Z – R equations. In the US, the relationship $Z = 300R^{1.4}$ is often used for convective precipitation, whereas in the UK, the equation $Z = 200R^{1.6}$ is used for stratiform precipitation. Automatic classification of stratiform and convective precipitation can potentially lead to better rainfall estimates.

Conventional SP weather radars can only measure the radar reflectivity (at horizontal or vertical polarization) to derive precipitation intensity. However, DP radars can provide more detailed information related to the characteristics of precipitation particles such as shape, size, spatial orientation and discrimination of thermodynamic phase [33, 77]. This is because raindrops have an oblate spheroidal shape being their maximal dimensions horizontally oriented, and that the degree of oblateness depends on the raindrop size [8].

DP radars can measure Z_{dr} and K_{dp}. Z_{dr} is a measure of the size of the raindrops and when combined with Z may improve the estimation of precipitation. K_{dp} is almost linearly related to the liquid water content and it also provides the possibility of better estimates of rainfall rates in heavy precipitation [8]. Therefore, algorithms of the forms $R = cZ^{\alpha}10^{\beta Zdr}$ and $R = cK_{dp}^{\beta}$ have been proposed in the literature. The $R - K_{dp}$ algorithm is useful in heavy precipitation and it has the advantage that K_{dp} is immune to attenuation. The coefficients of these algorithms can be obtained using scattering simulations of the radar measurements assuming a wide range of DSDs. Another algorithm of the form $Z = aR^{1.5}$ has been proposed, where the parameter a is a function of the DSD concentration, which can be calculated in real-time using DP radar measurements [78]. A composite algorithm that uses an algorithm $Z - R$ in light rain, an algorithm $R(Z, Z_{dr})$ in moderate rain and an algorithm $R(K_{dp})$ in heavy rain have shown to provide more accurate rainfall rates at C-band frequencies [79].

3 Adjusting Radar Rainfall with Raingauge Measurements

Weather radar provides precipitation estimates with good spatial and temporal resolutions, but as discussed previously, the rainfall estimates can be affected by different error sources. A raingauge network on the other hand may provide accurate rainfall measurements, but at individual point locations. However, raingauge measurements are not always available in mountainous areas and the measurements are also subject to several sources of errors. For instance, typical errors in tipping bucket raingauges include blockages, wetting and evaporation in the funnel, condensation errors, underestimation of high rain rates and wind effects [80]. Moreover, the accuracy of the raingauge measurement is also affected by spatial and temporal sampling uncertainties. The temporal sampling error is defined as the error resulting from repeated temporal gaps during the measurements, whereas the spatial sampling error is defined as the error resulting from approximating an areal estimate using point measurements [79, 81]. A number of spatial interpolation methods, geostatistical or non-geostatistical, are available for approximating an areal rainfall estimate using raingauge point measurements. Geostatistical methods (i.e. kriging) generally perform better than non-geostatistical methods [82–85]. Kriging for instance takes into account the spatial correlation of precipitation through the variogram (obtained from point rainfall observations) that helps to reconstruct the two-dimensional precipitation field. However, even a high-density raingauge network is unable to fully capture the true rainfall field at short timescales [86–89]. In order to exploit the strengths of both radar and raingauge measurement approaches, a number of radar–raingauge merging techniques have been developed. The advantages of merging radar and raingauge rainfall measurements are to produce a rainfall product that not only provides reliable distributed rainfall information, but also provide accurate measurements that are in agreement with the raingauge measurements. These merging methods range from non-statistical methods, such as mean field bias correction [90], spatial correction method [91] and range-dependent adjustment [92], to more complex statistical techniques. These statistical methods are based on univariate and multivariate geostatistical analysis, such as co-kriging [14], kriging with radar-based error (KRE) [93, 94] and kriging with external drift (KED) [83, 95]. For example, the mean field bias (MFB) correction is a simple and effective method, which was developed by Smith and Krajewski [90] for adjusting radar-based quantitative precipitation estimates based on raingauge information and it is widely used in several studies [31, 96–98]. The assumption is that the radar estimates are affected by a uniform systematic error. This error may be due to a poor electronic calibration or an erroneous coefficient in the $Z - R$ relationship due to variations of the DSD [9, 14, 99]. This systematic bias between radar and raingauge rainfall can be adjusted using raingauge information. The bias adjustment is based on estimation of a single multiplicative factor as the ratio of the accumulated raingauge rainfall and the radar rainfall. This simple MFB correction improves radar rainfall estimation considerably [100] and it is often used operationally [101, 102]. However, bias adjustment of radar rainfall may be subjected to sampling errors due to the fact that the raingauge network cannot represent the areal precipitation accurately especially during high variability of precipitation (e.g. during convective precipitation). There are however some techniques that account for the sampling errors to estimate the real-time mean field bias using a Kalman filter technique [90, 96, 97, 103]. Merging radar rainfall with raingauge measurement using KED is an effective method to combine both measurements. In KED, the rainfall predictions are modelled as a drift term plus a residual term. The drift term is an unknown linear function defined externally through an auxiliary variable (e.g. radar rainfall). A full description of the KED method is presented in Wackernagel [104], Haberlandt [83], and Verworn and Haberlandt [105]. The variogram is an important function in geostatistical interpolation [106–108]. The spatial characteristic of the rainfall field contained in the variogram is influenced by the characteristics of the storm, density of raingauge network and rainfall accumulation period. The variogram used in kriging methods can either be parametric or non-parametric and it is calculated independently for each time step. The predefined model variogram represent the spatial variability of the rainfall distribution, and therefore the suitability of the model function and parameter values used to estimate the variogram has impact on the quality of the final merged rainfall product. However, a non-parametric automatic procedure for estimating a spatial variability model does not require any prior assumption about the correlation of the observations. This non-parametric automatic methodology based on Fast Fourier Transform (FFT) was initially proposed in Yao and Journel [109] to estimate spatial variability models and further developed by Velasco-Forero et al. [110] to estimate rainfall fields by merging radar and raingauge data. It is worth to note that the variogram is a function of both distance and direction (e.g. anisotropic) in the non-parametric method. This non-parametric spatial variability model is particularly appropriate for real-time applications of radar observations for operational purpose.

The performance of any rainfall interpolation method highly relies on the density of the raingauge network [96, 97, 111, 112]. An accurate estimation of the true rainfall field often requires a high density raingauge network [81, 113, 114]. Simple adjustment methods (such as mean field bias correction) are less sensitive to the raingauge network density, while the improvement by geostatistical methods (e.g. KED) increases with a more dense raingauge network [115]. Moreover, the improvement of rainfall estimation by the different radar–raingauge merging techniques may vary between different accumulation timescales. For instance, Berndt et al. [116] examined the effect of accumulation timescales on the performance of different merging techniques at different accumulation timescales from 10 min to 6 h. Their results showed that the performance of the radar–raingauge merging

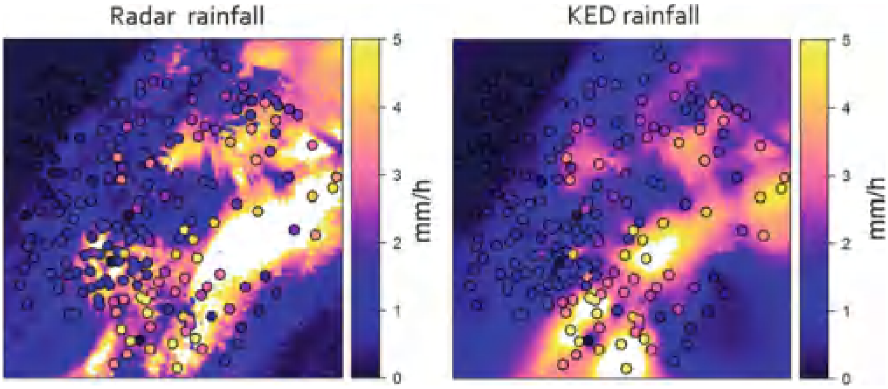


Fig. 5 Radar rainfall versus KED rainfall; the circles represent the raingauge measurements [118]

methods improves for large accumulation times. The improvement of precipitation estimation by the radar–raingauge merging techniques also varies between seasons [105, 111] and between storms [115, 117]. Since the radar–raingauge merging method is the final step for radar-based precipitation estimation, minimizing all the sources of error in radar rainfall before applying a radar–raingauge merging technique is important to improve precipitation estimates. Figure 5 shows a comparison between the original radar rainfall field and the KED rainfall product. The raingauge observations are also shown in both rainfall fields. As shown, the KED rainfall field shows the spatial distribution of precipitation from radar and the accuracy of point observations from raingauges measurements.

4 Applications of Weather Radar

Precipitation observations are made for a variety of reasons, such as real-time flood forecasting [119], weather forecasting and extreme weather warnings [120], climate modelling [121, 122], hydrological modelling [123], agricultural meteorology [124], and for research in meteorology and climatology [125, 126]. Moreover, precipitation data is also important for many design calculations, such as for sewer system design [127], assessment of combined sewer overflows [128], flood risk assessment, river discharges [129] and river water quality [130]. This section mainly discusses applications related to short-term precipitation forecasting with radar and hydrological uses of weather radar.

4.1 Radar-Based Precipitation Forecasting

Precipitation forecasts can be produced either by Numerical Weather Prediction (NWP) models or by using a sequence of radar rainfall scans. NWP models have a better performance over longer timescales as they dynamically resolve the large-scale atmospheric processes. Radar-based precipitation forecasting is known as precipitation nowcasting. Nowcasting models are based on the extrapolation of radar rainfall scans to track the motion of precipitation cells with a forecasting lead time of a few hours. Radar-based precipitation nowcasting has a higher performance than NWP forecasts for the first few hours of the forecasts, but NWP forecasts have a better performance at longer forecasting lead times. Radar nowcasting can be however very valuable for flash flood forecasting in urban areas or hydrological forecasting in large catchments.

Nowcasting aims to tracking the movement of storms to extrapolate the radar rainfall field into the future with a forecasting lead time of a few hours [131–133]. Radar-based nowcasting methods include Tracking radar echoes by correlation (TREC and COTREC methods), tracking the centroids of rain cells, use of wind fields from NWP forecasts to advect the precipitation field, the Variational Echo Tracking (VET) method and optical flow techniques. In the TREC method, the advection vector for each block is determined using a correlation method, but gaps appear where blocks diverge. The COTREC algorithm minimizing the divergence of the velocities of adjacent blocks to avoid the issues observed in the TREC method. The Variational Echo Tracking (VET) method is similar to the COTREC method, but also takes into account radial velocities from Doppler radar. Bowler et al. [131] developed a method to compute the advection field using optical flow techniques that has shown better performance in cases involving embedded convection. This method assumes that features in a sequence of radar scans only change shape, but do not change in size or intensity. Radar echo-tracking techniques assume Lagrangian persistence along the direction and speed of movement of storms. For forecasts beyond 5 min, nowcasting systems that rely on Lagrangian persistence are subjected to errors because storms evolve and change direction [87, 88]. Larger scale precipitation features have higher Lagrangian temporal autocorrelation and may persist longer than smaller scale features in the forecast. In fact, the predictability of precipitation systems depends on the size of their scale, with small-scale features being less persistent and less predictable [134]. Also, smaller scale precipitation features evolve faster than larger-scale features. For instance, isolated thunderstorms smaller in size (e.g. 10 km) will undergo considerable evolution over a short-time scale (e.g. 30 min) [135], whereas larger storm systems (e.g. 100 km in size) will evolve over several hours. This evolution is difficult to forecast and consequently the skill of the forecast decreases rapidly with lead time. Large-scale precipitation patterns can be forecasted by extrapolation whereas locally generated precipitation (e.g. due to convection) is less predictable. Convective storm initiation is therefore an important area of research in radar nowcasting.

Radar-based extrapolation techniques generally do not account for precipitation growth and decay [136]. However, precipitation systems evolve with time, and therefore growth and decay of precipitation processes become important resulting in a decrease of forecasting skill with lead time. In fact, uncertainties in radar-based forecasts are due to errors in the original radar rainfall field, uncertainties due to the temporal evolution of the velocity field (i.e. assumption that the rainfall trajectories do not change with time), uncertainties due to the fact that growth and decay of precipitation is not accounted for and uncertainties related to the type of nowcasting system used to produce the forecasts [137, 138]. It is therefore common practice for nowcasting models to account for the uncertainty in the predictions using ensemble forecasts, which is a set of equally likely forecasts. Ensemble radar-based forecasts can be generated by adding spatially correlated noise to the deterministic forecast. The ensemble of forecasts have in common the more predictable large-scale precipitation patterns but will differ in the small-scale patterns that are less predictable. More details on the implementation of ensemble forecasts can be found in Seed [139] and Berenguer et al. [140].

Although NWP models perform better than radar nowcasting for longer time scales, NWP models do not generally capture well the initial precipitation conditions in comparison to radar-based precipitation forecasts. Consequently, in order to combine the strengths of NWP forecasts and radar nowcasts, new blending methods have been developed such as the Short-Term Ensemble Prediction System (STEPS) [141]. With the increase in computer power, new NWP-based nowcasting approaches based on 4D-Var data assimilation have been developed and look promising [142]. However, for very short-term forecasts (60-min or so), radar-based nowcasting can provide better forecasts than NWP-based approaches. This is particularly important for radar rainfall applications in urban catchments.

4.2 Hydrological Applications

Hydrological forecasting is one of the most important applications of radar rainfall observations [14, 143–145]. The hydrological processes in river catchments can be modelled using rainfall–runoff models (also known as hydrological models). These models have a variety of applications that include flood forecasting. Depending on the application, hydrological models can have different levels of complexity and can be classified into lumped, semi-distributed and distributed models. In the UK, the national flood forecasting system platform uses several hydrological models including lumped (the probability-distributed model) and distributed (the grid-to-grid model) models [146, 147]. These models use radar rainfall measurements and forecasts (nowcasts, NWP forecasts or a combination of both) to simulate river flows for any catchment in the UK. River flow simulations and forecasts are used to issue flood warnings several hours in advance. In fact, radar-based hydrological forecasts are extremely important especially during flash floods. Hydrological forecasts with longer lead times are also available when radar nowcasts are blended with NWP forecasts. It is obvious that the quality of the hydrological forecasts depends on the quality of the rainfall measurements [148] and also on the quality of the hydrological models used to simulate the hydrological processes in a particular catchment.

On the other hand, urban catchments usually have shorter response times [149–151] and weather radar provides unique information on the dynamics of precipitation events in space and time at high spatial and temporal resolutions, which is very difficult to obtain through a network of raingauges. In fact, urban catchments require rainfall measurements with higher temporal and spatial resolutions [152, 153]. For example, Berne et al. [154] suggested that hydrological applications for an urban catchment with a large area (e.g. 1,000 ha) require rainfall measurements at 5 min/3 km resolutions, whereas smaller urban areas (e.g. 100 ha) require rainfall measurements with resolutions of 3 min/2 km. Ocho-Rodriguez et al. [155] investigated the impact of the spatial and temporal resolution of precipitation in the hydrodynamic response of urban catchments concluding that the temporal resolution of precipitation affects the modelling results more strongly than variations in rainfall spatial resolution. They concluded that resolutions of 1 min/1 km appear to be sufficient for urban hydrodynamic modelling. Although these resolutions are feasible with small X-band weather radar systems, common operational radar networks are not able to achieve the temporal resolutions required for small urban catchments. For instance, the operational weather radar network in the UK provides rainfall measurements at 5 min/1 km resolutions over the UK. So, it is evident that improvements in radar temporal resolution are required to satisfy urban hydrological applications in particular for smaller urban catchments. This could be achieved by using nowcasting models to interpolate 5 min radar rainfall measurements to produce measurements at 1 min temporal resolutions or by performing additional low-elevation scans within the 5 min radar scanning strategy.

5 Concluding Comments

Precipitation is the main driver of the hydrological cycle and therefore precipitation is a key input to hydrological models. Raingauges and weather radars are the most widely used instruments to measure precipitation. Raingauge measurements are traditionally used as the main input to rainfall–runoff models. In addition, they are also used for calibrating and validating radar rainfall algorithms [99, 156]. However, operational raingauge networks are often very sparse and unable to fulfill the density requirements for real-time hydrological modelling. Rainfall events with high variability in space and time may not be represented accurately by a raingauge network. The greatest benefit of weather radar is its potential to estimate rainfall rates at high spatiotemporal resolution (e.g. 1 km/5 min) in real-time and over a large area. Although radar rainfall measurements can be affected by different error sources, there are different algorithms to control the quality of radar rainfall that enable its quantitative use for hydrological and meteorological purposes. Polarimetric weather radars bring several benefits including improvements in radar data quality, identification of hydrometeors, attenuation correction and radar rainfall estimation. Several methods to merge radar rainfall with raingauge measurements have also been developed in the literature. The merging of radar rainfall and raingauge measurements can bring the benefits of both instruments, that is, the accuracy of point raingauge observations and the spatial distribution of precipitation from radar measurements.

Weather radars have also demonstrated a huge potential for real-time flood forecasting applications. Weather radar enables the monitoring of extreme weather events. Radar rainfall can be used to produce short-term precipitation forecasts using nowcasting models. The combination of radar measurements and forecasts with other models (e.g. hydrological models or flood models) enables the forecasting of high-water levels in rivers or potential areas likely to be affected by flooding.

References

1. Fabry F (2015) Radar meteorology: principles and practice. Cambridge University Press, Cambridge
2. Raghavan S (2013) Radar meteorology, vol 27. Springer, Dordrecht
3. Rinehart RE (1997) Radar for meteorologists. Rinehart Publishing, Grand Forks
4. Atlas D, Ulbrich CW (1977) Path-and area-integrated rainfall measurement by microwave attenuation in the 1-3 cm band. J Appl Meteorol 16(12):1322–1331
5. Marshall JS, Palmer WMK (1948) The distribution of raindrops with size. J Meteorol 5 (4):165–166
6. Marshall J, Hitschfeld W, Gunn K (1955) Advances in radar weather. Adv Geophys 2:1–56
7. Battan LJ (1973) Radar observation of the atmosphere. University of Chicago Press, Chicago
8. Bringi V, Chandrasekar V (2001) Polarimetric Doppler weather radar: principles and applications. Cambridge University Press, Cambridge
9. Borga M (2002) Accuracy of radar rainfall estimates for streamflow simulation. J Hydrol 267 (1):26–39
10. Browning K, Collier C (1989) Nowcasting of precipitation systems. Rev Geophys 27 (3):345–370
11. Fabry F, Austin G, Tees D (1992) The accuracy of rainfall estimates by radar as a function of range. Q J R Meteorol Soc 118(505):435–453
12. Hunter SM (1996) WSR-88D radar rainfall estimation: capabilities, limitations and potential improvements. Natl Weather Dig 20(4):26–38
13. Joss J, Waldvogel A, Collier C (1990) Precipitation measurement and hydrology. In: Radar in meteorology. Springer, Berlin, pp 577–606
14. Krajewski WF, Smith J (2002) Radar hydrology: rainfall estimation. Adv Water Resour 25 (8):1387–1394
15. Smith JA, Seo DJ, Baeck ML, Hudlow MD (1996) An intercomparison study of NEXRAD precipitation estimates. Water Resour Res 32(7):2035–2045
16. Villarini G, Krajewski WF (2010) Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall. Surv Geophys 31(1):107–129
17. Wilson JW, Brandes EA (1979) Radar measurement of rainfall – a summary. Bull Am Meteorol Soc 60(9):1048–1058
18. Young B, Nelson BR, Bradley AA, Smith JA, Peters-Lidard CD, Kruger A, Baeck ML (1999) An evaluation of NEXRAD precipitation estimates in complex terrain. J Geophys Res 104 (D16):19691–19703
19. Zawadzki I (1984) Factors affecting the precision of radar measurements of rain. In: Conference on radar meteorology, 22nd, Zurich, Switzerland, pp 251–256
20. Whiton RC, Smith Jr P, Harbuck A (1977) Calibration of weather radar systems using the sun as a radio source. In: 17th conference on radar meteorology, pp 60–65
21. Altube P, Bech J, Argemí O, Rigo T (2015) Quality control of antenna alignment and receiver calibration using the sun: adaptation to midrange weather radar observations at low elevation angles. J Atmos Ocean Technol 32(5):927–942

22. Holleman I, Huuskonen A, Kurri M, Beekhuis H (2010) Operational monitoring of weather radar receiving chain using the sun. J Atmos Ocean Technol 27(1):159–166
23. Huuskonen A, Kurri M, Hohti H, Beekhuis H, Leijnse H, Holleman I (2014) Radar performance monitoring using the angular width of the solar image. J Atmos Ocean Technol 31 (8):1704–1712
24. Wolff DB, Marks DA, Petersen WA (2015) General application of the relative calibration adjustment (rca) technique for monitoring and correcting radar reflectivity calibration. J Atmos Ocean Technol 32(3):496–506
25. Ribaud J-F, Bousquet O, Coquillat S, Al-Sakka H, Lambert D, Ducrocq V, Fontaine E (2016) Evaluation and application of hydrometeor classification algorithm outputs inferred from multi-frequency dual-polarimetric radar observations collected during hymex. Q J R Meteorol Soc 142(S1):95–107
26. Gorgucci E, Scarchilli G, Chandrasekar V (1992) Calibration of radars using polarimetric techniques. IEEE Trans Geosci Remote Sens 30(5):853–858
27. Scarchilli G, Gorgucci V, Chandrasekar V, Dobaie A (1996) Self-consistency of polarization diversity measurement of rainfall. IEEE Trans Geosci Remote Sens 34(1):22–26
28. Anagnostou EN, Krajewski WF, Seo D-J, Johnson ER (1998) Mean-field rainfall bias studies for WSR-88D. J Hydrol Eng 3(3):149–159
29. Kitchen M, Blackall R (1992) Representativeness errors in comparisons between radar and gauge measurements of rainfall. J Hydrol 134(1–4):13–33
30. Anagnostou EN, Morales CA, Dinku T (2001) The use of trmm precipitation radar observations in determining ground radar calibration biases. J Atmos Ocean Technol 18(4):616–628
31. Harrison D, Driscoll S, Kitchen M (2000) Improving precipitation estimates from weather radar using quality control and correction techniques. Meteorol Appl 7(02):135–144
32. Sugier J, du Chatelet JP, Roquain P, Smith A (2002) Detection and removal of clutter and anaprop in radar data using a statistical scheme based on echo fluctuation. Proc ERAD 2002:17–24
33. Doviak RJ, Zrnic DS (2014) Doppler radar and weather observations. Academic Press, San Diego
34. Steiner M, Smith JA (2002) Use of three-dimensional reflectivity structure for automated detection and removal of nonprecipitating echoes in radar data. J Atmos Ocean Technol 19 (5):673–686
35. Moszkowicz S, Ciach GJ, Krajewski WF (1994) Statistical detection of anomalous propagation in radar reflectivity patterns. J Atmos Ocean Technol 11(4):1026–1034
36. Rico-Ramirez MA, Cluckie ID (2008) Classification of ground clutter and anomalous propagation using dual-polarization weather radar. IEEE Trans Geosci Remote Sens 46 (7):1892–1904
37. Berenguer M, Sempere-Torres D, Corral C, Sánchez-Diezma R (2006) A fuzzy logic technique for identifying nonprecipitating echoes in radar scans. J Atmos Ocean Technol 23 (9):1157–1180
38. Cho Y-H, Lee GW, Kim K-E, Zawadzki I (2006) Identification and removal of ground echoes and anomalous propagation using the characteristics of radar echoes. J Atmos Ocean Technol 23(9):1206–1222
39. Hubbert J, Dixon M, Ellis S (2009) Weather radar ground clutter. Part II: real-time identification and filtering. J Atmos Ocean Technol 26(7):1181–1197
40. Grecu M, Krajewski WF (1999) Detection of anomalous propagation echoes in weather radar data using neural networks. IEEE Trans Geosci Remote Sens 37(1):287–296
41. Haykin S, Deng C (1991) Classification of radar clutter using neural networks. IEEE Trans Neural Netw 2(6):589–600
42. Gourley JJ, Tabary P, Parent du Chatelet J (2007) A fuzzy logic algorithm for the separation of precipitating from nonprecipitating echoes using polarimetric radar observations. J Atmos Ocean Technol 24(8):1439–1451

43. Liu H, Chandrasekar V (2000) Classification of hydrometeors based on polarimetric radar measurements: development of fuzzy logic and neuro-fuzzy systems, and in situ verification. J Atmos Ocean Technol 17(2):140–164
44. Park HS, Ryzhkov A, Zrnic′ D, Kim K-E (2009) The hydrometeor classification algorithm for the polarimetric WSR-88D: description and application to an mcs. Weather Forecast 24 (3):730–748
45. Hall W, Rico-Ramirez MA, Krämer S (2017) Offshore wind turbine clutter characteristics and identification in operational c-band weather radar measurements. Q J R Meteorol Soc 143 (703):720–730
46. Chandrasekar V, Keränen R, Lim S, Moisseev D (2013) Recent advances in classification of observations from dual polarization weather radars. Atmos Res 119:97–111
47. Bringi V, Chandrasekar V, Balakrishnan N, Zrnic D (1990) An examination of propagation effects in rainfall on radar measurements at microwave frequencies. J Atmos Ocean Technol 7 (6):829–840
48. Hitschfeld W, Bordan J (1954) Errors inherent in the radar measurement of rainfall at attenuating wavelengths. J Meteorol 11(1):58–67
49. Meneghini R, Eckerman J, Atlas D (1983) Determination of rain rate from a spaceborne radar using measurements of total attenuation. IEEE Trans Geosci Remote Sens 1:34–43
50. Testud J, Le Bouar E, Obligis E, Ali-Mehenni M (2000) The rain profiling algorithm applied to polarimetric weather radar. J Atmos Ocean Technol 17(3):332–356
51. Rico-Ramirez MA (2012) Adaptive attenuation correction techniques for C-band polarimetric weather radars. IEEE Trans Geosci Remote Sens 50(12):5061–5071
52. Bringi VN, Keenan T, Chandrasekar V (2001) Correcting c-band radar reflectivity and differential reflectivity data for rain attenuation: a self-consistent method with constraints. IEEE Trans Geosci Remote Sens 39(9):1906–1915
53. Kitchen M, Jackson P (1993) Weather radar performance at long range-simulated and observed. J Appl Meteorol 32(5):975–985
54. Joss J, Lee R (1995) The application of radar–gauge comparisons to operational precipitation profile corrections. J Appl Meteorol 34(12):2612–2630
55. Koistinen J (1991) Operational correction of radar rainfall errors due to the vertical reflectivity profile. In: Preprints, 25th international conference on radar meteorology, vol 9194. Americal Meteorolgical Society, Paris
56. Collier C (1986) Accuracy of rainfall estimates by radar, part I: calibration by telemetering raingauges. J Hydrol 83(3–4):207–223
57. Kitchen M, Brown R, Davies A (1994) Real-time correction of weather radar data for the effects of bright band, range and orographic growth in widespread precipitation. Q J R Meteorol Soc 120(519):1231–1254
58. Andrieu H, Creutin JD (1995) Identification of vertical profiles of radar reflectivity for hydrological applications using an inverse method. Part I: formulation. J Appl Meteorol 34 (1):225–239
59. Seo D-J, Breidenbach J, Fulton R, Miller D, O′Bannon T (2000) Real-time adjustment of range-dependent biases in wsr-88d rainfall estimates due to nonuniform vertical profile of reflectivity. J Hydrometeorol 1(3):222–240
60. Fabry F, Zawadzki I (1995) Long-term radar observations of the melting layer of precipitation and their interpretation. J Atmos Sci 52(7):838–851
61. Germann U, Joss J (2002) Mesobeta profiles to extrapolate radar precipitation measurements above the alps to the ground level. J Appl Meteorol 41(5):542–557
62. Gourley JJ, Calvert CM (2003) Automated detection of the bright band using wsr-88d data. Weather Forecast 18(4):585–599
63. Marzano FS, Vulpiani G, Picciotti E (2004) Rain field and reflectivity vertical profile reconstruction from c-band radar volumetric data. IEEE Trans Geosci Remote Sens 42 (5):1033–1046

64. Vignal B, Andrieu H, Creutin JD (1999) Identification of vertical profiles of reflectivity from volume scan radar data. J Appl Meteorol 38(8):1214–1228
65. Zhang J, Langston C, Howard K (2008) Brightband identification based on vertical profiles of reflectivity from the wsr-88d. J Atmos Ocean Technol 25(10):1859–1872
66. Bellon A, Lee GW, Kilambi A, Zawadzki I (2007) Real-time comparisons of vpr-corrected daily rainfall estimates with a gauge mesonet. J Appl Meteorol Climatol 46(6):726–741
67. Bellon A, Lee GW, Zawadzki I (2005) Error statistics of vpr corrections in stratiform precipitation. J Appl Meteorol 44(7):998–1015
68. Smyth T, Illingworth A (1998) Radar estimates of rainfall rates at the ground in bright band and non-bright band events. Q J R Meteorol Soc 124(551):2417–2434
69. Vivekanandan J, Ellis S, Oye R, Zrnic D, Ryzhkov A, Straka J (1999) Cloud microphysics retrieval using s-band dual-polarization radar measurements. Bull Am Meteorol Soc 80 (3):381–388
70. Rico-Ramirez MA, Cluckie I, Han D (2005) Correction of the bright band using dual-polarisation radar. Atmos Sci Lett 6(1):40–46
71. Hall W, Rico-Ramirez MA, Krämer S (2015) Classification and correction of the bright band using an operational c-band polarimetric radar. J Hydrol 531:248–258
72. Austin PM (1987) Relation between measured radar reflectivity and surface rainfall. Mon Weather Rev 115(5):1053–1070
73. Doviak RJ (1983) A survey of radar rain measurement techniques. J Clim Appl Meteorol 22 (5):832–849
74. Carbone RE, Nelson LD (1978) The evolution of raindrop spectra in warm-based convective storms as observed and numerically modeled. J Atmos Sci 35(12):2302–2314
75. Cataneo R, Stout GE (1968) Raindrop-size distributions in humid continental climates, and associated rainfall rate-radar reflectivity relationships. J Appl Meteorol 7(5):901–907
76. Smith JA, Krajewski WF (1993) A modeling study of rainfall rate-reflectivity relationships. Water Resour Res 29(8):2505–2514
77. Herzegh PH, Jameson AR (1992) Observing precipitation through dual-polarization radar measurements. Bull Am Meteorol Soc 73(9):1365–1374
78. Illingworth A (2004) Improved precipitation rates and data quality by using polarimetric measurements. In: Weather radar. Springer, Berlin, pp 130–166
79. Bringi V, Rico-Ramirez MA, Thurai M (2011) Rainfall estimation with an operational polarimetric C-band radar in the United Kingdom: comparison with a gauge network and error analysis. J Hydrometeorol 12(5):935–954
80. Upton G, Rahimi A (2003) On-line detection of errors in tipping-bucket raingauges. J Hydrol 278(1):197–212
81. Villarini G, Mandapaka PV, Krajewski WF, Moore RJ (2008) Rainfall and sampling uncertainties: a rain gauge perspective. J Geophys Res Atmos 113:D11
82. Goovaerts P (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J Hydrol 228(1):113–129
83. Haberlandt U (2007) Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. J Hydrol 332(1):144–157
84. Naoum S, Tsanis I (2004) Ranking spatial interpolation techniques using a GIS-based DSS. Global Nest 6:1), 1–1),20
85. Vicente Serrano SM, Sánchez S, Cuadrat JM (2003) Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature. Clim Res 24(2):161–180
86. Ciach GJ, Krajewski WF (2006) Analysis and modeling of spatial correlation structure in small-scale rainfall in Central Oklahoma. Adv Water Resour 29(10):1450–1463
87. Germann U, Galli G, Boscacci M, Bolliger M (2006) Radar precipitation measurement in a mountainous region. Q J R Meteorol Soc 132(618):1669–1692
88. Germann U, Zawadzki I, Turner B (2006) Predictability of precipitation from continental radar images. Part IV: limits to prediction. J Atmos Sci 63(8):2092–2108

