Data-driven water distribution system analysis – exploring challenges and potentials from smart meters and beyond

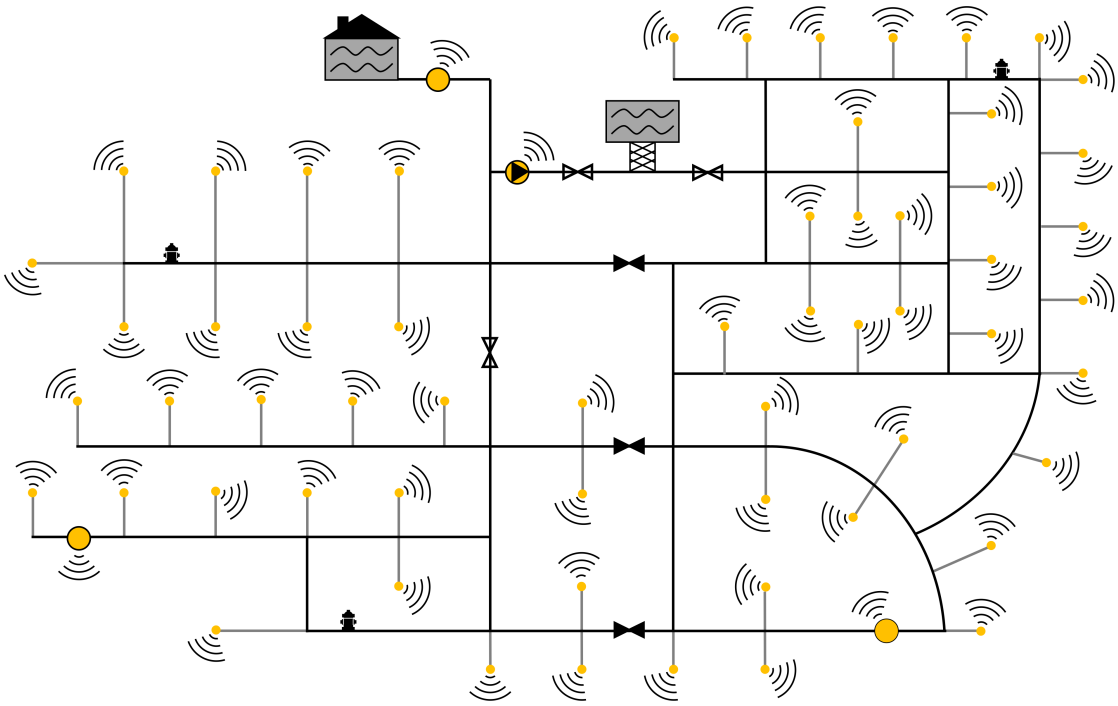
Kirstein, Jonas Kjeld

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Kirstein, J. K. (2020). Data-driven water distribution system analysis – exploring challenges and potentials from smart meters and beyond. Technical University of Denmark.

Data-driven water distribution system analysis – exploring challenges and potentials from smart meters and beyond



Jonas Kjeld Kirstein

PhD Thesis
February 2020

DTU Environment
Department of Environmental Engineering

Data-driven water distribution system analysis – exploring challenges and potentials from smart meters and beyond

Jonas Kjeld Kirstein

PhD Thesis
February 2020

DTU Environment
Department of Environmental Engineering
Technical University of Denmark

Data-driven water distribution system analysis – exploring challenges and potentials from smart meters and beyond

Jonas Kjeld Kirstein

PhD Thesis, February 2020

The synopsis part of this thesis is available as a pdf file for downloading from the DTU research database ORBIT: http://www.orbit.dtu.dk.

Address: DTU Environment
Department of Environmental Engineering
Technical University of Denmark
Miljoevej, building 113
2800 Kgs. Lyngby
Denmark

Phone reception: +45 4525 1600
Fax: +45 4593 2850

Homepage: http://www.env.dtu.dk
E-mail: reception@env.dtu.dk

Preface

This PhD thesis was conducted at the Department of Environmental Engineering at the Technical University of Denmark (DTU). The work was guided by main supervisor Assoc. Prof. Martin Rygaard, co-supervisor Assoc. Prof. Morten Borup and co-supervisor Klavs Høgh from NIRAS A/S. The PhD study was supported by the LEAKman project and partners.

This thesis is based on six papers and these will be referred to in the text by their paper number in Roman numerals (I–VI).