89. Peleg N, Ben-Asher M, Morin E (2013) Radar subpixel-scale rainfall variability and uncertainty: lessons learned from observations of a dense rain-gauge network. Hydrol Earth Syst Sci 17(6):2195–2208
90. Smith JA, Krajewski WF (1991) Estimation of the mean field bias of radar rainfall estimates. J Appl Meteorol 30(4):397–412
91. Brandes EA (1975) Optimizing rainfall estimates with the aid of radar. J Appl Meteorol 14 (7):1339–1345
92. Koistinen J, Michelson DB (2002) Baltex weather radar-based precipitation products and their accuracies. Boreal Environ Res 7(3):253–263
93. Ehret U, Götzinger J, Bárdossy A, Pegram GG (2008) Radar-based flood forecasting in small catchments, exemplified by the Goldersbach catchment, Germany. Int J River Basin Manag 6 (4):323–329
94. Sinclair S, Pegram G (2005) Combining radar and rain gauge rainfall estimates using conditional merging. Atmos Sci Lett 6(1):19–22
95. Delrieu G, Wijbrans A, Boudevillain B, Faure D, Bonnifait L, Kirstetter P-E (2014) Geostatistical radar–raingauge merging: a novel method for the quantification of rain estimation accuracy. Adv Water Resour 71:110–124
96. Chumchean S, Seed A, Sharma A (2006) Correcting of real-time radar rainfall bias using a Kalman filtering approach. J Hydrol 317(1):123–137
97. Chumchean S, Sharma A, Seed A (2006) An integrated approach to error correction for real-time radar-rainfall estimation. J Atmos Ocean Technol 23(1):67–79
98. Fulton RA, Breidenbach JP, Seo D-J, Miller DA, O′Bannon T (1998) The WSR-88D rainfall algorithm. Weather Forecast 13(2):377–395
99. Anagnostou EN, Krajewski WF (1998) Calibration of the WSR-88D precipitation processing subsystem. Weather Forecast 13(2):396–406
100. Wilson JW (1970) Integration of radar and raingage data for improved rainfall measurement. J Appl Meteorol 9(3):489–497
101. Collier C (1986) Accuracy of rainfall estimates by radar, part II: comparison with raingauge network. J Hydrol 83(3–4):225–235
102. Harrison DL, Scovell RW, Kitchen M (2009) High-resolution precipitation estimates for hydrological uses. In: Proceedings of the institution of civil engineers-water management, vol 162. Thomas Telford Ltd, London, pp 125–135
103. Seo D-J, Breidenbach J, Johnson E (1999) Real-time estimation of mean field bias in radar rainfall data. J Hydrol 223(3):131–147
104. Wackernagel H (1996) Multivariate geostatistics: an introduction with applications. Int J Rock Mech Min Sci Geomech Abstr 8:363A
105. Verworn A, Haberlandt U (2011) Spatial interpolation of hourly rainfall – effect of additional information, variogram inference and storm properties. Hydrol Earth Syst Sci 15(2):569–584
106. Holawe F, Dutter R (1999) Geostatistical study of precipitation series in Austria: time and space. J Hydrol 219(1):70–82
107. Skøien JO, Blöschl G, Western A (2003) Characteristic space scales and timescales in hydrology. Water Resour Res 39(10):1304
108. Van De Beek C, Leijnse H, Torfs P, Uijlenhoet R (2011) Climatology of daily rainfall semi-variance in the Netherlands. Hydrol Earth Syst Sci 15(1):171
109. Yao T, Journel AG (1998) Automatic modeling of (cross) covariance tables using fast fourier transform. Math Geol 30(6):589–615
110. Velasco-Forero CA, Sempere-Torres D, Cassiraga EF, Gómez-Hernández JJ (2009) A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data. Adv Water Resour 32(7):986–1002
111. Goudenhoofdt E, Delobbe L (2009) Evaluation of radar-gauge merging methods for quantitative precipitation estimates. Hydrol Earth Syst Sci 13(2):195–203

112. Schuurmans J, Bierkens M, Pebesma E, Uijlenhoet R (2007) Automatic prediction of high-resolution daily rainfall fields for multiple extents: the potential of operational radar. J Hydrometeorol 8(6):1204–1224
113. Ballester J, Moré J (2007) The representativeness problem of a station net applied to the verification of a precipitation forecast based on areas. Meteorol Appl 14(2):177–184
114. Cherubini T, Ghelli A, Lalaurette F (2002) Verification of precipitation forecasts over the alpine region using a high-density observing network. Weather Forecast 17(2):238–249
115. Nanding N, Rico-Ramirez MA, Han D (2015) Comparison of different radar-raingauge rainfall merging techniques. J Hydroinf 17(3):422–445
116. Berndt C, Rabiei E, Haberlandt U (2014) Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios. J Hydrol 508:88–101
117. Schiemann R, Erdin R, Willi M, Frei C, Berenguer M, Sempere-Torres D (2011) Geostatistical radar-raingauge combination with nonparametric correlograms: methodological considerations and application in Switzerland. Hydrol Earth Syst Sci 15(5):1515–1536
118. Cecinati F, Wani O, Rico-Ramirez MA (2017) Comparing approaches to deal with non-gaussianity of rainfall data in kriging-based radar-gauge rainfall merging. Water Resour Res 53(11):8999–9018
119. Courty LG, Rico-Ramirez MÁ, Pedrozo-Acuña A (2018) The significance of the spatial variability of rainfall on the numerical simulation of urban floods. Water 10(2):207
120. Kallos G, Pytharoulis I (2005) Short-term predictions (weather forecasting purposes). In: Encyclopedia of hydrological sciences. Wiley, Chichester
121. Lüthi D, Cress A, Davies H, Frei C, Schär C (1996) Interannual variability and regional climate simulations. Theor Appl Climatol 53(4):185–209
122. Widmann M, Schär C (1997) A principal component and long-term trend analysis of daily precipitation in Switzerland. Int J Climatol 17(12):1333–1356
123. Beven KJ (2005) Rainfall-runoff modeling: introduction. In: Encyclopedia of hydrological sciences. Wiley, Chichester
124. Sevruk B (2005) Rainfall measurement: gauges. In: Encyclopedia of hydrological sciences. Wiley, Chichester
125. Flossmann AI (2006) Models of clouds, precipitation and storms. In: Encyclopedia of hydrological sciences. Wiley, Chichester
126. Frei C, Schär C (1998) A precipitation climatology of the Alps from high-resolution raingauge observations. Int J Climatol 18(8):873–900
127. Vaes G (1999) The influence of rainfall and model simplification on combined sewer system design. PhD, KU Leuven, Belgium
128. Vaes G, Berlamont J (2000) Selection of appropriate short rainfall series for design of combined sewer systems. In: International conference on urban drainage on internet, Czech Republic
129. Krishnappan BG, Altinakar MS (2006) Numerical modeling of unsteady flows in rivers. In: Encyclopedia of hydrological sciences. Wiley, Chichester
130. Meybeck M, Peters NE, Chapman DV (2005) Water quality. In: Encyclopedia of hydrological sciences. Wiley, Chichester
131. Bowler NE, Pierce CE, Seed A (2004) Development of a precipitation nowcasting algorithm based upon optical flow techniques. J Hydrol 288(1):74–91
132. Collier CG, Kzyzysztofowicz R (2000) Quantitative precipitation forecasting. J Hydrol 239 (1):1–2
133. Liguori S, Rico-Ramirez MA, Schellart A, Saul A (2012) Using probabilistic radar rainfall nowcasts and NWP forecasts for flow prediction in urban catchments. Atmos Res 103:80–95
134. Liguori S, Rico-Ramirez MA (2014) A review of current approaches to radar-based quantitative precipitation forecasts. Int J River Basin Manag 12(4):391–402
135. Wilson JW, Megenhardt DL (1997) Thunderstorm initiation, organization, and lifetime associated with Florida boundary layer convergence lines. Mon Weather Rev 125 (7):1507–1525

136. Golding B (1998) NIMROD: a system for generating automated very short range forecasts. Meteorol Appl 5(01):1–16
137. Codo M, Rico-Ramirez M (2018) Ensemble radar-based rainfall forecasts for urban hydrological applications. Geosciences 8(8):297
138. Foresti L, Seed A (2014) The effect of flow and orography on the spatial distribution of the very short-term predictability of rainfall from composite radar images. Hydrol Earth Syst Sci 18(11):4671–4686
139. Seed A (2003) A dynamic and spatial scaling approach to advection forecasting. J Appl Meteorol 42(3):381–388
140. Berenguer M, Sempere-Torres D, Pegram GG (2011) Sbmcast – an ensemble nowcasting technique to assess the uncertainty in rainfall forecasts by lagrangian extrapolation. J Hydrol 404(3):226–240
141. Bowler NE, Pierce CE, Seed AW (2006) STEPS: a probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. Q J R Meteorol Soc 132(620):2127–2155
142. Simonin D, Pierce C, Roberts N, Ballard SP, Li Z (2017) Performance of met office hourly cycling nwp-based nowcasting for precipitation forecasts. Q J R Meteorol Soc 143 (708):2862–2873
143. Creutin J-D, Borga M (2003) Radar hydrology modifies the monitoring of flash-flood hazard. Hydrol Process 17(7):1453–1456
144. Delrieu G, Braud I, Berne A, Borga M, Boudevillain B, Fabry F, Freer J, Gaume E, Nakakita E, Seed A et al (2009) Weather radar and hydrology. Adv Water Resour 32 (7):969–974
145. Moore RJ, Cole SJ, Illingworth AJ (2012) Weather radar and hydrology, vol 351. International Association of Hydrological Sciences, Melbourne
146. Cole SJ, Moore RJ (2008) Hydrological modelling using raingauge-and radar-based estimators of areal rainfall. J Hydrol 358(3):159–181
147. Price D, Hudson K, Boyce G, Schellekens J, Moore RJ, Clark P, Harrison T, Connolly E, Pilling C (2012) Operational use of a grid-based model for flood forecasting. Proc Inst Civ Eng 165(2):65–77
148. Collier C (2007) Flash flood forecasting: what are the limits of predictability? Q J R Meteorol Soc 133(622):3–23
149. Austin G, Austin L (1974) The use of radar in urban hydrology. J Hydrol 22(1–2):131–142
150. Goormans T, Willems P (2012) Using local weather radar data for sewer system modeling: case study in Flanders, Belgium. J Hydrol Eng 18(2):269–278
151. Schellart A, Liguori S, Krämer S, Saul A, Rico-Ramirez MA (2014) Comparing quantitative precipitation forecast methods for prediction of sewer flows in a small urban area. Hydrol Sci J 59(7):1418–1436
152. Berndtsson R, Niemczynowicz J (1988) Spatial and temporal scales in rainfall analysis – some aspects and future perspectives. J Hydrol 100(1):293–313
153. Ogden F, Sharif H, Senarath S, Smith J, Baeck M, Richardson J (2000) Hydrologic analysis of the Fort Collins, Colorado, flash flood of 1997. J Hydrol 228(1):82–100
154. Berne A, Delrieu G, Creutin J-D, Obled C (2004) Temporal and spatial resolution of rainfall measurements required for urban hydrology. J Hydrol 299(3):166–179
155. Ochoa-Rodriguez S, Wang L-P, Gires A, Pina RD, Reinoso-Rondinel R, Bruni G, Ichiba A, Gaitan S, Cristiano E, van Assel J, Kroll S, Murl-Tuyls D, Tisserand B, Schertzer D, Tchiguirinskaia I, Onof C, Willems P, ten Veldhuis M-C (2015) Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: a multi-catchment investigation. J Hydrol 531:389–407
156. Anagnostou EN, Krajewski WF (1999) Real-time radar rainfall estimation. Part I: algorithm formulation. J Atmos Ocean Technol 16(2):189–197

Satellite Remote Sensing of Soil Moisture for Hydrological Applications: A Review of Issues to Be Solved

Lu Zhuo

Contents

Abstract Accurate soil moisture indicator is critically important for hydrological applications such as water resource management and hydrological modelling. Modern satellite remote sensing has shown a huge potential for providing soil moisture measurements at a large scale. However its effective utilisation in the aforementioned areas still needs comprehensive research. This chapter focuses on exploring the advances and potential issues in the current application of satellite soil moisture observations in hydrological modelling. It has been proposed that hydrological application of soil moisture data requires the data relevant to hydrology. In order to meet the requirement, the following two research tasks are suggested: the first is to carry out comprehensive assessments of satellite soil moisture observations for hydrological modelling, not merely based on evaluations against point-based in situ measurements; the second is that a soil moisture product (e.g. soil moisture deficit) directly applicable to hydrological modelling should

L. Zhuo (*)
WEMRC, Department of Civil Engineering, University of Bristol, Bristol, UK
e-mail: lu.zhuo@bristol.ac.uk

Andrea Scozzari, Steve Mounce, Dawei Han, Francesco Soldovieri, and Dimitri Solomatine (eds.), ICT for Smart Water Systems: Measurements and Data Science, Hdb Env Chem (2021) 102: 259–282, https://doi.org/10.1007/698_2019_394,
© Springer Nature Switzerland AG 2019, Published online: 28 September 2019

be developed. Only fully accomplishing these two steps will push forward the utilisation of satellite soil moisture in hydrological modelling to a greater extent.

Keywords Flood forecasting, Hydrological modelling, Review, Satellite remote sensing, Soil moisture, Soil moisture and ocean salinity (SMOS)

1 Introduction

Although soil moisture comprises only 0.01% of the total amount of water on the Earth [1], the existence of soil moisture is significant for many application areas such as agriculture, meteorology, climate investigations and natural hazards predictions. In hydrology, soil moisture is a significant state variable in real-time flood forecasting [2]. Over the past decades, numerous hydrological models have been developed, representing more or less accurately the main hydrological processes involved at a catchment scale [3]. The challenge in forecasting floods in a reliable way stems mainly from the error accumulation of the models, particularly during unusual hydrological events and after a long period of dryness. Solutions have thus been introduced to enhance flood forecasting by matching the model with the current observations prior to its use in forecasting mode – termed as updating or data assimilation [4]. Since hydrological models are highly sensitive to the state change of the soil moisture [3], a better soil moisture observation over a catchment should improve the forecasting performance via correcting the trajectory of the model [5]. Nevertheless it is very challenging to accurately monitor soil moisture that varies both spatially and temporally. Conventional in situ networks are expensive and impractical in large areas, and they are still too sparse to represent the spatial soil moisture distribution [6–11]. Model-based estimates such as those from land surface models (LSMs) are another source of soil moisture data, but they are uncertain due to imperfect parameterisation, meteorological forcing data and time drift problems (e.g. accumulation of errors) [12–14].

Alternatively, modern satellite remote sensing has shown potential for providing soil moisture measurements at a large scale [15]. However, satellite soil moisture products are calibrated mainly by in situ measurements, so they are not directly relevant to hydrology [8, 16, 17]. Moreover with all orbiting sensors, only the surface layer soil moisture can be acquired. It has been shown in many studies that the soil penetration depth is around 0.1–0.2 times the sensor wavelength, where the longest wavelength is only about 21 cm (L-band, with a penetrating depth ~5 cm) [18, 19]. Conversely operational hydrological models (most often the conceptual hydrological models) consider a much deeper surface soil depth (up to 2 m), which also varies across a catchment.

Clearly there is a mismatch between the satellite-retrieved soil moisture and the hydrological model-simulated soil moisture, which has caused a commensurate issue for the full utilisation of remotely sensed soil moisture products in operational hydrology. Although many studies have been carried out on the evaluation of satellite soil moisture observations for hydrological modelling [6–8, 20–22], with some correlations explored between the satellite soil moisture datasets and the hydrological models′ soil moisture state variables, their results have limited success and could be improved further. One possible way is by analysing the fundamental differences between the hydrological model-simulated soil moisture and the satellite-measured soil moisture, so that the satellite observations could be enhanced. The motivation of this study is to review the existing literature and explore the potential issues in current satellite soil moisture application in hydrological modelling, which is topical and timely.

2 Soil Moisture Measuring Methods

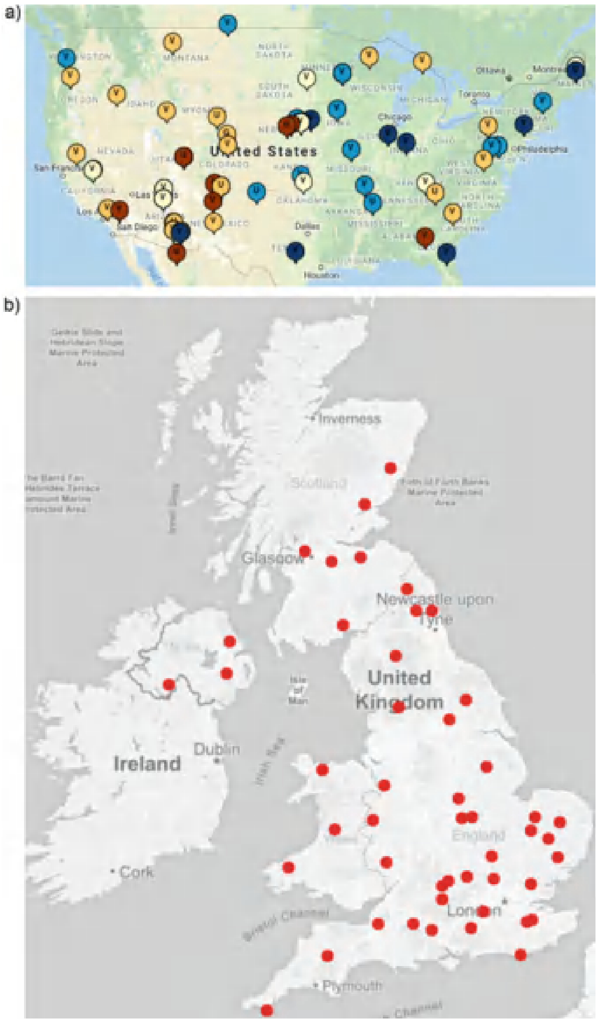
First, it is necessary to give a brief introduction of the existing main soil moisture measuring methods, so that basic knowledge about soil moisture could be better understood. The following are based on two major categories: in situ and satellite remote sensing.

2.1 In Situ Instruments

There are several techniques for in situ soil moisture measurements. The most widely used ones are tensiometer, neutron probe, and time domain reflectometry (TDR). Tensiometer is widely used in irrigation scheduling to help farmers to identify the optimal time for irrigation [23]. It is also useful for plant studies. However this approach is not suitable for sandy soil due to a limited range of bar (0–0.8) and is not electronically stable for automatic operation [24]. For neutron probe, although it is able to measure soil moisture at multi-depths fairly quickly and automatically, it is not capable of giving reliable estimation at shallow depths because some neutrons can escape from the soil surface into the air. Moreover, since the device includes radioactive material, its operation requires extremely strict training and inspection processes [25]. For TDR, the advantages of this method are its high accuracy, fast response, free from radiation hazard and automatic soil moisture estimation [26]. However the calibration of the sensor can be difficult and expensive, and the instrument is easy to corrode [27].

Apart from the aforementioned point-based soil moisture estimation methods, a novel technique capable of measuring area-averaged soil moisture has been introduced. It is called COsmic-ray Soil Moisture Observing System (COSMOS). The working mechanism of COSMOS is it measures the cosmic-ray neutrons above the ground, whose intensity is primarily dependent on soil moisture [28]. One COSMOS sensor can cover a horizontal effective area of about 600 m diameter [29] and the measurement depth from about 12 cm (wet soil) to about 76 cm (dry soil) [28, 30]. The COSMOS has been mainly installed in the USA and the UK as shown in Fig. 1 and elsewhere around the globe (e.g. Australia, Germany, India, etc.). Although the COSMOS networks are only sparsely available globally, there have already been some studies carried out using the data: for example, in [33], it showed COSMOS data could be used to test and diagnose hydro-meteorological models′ performance, indicating the potential for its model assimilation capabilities, as well as demonstrating the data application in remote sensed soil moisture evaluations; [34] explored the usage of COSMOS in landscape monitoring in a mixed agricultural land use system in northeast Austria; and [35] had used the data for hydrological processes investigation at catchment scales and found a good agreement between the COSMOS data and the distributed soil moisture sensor network. Since COSMOS is still under development, it has high uncertainty of soil moisture estimation, for

Fig. 1 COSMOS locations at (a) the USA [31] and (b) the UK [32]



example, the measured neutron intensity can be affected by variations in the atmospheric water vapour that, when not corrected for, ultimately can lead to bias in the derived soil moisture [36]. Therefore correction procedures must be carried out such as those studies by [33, 36–38].

2.2 Satellite Remote Sensing

Compared with in situ methods, satellite remote sensing provides soil moisture observations globally and at larger footprints, so it is more suitable for hydrological usages. A considerable number of studies have shown that near-surface soil moisture (~5 cm) can be measured by many remote sensing techniques including optical, thermal infrared and microwave [39, 40]. The major differences among them are the region of electromagnetic spectrum employed, the power of the corresponding electromagnetic energy, the signal received by the sensors and the relationship between the retrieved signal and the soil moisture [10, 11, 40]. Table 1 lists the advantages and limitations of each technique for surface soil moisture measurement and their characteristics [44].

Only a brief description of each technique is introduced as follows, and interested readers are encouraged to read further details from the references provided.

2.2.1 Optical

Optical satellites measure the reflected radiation of the Sun from Earth′s surface, known as the reflectance [45]. Its correlation with the soil moisture has long been recognised [46]. Although there are a large number of optical sensors currently serving in orbit, relatively fewer studies have been carried out regarding their application in soil moisture assessment [47]. This is partially because the optical sensors can only detect the reflectance or emittance at the top few millimetres of Earth surface. Compared with the longer microwave wavelength, the optical signal is highly affected by cloud contamination and vegetation cover. Furthermore the received soil reflectance is not solely affected by the soil moisture but also influenced by mineral composition, organic matter, soil texture and observation conditions, which makes this technique less popular for soil moisture estimation [42, 48]. Therefore the optical technique is normally applicable only under restricted conditions for soil moisture determination (e.g. with specific soil types, bare soil and climate dominated by clear sky) [47, 49].

2.2.2 Thermal Infrared

Thermal infrared satellites measure Earth radiative temperature, which is then converted to soil moisture either singularly or by combination with the vegetation index information obtained from the optical wavebands (e.g. Normalised Difference

Table 1 Summary of remote sensing techniques for soil moisture measurements

Spectrum domain	Physical processes	Primary information	Merits	Demerits
Optical – visible, near-infrared, short-wave infrared (0.4–2.5 μm)	It is related to surface soil moisture as a function of spectral absorption features; for bare soil, drier soil generally has higher soil reflectance	Soil reflectance	High spatial resolution Wide coverage Multiple satellites available	Limited surface penetration (~1 mm) Limited capability of passing through cloud and attenuated by Earth′s atmosphere Infrequent revisit time Many other noise sources Strongly perturbed by meteorological conditions and vegetation coverage Physical processes not well understood
Thermal infrared (3.5–14 μm)	Soil moisture can increase both specific heat and thermal conductivity of the soil, hence soil temperature varies; for bare soil, fluctuations in land surface temperature are mainly affected by the variations of surface soil moisture	Land surface temperature	High spatial resolution Wide coverage Multiple satellites available Physical processes well understood	Limited surface penetration (~1 mm) Limited capability of passing through cloud and attenuated by Earth′s atmosphere Infrequent revisit time Strongly perturbed by meteorological conditions and vegetation coverage
Microwave: passive (1–30 cm)	Microwave emissivity is related to soil moisture at Earth surface, due to the big dielectric constant difference between dry soil (<5) and water (~80); for bare soil, wetter soil normally shows lower brightness temperature (i.e. less emissivity)	Brightness temperature Dielectric properties	Wide coverage Multiple satellites recently available Low atmospheric noise Moderate surface penetration (up to 5 cm) Physical processes well understood	Low spatial resolution (~ 30 km) Perturbed mainly by surface roughness, vegetation coverage and incidence angle
Microwave: active (1–30 cm)	Based on the empirical relationships that relate the radar measured backscattering	Backscattering coefficient Dielectric properties	High spatial resolution Multiple satellites	Limited swath width Perturbed mainly by surface roughness,

(continued)

Table 1 (continued)

Spectrum domain	Physical processes	Primary information	Merits	Demerits
	coefficient to volumetric soil moisture, which is linked to the dielectric constant difference between dry soil and water; for bare soil, wetter surface soil has higher backscattering coefficient		recently available Low atmospheric noise Moderate surface penetration (up to 5 cm) Physical processes well understood	vegetation coverage and incidence angle

Note: The table is based on [10, 41–43]

Vegetation Index (NDVI) from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite). Since soil water content governs the thermal properties (i.e. the soil thermal conductivity and the soil heat capacity) of the soil, it should be expected that regions with wetter soil are usually cooler during the day and warmer at night [50]. Driven by this concept, a considerable number of studies have shown good accuracy of soil moisture measurements by this technique, such as through the simple thermal inertia approach [51] and the 'Universal Triangle' method [52, 53]. While these approaches are powerful and have thorough physical meanings, they are still restricted by various factors, similar to those in the optical wavebands. Therefore their accuracy varies across time and meteorological conditions (e.g. wind speed, air temperature and humidity) [54, 55].

2.2.3 Passive and Active Microwaves

The primary theory of microwave soil moisture estimation is based on the large contrast between the dielectric properties of water (~80) and dry soil (<5). Therefore when the soil becomes moist, the dielectric constant of the soil-water mixture rises, and this emission fluctuation is recorded by microwave sensors [56, 57]. For passive sensors, the retrieved emission from Earth surface is proportional to the product of surface temperature and surface emissivity, which is commonly referred to as the microwave brightness temperature [58]. For active sensors, a microwave pulse is first sent and then received. The power of the two signals is then compared to determine the backscattering coefficient of the surface, which has been proven to be sensitive to soil moisture [59]. For both sensor types, the measurement efficacy is related to wavelength, where longer wavelengths (>10 cm) penetrate deeper into the soil and have more ability to pass through cloud and some vegetation cover (such as the Soil Moisture and Ocean Salinity (SMOS) satellite with the

L-band wavelength (21 cm), which is able to probe about 5 cm into the ground) [57]. Comparatively, microwave bands have more advantages in soil moisture estimation than other spectral bands. With the modern microwave satellites such as the Special Sensor Microwave/Imager (SSM/I) passive microwave radiometer (19.35–85.5 GHz; [60]), on board the Defense Meteorological Satellite Program (DMSP) series satellites since 1987, the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) (from 6.9 to 89.0 GHz; [61]) which operated on the AQUA satellite between 2002 and 2011, the Advanced Scatterometer (ASCAT) (radar instrument measuring radar backscatter at 5.255 GHz; [62]) on board the Meteorological Operational (METOP) satellite series since 2006, the SMOS (Fig. 2; 1.4 GHz) launched in 2009 [58], the Aquarius (L-band radiometer with 1.413 GHz and scatterometer with 1.26 GHz; [64]) aboard the Argentine Satelite de Aplicaciones Cientificas-D (SAC-D) spacecraft from 2011, and the Soil Moisture Active/Passive mission (SMAP (Fig. 3); 1.20–1.41 GHz; [59]) which was just launched in early 2015, it is anticipated that more advanced soil moisture measurements would be available.

2.2.4 Satellite Missions

Satellites have been monitoring the global soil moisture variations for about 40 years [66], with missions operated by different space agencies globally. The various satellite missions are designed to measure soil moisture at different temporal and spatial scales so that a comprehensive view of Earth′s hydrological processes can be gained [39]. Table 2 gives an overview of the recent satellite missions that have been used for soil moisture monitoring.

Fig. 2 SMOS in orbit [63]

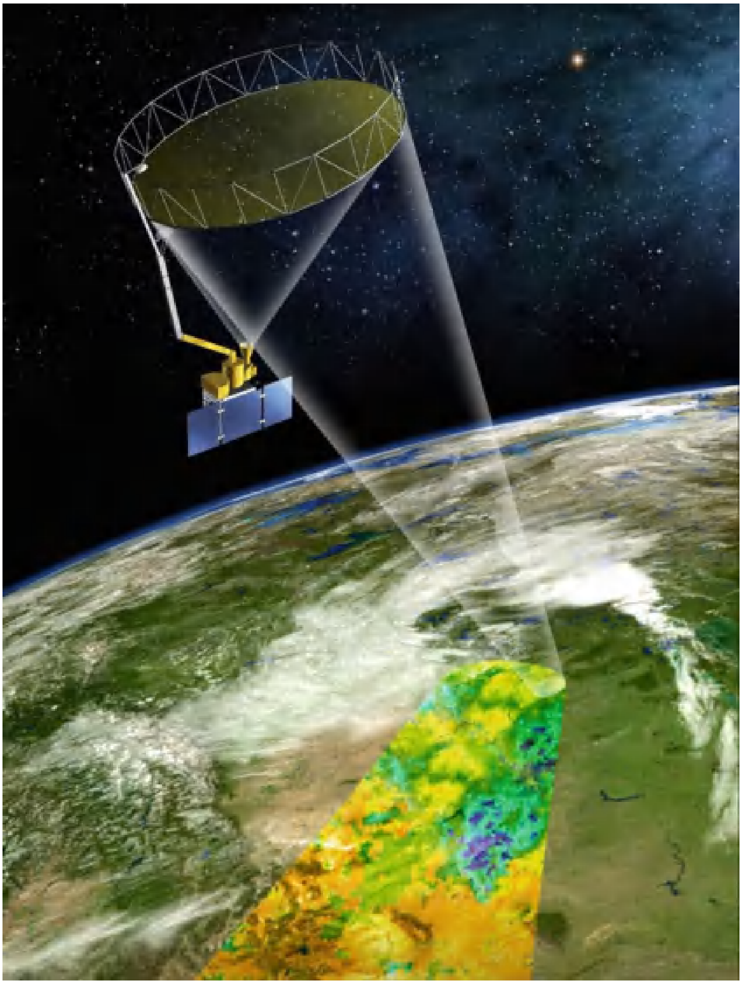


Fig. 3 SMAP in the space [65]

Here a brief description on SMOS and SMAP satellites is given, as those two are dedicated soil moisture missions for long-term soil moisture monitoring.

SMOS is the first dedicated soil moisture satellite using fully polarised passive microwave signals at the L-band, with a spatial resolution of 35–50 km [58, 72]. SMOS retrieves the thermal emission from Earth in both horizontal and vertical polarisations with a wide range of incidence angles from 0° to 60°. It has a Y-shaped antenna structure, which comprises 69 small antennas (a diameter of 16.5 cm) and 4.5-m-long arms to perform interferometry and synthesise an aperture of ~7.5 m [73, 74]. The projection of the synthesised beam on Earth surface is generally presented as an ellipse whose axis ratio and orientation depend on the observed point position [74]. As satellite progresses, any given location on Earth′s

Table 2 Summary of selected satellite missions for soil moisture estimations

	AMSR-E	ASCAT	SMOS	AMSR-2	SMAP
Data type and frequency	Passive microwave (C-band)	Active microwave (C-band)	Passive microwave (L-band)	Passive microwave (8 channels from 6.93–89.0 GHz)	Active and passive microwave (L-band)
Incidence angle (°)	55	25–65	0–55	55	40
Spatial resolution	~50 km	12.5 km, 25 km	35–50 km	15–2,170 km depends on band (GHz) and polarisation	3–40 km
Sampling depth	~0–1 cm	0.5–2 cm	~0–5 cm	~0–1 cm	~0–5 cm
Temporal resolution	<2 days	3 days	1–3 days	Twice a day	3 days
Mission period	2002–2011	Since 2007	Since 2010	Since 2013	Since 2015
Soil moisture accuracy (m^3/m^3)	≥0.04	0.03–0.07	≥0.04	≥0.04	≥0.04
Space agency	Japan Aerospace Exploration Agency (JAXA)	ESA	ESA	JAXA	NASA
Data holding website	1	2	3	4	5
Reference	[61]	[67]	[68]	[69]	[70]

Note: SMAP radar can no longer return data; however the mission continues to produce high-quality soil moisture measurements [71]

1: http://nsidc.org/data

2: http://www.eumetsat.int/website/home/Data/index.html

3: https://earth.esa.int/web/guest/data-access;jsessionid=BC8979A5A504D0E003961303BE9FB4A6.jvm2

4: https://earthdata.nasa.gov/earth-observation-data/near-real-time/download-nrt-data/amsr2-nrt

5: https://nsidc.org/data/smap/smap-data.html

surface is scanned a number of times at various incidence angles, depending on the location with respect to the satellite subtrack: the further away, the fewer the angular acquisitions [72]. SMOS offers a global coverage at the equator crossing the times of 6 a.m. (ascending) and 6 p.m. (descending), both at the local solar time [75].

SMAP is one of the first Earth observation satellites developed by the National Aeronautics and Space Administration (NASA). It is able to monitor global soil moisture at the land surface, as well as distinguish freeze/thaw land surface state. In SMAP, a 40° constant incidence angle is adopted which enables the satellite to scan the whole Earth in 2–3 days with ascending and descending overpasses at 6 p.m. and 6 a.m. (local solar time), respectively [59, 76]. SMAP includes a 6 m diameter conically scanning, deployable mesh reflector antenna which is shared by both the radiometer and the radar. One of the advantages of SMAP over SMOS is it combines an L-band radar with an L-band radiometer integrating the strengths of both active and passive remote sensing for improved soil moisture monitoring. As a result, the resolution of its soil moisture products is supposed to be high (~3 km). However the satellite′s radar instrument stopped working on July 7, 2015, due to a problem in the radar′s high-power amplifier [77]. Currently the SMAP can only retrieve soil moisture products from its radiometer with a resolution around 36 km [78]. Another advantage of SMAP is it includes a special flight hardware which is useful in detecting and filtering radio-frequency interference (RFI), so in many RFI-affected regions, the data loss problem can be prevented [79].

3 Hydrological Evaluation of Satellite Soil Moisture

As aforementioned, satellite remote sensing techniques are a major tool in retrieving soil moisture information on a large scale [11] and are able to provide soil moisture observations globally [58]. In particular, the data acquired by microwave sensors, both active and passive, have been employed to provide detailed soil moisture variability in recent years [80]. Due to SMOS′s longer period of data records, numerous studies have been carried out. Therefore, this paper focuses on discussing the issues related to its hydrological applications only (nevertheless, the discussions are general and applicable to other similar satellites).

SMOS soil moisture is calculated from the multi-angular and fully polarised L-band passive microwave measurements [75]. A number of studies have reported SMOS soil moisture retrieval, downscaling, and its validation against point-based in situ measurements over different regions [20, 68, 81–86]. However, in situ measurements are not directly relevant to hydrological modelling because they cannot be directly placed into a state variable of a hydrological model. On the other hand, some attempts have been made on hydrological evaluations of SMOS soil moisture, such as the ones carry out by [8, 11, 20–22]. The results show the SMOS soil moisture is not accurate enough for direct hydrological modelling usage, and additional work such as using separated algorithms for high- and low-vegetated seasons is needed for improved performance [8].

In comparison with shorter-wavelength satellites such as AMSR-E, SMOS generally provides more accurate soil moisture information. However some studies demonstrate that AMSR-E is actually more accurate than SMOS over certain regions [87–89]. From the temporal availability point, SMOS soil moisture observations are significantly less available than AMSR-E′s, as shown in Fig. 4.

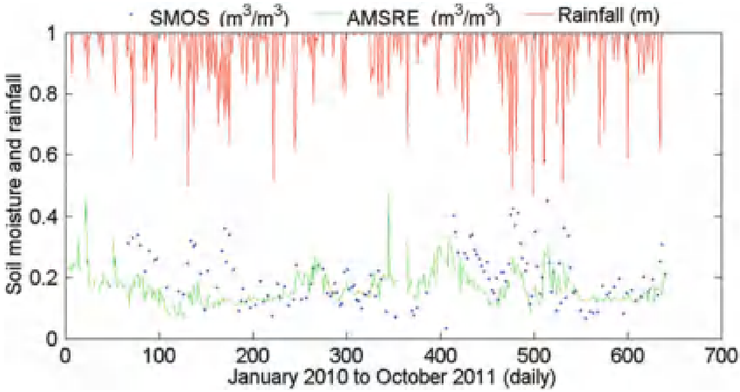


Fig. 4 Time series of SMOS and AMSR-E soil moisture observations with rainfall, in a cropland study area (i.e. Pontiac located in Mid-Illinois of the USA [90])

4 SMOS Descending and Ascending Overpasses

SMOS makes both ascending and descending overpasses; however the performance of those retrievals remains unclear [68, 91–93]. Based on the literature review, previous studies mainly focused on the downscaling, assimilation and evaluation of the SMOS ascending overpass in order to minimise the observation error caused by the daytime soil-drying effect and the impact of vertical soil-vegetation temperature gradients [8, 20, 61, 68, 85]. It is expected that satellite soil moisture measurements are more accurate in the hours near dawn when the soil profile has the most time to return to an equilibrium state from the previous day′s fluxes [94]. Hence, based on this hypothesis, it is more likely to be true that ascending soil moisture measurements would have better performance than their descending counterparts [68]. In addition, based on evaporation demand, it is expected that soil would be wetter at night and drier during the day; in other words, the ascending pass should hold higher soil moisture values than the descending pass if there is no rainfall during the day [83]. However it is found by [22] the SMOS descending orbit shows a stronger potential for improved hydrological predictions in a medium-sized cropland catchment. This outcome contradicts the previous hypothesis from other studies that ascending soil moisture measurements should have better performance than their descending counterparts. Additionally in [22], it is explored that SMOS retrievals from the descending overpass are consistently wetter (about 11.7% by volume) than the ascending retrievals (Fig. 5), which is again not expected. It is explained by the authors that the results could be partly explained by the RFI from the North Warning System radars across northern Canada (formerly called the Distant Early Warning (DEW) Line) [83], which preferentially affects the ascending retrievals in the study area because of the acquisition time and swath area [91]. The RFI increases the

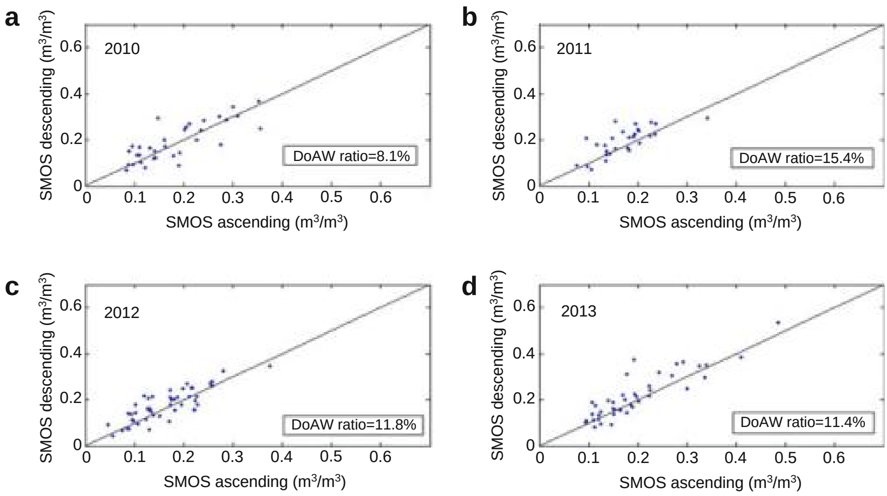


Fig. 5 Scatterplots of the SMOS descending soil moistures against the ascending retrievals, with DoAW ratio presented in each year [22]. Note: DoAW ratio stands for descending-over-ascending-wetting ratio (i.e. the difference between the descending and ascending soil moistures divided by the ascending soil moisture)

brightness temperature and hence artificially reduces the measured soil moisture [83]. But the RFI effect may not be the sole reason; therefore it is encouraged by the community to carry out more research on this topic, with an extended spectrum of catchment types, geographical locations and satellite products, so that more evidential reasons could be revealed and a look-up table might be built.

5 Error Distribution Modelling of SMOS Soil Moisture Measurements

Since satellite soil moisture measurements can be affected by several error sources (e.g. algorithms, sensors and physical processes) [95]. Quantification of such uncertainties is particularly important for applying the soil moisture datasets in real-time flood forecasting systems [96]. More importantly, this is the foundation to the optimal modelling performance in using such soil moisture datasets. Although there are many studies on exploring the uncertainty of satellite soil moisture estimates in hydrological applications, they are mainly represented as summary statistics (such as root-mean-square error (RMSE), Nash-Sutcliffe efficiency (NSE)) [6, 8, 16, 20, 21, 81–85, 97–99], and there is a lack of attention on the error distribution model (such as probability density function, spatial and temporal correlation, nonstationarity).

Proper identification of satellite soil moisture uncertainty in hydrological modelling is relevant for flow ensemble studies (e.g. error propagation). For example, if the observed flow falls outside the forecasted ensembles, then further revisions are required in the formulation of the hydrological model, its states or inputs. However, if the chosen error distribution model is wrong (i.e. flow uncertainty bands become too wide or too narrow), it can lead to false conclusions regarding the adequacy of the input datasets, the hydrological model, and its parameters. Furthermore, understanding the uncertainty features of remotely sensed soil moisture is also useful in controlling and correcting the soil moisture status in a hydrological model after dry periods, so that error accumulation impact can be reduced. Therefore, error distribution modelling of satellite soil moisture measurements is vital to the data application in the hydrological community.

A study by [11] has attempted for the first time in modelling satellite soil moisture error distribution in hydrological applications. It uses the SMOS soil moisture product [58] and a hydrological model called Xinanjiang (XAJ) [100] as a case study. In this study four commonly used probability distributions (Gaussian, extreme value (EV), general extreme value (GEV), and logistic) are adopted to describe the uncertainties of satellite soil moisture data, which are extensively evaluated by using the chi-square statistical test and the bootstrapping resampling technique. From the analysed results (Fig. 6), it is concluded that GEV is the best curve in describing the uncertainty of the SMOS soil moisture estimates. During its second-order error distribution modelling, Gaussian is the most suitable curve for describing the uncertainty of the GEV error distribution model. These results are rather useful for satellite soil moisture data assimilation in operational hydrology, because in a hydrological model, the soil moisture input can be described by ensembles with stochastic elements and the usage of error distribution modelling allows us to better understand the system [101, 102]. By analysing the error distribution models of the input dataset, a decision can be made based on a range of possible outcomes instead of a fixed dataset; this is rather important in water resource management [102].

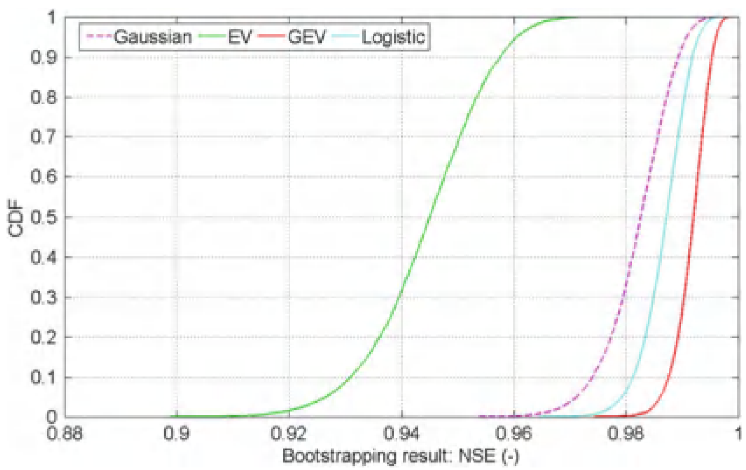


Fig. 6 The performance (Nash-Sutcliffe efficiency) of error distribution modelling for satellite soil moisture applied in hydrological modelling, with the bootstrapping resampling technique [11]

In the future research of this area, more detailed studies such as the spatial and temporal dependence analysis should be conducted. Studies are also needed to consider soil moisture information from other satellite missions over a wider range of catchment conditions with different hydrological models in order to find generalisation patterns of the error distribution models (this is especially important for ungauged catchments).

6 The Need for New Hydrological Soil Moisture Product Development

Although there have been significant investments by various organisations such as the European Space Agency (ESA), NASA and United States Department of Agriculture (USDA) in a wide range of soil moisture observational programs (e.g. satellite missions such as ASCAT, SMOS and SMAP) and ground-based networks such as Soil Climate Analysis Network, US Surface Climate Observing Reference Networks and COSMOS, they are not sufficiently used in hydrology mainly because they are calibrated by in situ soil moisture measurements or airborne retrievals which have significant spatial mismatch (both horizontally and vertically) to catchment scales and are therefore less applicable to hydrological modelling [103].

Similar to many satellites and soil moisture estimation algorithms, SMOS uses the L-band Microwave Emission of the Biosphere (LMEB) model for the data retrieval purpose [104]. LMEB is applied to estimate L-band brightness temperatures (T_bs) for a set of physical parameters, soil composition, and moisture content and vegetation opacity [74]. In order to estimate soil moisture, the simulated T_bs are compared with those measured by SMOS using an iterative process to minimise the difference between them. This approach then requires in situ observation data for soil moisture evaluation [87, 97]. However most areas do not have in situ sensors because they are expensive to set up and impractical to maintain; and they are too sparse for catchment-scale studies [6–10]. Another problem of using this type of method is that by decoupling the effects of soil properties and vegetation cover can significantly reduce its soil moisture accuracy and hence its useful application [105, 106].

In order to retrieve accurate soil wetness information that can be directly used in a hydrological model and avoid aforementioned shortcomings, a need for a data-driven model is desirable, which can effectively link the inputs to the desired output and is not computationally intensive. Works carried out by [7, 21, 99, 107] are good foundations for future hydrological soil moisture product development. For example, in [99], three artificial intelligence techniques along with the generalised linear model are used to improve the spatial resolution of the SMOS-derived soil moisture. The land surface temperature data retrieved from MODIS satellite is used for the data downscaling, and SMD data calculated from a hydrological model called

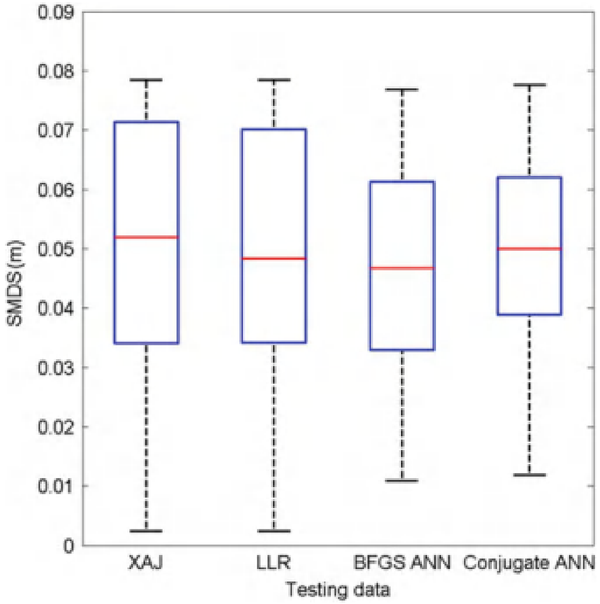


Fig. 7 The statistical plot of the hydrological model (XAJ) simulated Soil Moisture Deficit to Saturation (SMDS) and the algorithms estimated [17]

probability-distributed model (PDM) is selected for performance evaluation. The results show that all the downscaled soil moisture products surpass the original SMOS soil moisture estimation, which are more useful for hydrological modelling. Another study carried out by [17] describes a new approach to estimate hydrological soil moisture variables directly from the SMOS multi-angle brightness temperatures with both the horizontal and vertical polarisations. Local linear regression (LLR) and artificial neural networks (ANNs) with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) neural network training algorithm and the conjugate gradient training algorithm models are applied. The overall results indicate that the proposed methods especially the LLR approach (Fig. 7) have a huge potential to provide hydrologists with valuable information on the application of satellite brightness temperature for hydrological soil moisture estimation. It is noted that although many papers have been published using various data fusion techniques for soil moisture estimations [105, 108–113], the products they produced are not directly applicable for hydrological modelling. Therefore, they are not discussed in detail in this paper.

7 Discussion and Conclusions

Soil moisture is a key element in the hydrological cycle, regulating evapotranspiration, precipitation infiltration and overland flow. For hydrological applications, the antecedent wetness condition of a catchment is among the most significant factors for accurate flow generation processes. Additionally, an operational system requires reliable hydrological soil moisture state updates to reduce the time drift problem.

Through reviewing various satellite techniques for soil moisture estimation, hydrological evaluation of satellite soil moisture and building satellite soil moisture error model, a major problem in satellite soil moisture utilisation for hydrological modelling is concluded. It is that current satellite soil moisture products are mainly calibrated by in situ soil moisture observations. As a result, they are not directly relevant to catchment hydrological modelling. Therefore, a soil moisture product that can be directly linked with hydrological models is desired, and more studies about developing new hydrological soil moisture products as described in Sect. 6 are needed. Certainly, there are still many specific challenges remained, for example, currently the soil moisture retrievals are only available at a coarse spatial resolution which could lead to spatial mismatch problems (e.g. over- or under-representation of soil moisture condition of a catchment); therefore accurate retrieval up to a finer resolution may be necessary particularly for those small-sized catchment studies. This could be achieved by downscaling methods using machine learning techniques. Another issue is about the depth mismatch between the hydrological model soil layer and the satellite′s penetration. Currently for most operational hydrological models, it is difficult to have fixed soil depths. On the other hand, satellite-retrieved data also have unfixed soil depth, which depends on many factors such as vegetation and soil roughness. The important thing is to create a linkage between a hydrological model′s soil moisture state variable and the satellite data (e.g. build mathematical relationships). If a good linkage is built, the satellite soil moisture product can then be used effectively in hydrological model′s soil moisture state initiating and updating during real-time flood forecasting. Only breakthrough in those areas will lay a solid foundation for future data assimilation of soil moisture observations in the real-time flood forecasting. Furthermore, it is hoped this study will attract attention from the hydrological community on those problems and encourage more research to solve them in a wide range of geographical and climatic conditions.

References

1. Prigent C, Aires F, Rossow WB, Robock A (2005) Sensitivity of satellite microwave and infrared observations to soil moisture at a global scale: relationship of satellite observations to in situ soil moisture measurements. J Geophys Res D Atmos 110(D7)
2. Ochsner TE, Cosh MH, Cuenca RH, Dorigo WA, Draper CS, Hagimoto Y, Kerr YH, Njoku EG, Small EE, Zreda M (2013) State of the art in large-scale soil moisture monitoring. Soil Sci Soc Am J 77(6):1888–1919
3. Aubert D, Loumagne C, Oudin L (2003) Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall–runoff model. J Hydrol 280(1):145–161
4. Christian RJ (1997) Validation and intercomparison of different updating procedures for real-time forecasting. Nord Hydrol 28(2):65–84
5. Ottlé C, Vidal-Madjar D (1994) Assimilation of soil moisture inferred from infrared remote sensing in a hydrological model over the HAPEX-MOBILHY region. J Hydrol 158(3):241–264

6. Al-Shrafany D, Rico-Ramirez MA, Han D, Bray M (2013) Comparative assessment of soil moisture estimation from land surface model and satellite remote sensing based on catchment water balance. Meteorol Appl 21(3):521–534
7. Srivastava PK, Han D, Rico-Ramirez MA, Al-Shrafany D, Islam T (2013) Data fusion techniques for improving soil moisture deficit using SMOS satellite and WRF-NOAH land surface model. Water Resour Manag 27(15):5069–5087
8. Srivastava PK, Han D, Rico Ramirez MA, Islam T (2013) Appraisal of SMOS soil moisture at a catchment scale in a temperate maritime climate. J Hydrol 498:292–304
9. Walker JP, Willgoose GR, Kalma JD (2004) In situ measurement of soil moisture: a comparison of techniques. J Hydrol 293(1):85–99
10. Wang L, Qu JJ (2009) Satellite remote sensing applications for surface soil moisture monitoring: a review. Front Earth Sci China 3(2):237–247
11. Zhuo L, Dai Q, Islam T, Han D (2016) Error distribution modelling of satellite soil moisture measurements for hydrological applications. Hydrol Process. https://doi.org/10.1002/hyp.10789
12. Ridler ME, Madsen H, Stisen S, Bircher S, Fensholt R (2014) Assimilation of SMOS-derived soil moisture in a fully integrated hydrological and soil-vegetation-atmosphere transfer model in Western Denmark. Water Resour Res 50(11):8962–8981
13. Xia Y, Mitchell K, Ek M, Sheffield J, Cosgrove B, Wood E, Luo L, Alonge C, Wei H, Meng J (2012) Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1 intercomparison and application of model products. J Geophys Res D Atmos 117(D3)
14. Zhuo L, Han D, Dai Q, Islam T, Srivastava PK (2015) Appraisal of NLDAS-2 multi-model simulated soil moistures for hydrological modelling. Water Resour Manag 29(10):3503–3517
15. Engman ET, Chauhan N (1995) Status of microwave soil moisture measurements with remote sensing. Remote Sens Environ 51(1):189–198
16. Al-Shrafany D, Rico-Ramirez MA, Han D (2012) Calibration of roughness parameters using rainfall–runoff water balance for satellite soil moisture retrieval. J Hydrol Eng 17(6):704–714
17. Zhuo L, Han D, Dai Q (2016) Soil moisture deficit estimation using satellite multi-angle brightness temperature. J Hydrol 539:392–405
18. Oh Y (2000) Retrieval of the effective soil moisture contents as a ground truth from natural soil surfaces, geoscience and remote sensing symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 international. IEEE, pp 1702–1704
19. Ulaby FT, Dubois PC, van Zyl J (1996) Radar mapping of surface soil moisture. J Hydrol 184(1):57–84
20. Lacava T, Matgen P, Brocca L, Bittelli M, Pergola N, Moramarco T, Tramutoli V (2012) A first assessment of the SMOS soil moisture product with in situ and modeled data in Italy and Luxembourg. IEEE Trans Geosci Rem Sens 50(5):1612–1622
21. Srivastava PK, Han D, Rico-Ramirez MA, O′Neill P, Islam T, Gupta M (2014) Assessment of SMOS soil moisture retrieval parameters using tau–omega algorithms for soil moisture deficit estimation. J Hydrol 519:574–587
22. Zhuo L, Dai Q, Han D (2015) Evaluation of SMOS soil moisture retrievals over the central United States for hydro-meteorological application. Phys Chem Earth Parts A/B/C 83–84:146–155
23. Hensley D, Deputy J (1999) Using tensiometers for measuring soil water and scheduling irrigation. Landscape 50(10)
24. Zazueta FS, Xin J (1994) Soil moisture sensors. Soil Sci 73:391–401
25. University of California (2015) Neutron moisture meters. Division of Agriculture and Natural Resources. http://ucmanagedrought.ucdavis.edu/PDF/DROUGHT_WEB_NEUTRON_PRB.pdf
26. Jones SB, Wraith JM, Or D (2002) Time domain reflectometry measurement principles and applications. Hydrol Process 16(1):141–153

27. Mukhopadhyay SC (2013) Intelligent sensing, instrumentation and measurements. Time domain reflectometry (TDR) soil moisture sensor, volume 5 of smart sensors, measurement and instrumentation. Springer, Berlin. 175 pp
28. Zreda M, Desilets D, Ferré T, Scott RL (2008) Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. Geophys Res Lett 35(21)
29. Desilets D, Zreda M (2013) Footprint diameter for a cosmic-ray soil moisture probe: theory and Monte Carlo simulations. Water Resour Res 49(6):3566–3575
30. Iwema J, Rosolem R, Baatz R, Wagener T, Bogena H (2015) Investigating temporal field sampling strategies for site-specific calibration of three soil moisture–neutron intensity parameterisation methods. Hydrol Earth Syst Sci Discuss 12(2):2349–2389
31. UA (2018) Location of COSMOS probes. University of Arizona. http://cosmos.hwr.arizona.edu/Probes/probemap.php
32. CEH (2018) COSMOS-UK. https://cosmos.ceh.ac.uk/network
33. Evans J, Ward H, Blake J, Hewitt E, Morrison R, Fry M, Ball L, Doughty L, Libre J, Hitt O (2016) Soil water content in southern England derived from a cosmic-ray soil moisture observing system–COSMOS-UK. Hydrol Process 30(26):4987–4999
34. Franz TE, Wahbi A, Vreugdenhil M, Weltin G, Heng L, Oismueller M, Strauss P, Dercon G, Desilets D (2016) Using cosmic-ray neutron probes to monitor landscape scale soil water content in mixed land use agricultural systems. Appl Environ Soil Sci 2016:4323742
35. Schreiner-McGraw AP, Vivoni ER, Mascaro G, Franz TE (2016) Closing the water balance with cosmic-ray soil moisture measurements and assessing their relation to evapotranspiration in two semiarid watersheds. Hydrol Earth Syst Sci 20(1)
36. Rosolem R, Shuttleworth WJ, Zreda M, Franz T, Zeng X, Kurc S (2013) The effect of atmospheric water vapor on neutron count in the cosmic-ray soil moisture observing system. J Hydrometeorol 14(5):1659–1671
37. Baatz R, Bogena H, Hendricks Franssen HJ, Huisman J, Montzka C, Vereecken H (2015) An empirical vegetation correction for soil water content quantification using cosmic ray probes. Water Resour Res 51(4):2030–2046
38. Hawdon A, McJannet D, Wallace J (2014) Calibration and correction procedures for cosmic-ray neutron soil moisture probes located across Australia. Water Resour Res 50(6):5029–5043
39. Petropoulos GP, Ireland G, Barrett B (2015) Surface soil moisture retrievals from remote sensing: current status, products & future trends. Phys Chem Earth Parts A/B/C 83–84:36–56. https://doi.org/10.1016/j.pce.2015.02.009
40. Walker JP (1999) Estimating soil moisture profile dynamics from near-surface soil moisture measurements and standard meteorological data. PhD thesis, The University of Newcastle, New South Wales, Australia
41. Engman ET, Gurney RJ (1991) Remote sensing in hydrology. Chapman and Hall Ltd., London
42. Moran MS, Peters-Lidard CD, Watts JM, McElroy S (2004) Estimating soil moisture at the watershed scale with satellite-based radar and land surface models. Can J Remote Sens 30(5):805–826
43. Zhuo L, Han D (2016) Could operational hydrological models be made compatible with satellite soil moisture observations? Hydrol Process. https://doi.org/10.1002/hyp.10804
44. Zhuo L, Dai Q, Islam T, Han D (2016) Error distribution modelling of satellite soil moisture measurements for hydrological applications. Hydrol Process 30(13):2223–2236
45. Sadeghi A, Hancock G, Waite W, Scott H, Rand J (1984) Microwave measurements of moisture distributions in the upper soil profile. Water Resour Res 20(7):927–934
46. Ångström A (1925) The albedo of various surfaces of ground. Geogr Ann 7:323–342. https://doi.org/10.2307/519495
47. Muller E, Decamps H (2001) Modeling soil moisture–reflectance. Remote Sens Environ 76(2):173–180
48. Musick HB, Pelletier RE (1988) Response to soil moisture of spectral indexes derived from bidirectional reflectance in thematic mapper wavebands. Remote Sens Environ 25(2):167–184

49. Wang L, Qu J, Zhang S, Hao X, Dasgupta S (2007) Soil moisture estimation using MODIS and ground measurements in eastern China. Int J Remote Sens 28(6):1413–1418
50. Van D, Engman E (1985) Partial area hydrology and remote sensing. J Hydrol 81(3):211–251
51. Schmugge T (1978) Remote sensing of surface soil moisture. J Appl Meteorol 17(10):1549–1557
52. Carlson TN, Gillies RR, Perry EM (1994) A method to make use of thermal infrared temperature and NDVI measurements to infer surface soil water content and fractional vegetation cover. Remote Sens Rev 9(1–2):161–173
53. Gillies R, Kustas W, Humes K (1997) A verification of the 'triangle' method for obtaining surface soil water content and energy fluxes from remote measurements of the normalized difference vegetation index (NDVI) and surface e. Int J Remote Sens 18(15):3145–3166
54. Czajkowski KP, Goward SN, Stadler SJ, Walz A (2000) Thermal remote sensing of near surface environmental variables: application over the Oklahoma Mesonet. Prof Geogr 52(2):345–357
55. Smith R, Choudhury B (1991) Analysis of normalized difference and surface temperature observations over southeastern Australia. Remote Sens (Basel) 12(10):2021–2044
56. Dobson MC, Ulaby FT, Hallikainen MT, El-Rayes MA (1985) Microwave dielectric behavior of wet soil-part II: dielectric mixing models. IEEE Trans Geosci Rem Sens GE-23(1):35–46
57. Njoku EG, Kong JA (1977) Theory for passive microwave remote sensing of near-surface soil moisture. J Geophys Res 82(20):3108–3118
58. Kerr YH, Waldteufel P, Wigneron J-P, Martinuzzi J, Font J, Berger M (2001) Soil moisture retrieval from space: the soil moisture and ocean salinity (SMOS) mission. IEEE Trans Geosci Rem Sens 39(8):1729–1735
59. Entekhabi D, Njoku EG, Neill PE, Kellogg KH, Crow WT, Edelstein WN, Entin JK, Goodman SD, Jackson TJ, Johnson J (2010) The soil moisture active passive (SMAP) mission. Proc IEEE 98(5):704–716
60. Paloscia S, Macelloni G, Santi E, Koike T (2001) A multifrequency algorithm for the retrieval of soil moisture on a large scale using microwave data from SMMR and SSM/I satellites. IEEE Trans Geosci Remote Sens 39(8):1655–1661
61. Njoku EG, Jackson TJ, Lakshmi V, Chan TK, Nghiem SV (2003) Soil moisture retrieval from AMSR-E. IEEE Trans Geosci Rem Sens 41(2):215–229
62. Wagner W, Hahn S, Kidd R, Melzer T, Bartalis Z, Hasenauer S, Figa-Saldaña J, de Rosnay P, Jann A, Schneider S (2013) The ASCAT soil moisture product: a review of its specifications, validation results, and emerging applications. Meteorol Z 22(1):5–33
63. ESA (2009) ESA space in images, SMOS IN ORBIT. http://www.esa.int/spaceinimages/Images/2009/09/SMOS_in_orbit3
64. Bindlish R, Jackson T, Cosh M, Zhao T, O'neill P (2015) Global soil moisture from the Aquarius/SAC-D satellite: description and initial assessment. IEEE Geosci Remote Sens Lett 12(5):923–927
65. NASA (2015) NASA opens social media accreditation for SMAP mission launch. http://www.nasa.gov/smap-nasasocial
66. Wagner W, Naeimi V, Scipal K, de Jeu R, Martínez-Fernández J (2007) Soil moisture from operational meteorological satellites. Hydrgeol J 15(1):121–131
67. Bartalis Z, Wagner W, Naeimi V, Hasenauer S, Scipal K, Bonekamp H, Figa J, Anderson C (2007) Initial soil moisture retrievals from the METOP-A advanced scatterometer (ASCAT). Geophys Res Lett 34(20)
68. Jackson TJ, Bindlish R, Cosh MH, Zhao T, Starks PJ, Bosch DD, Seyfried M, Moran MS, Goodrich DC, Kerr YH (2012) Validation of soil moisture and ocean salinity (SMOS) soil moisture over watershed networks in the US. IEEE Trans Geosci Rem Sens 50(5):1530–1543
69. Cho E, Moon H, Choi M (2015) First assessment of the advanced microwave scanning radiometer 2 (AMSR2) soil moisture contents in northeast Asia. J Meteorol Soc Jpn Ser II 93(1):117–129
70. McNairn H, Jackson TJ, Wiseman G, Belair S, Berg A, Bullock P, Colliander A, Cosh MH, Kim S-B, Magagi R (2015) The soil moisture active passive validation experiment 2012