I **Kirstein, J.K.,** Høgh, K., Rygaard, M. & Borup, M. (2019). Effect of data sampling resolution of smart meter readings in water distribution network simulations. *Manuscript in preparation.*

II **Kirstein, J.K.,** Høgh, K., Rygaard, M. & Borup, M. (2019). A semi-automated approach to validation and error diagnostics of water network data. *Urban Water Journal,* **16**(1)**,** 1–10, doi:10.1080/1573062X.2019.1611884

III **Kirstein, J.K.,** Liu, S., Høgh, K., Borup, M. & Rygaard, M. (2019). Valve status identification by temperature modelling in water distribution networks. *Manuscript in preparation.*

IV **Kirstein, J.K.,** Høgh, K., Rygaard, M. & Borup, M. (2019). Using smart meter temperature and consumption data for water distribution system analysis. *Manuscript in preparation.*

V Hubeck-Graudal, H., **Kirstein, J.K.,** Ommen, T., Rygaard, M. & Elmegaard, B. (2019). Drinking water supply as low-temperature source in the district heating system: a case study for the city of Copenhagen. *Submitted*.

VI Lund, N.S.V., **Kirstein, J.K.,** Mikkelsen, P.S., Madsen, H., Mark, O. & Borup, M. (2019). Using smart meter water consumption data and in-sewer flow observations for model based sewer system analysis. *Submitted.*

In this online version of the thesis, Papers **I–VI** are not included but can be obtained from electronic article databases e.g. via www.orbit.dtu.dk or on request from DTU Environment, Technical University of Denmark, Miljoevej, Building 113, 2800 Kgs. Lyngby, Denmark, info@env.dtu.dk.

In addition, the following publications and conference contributions, not included in this thesis, were also concluded during this PhD study:

a) **Kirstein, J.K.,** Høgh, K., Borup, M. & Rygaard, M. (2016). Fra big data til smart data: driftsoptimering med højopløste sektionsdata i vandforsyningen. Casestudie: Halsnæs Forsyning. Dansk Vand Konference 2016, November 8-9, Århus, Denmark. *Abstract & oral presentation.*

b) **Kirstein, J.K.,** Borup, M., Rygaard, M. & Høgh, K. (2016). Målerdata kan gemme på gratis informationer for forsyningerne: Et case studie fra Halsnæs Forsyning baseret på højopløste data. *DanskVand*, **84** (6), 50–51. *Article.*

c) **Kirstein, J. K.,** Høgh, K., Borup, M. & Rygaard, M. (2018). Is your data correct? Validating and improving data collected in smart water networks. Nordic Drinking Water Conference 2018, June 11–13, Oslo, Norway. *Abstract & oral presentation.*

d) Høgh, K. & **Kirstein, J.K.** (2018). ICT Frameworks – Moving Towards Smart Water Networks. IWA World Water Congress & Exhibition 2018, 16–21 September, Tokyo, Japan. *Abstract.*

e) **Kirstein, J.K.,** Høgh, K., Borup, M. & Rygaard, M. (2018). Identifikation af ventilindstillinger fra temperaturmålinger. Dansk Vand Konference 2018, November 13–14, Århus, Denmark. *Abstract & oral presentation.*

f) **Kirstein, J.K.,** Høgh, K., Borup, M. & Rygaard, M. (2019). Valve status identification by temperature modelling in water distribution networks. 13th Danish Water Forum Conference, January 31, Copenhagen, Denmark. *Abstract & oral presentation.*

g) **Kirstein, J.K.,** Høgh, K., Rygaard, M. & Borup, M. (2019). Valve status detection using smart meter temperature and flow. 17th International Computing & Control for the Water Industry Conference, September 1–4, Exeter, UK. *Abstract & oral presentation.*

h) Lund, N.S.V., Borup, M., **Kirstein, J.K.,** Mark, O., Madsen, H. & Mikkelsen, P.S. (2019). Lessons learned from comparing smart meter water consumption data with measured wastewater flow in the drainage system. 17th International Computing & Control for the Water Industry Conference, September 1–4, Exeter, UK. *Abstract.*

Furthermore, a meter data validation tool based on the tests presented in Paper II was made available online at https://leakagemanagement.net/meter-validate.

Acknowledgements

This PhD project was conducted at DTU Environment at the Technical University of Denmark in collaboration with the LEAKman project partners.

First, I should like to thank my two main supervisors, Martin Rygaard and Morten Borup. Their doors were always open to me during the project irrespective of timing, question or topic. They guided me when needed and I greatly appreciated their support when the project tended to take second place to my personal life. I will miss our supervisor meetings, at which they were at times entertainingly contradictory, ultimately increasing our mutual knowledge of water distribution system analyses.