(SMAPVEX12): Prelaunch calibration and validation of the SMAP soil moisture algorithms. IEEE Trans Geosci Rem Sens 53(5):2784–2801
71. NASA JPL (2015) NASA soil moisture radar ends operations, Mission Science Continues. NASA. http://www.jpl.nasa.gov/news/news.php?feature=4710
72. Kerr YH, Waldteufel P, Wigneron J-P, Delwart S, Cabot FO, Boutin J, Escorihuela M-J, Font J, Reul N, Gruhier C (2010) The SMOS mission: new tool for monitoring key elements of the global water cycle. Proc IEEE 98(5):666–687
73. McMullan K, Brown M, Martín-Neira M, Rits W, Ekholm S, Marti J, Lemanczyk J (2008) SMOS: the payload. IEEE Trans Geosci Rem Sens 46(3):594–605
74. Rodriguez-Fernandez N, Aires F, Richaume P, Kerr YH, Prigent C, Kolassa J, Cabot F, Jiménez C, Mahmoodi A, Drusch M (2015) Soil moisture retrieval using neural networks: application to SMOS. IEEE Trans Geosci Rem Sens 53(11)
75. Kerr YH, Waldteufel P, Richaume P, Wigneron J-P, Ferrazzoli P, Mahmoodi A, Al Bitar A, Cabot F, Gruhier C, Juglea SE (2012) The SMOS soil moisture retrieval algorithm. IEEE Trans Geosci Rem Sens 50(5):1384–1403
76. Piepmeier J, Mohammed P, De Amici G, Kim E, Peng J, Ruf C, Hanna M, Yueh S, Entekhabi D (2016) Soil moisture active passive (SMAP) project algorithm theoretical basis document SMAP L1B radiometer data product: L1B_TB
77. Chan SK, Bindlish R, O'Neill PE, Njoku E, Jackson T, Colliander A, Chen F, Burgin M, Dunbar S, Piepmeier J (2016) Assessment of the SMAP passive soil moisture product. IEEE Trans Geosci Remote Sens 54(8):4994–5007
78. Al-Yaari A, Wigneron J-P, Kerr Y, Rodriguez-Fernandez N, O'Neill P, Jackson T, De Lannoy G, Al Bitar A, Mialon A, Richaume P (2017) Evaluating soil moisture retrievals from ESA′s SMOS and NASA′s SMAP brightness temperature datasets. Remote Sens Environ 193:257–273
79. Mohammed PN, Aksoy M, Piepmeier JR, Johnson JT, Bringer A (2016) SMAP L-band microwave radiometer: RFI mitigation prelaunch analysis and first year on-orbit observations. IEEE Trans Geosci Remote Sens 54(10):6035–6047
80. Calvet J-C, Wigneron J-P, Walker J, Karbou F, Chanzy A, Albergel C (2011) Sensitivity of passive microwave observations to soil moisture and vegetation water content: L-band to W-band. IEEE Trans Geosci Rem Sens 49(4):1190–1199
81. Albergel C, de Rosnay P, Gruhier C, Muñoz-Sabater J, Hasenauer S, Isaksen L, Kerr Y, Wagner W (2012) Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. Remote Sens Environ 118:215–226
82. Albergel C, Zakharova E, Calvet J-C, Zribi M, Pardé M, Wigneron J-P, Novello N, Kerr Y, Mialon A, Fritz N-E-D (2011) A first assessment of the SMOS data in southwestern France using in situ and airborne soil moisture estimates: the CAROLS airborne campaign. Remote Sens Environ 115(10):2718–2728
83. Collow TW, Robock A, Basara JB, Illston BG (2012) Evaluation of SMOS retrievals of soil moisture over the central United States with currently available in situ observations. J Geophys Res Atmos 117(D9)
84. Panciera R, Walker JP, Kalma JD, Kim EJ, Hacker JM, Merlin O, Berger M, Skou N (2008) The NAFE'05/CoSMOS data set: toward SMOS soil moisture retrieval, downscaling, and assimilation. IEEE Trans Geosci Rem Sens 46(3):736–745
85. Piles M, Camps A, Vall-Llossera M, Corbella I, Panciera R, Rudiger C, Kerr YH, Walker J (2011) Downscaling SMOS-derived soil moisture using MODIS visible/infrared data. IEEE Trans Geosci Rem Sens 49(9):3156–3166
86. Singh VP (2012) Computer models of watershed hydrology. Water Resources Publications, Littleton
87. Al-Yaari A, Wigneron J-P, Ducharne A, Kerr Y, de Rosnay P, de Jeu R, Govind A, Al Bitar A, Albergel C, Muñoz-Sabater J (2014) Global-scale evaluation of two satellite-based passive microwave soil moisture datasets (SMOS and AMSR-E) with respect to Land Data Assimilation System estimates. Remote Sens Environ 149:181–195

88. Louvet S, Pellarin T, Al Bitar A, Cappelaere B, Galle S, Grippa M, Gruhier C, Kerr Y, Lebel T, Mialon A (2015) SMOS soil moisture product evaluation over West-Africa from local to regional scale. Remote Sens Environ 156:383–394
89. Wanders N, Karssenberg D, Bierkens M, Parinussa R, de Jeu R, van Dam J, de Jong S (2012) Observation uncertainty of satellite soil moisture products determined with physically-based modeling. Remote Sens Environ 127:341–356
90. Zhuo L, Han D (2017) Hydrological evaluation of satellite soil moisture data in two basins of different climate and vegetation density conditions. Adv Meteorol 2017. https://doi.org/10.1155/2017/1086456
91. Dente L, Su Z, Wen J (2012) Validation of SMOS soil moisture products over the maqu and twente regions. Sensors 12(8):9965–9986
92. Rowlandson TL, Hornbuckle BK, Bramer LM, Patton JC, Logsdon SD (2012) Comparisons of evening and morning SMOS passes over the Midwest United States. IEEE Trans Geosci Rem Sens 50(5):1544–1555
93. Sanchez N, Martinez-Fernandez J, Scaini A, Perez-Gutierrez C (2012) Validation of the SMOS L2 soil moisture data in the REMEDHUS network (Spain). IEEE Trans Geosci Rem Sens 50(5):1602–1611
94. Jackson TJ (1980) Profile soil moisture from surface measurements. J Irrig Drain Div 106(2):81–92
95. Dorigo W, Scipal K, Parinussa R, Liu Y, Wagner W, De Jeu R, Naeimi V (2010) Error characterisation of global active and passive microwave soil moisture datasets. Hydrol Earth Syst Sci 14(12):2605–2616
96. Brocca L, Melone F, Moramarco T, Wagner W, Naeimi V, Bartalis Z, Hasenauer S (2010) Improving runoff prediction through the assimilation of the ASCAT soil moisture product. Hydrol Earth Syst Sci Discuss 7(4):4113–4144
97. Al-Bitar A, Leroux D, Kerr YH, Merlin O, Richaume P, Sahoo A, Wood EF (2012) Evaluation of SMOS soil moisture products over continental US using the SCAN/SNOTEL network. IEEE Trans Geosci Rem Sens 50(5):1572–1586
98. Crow WT, Van Loon E (2006) Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture. J Hydrometeorol 7(3):421–432
99. Srivastava PK, Han D, Ramirez MR, Islam T (2013) Machine learning techniques for downscaling SMOS satellite soil moisture using MODIS land surface temperature for hydrological application. Water Resour Manag 27(8):3127–3144
100. Zhao R (1980) The Xinanjiang model. Hydrological forecasting proceedings Oxford symposium, vol 129, IASH, pp 351–356
101. Beven K (2006) On undermining the science? Hydrol Process 20(14):3141–3146
102. Jain SK, Singh VP (2003) Water resources systems planning and management. Elsevier, Amsterdam
103. Pierdicca N, Pulvirenti L, Bignami C, Ticconi F (2013) Monitoring soil moisture in an agricultural test site using SAR data: design and test of a pre-operational procedure. IEEE J Sel Top Appl Earth Obs Remote Sens 6(3):1199–1210
104. Wigneron J-P, Kerr Y, Waldteufel P, Saleh K, Escorihuela M-J, Richaume P, Ferrazzoli P, De Rosnay P, Gurney R, Calvet J-C (2007) L-band microwave emission of the biosphere (L-MEB) model: description and calibration against experimental data sets over crop fields. Remote Sens Environ 107(4):639–655
105. Ahmad S, Kalra A, Stephen H (2010) Estimating soil moisture using remote sensing data: a machine learning approach. Adv Water Resour 33(1):69–80
106. Woodhouse IH, Hoekman DH (2000) A model-based determination of soil moisture trends in Spain with the ERS-scatterometer. IEEE Trans Geosci Rem Sens 38(4):1783–1793
107. Zhuo L, Han D (2016) Multi-source hydrological soil moisture state estimation using data fusion optimisation. Hydrol Earth Syst Sci Discuss. https://doi.org/10.5194/hess-2016-478
108. Elshorbagy A, Parasuraman K (2008) On the relevance of using artificial neural networks for estimating soil moisture content. J Hydrol 362(1):1–18

109. Gill MK, Asefa T, Kemblowski MW, McKee M (2006) Soil moisture prediction using support vector machines. J Am Water Resour Assoc 42(4):1033–1046
110. Jiang H, Cotton WR (2004) Soil moisture estimation using an artificial neural network: a feasibility study. Can J Remote Sens 30(5):827–839
111. Kashif Gill M, Kemblowski MW, McKee M (2007) Soil moisture data assimilation using support vector machines and ensemble Kalman filter. J Am Water Resour Assoc 43(4):1004–1015
112. Prakash R, Singh D, Pathak NP (2012) A fusion approach to retrieve soil moisture with SAR and optical data. IEEE J Sel Top Appl Earth Obs Remote Sens 5(1):196–206
113. Wagner W, Dorigo W, de Jeu R, Fernandez D, Benveniste J, Haas E, Ertl M (2012) Fusion of active and passive microwave observations to create an essential climate variable data record on soil moisture. ISPRS Ann Photogramm Remote Sens Spat Inf Sci 7:315–321

Spectroscopic Methods for Online Water Quality Monitoring

Joep van den Broeke and Ton Koster

Contents

Abstract Sensors deployed in smart water systems need to meet a number of criteria, first and foremost robustness of performance, autonomous operation and low maintenance. Solid-state, optical sensors are at the forefront in the development of the next generation of sensors for smart water systems. A range of optical sensor technologies is currently in use for (near) real-time water quality analysis and between them can cover most of the relevant quality parameters. Technologies used in the water industry include UV/Vis absorbance, fluorescence and NIR absorbance spectroscopy. Spectroscopic methods with potential for broader application include Raman spectroscopy and laser-induced breakdown spectroscopy. All approaches share the following properties: fully solid-state hardware, no reagents or other consumables required for their operation, and automatic interpretation of the sensor data performed

J. van den Broeke (*)
KWR Water Research Institute, Nieuwegein, The Netherlands
e-mail: joep.van.den.broeke@kwrwater.nl

T. Koster
Enschede, The Netherlands

Andrea Scozzari, Steve Mounce, Dawei Han, Francesco Soldovieri, and Dimitri Solomatine (eds.), ICT for Smart Water Systems: Measurements and Data Science, Hdb Env Chem (2021) 102: 283–314, https://doi.org/10.1007/698_2019_391,
© Springer Nature Switzerland AG 2019, Published online: 31 August 2019

by increasingly advanced chemometric solutions. This chapter provides a brief overview of the fundamentals behind these technologies and reviews their use in water quality monitoring applications.

Keywords Optical sensors, Smart water, Spectroscopy, Water quality

1 Introduction

Understanding and monitoring the quantity and quality of one of the world′s most precious resources, water, are essential. Currently, the most common method of analysis of water quality consists of (grab or composite) sampling followed by laboratory investigation. This approach fails to fully indicate the dynamics of water quality, since it only provides snapshots of specific points in time. Furthermore, due to delays in transportation, sample preparation and analysis, laboratory analysis only reveals a history of water quality, and not its current state.

Management of water and wastewater networks, meeting operational demands and regulatory compliance while simultaneously minimising costs, is becoming increasingly challenging. The water industry is beginning to recognise that further progress will be limited as long as assets are operated independently of each other and retrospectively with regard to changes in load and failures. An integrated approach to water management potentially offers major advantages to the water industry, hence the quickly expanding interest in smart water solutions.

The holistic approach that is at the cornerstone of smart water systems requires greater system knowledge and improved control. The use of real-time control will enable more flexible and efficient use of existing assets and will provide the ability to respond proactively to both short-term changes and longer-term challenges. Examples of areas where smart water solutions can achieve gains include in treatment processes, e.g. to enable more efficient operation, achieve compliance and reduce the carbon footprint; in water distribution systems, e.g. allowing reaction to operational problems and threats to public health and safety and reducing operational costs; in the sewerage network, e.g. to deal with (rain) events in real-time and moving from hydraulic to quality control; and in asset monitoring, e.g. to allow better forecasting and targeting of asset maintenance. For a brief introduction, see Peleg [1]; for a more detailed review, see Owen [2]. Additionally, there is a trend in the industry towards small-scale, distributed water systems [3]; although these systems remove the need for capital-intensive distribution and collection systems, they pose a special challenge to monitoring as there is no skilled workforce to supervise them and smaller volumes mean lower equipment costs are required. Without reliable automation and control solutions, small-scale systems will not be viable.

Real-time information about the water quality and quantity is the basis for smart water solutions. This information is collected by sensors distributed throughout the water network. Not all sensors are suited for use in smart water systems. The deployment of sensors in large numbers, possibly in locations that are difficult to access, such as in underground pipe systems, brings a number of requirements. Amongst the most important technical preconditions are robustness of performance, autonomous operation and low maintenance requirements. The latter means the instruments themselves are stable, e.g. free from drift and other effects that necessitate recalibrations, do not need (replacement of) consumables and are self-maintaining, e.g. that keep themselves clean and compensate for deviations. For a full discussion on the considerations concerning sensor selection and operations, please refer to van den Broeke et al. [4].

2 Spectroscopy

One type of instrument that is particularly suited for such demanding applications is the spectrometer. This chapter will describe the principles and applications of a number of optical methodologies that are in use as water quality sensors. This chapter is not intended as an exhaustive review, but provides a brief introduction in the application of various spectroscopic methods in water monitoring and includes references to more exhaustive texts.

The technologies described in this chapter have been selected because they are purely optical; they analyse the primary interaction of light with a sample matrix and its constituents. No additional aids are used to achieve selectivity or to enhance the signal, apart from sensitive photo detectors to collect the light after its interaction with the sample and subsequent data processing. Although other optical techniques are widely used, e.g. reagent-based photometric methods, the discussion herein focuses on techniques that meet, or have the potential to meet, the requirements set out above for sensors suitable for integration in smart water solutions. Most importantly, these technologies make use of solid-state, long-term stable components, do not require chemicals and are compatible with effective auto-cleaning techniques. Furthermore, with the development of ever smaller electric and optical components, they lend themselves to further miniaturisation and reduced power consumption that will be required for large-scale, autonomous deployment in distributed (water) networks.

The analysis of the interaction of light with matter, incidentally, is the oldest methodology for studying the chemical composition of samples. The history of spectroscopy begins with the publication of the studies on refraction of light by a prism by Isaac Newton in 1672. In this work, Newton proved that white light is composed of light of various colours. Subsequently it became clear that different chemicals absorb light of various colours and that this was a property that could be used to study their concentrations in a matrix. The step towards analysis of water samples was made in 1856 with the development of Nessler′s method, in which ammonia in water reacts with mercuric iodide and potassium, forming a reddish-brown colloid. The colour intensity of the reaction product depends on the initial concentration of ammonia in the sample.

With the development of photomultiplier tubes in the middle of the twentieth century, spectroscopy, especially absorption spectroscopy in the UV, visible and (near) infrared wavelength ranges, became firmly established in analytical laboratories, both for the analysis of whole samples as well as single components after separation by chromatography. By the mid-1990s, the online at-site spectrometer instrument had reached a mature development stage and is since seeing increasing use in real-time quality monitoring and process control. Applications include monitoring feed and product composition and quality in such diverse industries as pharmaceuticals, petrochemistry, food, as well as water quality monitoring. The most commonly used types of spectroscopy in such at-site, sometimes even in situ, devices are UV/Vis absorbance, near-infrared (NIR) absorbance and fluorescence spectroscopy. Further methodologies include refractive index measurement, Raman spectroscopy, laser-induced breakdown spectroscopy (LIBS) and image analysis.

3 Interaction of Light and Matter

All spectroscopic methods rely on the interaction of light with atoms and molecules. The interaction of light and matter can be described by two different models, one assuming light as a wave phenomenon and the other assuming light to consist of particles. The wave approach is most appropriate to describe such interactions as reflection, refraction and interference. For spectroscopic methods, the interaction of light with atoms and molecules can best be described using the particle approach, with the light particles being called photons. The important parameters of a photon are its energy E, wavelength λ and frequency f, which are related according to Eq. (1):

$$E = \frac{hc}{\lambda} = h\,f \tag{1}$$

where h is the Planck constant (6.63×10^{-34} Js). From Eq. (1), it follows that the energy of a photon is proportional to its frequency and reciprocal to its wavelength. For the discussion herein, the most important parameter is the wavelength, often expressed in units of [nm] or [μm]. Another parameter regularly used in spectroscopy is the reciprocal of the wavelength, the wavenumber $\bar{\nu}$, often expressed in [cm^{-1}].

Although the electromagnetic spectrum is broadly divided into eight regions, ranging from highly energetic γ-radiation to radio waves (Fig. 1), spectroscopy of aqueous samples is focused on the ultraviolet (UV) (200–400 nm), visible (Vis) (400–700 nm) and near-infrared (NIR) (750–1,400 nm) domain, with lower wavelengths corresponding to higher photon energies. The primary reasons for the prevalence of these domains in water analysis are the transparency of water to radiation at these wavelengths and the fact that spectrometers using these wavelengths do not require exotic materials or extreme operating conditions.

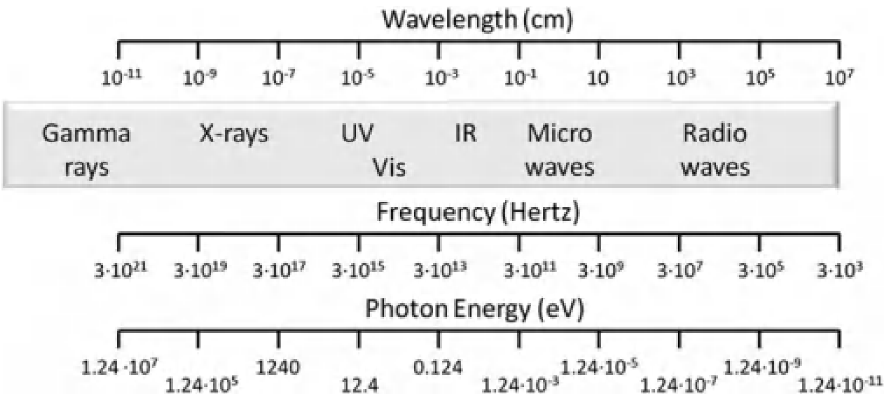


Fig. 1 The electromagnetic spectrum

Table 1 The fundamental light-matter interactions, which form the basis for most spectroscopic methods

Absorption	Matter absorbs light and undergoes a change in energy
Emission	Matter releases radiative energy, e.g. after excitation by light
Elastic scattering and reflection	Interaction between matter and light that changes the direction but not the energy of the photons
Inelastic scattering	Interaction between matter and light that changes the direction as well as the energy of the photons

UV/Vis, (N)IR and fluorescence spectroscopy are all based on absorption, in the case of fluorescence followed by emission (Table 1). Photons can be absorbed if the atom/molecule has energy states that differ by the same amount of energy corresponding to the photon energy. For ultraviolet and visible radiation, absorption of a photon results in one of the molecule′s valence electrons being excited to a higher energy state, while infrared absorption changes the vibrational energy of a molecular bond. The possible energy transitions that a molecule can undergo are discrete and determined by its molecular structure and by its environment in solution. As a result, absorption measurements as a function of the wavelength (reciprocal of the photon energy) result in spectra that are fingerprints for atoms/ molecules. Figure 2 shows a simplified view of a photon′s absorption and subsequent emission.

Although the possible energy transitions for one molecule are very precisely defined, in practice the signals observed in absorption spectroscopy and fluorescence are not as well defined. Whereas spectral lines for individual transitions may be visible in a vacuum, in a solution effects of temperature, inhomogeneities, solute-solvent interactions and in particular hydrogen bonding mean that each molecule may have slightly different vibrational levels. Especially in complex molecules, the possible energy transitions can be so close together that they cannot be distinguished and are observed as one overlapping signal (Fig. 3). The result is a number of closely

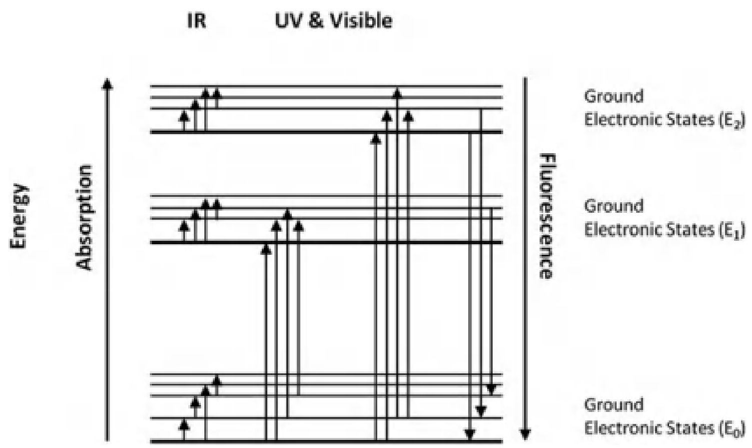


Fig. 2 Transitions between various electronic levels (E0, E1 and E2) and vibrational levels. Absorption of infrared radiation will cause a change in vibrational level, whereas absorption of UV or visible light will cause a change in the electronic energy level. Fluorescence is the falling back to a lower electronic energy level by emission of a photon, which will have higher wavelength (lower energy) than the photon that excited the molecule. The difference in energy is not lost, but converted to heat through vibrational relaxation. In case a molecule does not fluoresce, the excited electron falls back to its original level radiationless. Note that only some possible states are represented, with a typical molecule having many more electronic and vibrational energy levels

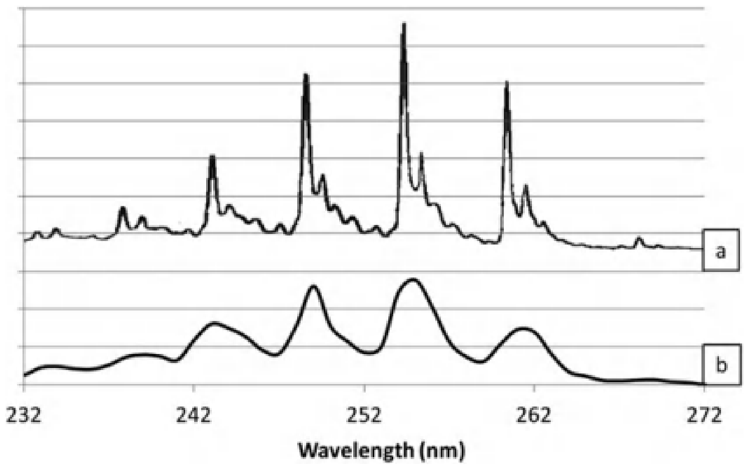


Fig. 3 Example of peak broadening in UV/Vis spectroscopy: (a) benzene in vapour phase, (b) benzene in hexane

spaced absorption bands that merge together to form a single broad absorption band as the spectrometer has insufficient resolution to resolve the individual peaks.

4 Signal Treatment

A water sample, whether natural water, drinking water or wastewater, will contain a wide range of substances in varying concentrations. The signals from all these chemicals are recorded simultaneously and in many cases overlap. In spectroscopy employed for smart water systems, the in situ and real-time nature of the monitoring requires that a sensor automatically extracts relevant information on the composition of the water from the superposed spectra. Effects that interfere with the signal, and therefore must be taken into account when analysing spectral data, include the absorption and scattering of light by particles and/or air bubbles, wear and tear of optical surfaces of the sensor (scratches, fouling, scaling), variations between measurements due to slight changes in the spectrum and intensity of the light source and variations in sensitivity as well as noise in the detector.

4.1 Data Validation

To turn raw data into useful information, modern spectrometers make use of mathematical algorithms. These clean up the signal and use correlations between the light intensity at various wavelengths and analytical parameters to calculate concentrations of specific (groups) of chemicals. Whether spectrometric data is used for the development of an automatic detection algorithm or for real-time autonomous data interpretation by the sensor system, the first step in analysis is always data validation. Because of the amount of data generated, and because of the fact that real-time process control requires making (near) real-time decisions, it is not possible to manually verify whether data are reliable and valid. For smart water systems, automatic validation is critical, ensuring only high-quality measurement results are used in the automated decision-making processes.

Data validation checks whether the sensor system was working properly during the measurement and whether the sample analysed was representative of the medium being monitored. If results are valid and representative, the data is considered reliable. Typical components in the data validation process include sensor status checks, noise analysis and detection of outliers, drift, gaps and steps in data. Tests to validate correct sensor operation include checks against realistic range, detection of constant values and signal-gradient monitoring [5] as well as hardware and software error messages. Once identified, anomalous or missing sensor data might be corrected for, e.g. by interpolation, data smoothing and averaging. More advanced validation methods include data forecasting [6] and the use of distributed algorithms in sensor networks where the results for one particular sensor can be inferred from those of its neighbours [7].

For spectral data averaging is a widely applied method: a set of subsequently recorded spectra is combined to average out fluctuations in the instrument and

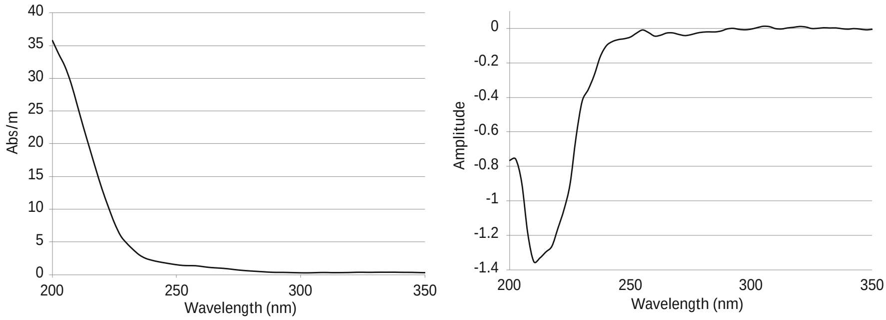


Fig. 4 Raw spectrum (left) and first-order (right) derivative of a UV/Vis absorption spectrum which shows the changes in the slope of the original spectrum

rapidly changing properties of the medium, such as fluctuations in light reflections and scattering by air bubbles and particles.

4.2 Transformations

When dealing with raw data that have a low signal to noise ratio, mathematical transformations can be used to extract useful information. The most widely used operation for removal of measurement noise is the Fourier transform (FT) [8]. This utilises the fact that a signal can be represented as a combination of periodic functions. If the noise and drift on time series data have a significantly different frequency compared to the signal, they can be filtered out. Typically, low-pass filtering is used to suppress noise and high-pass filtering to remove drift. Next to cleaning up the measurement signal, FT is also used in so-called Fourier transform infrared spectroscopy (FT-IR), a specific method of infrared spectroscopy with a very good signal to noise ratio and a high wavelength accuracy.

Another operation frequently used to optimise signal to noise ratios and help with (visual) identification of spectral features is derivatisation [9], as shown in Fig. 4. This removes spectral interferences such as the gentle absorption increase vs. wavelength caused by turbidity in UV/Vis spectra, the fouling of the optical surfaces and light scattering due to air bubbles in the medium, which predominantly result in offsets in the spectra. Most used is the first derivative, which in particular helps with visual identification of features such as shoulders on peaks. Practically, third of higher-order derivatives are not useful as the result of too high noise levels.

4.3 Chemometrics

Chemometrics is the chemical discipline that uses mathematics and statistics to design or select optimal experimental procedures and to obtain knowledge about a chemical system, such as a water matrix. In spectroscopy the primary uses for data analysis algorithms are grouping and classification and modelling relationships between different analytical data. Examples include the classification of samples, such as chemical compounds or materials, based on spectra, and the building of calibration models for calculation of concentrations of chemical constituents in a mixture, e.g. a water sample. The superposition of numerous single substance signals in real-world samples causes cross-sensitivities; when the concentration of an analyte is directly deduced from the signal at one individual wavelength, it will often respond to other, non-related, variations in the matrix. Chemometric models are required to extract information on the specific parameters from the spectra.

The most frequently used method to develop calibration models is indirect modelling using multivariate analysis [10]. In indirect modelling, a calibration model is built from a dataset containing spectral data and the concentrations of the parameters of interest. These concentration values are acquired through separate analytical methods, such as the standard methods for water quality analysis [11]. The multivariate approach has the advantage that interactions between analytes, or between analytes and the matrix, can be accounted for in the calibration model. Also, indirect modelling can deal with any correlations between target analytes. In water applications such correlations are often quite prevalent, such as, for example, the strong correlation between chemical oxygen demand (COD) and total suspended solids (TSS) in wastewater. The first step in the indirect modelling approach is the grouping of analytical data into clusters. An example of a projection method often used in spectroscopy is principal component analysis (PCA) [9]. PCA reduces the dimensionality of a set of variables while conserving the variability within the data as much as possible. In other words, PCA tries to explain the variance-covariance structure of the data using a new coordinate system that is lesser in dimension than the number of original variables; spectra typically consist of 200+ wavelength measurements (variables). The deduced principal components (PC) are new, uncorrelated, orthogonal variables that describe a maximum of variance in the dataset. Another method widely used in spectroscopy is partial least squares (PLS) regression [8]. Working in a similar way as PCA, PLS reduces a complex multidimensional dataset into a smaller number of components accounting for as much variation as possible while also modelling the Y-variables (the reference values). Because PLS is suited for cases with insufficient data to construct a model to predict all variability, it is especially popular in industrial applications where sufficiently complete datasets are often impossible to obtain. Both PCA and PLS are combined with cross-validation procedures and outlier tests to reach both high correlation quality and robustness of the model [8]. The result of the calibration procedure is a function describing how to combine a selection of wavelengths to calculate the target variables. The goodness of fit is described in the recovery

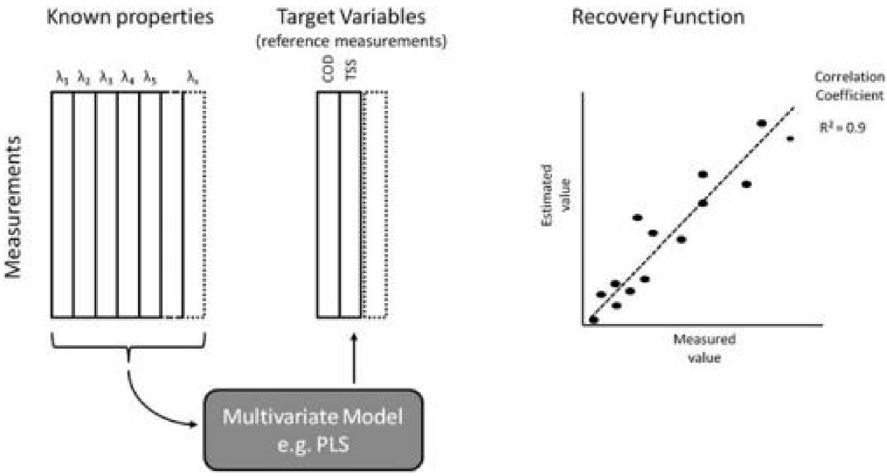


Fig. 5 Scheme of the multivariate calibration procedure

function (Fig. 5), which is obtained by plotting the reference measurements (actual targets) vs. the predicted values (estimated targets).

With increasing sensor complexity, connected sensors and the possible integration of metadata (e.g. weather data, water quantity information or even social media data), more advanced data treatment methodologies are required to find meaningful patterns. Machine learning can be used to find correlations in such big datasets [12]. Supervised machine learning algorithms, in which the computer is presented with example inputs and the desired outputs to these inputs, can be used to develop calibration models in cases where the datasets are too complex to handle by PCA or PLS. Non-supervised learning is used when looking for hidden patterns in the data. Not only can this be used to analyse big datasets but also for feature extraction (e.g. from 3D spectral datasets) and image analysis.

5 In Situ Spectroscopy for Water and Wastewater Analysis

5.1 UV/Vis Absorption Spectroscopy

In the water and wastewater industry, UV/Vis absorption spectroscopy is the most widely applied spectroscopic method for in situ and/or at-site, real-time monitoring. For an in-depth study of the principles of UV/Vis absorption spectroscopy and its application in water and wastewater analysis, Thomas and Burgess [9] provide a detailed treatise and Mesquita et al. [13] a comprehensive review of parameters analysed in wastewater systems.

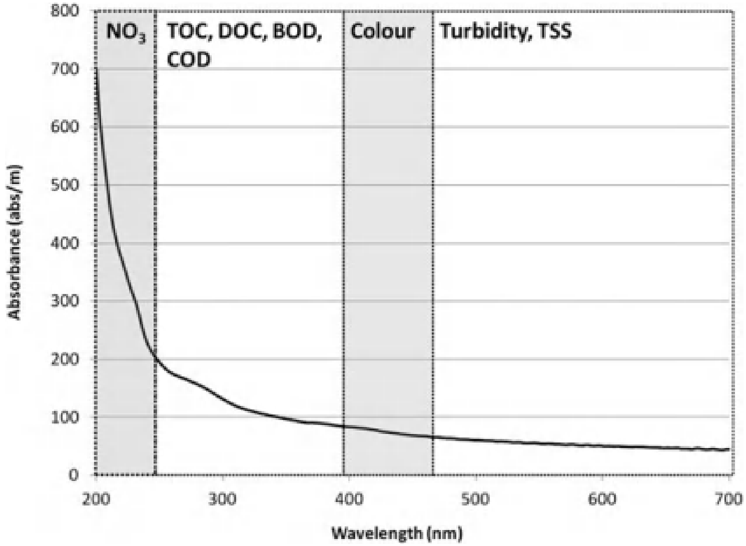


Fig. 6 Typical UV/Vis absorption spectrum of municipal wastewater. Areas used for determination of common UV/Vis parameters are indicated

UV/Vis absorption spectroscopy uses the linear dependence between the absorption measured and the concentration of the analytes, called the Lambert-Beer relationship, to determine parameter concentration values:

$$A = \varepsilon \times c \times L \tag{2}$$

$$A_{total}(\lambda) = \sum_i \varepsilon_i(\lambda) c_i L \tag{3}$$

where A is the absorption value, ε (λ) is the molar extinction coefficient of the molecule or ion at wavelength λ, c is its concentration and L is the distance the light travels through the sample. The absorption at a particular wavelength A_i of multiple species i in a complex matrix can be linearly added to give a total absorption value A_{total} (λ). For the purpose of water quality monitoring, spectrometer devices use a reference spectrum recorded in high-purity water which is subtracted from every measurement. Figure 6 shows an example of a typical UV/Vis absorption spectrum of wastewater after subtraction of the reference spectrum, illustrating the absorption of dissolved and suspended matter in the water.

Online UV/Vis spectroscopy started out as a method to estimate the level of organic matter in water by monitoring the light absorption at 254 nm. The use of 254 nm was the result of initial instruments using a low-pressure mercury light source, which has a strong emission line at 254 nm. The absorption at this wavelength correlates with the concentration of natural organic matter (NOM); as many organic compounds commonly found in water and wastewater (e.g. lignin, tannin, humic acids, fulvic acids, proteins, various aromatic substances) absorb UV light, this signal can be used as a surrogate measure of NOM.

These early devices had serious limitations; apart from the limited lifetime of the mercury lamp, the use of a single wavelength makes the measurement sensitive to cross-interference; a change in the composition of the matrix, such as caused by heavy rainfall, the influx of industrial wastewater or the daily and seasonal changes in composition can cause strong changes in the relationship between UV254 and the true parameter of interest. For improved correlation, devices using multiple wavelengths were developed. The following types of instruments are now widely available:

- Dual wavelength, with a second wavelength used to compensate for turbidity and suspended matter.
- Multiple discrete wavelengths using LEDs with emission spectra at various wavelengths. Specificity is achieved through basic algorithms.
- Full spectral instruments, measuring the entire UV (200–400 nm) or UV/Vis (200–700 nm) range with nanometre resolution allowing for advanced algorithms.

UV/Vis spectrometers are offered both as submersible in situ probes and flow-through devices that can be used to monitor a sidestream. The sensitivity of the instrument depends on the length of its measurement compartment; a longer path length gives higher sensitivity but also a reduced maximum concentration level at which the instrument can operate. Therefore, a device with an optical path that fits the application needs to be used. Typical path lengths available are in the order of 0.5–100 mm.

5.1.1 Sum Organic Parameters

Natural water as well as domestic and industrial wastewater consists of a mixture of various organic substances. Using UV/Vis spectroscopy, the sum of all the absorption signals in the mixture is measured (Eq. 3). The recorded spectra are typically broad and lack characteristic features, because they consist of overlapping spectra of the individual components in the sample. Determination of individual substances is possible only in few applications, with substances with highly characteristic signals in areas of the spectrum where absorption by other components is low. It is, however, possible to accurately calculate sum organic parameters from absorption spectra, even when individual components cannot be identified. Parameters that can reliably be derived from the spectrum include total organic carbon (TOC), chemical oxygen demand (COD) and biological oxygen demand (BOD). As the traditional analytical methods for these parameters are time-consuming, e.g. the 5-day BOD test, or require toxic chemicals such as perchromate, a purely optical method to obtain an accurate indication of their concentration levels in real-time is a powerful tool. Using mathematical turbidity compensation, e.g. based on the description of the optical properties of turbidity and suspended matter [14], it is possible to filter out the

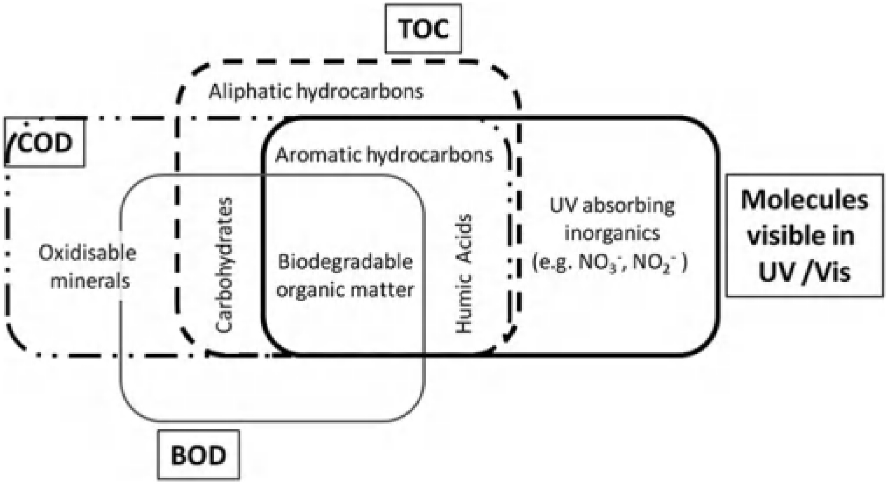


Fig. 7 The relationship between UV response and sum organic parameters

signal of the particulate matter and determine the dissolved organic carbon (DOC) and COD filtered (COD_f) concentrations. Although each of these sum parameters describes a different sub-group of the organic matter, using calibration models each can be estimated on the basis of the spectral information (Fig. 7).

Online monitoring of sum organic parameters is used in water and wastewater treatment, as well as environmental studies. In drinking water, TOC is a relevant quality parameter of water resources, allowing treatment process control such as coagulation [15] as well as optimisation of disinfection and prediction of disinfection by-products formation [16]. In finished drinking water, it is a quality indicator primarily related to aesthetics (taste and odour). In wastewater, COD and BOD are monitored in the wastewater treatment plant (WWTP) influent to determine pollution load, to optimise treatment and to protect the plant from overloading. In the plant effluent, COD can be monitored to determine treatment performance and for consent monitoring. In environmental studies, TOC is of particular interest in studies related to impact of human activities and climate change on lakes and rivers and the release of natural organic matter, e.g. from peat lands and boreal forests [17].

5.1.2 Nitrate and Nitrite

Although real-time UV/Vis spectroscopy is primarily suitable for the monitoring of sum organics, it is also particularly capable of measuring nitrate and nitrite concentrations. Both these ions have a strong absorption signal; in natural waters and drinking water, their signal dominates the 200–230 nm wavelength range, whereas in wastewater it is strong enough to allow reliable derivation from the spectral data using PLS calibration models [10]. Because the spectra of nitrate and nitrite are very similar, in most cases their combined concentration is determined. Using instruments with 1 nm or better spectral resolution, however, it has been possible to obtain concentrations of both ions individually [18]. Nitrate and nitrite monitoring is primarily used to monitor and control nutrient removal processes in WWTPs and to monitor the nutrient load in surface waters.

5.1.3 Colour

Online monitoring of colour is mainly of interest in drinking water treatment. Colour in water is caused by the absorption of visible light by dissolved and colloidal substances and by the scattering of light by suspended particles. Both organic compounds, such as humic acids, and inorganic compounds, such as iron, copper and manganese, can be responsible for colour in water. The colour of natural water is typically yellow to brown. Although colour in itself does not constitute a health risk, high colour is considered aesthetically displeasing, and therefore, limits are defined for colour in drinking water regulations. The most commonly used standard method expresses the colour intensity compared to a solution of a platinum-cobalt complex using Pt-Co units, also referred to as Hazen [11]. A distinction is made between “apparent colour” for samples which include suspended matter and “true colour” for samples that do not include suspended matter (after filtration through a 0.45 μm filter). Both can be determined using online spectrometer instruments, where the apparent colour value is obtained after applying a turbidity correction on the raw spectral data.

5.1.4 Turbidity and Suspended Solids

Non-dissolved matter and colloidal matter cause scattering of the light passing through a water sample. This scattering is referred to as turbidity and is observed as a cloudiness or haziness of the liquid. A number of standard methods have been defined to measure turbidity in water. The most common methods are US EPA method 180.1 and ISO 7027, which measure scattering of light at a 90° angle with a white (tungsten) and infrared (860 nm) light source, respectively. Instead of 90° scattering, UV/Vis spectrometer devices measure the attenuation of light at 180°. The extinction of the signal observed in this instrument layout is caused by the combination of scattering, blocking and shading by particles as well as absorption by dissolved and particulate matter. Because the effect of turbidity on the spectrum is predictable [14] and as in natural water and domestic wastewater the particulates are the prime absorbers at wavelengths longer than 450 nm, turbidity and total suspended solids can be derived from the absorption in the visible range of the spectrum. Turbidity is used to assess the treatability of water and as quality control in drinking water. For example, an increase in turbidity in the distribution network can be an indication of ingress of foreign water, e.g. wastewater, resuspension of sediments or the release of biofilm from pipe walls. In water treatment it can be used to monitor particle carry-over from (sand) filters, helping in the optimisation of the backwashing and cleaning regimes. In wastewater treatment, TSS or the closely related mixed liquor suspended solids (MLSS) are used to monitor sludge concentrations in aeration tanks and to control sludge recirculation and removal.

5.1.5 Other Direct Parameters

Various other substances have a strong and/or characteristic absorption signal that allows direct measurement. Those relevant for water and wastewater treatment include the treatment chemicals ozone (O_3) and permanganate (MnO_4^-) and pollutants such as BTEX (benzene, toluene, xylene), phenol, iron and chromium, which are indicators for contamination of water with hydrocarbons or industrial waste. In sewer systems, hydrogen sulphide (H_2S) can be measured. H_2S is formed under anoxic conditions and is both dangerous (it is highly toxic) and corrosive, leading to biogenic sulphuric acid corrosion of pipe materials.

5.1.6 Indirect Parameters

Not all parameters can be measured directly using UV/Vis spectroscopy. In many cases either the concentrations of the target analytes are too low to detect, or they do not absorb enough light at the wavelengths monitored. In some cases, however, the covariance of the invisible analytes with other, detectable components in the medium allows the building of a calibration model that exploits this relationship. Such calibration models for the indirect measurement of parameters have been reported for a wide range of parameters, including ammonium, total nitrogen and orthophosphate in wastewater, assimilable organic carbon (AOC) in drinking water and bacteria (E. coli) in surface waters and drinking water. As these models rely on a consistent relationship between the visible components and the invisible target analyte, they are often specific for a particular monitoring location, and their validity needs to be checked regularly.

Another type of indirect parameter is the process parameter. In this case, a calibration model is built between (variations) in the spectrum and the (desired) states in a treatment process. These parameters are used as real-time control inputs and facilitate optimisation of water treatment through reduction of chemical and energy consumption while safeguarding or improving treatment effectiveness. Examples of such process parameters include prediction of coagulation dose [15], prediction of chlorine demand and prediction of disinfection by-product formation [19].

One method which has been applied especially for the building of models for process parameters is differential spectroscopy, i.e. the subtraction of spectra measured before and after a process. The resulting “delta-spectrum” reflects the change in the composition of the water as a result of the process [9].

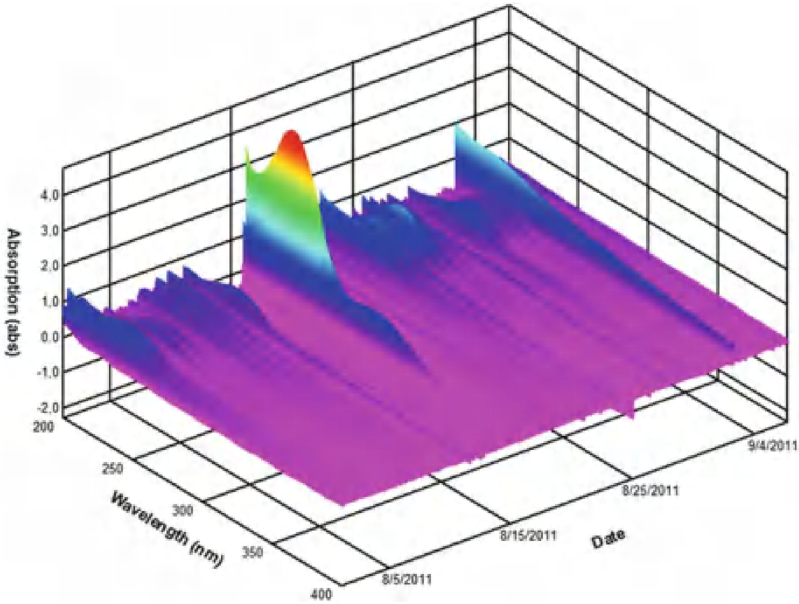


Fig. 8 3D absorption plot of a contamination event as it develops over time

5.1.7 Spectral Fingerprint and Contamination Alarm

Whereas differential spectroscopy for process control evaluates the changes in the water due to a particular (intentional) effect on the water, contamination alarm systems focus on detecting unexpected and often unpredictable changes (Fig. 8). The method used relies on automatic monitoring of the shape of the spectrum ("fingerprint") and comparing this shape to known conditions. When the fingerprint corresponds to a known undesirable state, or in case it does not correspond with a previously seen shape, the spectrometer reports an alarm. Alarm systems are available in many forms; the simplest versions monitor the absorption values at specific wavelengths. More advanced systems treat each spectrum as a vector (the number of dimensions being equal to the number of wavelengths) and perform a nearest neighbour analysis in the vector space [20]. With increasing computing power and storage capacity, more powerful statistical methods capable of handling bigger historical datasets are becoming available embedded in the sensors themselves, increasing real-time event detection capabilities.

The most typical applications for event detection are the monitoring of source waters for drinking water production (detection of harmful contaminations before the intake) and monitoring finished drinking water to safeguard its quality during storage and distribution (provide early warning in case of accidental or intentional contamination). The primary driver in these drinking water applications is protection of the public health. In wastewater early warning systems are used to monitor the influent of the WWTP. Here they can detect peak loads and warn for the presence of harmful (industrial) contaminations, which can cause failure of the biological treatment resulting in costs for plant recovery and economic damages due to noncompliance of the discharged effluent.

5.1.8 Remote Sensing

A recent addition to the field of UV/Vis spectroscopy for water monitoring is remote sensing. Remote sensing is primarily used to determine levels of chlorophyll, suspended solids and chromatic dissolved organic matter (CDOM) in lakes and seas. It uses either long-range satellite-based optical sensors; close-range sensors that can be hand-held, fixed or mounted on drones; or a combination of both. The sensors are radiometers, i.e. they measure reflected natural light. Satellites are used to map the water quality over large areas, whereas smaller land-based systems provide higher-resolution measurements and have the advantage that they can also be operated under cloud cover. These are used independently as well as for validation of optical satellite data. The big advantages of remote sensors are the fact that they work from a distance, removing the need to physically access a location, and their ability to cover a large area where in situ sensors only provide spot sampling.

5.2 Fluorescence Spectroscopy

Another method widely used for water quality monitoring is fluorescence spectroscopy. Light absorbed by a molecule excites it to a higher energy state. Generally, molecules fall back from this higher energy state to their equilibrium energy state through various non-radiative mechanisms (Fig. 2). With fluorescent materials, however, part of the energy is emitted through the emission of a photon. This photon has a longer wavelength (i.e. lower energy) than the excitation energy. The intensity of the light at the emitted wavelength (known as fluorescence) is a measure for the concentration of the fluorescent molecules.

A fluorescence spectrometer is an instrument consisting of the following main components: a light source, a measurement compartment and a detector. The light source sends a beam of light at a wavelength tuned to the molecules of interest through the sample in the measurement compartment. The detector measures the intensity of the fluorescence generated. Generally, a wavelength filter is placed in front of the detector, allowing only the fluorescence to fall on the detector. This filter makes the measurement specific and reduces interference from natural light. Different types of online fluorescence spectrometer instruments are available:

- Single excitation – emission pair (excitation at a specific wavelength, emission measured at a specific wavelength). Often the excitation and/or emission wavelengths can be changed, by the user or the manufacturer, via the selection of different filters and/or different LEDs.

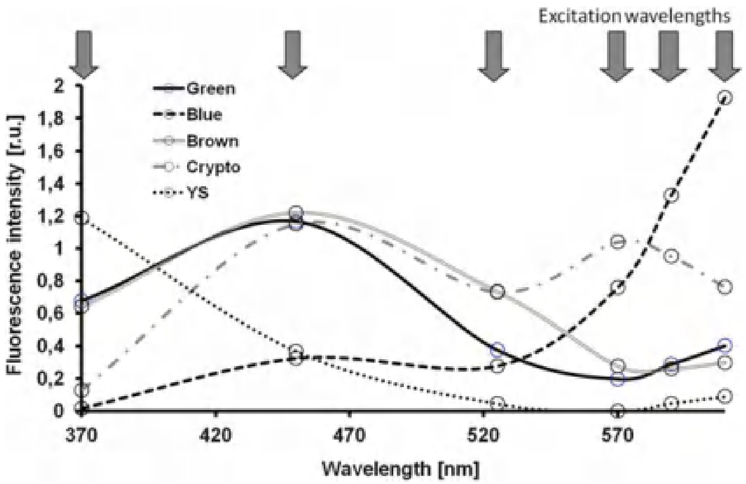


Fig. 9 Intensity of fluorescence emission of various types of algae plotted against the excitation wavelength (source: BBE Moldaenke, www.bbe-moldaenke.de)

  - Suitable for the measurement of one target substance (group).

- Multiple discrete excitation wavelengths (using multiple LEDs as light source in combination with a broadband detector).
  - Suitable for the measurement of multiple target substances or groups; requires more advanced calibration and signal processing for deconvolution of the measurement signal into individual parameters (Fig. 9). An example is fitting of known spectra of individual analytes to the measurement results; the best fit will produce coefficients which can be used to determine the concentration of each component used in the fit.
- Laser-induced fluorescence, in which a laser source excites the target molecules with higher quantum efficiency than achievable with broadband light sources. The stronger emission signal allows for remote sensing or spectral analysis using diode array detectors.
- Fluorescence excitation-emission matrices (EEMs) in which emission spectra are recorded across a range of excitation wavelengths. Typically used for fingerprinting complex mixtures such as natural organic matter in surface waters and wastewaters.

5.2.1 Algal Pigments

In algae, fluorescence is a natural by-product of the photosynthesis process. Although most light captured by the algal pigments is used for the photosynthesis, a small portion leaks out in the form of fluorescence. The intensity of the fluorescence can be used to estimate the concentration of algae and to track developments in the algal population, e.g. as early warning system for algal blooms. The central pigment responsible for photosynthesis is chlorophyll. However, as this is widely present in most photosynthetic organisms, the measurement of chlorophyll alone does only allow for monitoring total algal concentrations, but not classification. As some algae use auxiliary pigments next to chlorophyll A for the collection of photons, using multiple light sources to selectively excite the specific pigments, differentiation between classes becomes possible [21]. An example is the use of red light for the detection of the pigment phycocyanin that is present in cyanobacteria, which are a major cause for toxic algal blooms. Alternatively, the red fluorescence of the accessory pigment phycoerythrin is used to monitor some salt water cyanobacteria. The more advanced instruments attempt to distinguish between cyanobacteria, green algae and diatoms using spectral curve fitting methods [21], although adaptation of the calibration to the algae that are predominant in the waters analysed is often necessary.

It should be noted that in situ measurement of algal pigments does not provide quantitative information about cell concentrations or biovolumes, as signals strongly depend on the algae present, their physiological state and environmental factors such as brightness of the sunlight.

5.2.2 Dissolved Organic Matter

All natural waters as well as drinking waters contain natural organic matter (NOM). Common NOM compounds include proteins, polysaccharides and humic substances, which originate primarily from the breakdown products of plant material. Although NOM does not pose a risk to human health on its own, some NOM compounds are known to react with chlorine and chloramines to produce disinfection by-products (DBPs), some of which are carcinogenic and genotoxic. Monitoring the NOM levels in source waters is used to optimise water treatment and minimise DBP formation. In particular in surface waters with highly fluctuating compositions, e.g. strong seasonal influences, or high sensitivity to runoff during heavy rainfall, monitoring NOM is critical for water treatment performance.

NOM is also receiving attention in research related to climate change, with a particular focus on the release of NOM from boreal forests [22]. In this work, NOM levels and composition are used as indicators for changes in the biochemical cycles and mobilisation of organic matter (e.g. from permafrost) as a result global warming. Furthermore, changes in NOM may require adaptation of the water treatment systems to ensure continued supply of safe drinking water.

Monitoring of NOM is focusing on detection of humic and fulvic acids, both groups of substances with an aromatic character. Due to this aromatic character they are easily detected using fluorescence. This parameter is referred to as fluorescent dissolved organic matter (FDOM) and typically measured using fixed excitation-emission pair filter spectrometer devices. More detailed characterisation can be

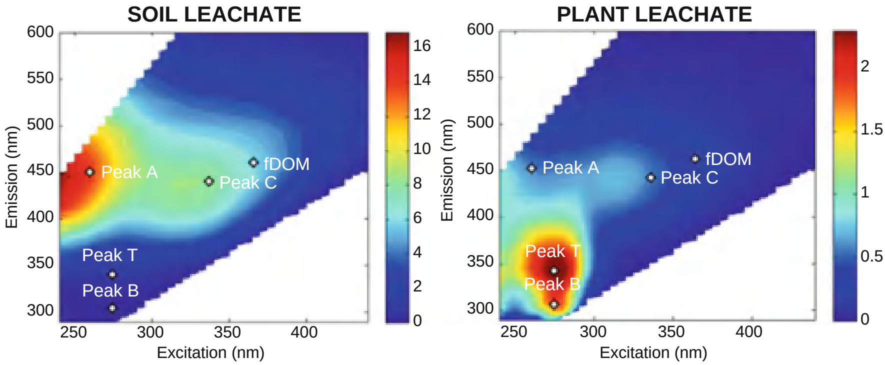


Fig. 10 An example excitation-emission matrix (EEM) showing the general locations of selected fluorescence peaks, with letter indication characteristic peaks (Reproduced from US Geological Survey, https://ca.water.usgs.gov/OMRL/OpticalProperties.html)

performed using EEM spectroscopy, but its use has been limited to measurement of discrete samples in static laboratory spectrometers.

5.2.3 Wastewater

Similarly to natural waters, municipal wastewater contains organic matter. A main role of wastewater treatment plants (WWTPs) is the removal of the majority of this organic matter. Still, the effluent of a WWTP is a complex mixture of dissolved effluent organic matter (dE$_f$OM), containing dissolved natural organic matter, soluble microbial products, endocrine disrupting compounds, pharmaceuticals and personal care product residues, disinfection by-products and more. Although current online sensors are not capable of monitoring the individual components in the wastewater effluent, it is possible to measure sum parameters. In Sect. 5.1.1 monitoring of BOD and COD using UV/Vis absorption spectroscopy was described. Fluorescence spectroscopy can also be used to monitor sum parameters, and using the ratio between characteristic peaks for natural organic matter (humic acids) and nonnatural organic matter (protein like), sewage contamination of surface waters can be detected. This is done by comparing peak C (humic like), with excitation at 350 nm and emission at 420–480 nm, and peak T (protein like), with excitation at 250 nm and emission at 340 nm (Fig. 10) [23]. Peak T has been found to correlate strongly with BOD in rivers and sewer systems. Furthermore, as it correlates strongly with the concentration of tryptophan, an amino acid derived from microbial matter, it is also used to estimate the levels of bacterial contamination in sewage-affected surface waters. For more advanced analysis, EEM can be combined with advanced statistical methods or automated characterisation and quantification of substance/contaminant classes [24].

5.2.4 Oil in Water

Contamination with mineral oils is a recurring issue in surface- and groundwater. Fluorescence sensors can be used to detect such contaminations. Monitoring the removal of mineral oils in industrial wastewater treatment, before discharge to sewer systems or surface waters, is another application. Mineral oils can be monitored with fluorescence as they typically consist of a mixture of aromatic and aliphatic hydrocarbons. It is the aromatic fraction which is detectable. Using fluorescence sensors employing fixed excitation-emission wavelength pairs in the UV range, either monocyclic aromatics (BTX) or polycyclic aromatic hydrocarbons (PAH) are measured. The presence of substances from these groups is used as an indicator for contamination for different types of mineral oil products, e.g. BTX as indicator for refined oil products (such as gasoline, diesel and kerosene), which are rich in these components.

More detailed analysis of the oil type can be done using fluorescence spectroscopy, where a high-resolution spectrometer is used as detector instead of the typical filter-covered photomultiplier or photodiode. Such instruments, often using laser-induced fluorescence to get sufficient signal to noise levels, can distinguish between different oil types. These laser-induced fluorescence (LIF) systems are primarily used in the offshore industry, refineries and applications where contamination with lubricating or cooling oils is common, e.g. industrial or bilge water in ships. The oils in these applications are low in aromatic contents and therefore difficult to detect using LED-induced fluorescence. A further issue with oil and hydrocarbon products is their poor miscibility with water, meaning a submersed sensor may not detect a contamination as is poorly mixed or floats on the surface of the water. To detect floating layers, remote sensing can be used; using a laser mounted above the water, the oil layer is illuminated, and the induced fluorescence is recorded. This allows detection of oil films down to 1 μm.

5.3 NIR

Infrared (IR) spectroscopy is similar to the previously described UV/Vis spectroscopy but uses a lightsource with a higher wavelength, i.e. lower-energy photons. The infrared spectrum is divided into near infrared (NIR) (750–2,500 nm), mid infrared (MIR) (2.5–16 μm) and far infrared (16–1,000 μm). For water quality analysis, the NIR spectrum is most widely used.

Because photon energies in IR are lower than in UV/Vis, it probes a different property of the molecule: instead of exciting electrons to a higher electronic energy level, it changes the vibrational state. IR radiation changes the vibrations of atomic bonds, with the behaviour of specific bonds depending on the atoms in the bond and their environment (both within the larger molecular environment as well as the physical environment, e.g. temperature, solution state). Similar to the electronic levels, these vibrational states are discrete, and an IR spectrum reflects these discrete energy transitions.

In the MIR the absorption bands are well defined, and it is possible to identify specific atomic bonds. However, radiation in the MIR region does not penetrate water well enough for direct measurement of water samples. The lower absorption of NIR radiation by water allows for enough sample thickness to achieve sensitive direct measurements. Absorption bands in the NIR represent the overtones of the fundamental bands in MIR, and these overtones are relatively weak and not clearly delineated; whereas the much sharper signals in the MIR range often allow identification of individual substances, NIR spectra often do not provide such detailed information.

The applications of NIR are therefore found in situations where simple mixtures are analysed. Examples include quality control in food and pharmaceutical industry, where the expected NIR spectrum is known. A NIR sensor, monitoring deviations between the measured spectrum and that of a pure product, is a tool in quality control. Multivariate techniques such as principal component analysis (PCA) and partial least squares (PLS) regression are used for development of calibration models.

In the water industry, NIR has been used for various experimental studies, e.g. for oil in water monitoring or studies on microalgae [25] and extracellular polymeric substances in wastewater processes [13]. Commercial NIR systems for in situ analysis of water samples primarily use reflection instead of transmission spectroscopy; in the transmission mode (e.g. as in UV/Vis spectroscopy), a beam of light passes through a sample, and attenuation of the light by the sample is measured. Reflection spectroscopy, however, analyses the light reflected by the top layer of the matrix, where the makeup of the incident light is modified by the processes of absorption and scattering. It is used for the analysis of contaminations that float on the surface of water, e.g. oil slicks, and on media not transparent enough for transmission spectroscopy. An example is the dewatered sludge from WWTPs [13, 26]. Parameters measured in sludge include dry matter, ammonia and organic matter. The application in sludge monitoring offers the possibility for smart process control, allowing dosing control of polymers used in dewatering and safeguarding optimal composition of sludge for subsequent digestion.

5.4 Further Optical Technologies with Potential for Online Use in Smart Water Systems

UV/Vis absorbance and fluorescence are the most widely used spectroscopic methods in the water industry. NIR also has a well-established place, especially in industrial applications. Next to these technologies, there are optical methods with potential for wider use in the near future. A selection is discussed in this section.

5.4.1 Raman Spectroscopy

As opposed to absorption of fluorescence, Raman spectroscopy studies the scattering of photons by molecules. When a sample is irradiated with monochromatic light, a small proportion of photons (0.0001%) are scattered undergoing a shift in frequency; this inelastic scattering of photons is known as the Raman effect. The difference in the frequencies of the input and scattered light corresponds to the quantised energy levels of the molecule studied. For a more complete treatise on the principles of Raman and instrumentation used, see Li et al. [27].