Second, I should like to thank my supervisor, Klavs Høgh for sharing his professional knowledge and experience, often highlighting the gap between academia and the 'real world'. Also, I should like to thank the VAF2 team at NIRAS A/S, in particular, Gaby, Gitte, Jesper, Adam, Andreas, Thomas, Lars, Jan, Anders, Janus, Nikolai, Rosa and Pia for help with all kinds of questions related to water supply and for always making me feel welcome at NIRAS.

Third, I should like to thank Professor Shuming Liu and his research team at Tsinghua University. They made me feel welcome from day one of my stay in China. I learned a lot from being a member of their group – from an academic as well as a personal point of view. I am sure this was not my last visit to China.

Next, I should like to thank Mette Fogtmann and Bo Lindhardt from Novafos and Per Grønvald, Henrik Horsholt Christensen and Thorkil Bartholdy Neergaard from Brønderslev Forsyning for sharing data and knowledge related to their water supply systems.

I should also like to thank the project partners of the LEAKman project, in particular Kamstrup A/S and HOFOR for their input and financial support to my PhD project.

I also want to praise and thank Frank for his help with my English during the project and, in particular, Nadia, for her never-ending support.

Last but not least, my final thanks go to my dad who, nine years ago, persuaded me to follow an education in environmental engineering instead of computer sciences. "You will end up in front of a computer anyway." Well, he was right!

Summary

The availability of clean drinking water at any time is often taken for granted. However, the reliable supply of drinking water is, unfortunately, challenged by various threats, such as global warming and aging infrastructures, which put an immense pressure on the drinking water systems as we know them today. Consequently, utilities, technology providers and researchers seek to identify optimised and new approaches to maintaining and improving the quality of the delivered water. In an age of digitalisation, data-driven approaches are becoming increasingly important, as they have demonstrated various benefits for the operation and design of water distribution networks. However, the increased collection and application of the data also pose a major challenge to the water sector. This PhD developed methods to help utilities in validating and applying their data in novel ways by analysing 'real world' data obtained from five Danish utilities, applied in six case studies. The PhD study was sectioned into: 1) data collection; 2) data validation and reconstruction; and 3) data application.

Data-collection devices such as smart meters are increasingly deployed throughout water distribution systems. When utilities introduce smart meters, the selection of sampling resolutions has a trade-off between the applicability of the collected data and transmission costs. Analysis of a district metered area with smart meters installed revealed that common sampling resolutions of between 1 and 24 hours are sufficient for water loss assessments as long as utilities have representative demand patterns of their network available. However, sampling resolutions < 1 hour are potentially important to obtain reliable water quality simulations.

Automatic validation and reconstruction processing of the collected data are of paramount importance for utilities. The PhD project developed a systematic approach for categorizing anomalies. Four categories were introduced, with Type 0 describing system anomalies, Types 1 and 2 describing sensor data containing a low data quality, and Type 3 covering sensor data anomalies storing information about actual – though unusual – events appearing in the water distribution network. To identify anomalies of Types 1 and 2, seven validation tests were developed. Analysis of pressure and flow data sets from three Danish utilities revealed a large proportion of anomalies, with on average 10% missing data and up to 35% anomalies of Types 1 and 2 in a utility's pressure data sets. These high numbers also emphasised the need for reconstruction processes to generate reliable data streams that are required in data applications. An example was presented whereby artificial neural networks were used to provide

missing data and to further validate dubious observations.

The collection of data from water distribution systems is not a new concept, but large amounts of data (such as temperature data) are often left unused due to a lack of evidence of successful applications. To show the benefits of temperature data, a temperature model and a hydraulic model were combined to identify the status and location of valves in the network. This novel approach and field tests in the network unexpectedly revealed various anomalies of Type 0 in the utility's asset database, ultimately casting doubt on the validity of the hydraulic model. As long as such anomalies prevail in the data sets, it is not possible to apply advanced data-driven applications successfully. Another issue in the case study included a low quantity of applicable temperature data. Smart meter temperature data, potentially available in each household, can be used to overcome this challenge. In another case study, the simulated temperature throughout a district metered area showed a satisfying resemblance to smart meter temperature data (average root mean square error of 0.9 ºC). This highlights the potential of using smart meter temperature data for more advanced applications, such as leakage detection and valve status detection.

Silo thinking is traditionally a common feature of the water sector, and the value of water supply data is thus often overlooked in external applications. In a case study, the effect of deploying heat pumps on the water distribution network mains was assessed as a supplement to the district heating system of Copenhagen. A net heat extraction potential of 20.7 MW was estimated. Moreover, this caused the share of users complying with an upper temperature limit of 12 ºC to increase from 41% to 81% during August. In another case study, smart meter water consumption data were linked to an urban drainage model to compare simulated wastewater flows with in-sewer observations. The in-sewer observations were found to be erroneous, and the smart meter data were deemed more valid in estimating dry weather flow than in-sewer observations.