Raman spectra are more distinct and less overlapped than UV/Vis/NIR absorption spectra. Therefore, Raman is complementary to absorption and fluorescence spectroscopy, as it offers a more selective signal. This allows for classification and in some cases even identification of target substances. It has successfully been applied to determine organic and inorganic analytes, including various metals, in a water matrix with examples including polycyclic aromatic hydrocarbons, pesticides, mercury and arsenic [28, 29]. In the case mixtures are analysed, if the components are known, the relative peak intensities can be used to generate quantitative information about the mixture′s composition. In the case of water applications, however, the number of components may be so high that also Raman spectra overlap, and identification of individual species becomes impossible. Furthermore, the low percentage of photons undergoing inelastic scattering means Raman detection limits are significantly higher than those achieved with UV/Vis/NIR and are insufficient for application in online water quality monitoring as discussed here.

Surface-enhanced Raman (SERS) is a technique where Raman scattering is measured after adsorption of the target molecules on a substrate. This substrate enhances the sensitivity, in some cases allowing analysis of single molecules or single cells. As such it has been used to identify bacterial cells. SERS offers potential for general bacteria classification and pathogen detection [30] and label-free detection of biotoxins [31].

One promising application for Raman spectroscopy is the characterisation of microplastics. Although sample pretreatment remains a challenge, the characterisation of single particles with Raman, e.g. in combination with a flow cytometer, is a promising development [32].

5.4.2 Laser-Induced Breakdown Spectroscopy

Whereas all technologies discussed thus far analyse sample composition on a molecular level, laser-induced breakdown spectroscopy (LIBS) is an elemental analysis technique. In LIBS, a pulsed laser is used to heat a very small spot of the sample to extremely high temperatures (in excess of 30,000 K). As a result, a small amount of the material is transformed into a plasma consisting of free electrons, ions and excited atoms. As the plasma cools and electrons fall down from high-energy to lower-energy atomic orbitals, light is emitted at wavelengths characteristic for the

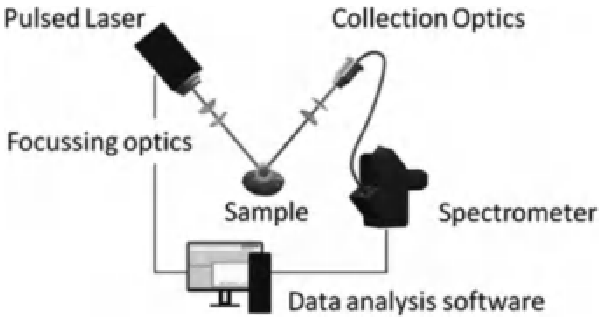


Fig. 11 Schematic representation of a LIBS measurement setup. The sample can be either solid or liquid

elements present in the plasma. The emission spectrum thus allows a rapid determination of the chemical composition of the sampled material (Fig. 11).

LIBS has been used since the 1970s but has only recently become available for rapid analysis outside of laboratory conditions. Although primarily used for the elemental analysis of raw materials (e.g. for production of pharmaceuticals and in mining), recently various applications in environmental and water analysis have also been described [33]. A potentially interesting application is the determination of ionic species, in particular heavy metals, which remain out or reach for other online sensor technologies. Sensitivity of LIBS is typically in the low ppm range, although the use of ultrashort laser pulses (femto seconds) and double pulsed lasers has been reported to offer the potential of sensitivity improvements. The sensitivity of LIBS, however, differs widely between various elements. Furthermore, for certain elements such as arsenic a controlled environment is required as emission lines are located at such short UV wavelengths that water vapour in the atmosphere will absorb all the emitted light.

5.4.3 Refractive Index

Refractive index is an optical property of a material that describes the propagation of light through it. Every substance has a specific refractive index. A mix of substances, such as a water matrix, can be described by a weighted sum of their individual refractive indices; all substances dissolved in a water matrix will contribute to the refractive index of that water matrix. A change in the composition of the matrix will result in a change in its refractive index. As such, monitoring the refractive index of a liquid provides information about the stability of the composition of the liquid; a change in composition will trigger a change in the refractive index. This is particularly useful for the monitoring of matrices with high stability, e.g. drinking water, or high-purity water in semiconductor industry; a rapid change in their composition indicates there is an issue with the water treatment or an intrusion of a contamination.

Although the refractive index itself does not provide information about the nature of the contamination, it can provide an early warning of a quality incident. Moreover, preliminary classification becomes possible when it is combined with one or more other measurements such as electrical conductivity or UV absorption spectroscopy. With refractive index being sensitive to all types of substances, both organic and inorganic, and other technologies being primarily sensitive to either of these groups of substances, a combination of sensors with intelligent data analysis software will be able to classify the nature of the agent [34].

A number of technical solutions exist for the accurate tracking of refractive index changes required to reach ppm and sub-ppm level sensitivity. These include the Mach-Zehnder interferometer and the optical ring resonator. A crucial factor for accurate refractive index measurement by these devices is the temperature compensation of the signal. Both have also been used in combination with surface coatings, such as antibodies, to increase their sensitivity and make them selective to specific molecules and even microorganisms [35].

5.4.4 Image Analysis

With the growing processing power and image analysis algorithms, various systems have now been introduced that perform fully automated scanning the particles of water samples. In such systems a set of optical properties of each particle is measured. This then allows the classification of the individual particles. Applications include counting and analysis of algae in surface water as well as monitoring the total number of bacteria in drinking and wastewater. An approach used in commercially available products for water analysis is 3D scanning with a microscope, collecting images at different depths in a sample and analysing the in-focus objects. Another approach uses flow cytometry to analyse individual particles, collecting scattering and fluorescence spectra of each particle that passes through a laser beam. Both methodologies assess a range of optical parameters for each particle, such as shape, size and optical density, in order to classify them. Although not capable of monitoring hygienic parameters in drinking water, which require single cell detection and identification in large volumes, these systems provide automatic culture- and reagent-free analysis of sum microbiological parameters, e.g. total cell count (TCC) and intact cell count (ICC), based on the characteristics of individual cells. Products of this type have been arriving on the market over the last 5 years and are used to monitor drinking water treatment processes and reservoirs, where an increase in cell counts can be indicative for failure of the treatment or contamination of a reservoir. They are also used to detect harmful algal blooms [36], for quality control in aquaculture and to monitor ballast water to prevent dispersion of nonindigenous organisms [37].

6 Discussion

In the previous sections, a variety of spectroscopic and related optical methods has been described, each with current applications in online water quality monitoring or the potential to be used in such applications. These technologies all share the sole use of interaction between light and matter as their principal measurement. The fully solid-state sensor hardware and the lack of reagents mean these sensors are potentially highly robust and potentially provide long-term performance stability. The current generation of sensors, however, is used in limited numbers and only very rarely in larger numbers as would be expected in smart sensor networks. The main reasons for their limited use are maintenance requirements, power requirements, instrument price and approach to data interpretation.

All optical instruments have an interface where the light used for the analysis crosses from the interior of the instrument into the sample and subsequently the same interface or a secondary, for collection of the light and guiding it to the detector. As only the interaction of the light with the sample is of interest, the optical interface should be fully transparent. However, when it is in contact with the sample, there is a risk of buildup of foreign material. Typical issues include scaling and (bio)fouling. It is therefore critical that optical systems deal with these issues if they are to be deployed in larger numbers, as otherwise the maintenance will be prohibitive. This issue is currently not solved in a satisfactory manner. Although manual cleaning intervals in drinking water applications are often satisfactory, in natural waters and especially in wastewaters, these sensors require frequent (weekly–monthly) maintenance. New methods to prevent contamination of optical surfaces (e.g. antifouling coatings) and/or methods to recognise and correct for fouling in the data processing are required to deal with this issue.

Virtually all spectrometers use an artificial light source. The majority of the advanced systems described in this chapter make use of incandescent lamps and arc lamps. Examples are the xenon, deuterium, deuterium/halogen, tungsten/halogen and mercury/argon lamps. Except for the xenon lamp, which is often used as a flash lamp, these lamps are used in continuous mode. As the lifetime of these lamps is measured in hundreds or thousands of hours, they need replacement. Furthermore, the power requirements for these lamps are such that a main power supply is required to operate the instruments. LED technology provides an alternative to these lamps. Although not all relevant wavelengths are currently available using LEDs, UV spectroscopy and fluorescence devices are making use of this technology. Although LEDs offer an advantage regarding cost and power consumption, allowing for long-term battery-powered operation, they suffer from decreasing brightness over time, which needs to be monitored and corrected for. As a result such instruments require occasional recalibration or replacements of LEDs, e.g. every 1–2 years.

Instrument prices for the devices described vary widely but are all in the 1,000+ euro range. The simplest LED-powered single or dual wavelength devices can be acquired for a few thousand euro, whereas the more advanced spectrometers may cost upwards of 30,000 euro. Even the costs for the simpler instruments prohibit their large-scale deployment. It has to be noted, however, that the production volumes of these devices are low, typically in the hundreds or at most a few thousand per year. Therefore, these are all specialty products, which are intrinsically less cost-effective to produce than mass products. With ongoing miniaturisation of components and increasing demand (e.g. in situ UV/Vis spectrometer sales volumes have been increasing steadily for the last 15 years), prices can be expected to decrease in the future.

Perhaps the biggest challenge to wider use of these instruments is the data interpretation. This is less of technological challenge than a conceptual one. Most spectroscopic methods, as described previously, do not provide information on specific compounds but on the general state of the sample and/or on substance groups. In the water industry, however, a substance specific look at water quality is deeply ingrained. Spectrometer devices are traditionally compared against laboratory methods and are in many cases found to be less sensitive and less specific and therefore disregarded. However, they provide a different type of information: continuous insight into water composition and a much broader coverage and descriptive power on the state and state changes of a medium than can be achieved with the traditional grab-sampling and laboratory approach. The complementary nature of the online approach is of prime importance; the online methods are not likely to replace the laboratory altogether, but they provide another level of information. It is especially this real-time information that allows the direct monitoring of water systems and allows for operational and control applications. Furthermore, the continuous monitoring of the state of the water and changes therein provides useful inputs for smart water systems, especially when combined with other data sources.

Although all the methods described have the potential to produce results (near) real-time, each methodology provides a different type of information. The established methodologies primarily provide information on classes of chemicals and a small number of selected substances. The techniques that so far remain in limited use or have only been demonstrated to have potential in academic research are either more generic offering a generic chemical status indicator (refractive index) and generic microbiological status indicator (image analysis) or are highly specific for molecules (Raman) or elements (LIBS). Table 2 provides a short overview of the different methods and their strengths and weaknesses.

The primary application of the technologies described is found in process control and early detection of incidents or process failures. These are applications where a rapid response is essential. This is where the current generation of online systems comes into its own. The possibility to monitor processes and, through better understanding, optimise operation and control means a positive return on investment can be achieved. The case for quality monitoring, e.g. for drinking water, is often more difficult to make as it primarily provides more insight but not necessarily any gains in operational efficiency or reduction in costs. Online monitoring of wastewater effluent for compliance purposes is, however, in some cases being applied as it can be used to determine the total contaminant load discharged as well as failure of the treatment. Monitoring for legislative purposes on individual substances remains

Table 2 Overview of spectroscopic methods and their general characteristics

Method	Generic properties	Advantages	Disadvantages
UV/Vis (absorption)	• Measures broad spectrum and uses calibration algorithm to extract desired information • Broad spectral signals • Primarily sensitive for organic substances with unsaturated bonds (e.g. aromatics) • Sensitivity down to high ppb level for substances with high absorption coefficients	• Good for sum parameters (BOD, COD, TOC, TSS) • Good at fingerprinting (distinguish between normal/abnormal conditions)	• Limited capability to identify individual compounds • Sensitivity varies widely per compound, with poor sensitivity for aliphatic hydrocarbons and most inorganics • Cross-sensitivity to substances with overlapping spectra can be an issue when not compensated through calibration algorithms
Fluorescence	• Uses combinations of excitation an and emission wavelength to measure specific groups of substances • Primarily sensitive for organic substances with unsaturated bonds (e.g. aromatics) • ppb level sensitivity	• Good for sum parameters (BOD, FDOM, oil in water) • Good at detecting algal pigments • High sensitivity for polycyclic aromatics (low ppb to high ppt levels) • Suitable for noncontact/remote applications and monitoring of floating layers	• Difficulty with identification of individual compounds • Sensitivity varies widely per compounds, with poor sensitivity for aliphatic hydrocarbons and most inorganics • Equipment not flexible: excitation and emission wavelengths fixed, not flexible in substances that can be detected
Near infrared (absorption)	• Measures broad spectrum and uses calibration algorithm to extract desired information • Measures molecular bonds (e.g. O-H) • Broad spectral signals • ppm level sensitivity	• Good at fingerprinting (distinguish between normal/abnormal conditions) • Reflection spectroscopy useful in inhomogeneous and nontransparent media	• Difficulty with identification of individual compounds • Cross-sensitivity to substances with overlapping spectra can be an issue when not compensated through calibration algorithms
Raman (scattering)	• Spectra with distinct signals	• Suited for both organic and inorganic analytes, including metals • Allows identification of individual substances • SERS offers possibility for pathogen detection	• Detection limits higher than with UV/Vis/NIR • Sensitivity varies widely per compound
LIBS (emission)	• Provides elemental information • Mid-ppm level sensitivity	• Capable of measuring all elements, from hydrogen to uranium • Capability to detect	• Sensitivity insufficient for low level detection of most contaminants

(continued)

Table 2 (continued)

Method	Generic properties	Advantages	Disadvantages
		independent of the nature of the analyte-plasma releases all atoms from their chemical environment. Use, e.g. to detect total concentration of an element, such as P, As or Cr	• Sensitivity varies with element and with sample matrix • Provides elemental composition of the sample, no information on chemical composition
Refractive index	• Measures physical property of sample, which is responsive to the chemical makeup of the sample • ppm level sensitivity	• Generic, sensitive to all types of chemicals, both organic and inorganic • Good for monitoring stability/variability of the sample composition • Consistent sensitivity, with only minor variations between different types of substances	• Generic, not possible to identify the cause for a change in signal
Image analysis	• Analysis of combinations of optical properties of particles • Targeting samples with cell concentrations in the 1,000/mL or higher	• Culture- and reagent-free analysis of microbiological properties, such as total cell counts • Combinations of optical properties sometimes allow for classification of cells	• Only suited for analysis of particles • Time-consuming when analysing properties of individual cells (e.g. flow cytometry)

limited to nitrate; spectroscopic measurement of nitrate is listed as a standard method [11]. Especially for trace contaminations, where individual substances concentrations are requested at the low ppb or even ppt levels, the current generation of spectrophotometric devices is lacking in sensitivity when applied in a water matrix. Typical sensitivities for each technology type are given in Table 2.

7 Outlook

Smart water systems are expected to revolutionise the water industry. Through the combination of information from various sources at all levels in the water system, more effective management will become possible. This will help reduce costs, increase safety and ensure that infrastructure will be able to cope with changing demands, such as urbanisation, climate change and decentralisation. In this data-driven approach, sensors are an essential link in the chain, as they provide the raw data upon which the whole system is built. As smart systems are expected to acquire information not only from traditional monitoring locations such as water treatment plants, but actually from the entire water system, there is a requirement for durable, autonomous, networked and affordable sensor technology. Optical sensor systems provide a good basis for this sensor generation: they are robust and low maintenance. In this chapter a selection of technologies has been described, which have proven they can provide valuable information on the composition and quality of water. Although the current application of these methodologies in real-time online sensing remains limited, such sensors, UV/Vis absorbance and fluorescence devices in particular, have become widely accepted and established in the water industry. Their use is on the rise, especially in process monitoring and control and as early warning systems. This is expected to gain further momentum as the performance and cost-effectiveness of these systems increase further.

References

1. Peleg A (2014) Investment in smart water networks on the rise. Water Technology
2. Owen DAL (2018) Smart water technologies and techniques – data capture and analysis for sustainable water management. Wiley, Oxford
3. Makropoulos CK, Butler D (2010) Distributed water infrastructure for sustainable communities. Water Resour Manag 24:2795–2816
4. van den Broeke J, Carpentier C, Moore C, Carswell L, Jonsson J, Sivil D, Rosen JS, Cade L, Mofidi A, Swartz C, Coomans N (2015) Compendium of sensors and monitors and their use in the global water industry. IWA Publishing, London
5. Sun S, Bertrand-Krajewski J-L, Lynggaard Jensen A, van den Broeke J, Edthofer F, do Céu Almeida M, Silva Ribeiro À, Menaia J (2011) Literature review for data validation methods. Deliverable 3.1.1, PREPARED (EC FP7 Programme project number 244232)
6. Hyndman RJ, Athanasopoulos G (2018) Online open access textbooks. www.otexts.org/fpp2/
7. Bettencourt LMA, Hagberg AA, Larkey LB (2007) Separating the wheat from the chaff: practical anomaly detection schemes in ecological applications of distributed sensor networks. In: Aspnes J, Scheideler C, Arora A, Madden S (eds) Lecture notes in computer science, vol 4549. Springer, Berlin
8. Otto M (1999) Chemometrics: statistics and computer application in analytical chemistry. Wiley, Weinheim
9. Thomas O, Burgess C (2017) UV-visible spectrophotometry of water and wastewater, 2nd edn. Elsevier, Amsterdam
10. Langergraber G, Fleischmann N, Hofstaedter F (2003) A multivariate calibration procedure for UV/Vis spectrometric quantification of organic matter and nitrate in wastewater. Water Sci Technol 47(2):63–71
11. Rice EW, Baird RB, Eaton AD (2017) Standard methods for the examination of water and wastewater. AWWA Catalog no. 10086, American Public Health Association, American Water Works Association and Water Environment Federation
12. Bishop CM (2006) Pattern recognition and machine learning. Springer, New York
13. Mesquita DP, Quintelas C, Amaral AL, Ferreira EC (2017) Monitoring biological wastewater treatment processes: recent advances in spectroscopy applications. Rev Environ Sci Biotechnol. https://doi.org/10.1007/s1115701794399
14. Huber E, Frost M (1998) Light scattering by small particles. J Water Supply Res Technol 47 (2):87–94

15. Fabris R, Chow C, Dexter R, Colton J, Knoblauch J, Drikas M (2013) Feed-forward coagulant control using online UV/Vis monitoring. Water Sci Technol Water Supply 13(2):420–426
16. Korshin GV, Li C-W, Benjamin MM (1997) The decrease of UV-absorbance as an indicator of TOX formation. Water Res 31(4):946–949
17. Erlandsson M, Futter MN, Kothawala DN, Köhler SJ (2012) Variability in spectral absorbance metrics across boreal lake waters. J Environ Chem 14(10):2643–2652
18. Rieger L, Vanrolleghem PA, Langergraber G, Kaelin D, Siegrist H (2008) Long-term evaluation of a spectral sensor for nitrite and nitrate. Water Sci Technol 57(10):1563–1569
19. Chow C, Fabris R, Dixon M (2008) Case studies using S::CAN on-line monitoring system. Research report 75, Water Quality Research Australia
20. van den Broeke J, Edthofer F, Weingartner A (2011) Clean data and reliable event detection – turning results from online sensors into information. In: Borchers U, Thompson KC (eds) Water contamination emergencies: monitoring, understanding and acting. RSC Publishing, Cambridge
21. Beutler M, Wiltshire KH, Meyer B, Moldaenke C, Lüring C, Meyerhöfer M, Hansen U-P, Dau H (2002) A fluorometric method for the differentiation of algal populations in vivo and in situ. Photosynth Res 72:39–53
22. Eikebrokk B, Vogt RD, Liltved H (2004) NOM increase in northern European source waters: discussion of possible causes and impacts on coagulation/contact filtration processes. Water Sci Technol Water Supply 4(4):47–54
23. Hansen AM, Fleck JA, Kraus TEC, Downing BD, von Dessonneck T, Bergamaschi BA (2018) Procedures for using the Horiba Scientific Aqualog® fluorometer to measure absorbance and fluorescence from dissolved organic matter. U.S. Department of the Interior, U.S. Geological Survey, Open file report 2018-1096
24. Markechova D, Tomkova M, Sadecka J (2013) Fluorescence excitation-emission matrix spectroscopy and parallel factor analysis in drinking water treatment: a review. Pol J Environ Stud 22(5):1289–1295
25. Gowen AA, Tsenkova R, Bruen M, O′donnell C (2012) Vibrational spectroscopy for analysis of water for human use and in aquatic ecosystems. Crit Rev Environ Sci Technol 42 (23):2546–2573
26. Garcia-Sanchez F, Galvez-Sola L, Nicolas JJM, Muelas-Domingo R, Nieves M (2017) Using near-infrared spectroscopy in agricultural systems. In: Kyprianidis K, Skvaril J (eds) Developments in near-infrared spectroscopy. IntechOpen, London
27. Li Z, Deen MJ, Kumar S, Selvaganapathy PR (2014) Raman spectroscopy for in-line water quality monitoring – instrumentation and potential. Sensors 14(9):17275–17303
28. Li Z, Wang J, Li D (2016) Applications of Raman spectroscopy in detection of water quality. Appl Spectrosc Rev 51(4):333–357
29. Collette TW, Williams TL (2002) The role of Raman spectroscopy in the analytical chemistry of potable water. J Environ Monit 4(1):27–34
30. Tanchou V (2014) Review of methods for the rapid identification of pathogens in water samples. Scientific and Technical Research series, Joint Research Centre, European Commission
31. Grow AE, Wood LL, Claycomb JL, Thompson PA (2003) New biochip technology for label-free detection of pathogens and their toxins. J Microbiol Methods 53(2):221–233
32. Araujo CF, Nolasco MM, Ribeiro AMP, Ribeiro-Claro PJA (2018) Identification of microplastics using Raman spectroscopy: latest developments and future prospects. Water Res 142:426–440
33. Gaudiuso R, Dell′Aglio M, DePascale O, Senesi GS, DeGiacomo A (2010) Laser induced breakdown spectroscopy for elemental analysis in environmental, cultural heritage and space applications. Sensors 10:7434–7468
34. Williamson F, van den Broeke J, Koster T, Klein Koerkamp M, Verhoef JW, Hoogterp E, de Graaf BR (2014) Online water quality monitoring in the distribution network. Water Pract Technol 9(4):575–585. https://doi.org/10.2166/wpt.2014.064

35. Ramachandran A, Wang S, Clarke J, Ja SJ, Goad D, Wald L, Flood EM, Knobbe E, Hryniewicz JV, Chu ST, Gill D, Chen W, King O, Little BE (2008) A universal biosensing platform based on optical micro-ring resonators. Biosens Bioelectron 23(7):939–944
36. Zhou Q, Chen W, Zhang H, Peng L, Liu L, Han Z, Wan N, Li L, Song L (2012) A flow cytometer based protocol for quantitative analysis of bloom-forming cyanobacteria (Microcystis) in lake sediments. J Environ Sci 24(9):1709–1716
37. Peperzak L, Zetsche E-M, Gollasch S, Artigas LF, Bonato S, Creach V, Dubelaar GBJ, Henneghien J, Hess-Erga O-K, Langelaar R, Larsen A, Maurer BN, Mosselaar A, Reavie ED, Rijkeboer M, Tobiesen A (2018) Comparing flow cytometry and microscopy in the quantification of vital aquatic organisms in ballast water. J Mar Eng Technol. https://doi.org/10.1080/20464177.2018.1525806

Quartz Crystal Microbalance Sensors: New Tools for the Assessment of Organic Threats to the Quality of Water

Bartolomeo Della Ventura, Marco Mauro, Raffaele Battaglia, and Raffaele Velotta

Contents

Abstract Water monitoring technologies are widely used for contaminant detection in a wide variety of water ecology applications such as water treatment plants and water distribution systems. A tremendous amount of research has been conducted over the past decades to develop robust and efficient techniques of contaminant detection with minimum operating cost and energy. Recent developments in spectroscopic techniques and biosensor approach have improved the detection sensitivities,

B. Della Ventura and R. Velotta (*)
Department of Physics "E. Pancini", Università di Napoli Federico II, Naples, Italy
e-mail: rvelotta@unina.it

M. Mauro and R. Battaglia
Novaetech S.r.l., Pompei, NA, Italy

Andrea Scozzari, Steve Mounce, Dawei Han, Francesco Soldovieri, and Dimitri Solomatine (eds.), ICT for Smart Water Systems: Measurements and Data Science, Hdb Env Chem (2021) 102: 315–342, https://doi.org/10.1007/698_2019_390,
© Springer Nature Switzerland AG 2019, Published online: 26 July 2019

quantitatively and qualitatively. The availability of in-situ measurements and multiple detection analyses has expanded water monitoring applications in various advanced techniques including successful development in hand-held sensing devices, which improves portability in real-time basis for the detection of contaminants, such as microorganisms, pesticides, heavy metal ions, and inorganic and organic components.

Keywords Biosensor, Contaminant, Functionalization, Gold electrode, Immunosensor, Organic threats, Pollutants, Quartz crystal microbalance (QCM), Transducer, Water

1 Introduction

The Safe Drinking Water Act defines the term “contaminant” as meaning any physical, chemical, biological, or radiological substance or matter in water. Therefore, the law defines “contaminant” very broadly as being anything other than water molecules. Drinking water may reasonably be expected to contain at least small amounts of some contaminants. Some drinking water contaminants may be harmful if consumed above certain levels, while others may be harmless. The presence of contaminants does not necessarily indicate that water poses a health risk. Waste production from agriculture, industrial sewage, and animal and human activities are affecting the boundaries between clean and waste water, causing a reduction in the fresh water available for humans. Water ecology provides services such as food production, nutrient cycling, habitat provision, flood regulation, water purification, and soil formation. Biological and chemical contaminants in tap and drinking water can initiate the evolution of contagious diseases.

The following are general categories of drinking water contaminants and some examples of each:

- Physical contaminants primarily impact the physical appearance or other physical properties of water. Examples of physical contaminants are sediments or organic material suspended in the water of lakes, rivers, and streams from soil erosion.
- Chemical contaminants are inorganic elements or compounds. These contaminants may be naturally occurring or due to human activities. Examples of chemical contaminants include nitrogen, bleach, salts, pesticides, metals, toxins produced by bacteria, and human or animal drugs.
- Biological contaminants are living organisms also referred to as microbes or microbial contaminants. Examples include bacteria, viruses, protozoan, and parasites.
- Radiological contaminants are chemical elements with an unbalanced number of protons and neutrons resulting in unstable atoms that can emit ionizing radiation. Examples of radiological contaminants include cesium, plutonium, and uranium.

Recently, analytical technologies in water monitoring have taken a variety of directions. There are several water monitoring techniques, including conventional instrumental analysis (laboratory-based analysis), sensor placement approach, model-based event detection, microfluidic devices, spectroscopic approach, and biosensors. Selecting a certain detection technique is strongly dependent on the purpose of the analysis, whether it requires quantitative, qualitative, or hybrid measurements. Biological and chemical sensors have been in great demand for use in water monitoring technology, and they appear to be suitable for device integration and commercialization.

Previously, the detection of water contaminants was often conducted manually in water laboratory facilities [1]. At the laboratory level, analyses are usually carried out by skillful personnel using high-end and cutting-edge technologies. Conventionally, multiple fermentation tube technique [2], filtration method [3], DNA amplification [4], fluorescence in-situ hybridization (FISH) techniques [5, 6], capillary electrophoresis [7, 8], field-flow fractionation [9], chromatography [10], mass spectrometry [11], and electrochemical-based device [12] are the most commonly used instruments. The overall benefits of laboratory-based analytical methods have been recognized since a long time, but recent studies have shown that they are not efficient for on-site monitoring applications. With the technological advancements in analytical chemistry, new techniques have been developed through the introduction of advanced spectroscopy [13] and water quality sensors [14–16].

High sensitivity and real-time monitoring of mass changes on the sensor crystal make quartz crystal microbalance (QCM) a very attractive technique for a large range of applications. The development of QCM systems for use in fluids or with viscoelastic deposits has dramatically increased the interest for this technique. A major advantage of the technique used for liquid systems is that it allows for a label-free detection of molecules. QCM is capable of measuring mass changes as small as a fraction of a monolayer of atoms. QCM crystals are becoming a good alternative analytical method in a great deal of applications such as biosensors, analysis of biomolecular interactions, study of bacterial adhesion at specific interfaces, pathogen and microorganism detection, study of polymer film–biomolecule or cell–substrate interactions, immunosensors, and extensive use in fluids and polymer characterization and electrochemical applications among others. QCM is used also in gaseous environments, e.g., as gas and humidity sensors and for the detection of aerosols [17], but its main capability consists in providing real-time monitoring of contaminants in process, recycle, and waste water; groundwater quality monitoring; detection of contaminants in streams, lakes, and water supplies; monitoring dumping in off-shore waterways [18].

2 Theory and Modeling of QCM Data

2.1 Sauerbrey′s Equation: Rigid Mass

The first quantitative analysis of using quartz crystal resonators as mass sensors was developed by Sauerbrey in 1959 [19].

The deposition of an additional mass causes a decrease in the resonance frequency of the quartz crystal resonators. The Sauerbrey′s equation provides a linear relationship between variations in the resonance frequency and the mass of a film present on the quartz crystal surface. This linear relationship is valid if the following assumptions are fulfilled:

- the film mass and thickness are much smaller than those of the quartz crystal;
- the film is uniform, rigid, and rigidly attached to the quartz crystal surface;
- the quartz crystal oscillation takes place in vacuum or air.

Sauerbrey′s equation is not valid when the deposited film is liquid because it does not follow the shear oscillations of the quartz crystal surface in a solid manner. Indeed, if just one face of the quartz crystal is immersed in a viscous liquid, there are mechanical dissipation phenomena that make the Sauerbrey′s relationship inapplicable.

The application of an alternate voltage to the electrodes deposited on the quartz crystal surface causes a shear deformation that propagates along the thickness (Fig. 1). The Sauerbrey′s relationship is obtained by solving the unidimensional equation for a transverse shear wave propagating along the direction of the crystal thickness. It is based on the idea that a film deposited on the quartz crystal surface increases its thickness, causing an increase in the stationary shear wavelength propagating along the quartz thickness.

The frequency values of the free standing shear waves are given by the following relationship:

$$f_N = \frac{N v_s}{2 h_s}$$

where h_s is the quartz thickness, v_s is the propagation speed of the shear wave, and $N = (1, 3, 5, \ldots)$ is an odd number. It is not trivial to point out that the only harmonics that can be excited are those that have an odd wave number.

Under the hypothesis reported above, it is possible to derive the Sauerbrey′s equation, which establishes a linear relationship between the variation of the

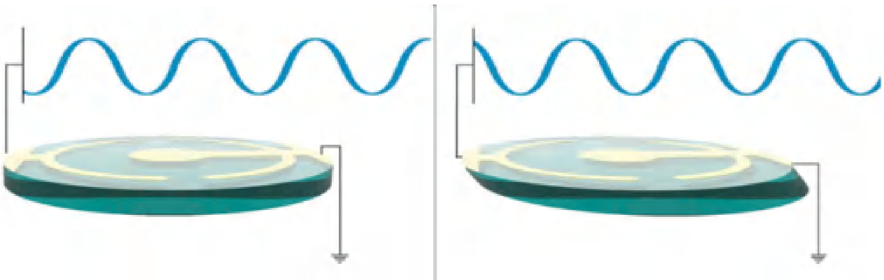


Fig. 1 The application of an alternate voltage between electrodes induces a shear deformation due to piezoelectric effect. Reprinted with permission from [20]

resonance frequency Δf_N and the deposition of an additional mass, Δm, rigidly connected to the quartz crystal surface:

$$\Delta f_N = -N \frac{2 f_N{}^2}{(\mu_q \rho_q)^{1/2}} \cdot \frac{\Delta m}{A}$$

where f_N is the unperturbed resonance frequency of the N-th mode, μ_q is the piezoelectric shear strength of the quartz crystal, ρ_q is the mass density of the quartz crystal, and A is the electrode surface. It is important to note that the coefficient of proportionality depends only on intrinsic characteristics of the quartz crystal resonators.

2.2 Sauerbrey′s Mass Sensitivity

Assuming that the rigid film is uniformly deposited on the quartz crystal surface, it is possible to define the sensitivity of the quartz crystal resonators S_N, measured in the CGS unit as $g^{-1}\ cm^2\ s^{-1}$, as the ratio between the frequency variation Δf_N and the variation of the surface mass density:

$$S_N = \frac{\Delta f_N}{(\Delta m / A)}$$

which can be written in the form:

$$S_N = -N \frac{2 f_N{}^2}{(\mu_q \rho_q)} = -\frac{2 f_N{}^2}{\rho_q v_s}$$

For an AT-cut quartz crystal, which is a specific cutting of original crystal stones characterized by a cut angle of 35° 15′ respect to the crystallographic Z-axis, the piezoelectric shear strength is $\mu_q = 2.947 \times 10^{11}$ g cm^{-1} s^{-2}, the mass density $\rho_q = 2.648$ g cm^{-3}, and speed of propagation of the shear wave is $v_s = (\mu_q/\rho_q)^{1/2} = 3.340 \times 10^5$ cm s^{-1}. Based on the equations above, for a quartz crystal oscillating at the unperturbed fundamental frequency of 10 MHz, a frequency shift of 1 Hz is caused by a mass deposited per unit area equal to 4.49×10^{-11} g.

The mass sensitivity of the quartz is not uniform over the entire surface of the quartz but has a maximum in the center and decreases as it approaches the edges of the electrodes. The experimental results [21] show that the spatial distribution of mass sensitivity on the quartz surface follows the distribution of the vibration amplitude. Both the sensitivity and the acceleration follow a Gaussian distribution and, in particular, the sensitivity is proportional to the square of the radial displacement.

2.3 Kanazawa: Gordon Equation: Quartz Crystal in Contact with a Liquid

The first pioneering physical model for the quantitative determination of the variation of the resonance frequency of a quartz crystal immersed in a liquid was developed by Kanazawa and Gordon [22]. This model is based on the assumption that the quartz crystal is a perfectly elastic solid, therefore not subject to mechanical energy losses by dissipation, and the liquid is a purely viscous fluid (or Newtonian fluid). Quartz crystal stable oscillation can be obtained when one side of the quartz is in contact with a liquid. However, the viscous effect of the liquid causes not only a large variation in the resonance frequency but also a loss in the Q quality factor, causing instability and total damping of the oscillation. It is possible to determine the physical behavior of the quartz crystal–liquid system, by considering the coupling between the elastic shear wave in the crystal and the one propagating within the viscous fluid. The resonance condition results from the choice of appropriate boundary conditions for the quartz crystal–liquid interface.

The resulting wave is composed of an undamped shear wave that propagates inside the crystal along the thickness direction and of a highly damped shear wave propagating within the liquid away from the crystal surface (Fig. 2). Propagation waves in the liquid may be written in terms of the instantaneous velocity of the liquid at a given position y:

$$v_x(y;t) = U_0 e^{-k(y-h_s)} \cos\left[k(y-h_s) - \omega t\right]$$

where U_0 is the wave amplitude at the separation surface, h_s is the quartz crystal thickness, and k is the propagation constant.

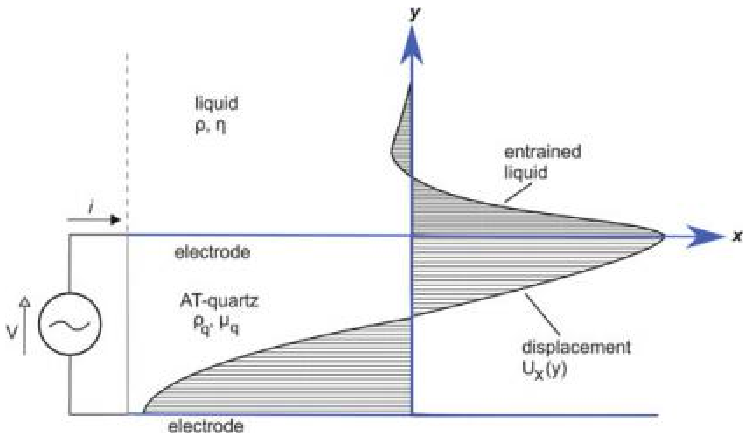


Fig. 2 The vibration of a quartz crystal in contact with a liquid medium consists of a stationary shear wave propagating in the quartz and a strongly damped acoustic wave that propagates within the liquid. Reprinted with permission from [23]

The characteristic length (or penetration length), λ, which describes the envelope of the damped oscillation, is equal to 1/k, the reciprocal of the propagation constant. The characteristic length can be written in terms of density ρ_L and viscosity η_L of the liquid as follows:

$$\lambda = \sqrt{\frac{\eta_L}{\pi f_0 \rho_L}}$$

For an AT-cut quartz crystal vibrating at 10 MHz, with one face in contact with pure water at $T = 20°C$, $\rho_L = 0.9982$ g cm^{-3}, and $\eta_L = 1.0022 \times 10^{-2}$ g cm^{-1} s^{-1}, the characteristic length is about 180 nm.

Assuming (1) the continuity of the velocity field at the separation surface (that is the quartz surface transverse speed is equal to that of the adjacent fluid) and (2) the force exerted by the liquid on the quartz surface is equal and opposite to the force that the quartz exerts on the fluid, the difference Δf between the resonance frequency of the unperturbed crystal f_0 and that in contact with the liquid is given by the Kanazawa–Gordon equation:

$$\Delta f = f_0^{3=2} \sqrt{\frac{\eta_L \rho_L}{\pi \mu_q \rho_q}}$$

According to this model, the quartz crystal does not transmit the vibration to the entire liquid above the surface, since the transverse displacement decays with exponential law with a characteristic decay length, λ, so that just a thin layer of liquid produces the response of the quartz crystal. The effective mass of liquid Δm_L can be calculated using the following relation:

$$\Delta m_L = \lambda \rho_L = \sqrt{\frac{\rho_L \eta_L}{\pi f_0}}$$

By considering the above equations, for an AT-cut quartz crystal vibrating at 10 MHz, with one face in contact with pure water at $T = 20°C$, $\rho_L = 0.9982$ g cm^{-3}, and $\eta_L = 1.0022 \times 10^{-2}$ g cm^{-1} s^{-1}, the frequency shift is $\Delta f = 2{,}020$ Hz and the effective mass of liquid is $\Delta m_L = 17 \times 10^{-6}$ g.

2.4 Small Load Approximation: The Electromechanical Model

A more general description of the response of the quartz crystal resonator in contact with a generic sample is given by the so-called small load approximation model [24, 25]. It can be derived using the electromechanical model of a quartz crystal resonator. In the approximation of small loads and small frequency variations close

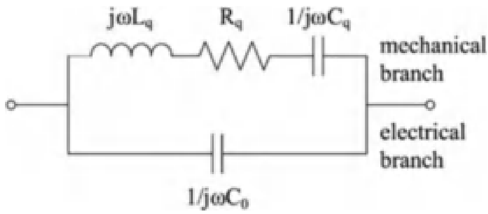


Fig. 3 Equivalent circuit for an unperturbed quartz crystal resonator according to the BVD model

Table 1 Electrical parameters of the BVD model as a function of the physical characteristics of an AT-cut quartz crystal

Parameter	Expression
C_0	$\frac{\varepsilon_{22}A}{h_s}$
C_q	$\frac{8Ae_{26}^2}{\pi^2 h_s G_q}$
L_q	$\frac{\rho_q h_s^3}{8Ae_{26}^2}$
R_q	$\frac{\pi^2 h_s \eta_q}{8Ae_{26}^2}$

where ε_{22} is the quartz dielectric constant, e_{26} is the piezoelectric constant depending on the quartz cutting angle, A is the electrode surface deposited on the quartz, and $G_q \sim 29.3 \times 10^9$ Pa is the shear modulus of an AT-cut quartz

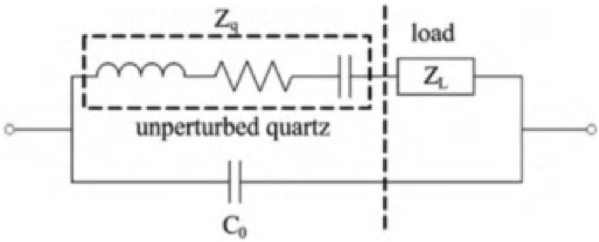


Fig. 4 Equivalent circuit for a loaded quartz crystal resonator, according to BVD model

to the resonance, the equivalent electrical circuit of the quartz crystal is known as the Butterworth-Van Dyke (BVD) circuit.

The BVD circuit combines a “mechanical branch” in parallel with an electrical branch (Fig. 3). The “mechanical branch” consists of three elements in series: an inductor L_q, which corresponds to the initial mass of the quartz crystal, a capacitor C_q, which represents the quartz mechanical elasticity, and a resistance R_q, which corresponds to the dissipation of mechanical energy, caused by effects of viscosity and friction. The electrical branch consists only of a capacitor, C_0, which mainly corresponds to the value of the electrical capacity between the electrodes deposited on the quartz crystal surface.

The electrical parameters of the BVD equivalent circuit based on the physical characteristics of an AT-cut quartz crystal resonators are given in Table 1 [26].

Using the BVD equivalent circuit, any load on the quartz crystal surface can be represented as a “load impedance” Z_L in series to the “mechanical branch” (Fig. 4).

The load impedance is in general equal to the ratio between the applied stress and the speed on the quartz crystal surface. Using the BVD equivalent circuit, it is possible to derive an important relationship that binds the resonance frequency and the dissipation variation as a function of the stress–speed ratio. So that if an explicit form of Z_L is known, it is possible to calculate both frequency and dissipation variations no matter what is the sample in contact with the quartz crystal surface.

In this model it is useful to introduce a complex resonance frequency defined as follows:

$$\tilde{f} = f + i\Gamma$$

where the real part f is the resonance frequency and the imaginary part Γ is "half bandwidth at half maximum" of the resonance. Indeed, Γ is related to the dissipation factor D, which is a dimensionless parameter defined as the ratio between the energy loss and stored in each cycle:

$$D \equiv \frac{E_{\text{dissipated}}}{2\pi E_{\text{stored}}}$$

through the following relation:

$$D = \frac{2\Gamma}{f}$$

Dissipation is an important physical observable because it is related to the viscoelastic properties of the sample in contact with the quartz crystal surface.

In the small load approximation one obtains the following relationship:

$$\frac{\Delta\tilde{f}}{f_f} = \frac{i}{\pi Z_q} Z_L = \frac{i}{\pi Z_q} \frac{\sigma}{\underline{u}}$$

where $Z_q = \rho_q v_q = (\rho_q \mu_q)^{1/2} = 8.8 \times 10^5$ g cm^{-2} s^{-1} is the acoustic impedance of an AT-cut quartz crystal, σ is the mechanical stress, and u' is the velocity on the quartz crystal surface, respectively. This relationship is decisive for modeling QCM data and will be used in the next paragraph to derive frequency and dissipation variations for a layered system evenly distributed over the quartz crystal surface. These equations are valid if the following conditions are fulfilled:

- the quartz crystal and the layered system are laterally homogeneous and infinite;
- the quartz crystal mechanical deformation is caused only by a transverse shear wave with a wave vector perpendicular to the surface of the crystal (thickness-shear mode). There are no compression and no flexural waves;

- the stress tensor is proportional to that of the deformations, that is a linear viscoelasticity relationship applies;
- the contribution due to piezoelectric stiffness can be neglected.

2.5 Semi-Infinite Viscoelastic Layer Newtonian Liquid

For a QCM in contact with a semi-infinite viscoelastic medium, there is a transverse shear wave inside the quartz crystal and a shear wave propagating through the liquid away from the quartz crystal surface. In addition, for a Newtonian liquid the imaginary part of the viscosity is null and the real part is constant and independent from the frequency. In the framework of the small load approximation, the complex frequency variation is given as follows:

$$\frac{\Delta \tilde{f}}{f_f} = \frac{1}{\pi Z_q} \frac{-1+i}{\sqrt{2}} \sqrt{2\pi n f_f \rho_L \eta_L}$$

By separating the real and imaginary part of the relation above, the resonance frequency variation is calculated as follows:

$$\Delta f_f = -f_f^{3/2} \sqrt{\frac{n \rho_L \eta_L}{\pi \mu_q \rho_q}}$$

which corresponds to the Kanazawa–Gordon equation, and the dissipation variation as follows:

$$\Delta D = 2 f_f^{1/2} \sqrt{\frac{n \rho_L \eta_L}{\pi \mu_q \rho_q}}$$

By considering the above equation, for an AT-cut quartz crystal vibrating at 10 MHz, with one face in contact with pure water at $T = 20°C$, $\rho_L = 0.9982$ g cm^{-3}, and $\eta_L = 1.0022 \times 10^{-2}$ g cm^{-1} s^{-1}, the dissipation shift is $\Delta D = 404 \times 10^{-6}$.

It is worth noting that frequency and dissipation variations scale as $\sqrt{n}$ with the QCM overtone number. By combining the equations above, it is shown that for a Newtonian liquid in contact with a QCM that

$$\Delta D = -2 \frac{\Delta f_f}{f_f}$$

no matter what liquid is in contact with the quartz crystal surface and which QCM overtone is interrogated.

2.6 Purely Inertial Layer: Sauerbrey′s Equation

For the case of QCM in contact with a purely inertial load, the viscoelasticity can be ignored. One can derive the frequency shift−mass proportionality of the Sauerbrey′s equation in the small load approximation. The stress induced by a very thin layer is only caused by inertia and it is given as follows:

$$\sigma = -\omega^2 u_0 m_f$$

where u_0 is the oscillation amplitude and m_f is the mass of the inertial layer. By inserting it in the small load approximation, the Sauerbrey equation is obtained again:

$$\frac{\Delta f}{f_f} = -2\frac{f_f}{Z_q} m_f$$

The imaginary part of the complex frequency is null because the viscoelasticity of the load has been neglected.

2.7 Viscoelastic Layer of Arbitrary Thickness

In the small load approximation model it is possible to derive a more general relation than the Sauerbrey′s equation, if the hypothesis of very thin deposited films is abandoned and viscoelastic films of an arbitrary thickness are considered. In this case the vibration consists of a transverse shear wave inside the quartz crystal and a shear wave transmitted and reflected within the viscoelastic layer.

In the approximation of a viscoelastic layer thickness much smaller than the length of the propagation wave, the small load approximation provides the following relationship:

$$\frac{\Delta \tilde{f}}{f_f} = -2\frac{f_f}{Z_q} m_f \left(1 + \frac{Z_q{}^2}{Z_f{}^2}\left(\pi\frac{m_f}{m_q}\right)^2\right)$$

where Z_f is the complex acoustic impedance of the viscoelastic layer and m_f is the mass of the viscoelastic layer. If $m_f << m_q$, the previous equation reduces to the Sauerbrey′s equation. Otherwise, the terms in bracket corresponds to the “visco-elastic correction” to the Sauerbrey′s equation.

2.8 Viscoelastic Layer in Liquid

The small load approximation is used to quantify the variation of the complex resonance frequency of a QCM in contact with a viscoelastic film and immersed in a liquid. In this case the wave is made of a transverse shear wave inside the quartz crystal, a shear wave transmitted in the viscoelastic layer and then reflected at the separation surfaces of the crystal–viscoelastic layer and viscoelastic layer–liquid, and finally a propagation wave that travels within the liquid far from the surface of the viscoelastic layer.

The small load approximation predicts the following relation for the variation of the complex resonance frequency:

$$\frac{\Delta \tilde{f}}{f_f} = \frac{i}{\pi Z_q}\left[Z_L + i2\pi f_f m_f\left(1 - \frac{{Z_L}^2}{{Z_f}^2}\right)\right]$$

The first term in the sum corresponds to the Kanazawa–Gordon equation (liquid contribution), the second term corresponds to the Sauerbrey′s equation (inertial mass layer load), and the third term is a viscoelastic correction caused by the liquid environment, also known in literature as the “missing mass effect” [27].

In most experimental setups, the resonance frequency is determined respect to a reference state where the quartz is already immersed in liquid, in this case the previous equation can be written as follows:

$$\frac{\Delta \tilde{f}}{f_f} = -\frac{2 f_f m_f}{Z_q}\left(1 - \frac{{Z_L}^2}{{Z_f}^2}\right)$$

Showing that in this case the liquid contribution leads to a smaller mass of the viscoelastic layer.

3 QCM Detection Scheme and Electronic Interfaces

The application of a quartz crystal as sensor requires the usage of an appropriate electronic interface. In the next paragraphs two different approaches are described in detail: (1) quartz oscillators and (2) network or impedance analysis.

A quartz crystal is a resonant element and stable vibrations can be ensured by using a simple oscillator driver. In this scheme, the output signal consists of an analog voltage whose frequency can be processed with very high accuracy by a digital system. Network or impedance analysis is based on the passive interrogation of the quartz crystal for monitoring the amplitude and phase response, in order to characterize the electrical parameters of the quartz.

3.1 Quartz Oscillators

The application of oscillator circuits as an electronic interface for QCM devices is one of the most commonly used methods for recovering frequency variations with a high accuracy. Since a quartz crystal is a resonant element, it can be driven at a stable amplitude with an appropriate electronic circuit. It should be noted that the quartz itself is an integral part of the oscillator circuit and particular attention must be paid in the circuit design. Application of quartz crystal oscillators in liquid phase or in contact with heavy-load layers causes a drastic decrease in both the quality factor Q and phase slope so that a suitable electronic configuration should be developed and electronic components selected.

The advantages of the oscillator circuit scheme consist mainly in the capability of continuous data monitoring and in the fact that frequency measurements can be made with very high accuracy. In addition, the integration capability and low-cost electronic circuitry makes this detection scheme ideal for air–vacuum applications and suitable for the most common chemical applications.

The general requirements to drive a quartz crystal at stable oscillations in a closed resonant loop are to maintain the loop gain equal to 1 and to have a total loop phase shift equal to zero or a multiples of 360° (Barkhausen criteria). With reference to the Pierce oscillator circuit in Fig. 5, the inverting amplifier causes a 180° phase shift and the phase condition of the Barkhausen criteria is ensured by the additional 180° phase shift introduced by the feedback network, which consists of R1, C1 (90° phase shift), and quartz, C2 (90° phase shift).

The main requirements for the application of a QCM based on quartz oscillator interfaces for sensing liquid samples were found by Barnes in 1991 [28]:

- the quartz oscillator should operate near its series resonance frequency, where the effects of the parallel capacitance on the frequency variations are minimized;
- one face of the resonator should be grounded for electrochemical or biological applications and for reducing parasitic capacitance effects;
- an automatic gain control (AGC) should be developed to control the loop gain to ensure stable oscillation for heavy-load and highly viscous samples;

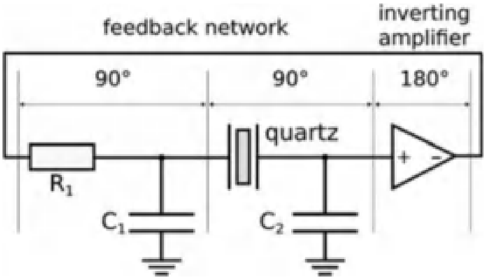


Fig. 5 Criteria for stable oscillation using a Pierce oscillator

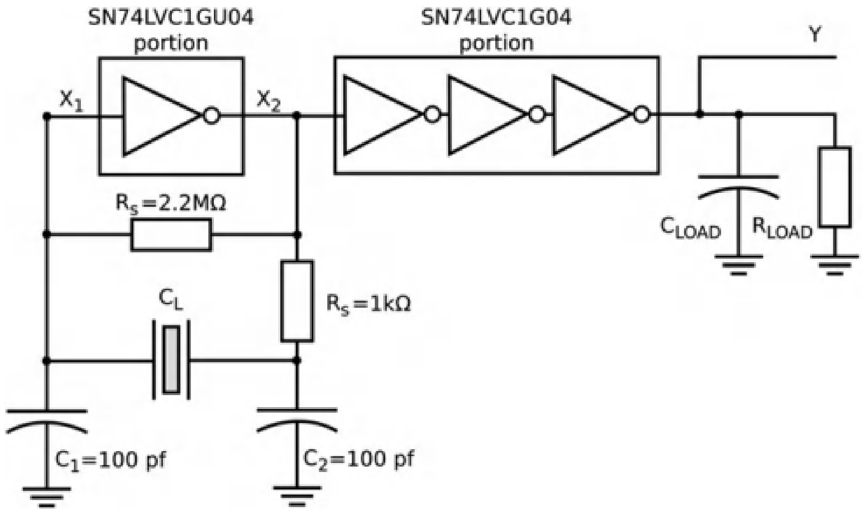


Fig. 6 Standard Pierce oscillator as the electronic interface for the openQCM device. Reprinted with permission from [20]

- the motional resistance, which is the value of the resistor in the motional branch of the BVD model, should be monitored;
- the parallel capacitance is a key parameter in determining the oscillation frequency.

Standard electronic oscillators have been widely used in QCM applications under various experimental setups. For example, openQCM, an open source QCM device designed for general purposes [20], uses a circuit based on the standard Pierce oscillator configuration. The IC oscillator incorporates an unbuffered inverter plus a standard buffered inverter into a single device, the first one is used as a linear amplifier for the crystal oscillator. The feedback network made by capacitors C1, C2, and the RF resistor ensures the phase shift for stable oscillations. The output of the oscillator circuit is a buffered square-wave output, whose frequency can be measured with a resolution of 0.1 Hz by using the microcontroller frequency counter, embedded in the openQCM device (Fig. 6).

The OpenQCM based on Pierce oscillator has been used in several scientific applications demonstrating the possibility of using this configuration of oscillators both in gas and in liquid. Researchers demonstrated the capability of using an array of openQCM electronic circuits for environmental sensing. Using an array of polymer-coated QCM sensors for selective gas sensing, they obtained results that are comparable to those of a high-end commercial QCM system [29] (Fig. 7).

Researchers also demonstrated the capability of using the openQCM oscillator as the electronic interface for QCM-based immunosensor for detecting small molecules [30]. In this case the quartz crystal electrode is antibody-functionalized using

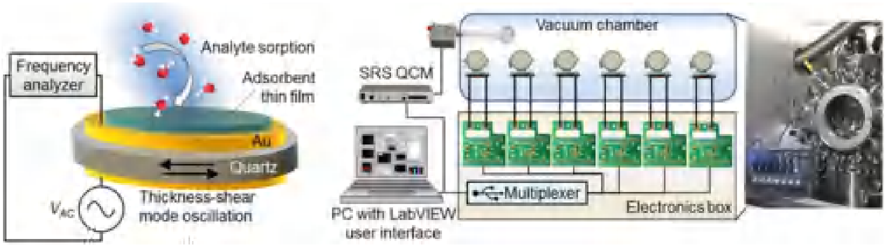


Fig. 7 Scheme of QCM operation using an array of six Pierce oscillators working in parallel for selective gas sensing. Reprinted with permission from [29]

Photochemical Immobilization Technique, which is capable of detecting pesticide like parathion in water with a limit of detection (LOD) approximately equal to 0.8 μg/L.

The Pierce standard oscillator has the disadvantage of not being able to have one face of the resonator grounded and does not guarantee stable oscillations for heavy loads or highly viscous media.

A variety of modifications to the standard oscillators have been proposed in QCM applications [31]. As suggested by Barnes, an oscillator integrating an Automatic Gain Control (AGC) system should be developed for QCM applications. The AGC system aims to keep constant the output amplitude of the oscillator. For this purpose, an input voltage is supplied to the system in order to modify the gain of the amplifier so that the signal is kept constant with respect to a reference. In addition, the AGC design claims also a proportionality between the supplied voltage and the variation of the motional resistance, which is related to the dissipation variation [32, 33].

3.2 Network or Impedance Analysis

The aim of the network or impedance analysis is to monitor the amplitude variations and the phase response of a quartz crystal resonator in order to completely characterize its electrical parameters. The network analysis principle is based on the passive interrogation of the quartz crystal sensor by sweeping the frequency around the resonance and analyzes its impedance/admittance behavior.

Unlike oscillator circuits, the network analysis has the advantage of passively interrogating the quartz sensor so that the circuit interface can′t alter the electrical parameters of the quartz resonator. Thanks to the passive operation of the quartz crystal; it is also possible to minimize parasitic influences and exclude their effects by means of a calibration procedure.

The main disadvantages of the standard network analysis are the high costs and the large dimensions, which do not consent the development of a portable and integrable QCM device. Furthermore, the passive quartz stimulation scheme does not allow for a multi-scan analysis of fundamental frequencies and overtones in a

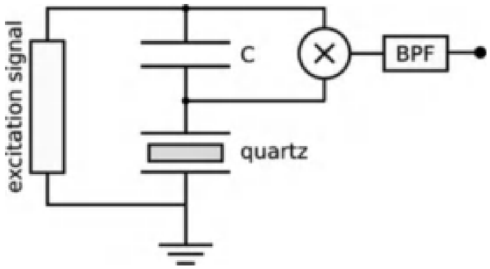


Fig. 8 Schematic diagram of the electronic interface, enhancing voltage divider-based network analyzer. Reprinted with permission from [34]

relatively short time, nor does it allow for the simultaneous interrogation of multiple sensors with the same interface.

Researchers have implemented various solutions by developing compact network analyzer electronic interfaces, in order to take advantage of the potentiality of the conventional bulky devices. Researchers developed a voltage divider-based network analyzer, whose basic principle is to drive the quartz crystal in voltage divider circuit made by the quartz itself and a series resistor of known value. By measuring the overall voltage of the divider, the voltage across the quartz crystal and phase shift between them, it is possible to calculate the unknown impedance of the quartz crystal.

An interesting approach, which enhances the voltage divider, was developed by Kankare and co-workers [35]. The schematic diagram of the electronic interface is shown in Fig. 8: the excitation signal fed to the voltage divider consists of a double-sideband suppressed carrier amplitude modulated signal whose carrier is swept around the resonance frequency range. By mixing the input excitation signal with the QCM output and removing high frequency and DC components using a band-pass filtering, the resulting output signal is formed by two coherent terms which contains information about both real and imaginary part of the surface load impedance. This strategy has special advantages compared to standard voltage divider techniques: (1) the output signal is mixed down to a low frequency region, which makes the signal acquisition and processing easier; (2) because the output signal is made of the difference of two coherent signals, any additive source of noise is automatically canceled; (3) the differential form of the output signal is advantageous in case of heavily loaded QCM resonators.

Recently a compact, reliable, and open source scalar network analyzer electronic interface for QCM has been developed by openQCM [20]. The device is capable of measuring simultaneously frequency and dissipation variations of the quartz crystal sensors. The electronic front-end mainly consists of a scalar network analyzer; the main block diagram is shown in Fig. 9. The scheme of measurement follows the principle of passive interrogation of the quartz sensor by sweeping around the resonance frequency. The actuation signal is generated using the AD9851 DDS/DAC frequency synthesizer, which can generate a sine wave with frequency from DC up to 72 MHz, with an output tuning resolution of about 0.04 Hz when clocked at 180 MHz. The output signal is read by AD8302 gain and phase detector, which is capable of measure the magnitude ratio, defined as gain, and phase

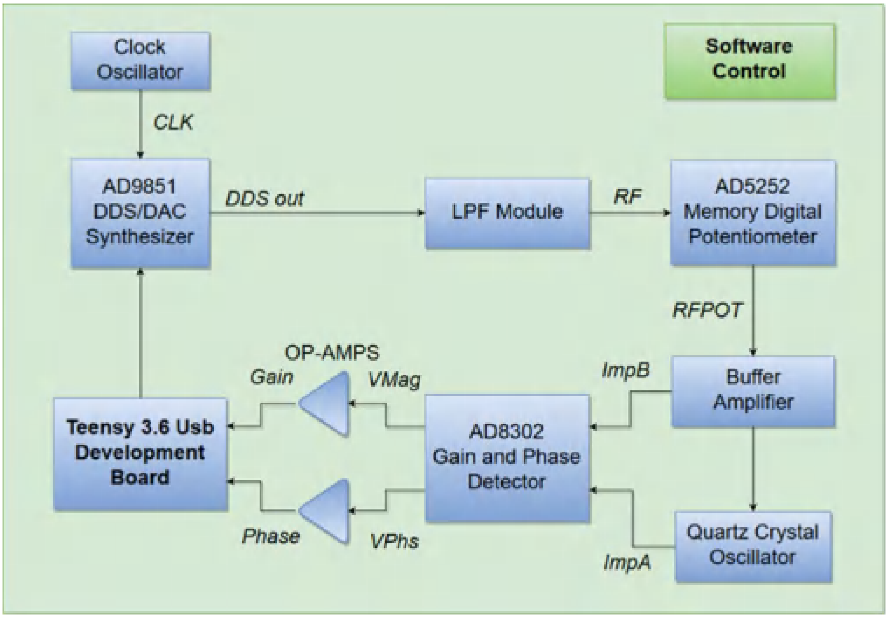


Fig. 9 Simplified block diagram of the scalar network analyzer electronic interface, developed for openQCM Q-1 with dissipation monitoring device. Reprinted with permission from [20]

difference between two signals, from low frequencies up to 2.7 GHz. The AD8302 measures the gain loss through the quartz crystal, referenced to the input signal, and simultaneously the phase lag between the response and actuation signal. The wide-range frequency output capability enables to interrogate the quartz sensor not only at the fundamental frequency but also at higher modes of vibration.

3.3 Functionalization Methods of the QCM Gold Surface

The sensitivity and specificity of QCM-based immunosensors is dependent on the immobilization of a recognition layer. Various strategies have been employed for the immobilization of antibodies on the crystal surface. Passive adsorption of protein/enzyme has been the most widely used method because the molecules can be attached to different interfaces with various mechanisms and forces acting between protein and surface, including hydrophobic, electrostatic, and van der Waals forces. Generally proteins tend to adhere onto hydrophobic surfaces due to the release of water dipoles from the hydrophobic surface to the bulk solution during protein adsorption [36]. In addition, the protein concentration plays an important role in the steady-state adsorption of proteins only on hydrophilic surfaces [37], not on hydrophobic ones [38].

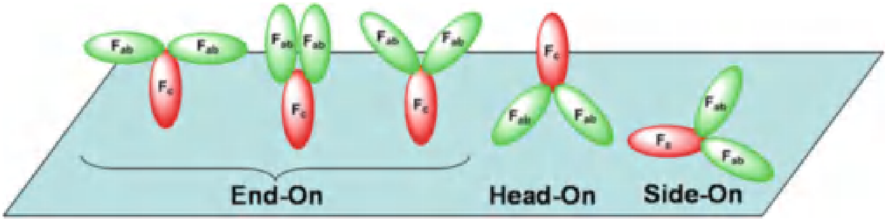


Fig. 10 Different possible orientations of IgG on a substrate. Reprinted with permission from [40]

Proteins can undergo conformational changes and unfold due to different forces acting during the adsorption process, affecting their biological activity. “Hard” proteins (lysozyme, β-lactoglobulin, α-chymotrypsin, etc.) upon adsorption retain their dissolved state without spreading. In contrast, “soft” proteins (bovine serum albumin (BSA), IgG, fibrinogen, or α-lactoglobulin) can spread on the surface upon adsorption. “Soft” proteins tend to adsorb onto various surfaces as compared to “hard” proteins that adsorb onto hydrophilic surfaces only when there is electrostatic attraction between protein molecules and the surface.

The surface roughness also affects adsorption and the morphology of adsorbed protein. For example, larger amounts are adsorbed to the hydrophobic surfaces as compared to the hydrophilic ones. Only on smooth hydrophilic or hydrophobic surfaces does collagen form elongated assemblies with small or high surface features, respectively [39]. Random orientation is often encountered with protein adsorption; that is, there are a few different possible orientations of IgG on the surface (Fig. 10).

Several strategies can be adopted to functionalize the surfaces covalently. One employs the initial generation of amino groups on the quartz surface by treatment with 3-aminopropyltriethoxysilane (APTES) followed by the activation of amine-functionalized surface with glutaraldehyde to generate aldehyde groups, which bind to the antibody through its amino groups. After the cleaning step, oxygen plasma treatment is done to generate OH groups on the surface, resulting in a reduction of the surface roughness and improving the formation of homogeneous layer. The gold surfaces are then carefully rinsed with DI water and dried under a stream of nitrogen gas. Finally, the substrates are dried in an oven at 110°C for 1 h to remove the moisture present on the surface. APTES solution is successfully prepared in pre-heated anhydrous toluene (100–120°C), and the cleaned gold surfaces are immersed in a solution of different concentrations of APTES for 12 h of silanization (incubation) time to reach the saturation for growing silane layer at room temperature in a nitrogen ambient.

APTES-modified gold surfaces are then washed in PBS and allowed to react with 2.5% (v/v) glutaraldehyde in PBS for 30 min at room temperature. This is followed by thoroughly rinsing the substrate with DI water to avoid non-specific adsorption of the antibody. The glutaraldehyde-activated surface is then reacted with 0.1 mg/mL of capture antibodies in PBS buffer along with 1% Tween 20 at room temperature for

15 min to get an antibody layer. At this point the gold surfaces are ready to detect the antigens [41]. In Fig. 11, a schematic representation of the functionalization of gold surfaces with the method above descripted is shown.

In addition, the crystallizable fragment (Fc)-binding proteins such as protein A, protein G, and protein A/G can be employed for an oriented immobilization of antibodies on the gold-coated QCM surface such that their antigen-binding sites (Fab region) are completely free for binding antigens. Protein A and Protein G are small proteins, derived from bacteria, which can specifically bind the Fc portion of antibodies allowing oriented systems to be obtained [42]. The IgG binding domain of Protein A, known as the Z-domain or ZZ-domain, is also used as a smaller synthetic option for Fc binding. This technique offers a method for truly obtaining oriented antibodies as binding can only occur via the Fc portion. Due to its effectiveness, protein A, or its derivatives, has been exploited with many surface immobilization strategies including biotin–streptavidin [43], SAMs [25, 26], EDC/NHS chemistry [44–46], glutaraldehyde [47], tyrosinase chemistry [47], non-natural amino acid insertion, gold-binding peptide or polystyrene affinity ligand fusion, and additional protein linkers [48]. It is worth to note that several issues may arise:

- the Protein A capture of the Fc is reversible, protein A has been reported to bind Fab regions and albumin (although to a much lesser extent);
- it is required that the Fc binding site of the Protein A is correctly oriented at the substrate to permit antibody binding [49].

Another method to immobilize the IgG antibodies onto gold surface is by using the interactions between thiols and gold, which are very strong and have been exploited for antibody binding by employing self-assembled layers of thiols and sulfides. One approach for site-oriented immobilization of immunoglobulins onto the gold supports consists in using the native immunoglobulin thiol groups, which are free after the splitting of the intact antibody into two half-IgG fragments without the destruction of the binding site of the antibody (Fig. 12). The half-IgG fragments can be immobilized onto gold supports by simple adsorption. The proposed approach is advantageous over existing methods because the immobilized antibodies maintain both high antigen-binding constants and high stability.

Another very elegant technique called photochemical immobilization technique (PIT) consists in a light-assisted approach for Ab immobilization and was adopted by Della Ventura et al. [51, 52]. Disulfide bonds are broken upon absorption of UV light by nearby aromatic amino acids, yielding reactive thiol groups that are effective for oriented binding onto gold electrodes. A working scheme of this technique is shown in Fig. 13.

The authors affirm that PIT preserves the native structure and the functional properties of the immobilized proteins while favoring the proper orientation of the biomolecule on the support. This is achieved by avoiding any chemical and thermal treatment. One of the advantages of PIT relies in the wide field of application since the closely spaced triad of residues Trp/Cys-Cys is present in all members of the immunoglobulin superfamily. Every IgG has 12 intradomain disulfide bridges near a

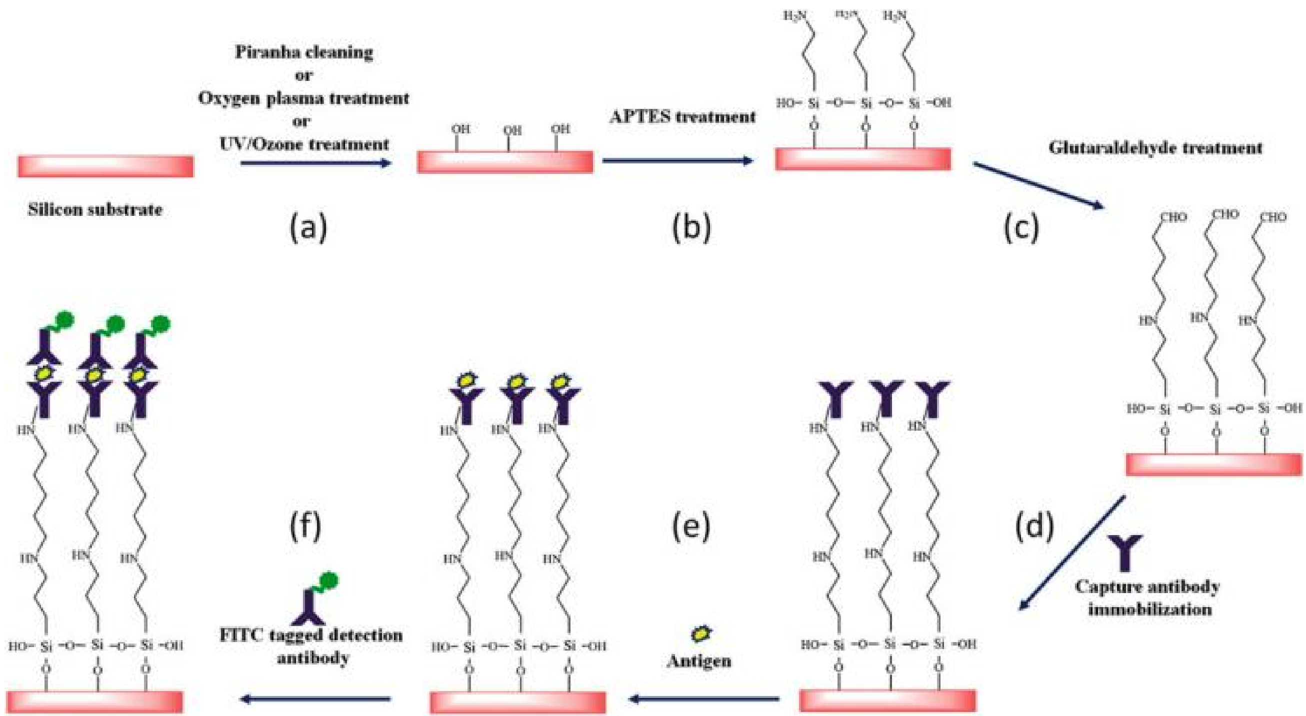


Fig. 11 Schematic representation of detection of biomolecules on bio-chemically treated silicon substrates using APTES and glutaraldehyde linker. The method involves the following: (a) silicon substrates cleaning (piranha cleaning, oxygen plasma treatment, or UV/ozone treatment); (b) generation of silane layer using optimized APTES (2%, v/v) solution; (c) treatment with glutaraldehyde solution; (d) immobilization of capture antibodies; (e) addition of antigen biomolecules; and (f) detection of antigens using FITC-tagged detection antibodies. Reprinted with permission from [41]

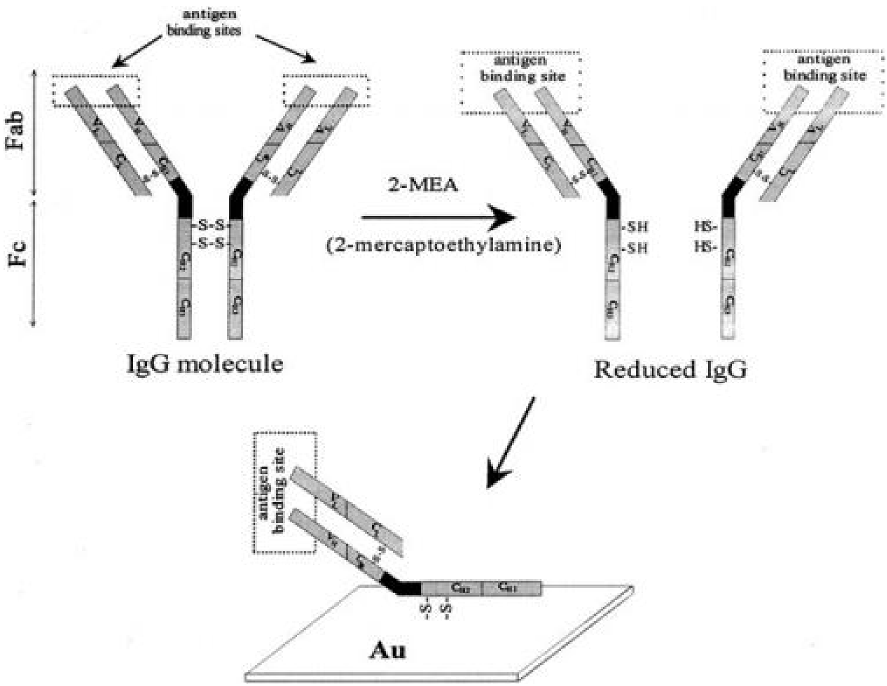


Fig. 12 The half-IgG fragments produced by reaction with 2-MEA immobilized onto gold supports by simple adsorption. The fragment antigen binding is free to recognize the analytes. Reprinted with permission from [50]

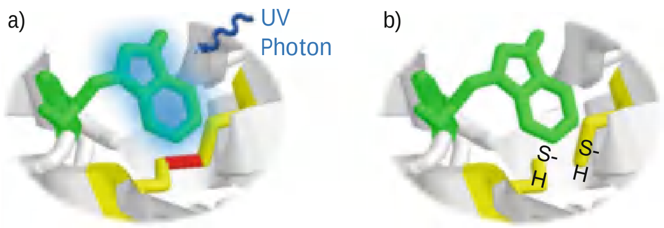


Fig. 13 (a) The protein solution is irradiated. (b) One UV photon is absorbed by a tryptophan side chain, which transfers the energy to the near cysteines. The disulfide bridge opens and the thiol groups so produced can effectively interact with the gold surface

tryptophan residue, one of them being present in every domain of the protein. It is likely to induce the opening of these disulfide bonds through UV irradiation of the near aromatic residue. With the breaking of these disulfide bridges, there is an increase of free thiol groups, which can react with a gold or thiol-rich surface.

3.4 Detection Step

QCM-based biosensor represents a new technology for the rapid detection, ease of use, low cost, online monitoring, shorter analysis time to measure pathogens, toxins, pesticides, and any analyte that can be recognized by aptamers, antibodies, or complementary DNA strand. Nevertheless it is worth to highlight that the dimensions of the analyte to recognize play a key role in the detection step because in case of large molecules no strategy is needed and a direct frequency shift will be shown. On contrary, when the molecules to detect are of medium or small size, it is indispensable to use different strategies to ballast them and to improve the limit of detection (LOD).

Starting from the detection of large molecules, *Escherichia coli* O157:H7 (*E. coli* O157:H7) is one of the most studied water contaminants because it is a dangerous pathogen. It causes serious illnesses such as bloody diarrhea, bloody feces, anemia, and kidney failure. Hence, an establishment of rapid and sensitive methods for *E. coli* O157:H7 detection is strongly needed to control this pathogenic bacterium in water supplies or food. Traditional methods for testing of *E. coli* O157:H7 include plating and culturing, enumeration methods, and biochemical testing. Although the detection limits for these methods are very low (about a few colony-forming units (CFU)/ml), the testing time is time-consuming (from 1 day to 1 week). Besides, some new techniques for rapid detection of this bacteria have been developed including immunoassays, polymerase chain reaction (PCR) [8], DNA microarrays, and immunomagnetic separations.

It has been shown that sensitivity and selectivity of these methods are good and detection time for these methods is from about 2 h to 24 h. However, these methods have a disadvantage in that they are expensive or complicated due to the use of laboratories equipped with specific instruments and chromospheres. Therefore, they are not suitable for rapid test of *E. coli* O157:H7 bacteria. Nurliyana et al. [53] developed a QCM-based biosensor anchoring the antibodies onto gold surface with SAMs method. The LOD they were able to reach was 10^2–10^3 CFU/mL of *E. coli* O157:H7 observing frequency shifts of about 15 Hz and 34 Hz, respectively. This result allows QCM system to be used for qualitative and quantitative analysis of cell concentration in solution. Fulgione et al. [54] reached an LOD of 10^0 CFU/mL in the detection of *Salmonella typhimurium* in chicken meat with a relatively simple protocol which requires a pre-enrichment step lasting only 4 h at 37°C.

The reliability of the proposed immunosensor has been demonstrated through the validation of the experimental results with ISO standardized culture method which takes up to 10 days to provide a reliable response. In order to further improve the LOD of bacteria, some authors developed a quartz crystal microbalance immunosensor for detection of *E. coli* O157:H7 by self-assembling of protein A and affinity-purified anti-*E. coli* O157:H7 antibodies on the gold electrode of an AT-cut piezoelectric quartz crystal. To enhance the sensitivity of the QCM immunosensor, nanoparticle–antibody conjugates, which were prepared using streptavidin-conjugated nanoparticles (145 nm diameter) and biotinylated anti-*E. coli* antibodies, were used

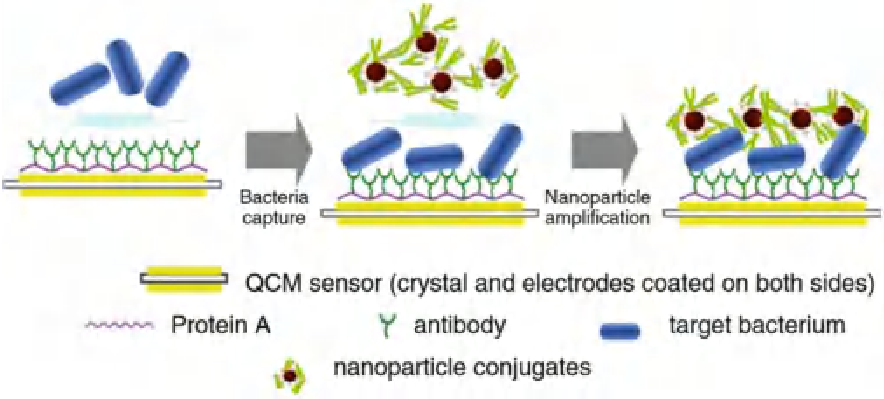


Fig. 14 Use of the nanoparticle–antibody conjugates for signal amplification in the detection of E. coli O157:H7 with QCM immunosensor. Reprinted with permission from [55]

for signal amplification. After the binding of E. coli O157:H7 cells with the antibodies immobilized on the electrode, nanoparticle–antibody conjugates were introduced as mass amplifiers (Fig. 14). Compared to the direct detection, the binding of the nanoparticle conjugates further resulted in a decrease in resonant frequency and an increase in resonant resistance, and the detection sensitivity was improved by lowering the detection limit until to 10^1 CFU/mL [55].

The same strategy has been used to detect water contaminants with small/medium dimensions such as organic compounds. An example is the detection of parathion, a pesticide ($M = 292$ Da) for which a signal amplification procedure is desirable since the signal they induce in the transducer, and specifically in a QCM, is undetectable. The authors extend the application of such a method to small analytes by showing that once the working surface of a QCM has been properly functionalized, a limit of detection lower than 1 ppb is reached for parathion. The strategy adopted to enhance the sensitivity of the QCM-based immunosensors is like sketched in Fig. 14. Thanks to PIT the Abs are covalently bound to the QCM gold surface with antigen-binding site exposed to the fluid. The solution containing the parathion is subsequently conveyed to the cell, and the analyte is recognized by the Abs. At this stage, no appreciable signal is detected in view of small mass of the parathion, but the following interaction with a ballast constituted by functionalized Au-NPs allows one to detect the presence of small molecules. In this scheme the same Abs used to functionalize the gold surface of the QCM are tethered to the Au-NPs. In Fig. 15 the output of the QCM is shown, which includes the functionalization (steps I–III) and the measurement (steps IV–VI) of parathion at 290 μg/L [30].

Gold nanoparticles have been used also to detect Hg^{2+} by Zhong et al. [56]. The authors realized a short mercury-specific aptamer (MSA) along with gold nanoparticles (Au-NPs) to determine Hg(II) ion by a combination of a QCM-based sensor and a flow system. The MSA binds specifically to Hg(II), and the Au-NPs can

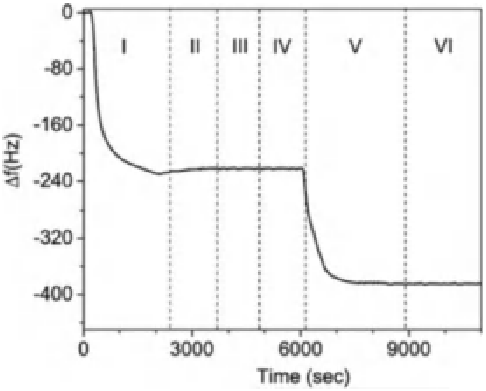


Fig. 15 I: Functionalization by PIT (Abs are tethered to the probe surface); II: the washing by PBS solution (1×) removes unspecific bonds giving rise to a small increase in the frequency; III: BSA is conveyed to the cell and no frequency change is observed warranting the full coverage of the surface: IV: parathion is injected, but no frequency change is observed because of its small mass; V: the injection of the Au-NPs complexed with Abs against parathion yields a huge frequency change; VI: the eventual washing with PBS (1×) does not change the frequency since all the bonds are specific

amplify the signal to enhance sensitivity. Specifically, the short thiolated MSAs are immobilized on the surface of the QCM as the capture probe, and the MSAs are linked to the Au-NPs as the linking probe. The two components can form a sandwich structure of the T-Hg(II)-T type in the presence of Hg(II) ions. This leads to change in the mass on the QCM and a change in the resonance frequency. Hg(II) can be determined with a detection limit of 0.24 ± 0.06 nM, which is three orders of magnitude better than previous methods. The sensor can be regenerated by disrupting the T-Hg(II)-T base pairs with a solution of cysteine. The water can contain also phosphates, nitrates, ammonia, and their mixtures that can leach into the groundwater. Those chemicals cause a number of health issues, such as kidney problems and cancer. These risks are high for pregnant and nursing women, infants, and the elderly. Those contaminants are not so easy to detect with QCM, and strategies are necessary to recognize them. Ayad et al. [57] used a thin-film polyaniline (PANI) as the sensing materials. The adsorption of the compound onto the PANI-coated QCM resulted in increasing the frequency shift due to hydrophilic nature of the film and electrostatic interactions.

However, QCM might be appropriate for the development of potential bio-analytical systems that are based on the use of various sensing technologies into a single system. It is possible to integrate QCM with SPR (surface plasmon resonance) and with electrochemical detection to extend the range of applications. The reuse of the crystal is another concern for low-cost applications. Following various cleaning procedures, it is possible to regenerate the electrodes for 3–4 times. QCMs have been widely used in a broad range of analytical applications as the sensing procedure is simple, cost-effective, non-hazardous, real time, and less time-consuming. QCM results to be an appropriate sensing platform for the online monitoring of analytes in water.

4 Conclusions

The availability of in-situ measurements and multiple detection analyses has expanded water monitoring applications in various advanced techniques including successful development in hand-held sensing devices. High sensitivity and real-time monitoring of contaminants in water is offered by quartz crystal microbalance (QCM) that is a very attractive technique for a large range of applications. A major advantage of the technique used for liquid systems is that it allows for a label-free detection of molecules. QCM is capable of measuring mass changes as small as a fraction of a monolayer of atoms. QCM crystals are becoming a good alternative analytical method in a great deal of applications such as biosensors, analysis of biomolecular interactions, study of bacterial adhesion at specific interfaces, and pathogen and microorganism detection.

References

1. Richardson SD, Thruston AD, Collette TW, Patterson KS, Lyklns BW, Majetich G, Zhang Y (1994) Multispectral identification of chlorine dioxide disinfection byproducts in drinking water. Environ Sci Technol 28(4):592–599
2. Archibald RG (1927) An unusual case of infection with a non-motile strain of B. typhosus in an individual previously inoculated with T. A. B. vaccine. J Trop Med Hyg 30:125–126
3. Slanetz LW, Bent DF, Bartley CH (1955) Use of the membrane filter technique to enumerate enterococci in water. Public Health Rep 70(1):67
4. Kabler P (1954) Water examinations by membrane filter and most probable number procedures. Am J Public Health Nations Health 44(3):379–386
5. Figueiró CSM, Bastos de Oliveira D, Russo MR, Caires ARL, Rojas SS (2018) Fish farming water quality monitored by optical analysis: the potential application of UV–Vis absorption and fluorescence spectroscopy. Aquaculture 490:91–97
6. Savichtcheva O, Okayama N, Ito T, Okabe S (2005) Application of a direct fluorescence-based live/dead staining combined with fluorescence in situ hybridization for assessment of survival rate of Bacteroides spp. in drinking water. Biotechnol Bioeng 92(3):356–363
7. Bosserhoff AK, Hellerbrand C (2009) Capillary electrophoresis. Molecular diagnostics: second edition. Elsevier, Amsterdam, pp 59–73
8. Kenndler E, Maier NM (2018) Capillary electrophoresis in organic solvents. Capillary electromigration separation methods. Elsevier, Amsterdam, pp 69–111
9. El Hadri H, Gigault J, Chéry P, Potin-Gautier M, Lespes G (2014) Optimization of flow field-flow fractionation for the characterization of natural colloids field-flow fractionation. Anal Bioanal Chem 406(6):1639–1649
10. Lasagna C, Raffo E, Foppiano D (2011) Potential presence of trihalomethanes in water intended for human consumption. J Biol Res 84(1):111–113

11. Yuan T, Hu JY, Ong SL, Ng WJ (2003) Simultaneous analysis of five selected phenolic xenoestrogens in water sample using solid phase extraction-gas chromatography-mass spectrometry. Chromatographia 58(9):643–648. https://doi.org/10.1365/s10337-003-0067-x
12. Byrne B, Stack E, Gilmartin N, O′Kennedy R (2009) Antibody-based sensors: principles, problems and potential for detection of pathogens and associated toxins. Sensors (Switzerland) 9(6):4407–4445
13. Jokisaari JR, Hachtel JA, Hu X, Mukherjee A, Wang C, Konecna A, Lovejoy TC, Dellby N, Aizpurua J, Krivanek OL, Idrobo JC, Klie RF (2018) Vibrational spectroscopy of water with high spatial resolution. Adv Mater 30(36):1802702
14. Della Ventura B, Iannaccone M, Funari R, Pica Ciamarra M, Altucci C, Capparelli R, Roperto S, Velotta R (2017) Effective antibodies immobilization and functionalized nanoparticles in a quartz-crystal microbalance-based immunosensor for the detection of parathion. PLoS One 12(2):e0171754
15. Lemarchand K, Lebaron P (2002) Influence of mutation frequency on the persistence of Salmonella enterica serotypes in natural waters. FEMS Microbiol Ecol 41(2):125–131
16. Park M-K, Oh J-H (2012) Rapid detection of E. coli O157:H7 on turnip greens using a modified gold biosensor combined with light microscopic imaging system. J Food Sci 77(2):M127–M134
17. Cassiède M, Daridon JL, Paillol JH, Pauly J (2010) Impedance analysis for characterizing the influence of hydrostatic pressure on piezoelectric quartz crystal sensors. J Appl Phys 108 (3):034505
18. Buszewski B, Grzywinski D, Ligor T, Stacewicz T, Bielecki Z, Wojtas J (2013) Detection of volatile organic compounds as biomarkers in breath analysis by different analytical techniques. Bioanalysis 5(18):2287–2306
19. Sauerbrey G (1959) Verwendung von Schwingquarzen zur Wägung dünner Schichten und zur Mikrowägung. Z Phys 155(2):206–222
20. OpenQCM. The first open source quartz crystal microbalance [Online]. https://openqcm.com
21. Mecea VM (2005) From quartz crystal microbalance to fundamental principles of mass measurements. Anal Lett 38(5):753–767
22. Kanazawa KK, Gordon JG (1985) Frequency of a quartz microbalance in contact with liquid. Anal Chem 57(8):1770–1771
23. Martin SJ, Frye GC, Ricco AJ, Senturia SD (1993) Effect of surface roughness on the response of thickness-shear mode resonators in liquids. Anal Chem 65(20):2910–2922
24. Johannsmann D (1999) Viscoelastic analysis of organic thin films on quartz resonators. Macromol Chem Phys 200:501–516
25. Janshoff A, Steinem C, Sieber M, El Bayâ A, Schmidt MA, Galla HJ (1997) Quartz crystal microbalance investigation of the interaction of bacterial toxins with ganglioside containing solid supported membranes. Eur Biophys J 26(3):261–270
26. Bottom VE (1982) Introduction to quartz crystal unit design. Van Nostrand Reinhold, New York
27. Voinova MV, Jonson M, Kasemo B (2002) ′Missing mass′ effect in biosensor′s QCM applications. Biosens Bioelectron 17(10):835–841
28. Barnes C (1991) Development of quartz crystal oscillators for under-liquid sensing. Sensors Actuators A Phys 29(1):59–69
29. Muckley ES, Anazagasty C, Jacobs CB, Hianik T, Ivanov IN (2016) Low-cost scalable quartz crystal microbalance array for environmental sensing. Organic sensors and bioelectronics IX, vol 9944. International Society for Optics and Photonics, San Diego, p 99440Y
30. Funari R, Della Ventura B, Carrieri R, Morra L, Lahoz E, Gesuele F, Altucci C, Velotta R (2015) Detection of parathion and patulin by quartz-crystal microbalance functionalized by the photonics immobilization technique. Biosens Bioelectron 67:224–229. https://doi.org/10.1016/j.bios.2014.08.020
31. Arnau A (2008) A review of interface electronic systems for AT-cut quartz crystal microbalance applications in liquids. Sensors 8(1):370–411

32. IPCC (2013) Climate change 2013: the physical science basis. Summary for policymakers. In: IPCC fifth assessment report, pp 1–29
33. Hayward GL, Chu GZ (1994) Simultaneous measurement of mass and viscosity using piezoelectric quartz crystals in liquid media. Anal Chim Acta 288(3):179–185
34. Chagnard C, Gilbert P, Watkins AN, Beeler T, Paul DW (1996) An electronic oscillator with automatic gain control: EQCM applications. Sensors Actuators B Chem 32(2):129–136
35. Kankare J, Loikas K, Salomäki M (2006) Method for measuring the losses and loading of a quartz crystal microbalance. Anal Chem 78(6):1875–1882
36. Rabe M, Verdes D, Seeger S (2011) Understanding protein adsorption phenomena at solid surfaces. Adv Colloid Interf Sci 162(1–2):87–106
37. Bremer MGEG, Duval J, Norde W, Lyklema J (2004) Electrostatic interactions between immunoglobulin (IgG) molecules and a charged sorbent. Colloids Surf A Physicochem Eng Asp 250(1–3):29–42
38. van der Veen M, Stuart MC, Norde W (2007) Spreading of proteins and its effect on adsorption and desorption kinetics. Colloids Surfaces B Biointerfaces 54(2):136–142
39. Denis FA, Hanarp P, Sutherland DS, Gold J, Mustin C, Rouxhet PG, Dufrêne YF (2002) Protein adsorption on model surfaces with controlled nanotopography and chemistry. Langmuir 18 (3):819–828
40. Vashist SK, Lam E, Hrapovic S, Male KB, Luong JHT (2014) Immobilization of antibodies and enzymes on 3-aminopropyltriethoxysilane-functionalized bioanalytical platforms for biosensors and diagnostics. Chem Rev 114(21):11083–11130. https://doi.org/10.1021/cr5000943
41. Gunda NSK, Singh M, Norman L, Kaur K, Mitra SK (2014) Optimization and characterization of biomolecule immobilization on silicon substrates using (3-aminopropyl)triethoxysilane (APTES) and glutaraldehyde linker. Appl Surf Sci 305:522–530
42. Iijima M, Yoshimoto N, Niimi T, Maturana AD, Kuroda S (2013) Nanocapsule-based probe for evaluating the orientation of antibodies immobilized on a solid phase. Analyst 138(12):3470
43. Chung JW, Park JM, Bernhardt R, Pyun JC (2006) Immunosensor with a controlled orientation of antibodies by using NeutrAvidin–protein A complex at immunoaffinity layer. J Biotechnol 126(3):325–333
44. Crivianu-Gaita V, Aamer M, Posaratnanathan RT, Romaschin A, Thompson M (2016) Acoustic wave biosensor for the detection of the breast and prostate cancer metastasis biomarker protein PTHrP. Biosens Bioelectron 78:92–99
45. Makaraviciute A, Ramanavicius A, Ramanaviciene A (2015) Development of a reusable protein G based SPR immunosensor for direct human growth hormone detection in real samples. Anal Methods 7(23):9875–9884
46. Verma SK, Amoah A, Schellhaas U, Winterhalter M, Springer S, Kolesnikova TA (2016) ′To catch or not to catch′: microcapsule-based sandwich assay for detection of proteins and nucleic acids. Adv Funct Mater 26(33):6015–6024
47. Ahmed S, Lutes A, Barbaril T (2006) Specific capture of target proteins by oriented antibodies bound to tyrosinase-immobilized protein A on a polyallylamine affinity membrane surface. J Memb Sci 282(1–2):311–321
48. Miyao H, Ikeda Y, Shiraishi A, Kawakami Y, Sueda S (2015) Immobilization of immunoglobulin-G-binding domain of protein A on a gold surface modified with biotin ligase. Anal Biochem 484:113–121
49. Liu Y, Yu J (2016) Oriented immobilization of proteins on solid supports for use in biosensors and biochips: a review. Microchim Acta 183(1):1–19
50. Karyakin AA, Presnova GV, Rubtsova MY, Egorov AM (2000) Oriented immobilization of antibodies onto the gold surfaces via their native thiol groups. Anal Chem 72(16):3805–3811
51. Della Ventura B, Schiavo L, Altucci C, Esposito R, Velotta R (2011) Light assisted antibody immobilization for bio-sensing. Biomed Opt Express 2(11):3223–3231
52. Funari R, Della Ventura B, Altucci C, Offenhäusser A, Mayer D, Velotta R (2016) Single molecule characterization of UV-activated antibodies on gold by atomic force microscopy. Langmuir 32(32):8084–8091

53. Nurliyana MR, Sahdan MZ, Wibowo KM, Muslihati A, Saim H, Ahmad SA, Sari Y, Mansor Z (2018) The detection method of escherichia coli in water resources: a review. J Phys Conf Ser 995(1)
54. Fulgione A, Cimafonte M, Della Ventura B, Iannaccone M, Ambrosino C, Capuano F, Proroga YTR, Velotta R, Capparelli R (2018) QCM-based immunosensor for rapid detection of Salmonella Typhimurium in food. Sci Rep 8(1):16137
55. Liu F, Li Y, Su X-L, Slavik M, Ying Y, Wang J (2007) QCM immunosensor with nanoparticle amplification for detection of Escherichia coli O157:H7. Sens & Instrumen Food Qual 1:161–168
56. Dong Z-M, Zhao G-C (2012) Quartz crystal microbalance aptasensor for sensitive detection of mercury(II) based on signal amplification with gold nanoparticles. Sensors 12(6):7080–7094
57. Ayad MM, El-Hefnawey G, Torad NL (2008) Quartz crystal microbalance sensor coated with polyaniline emeraldine base for determination of chlorinated aliphatic hydrocarbons. Sensors Actuators B Chem 134(2):887–894