Overall, the project showed that many anomalies prevailing in the utilities' asset databases and sensor data are first discovered through the application of the data. As long as utilities cannot maintain a high level of data reliability, it is doubtful whether more sensors will increase utilities' understanding of their systems. The true potential of the data will not be unlocked until a high level of data reliability is secured. In the coming years, utilities, technology providers and researchers should together identify methods for reducing the uncertainties prevailing in asset and sensor data, making it possible for the sector to reach higher levels of digital maturity.

Dansk sammenfatning

Tilgængelighed af rent drikkevand til enhver tid tages ofte som en selvfølge. Desværre er den høje forsyningssikkerhed truet af en række udfordringer, såsom global opvarmning og aldrende infrastruktur, hvilket lægger et enormt pres på drikkevandssystemerne, som vi kender dem i dag. Forsyningsselskaber, teknologileverandører og forskere arbejder derfor på at identificere optimerede og nye tilgange til at opretholde og forbedre kvaliteten af det leverede vand. I digitaliseringens tidsalder bliver datadrevne fremgangsmåder mere og mere vigtige, da disse har vist sig fordelagtige inden for drift og design af vandledningsnet. Den øgede indsamling og anvendelse af data udgør dog også en stor udfordring for vandsektoren. Ph.d.-afhandlingen udviklede derfor nye metoder til at hjælpe forsyninger med at øge valideringen og anvendelsen af deres data. Dette blev gjort ved at analysere rigtige data fra fem danske forsyningsselskaber fordelt over seks casestudier. Ph.d. afhandlingen blev opdelt i tre trin: 1) dataindsamling, 2) datavalidering og datarekonstruktion, og 3) brug af dataene.

Forsyninger installerer i stigende grad dataindsamlende enheder, såsom intelligente målere (*smart meters)*. Når forsyningerne vælger at implementere smart meters, skal fordelene og ulemperene imellem en fin dataopløsning (større brugbarhed af de indsamlede data) og transmissionsomkostningerne opvejes mod hinanden. En analyse af et målerdistrikt med installerede smart meters viste, at typiske prøveopløsninger imellem 1 og 24 timer er tilstrækkelige til at vurdere vandtab, så længe forsyningsselskabet har repræsentative forbrugskurver fra netværket. En dataopløsning på < 1 time er dog potentielt nødvendigt for at opnå pålidelige vandkvalitetssimuleringer.

Automatiske validerings- og rekonstruktionsprocesser for de indsamlede data bør være af høj prioritet for forsyningerne. Ph.d.-projektet udviklede en systematisk tilgang til at kategorisere anomalier. Der blev introduceret fire kategorier, hvor Type 0 repræsenterede systemanomalier, Type 1 og 2 beskrev sensordata med en lav kvalitet, og Type 3 beskrev anomalier i sensordata baseret på rigtige, men usædvanlige, begivenheder i ledningsnettet. Der blev udviklet syv forskellige valideringstest til at bestemme Type 1 og 2 anomalier. En analyse af tryk- og flowdatasæt fra tre danske forsyningsselskaber afslørede et stort antal af anomalier, med et gennemsnit på 10% manglende data og op mod 35% anomalier i en af forsyningernes trykdata. Disse høje tal understreger behovet for rekonstruktionsprocesser til at generere de konsistente datastrømme, der kræves til en pålidelig anvendelse af data. Neurale netværk blev anvendt som et eksempel på en metode til at rekonstruere manglende data og yderligere

validere tvivlsomme observationer.

Indsamling af data fra vandforsyningssystemer er ikke et nyt koncept. Desværre lagres store mængder af data (såsom temperaturdata) ofte uden at blive brugt fordi der mangler eksempler på vellykkede anvendelser af dataene. For at demonstrere fordelene ved temperaturdata blev en temperaturmodel og en hydraulisk model kombineret for at identificere indstillingen og placeringen af ventiler i et netværk. Denne nyudviklede metode samt forsøg i felten førte utilsigtet til identifikationen af forskellige Type 0 anomalier i forsyningens ledningsdatabase. Metoden har dermed sået tvivl om pålideligheden af den hydrauliske model. Så længe anomalier gemmer sig i sensor- og ledningsdatabaser, er det ikke muligt succesfuldt at udføre mere avancerede datadrevne analyser. Et andet problem i casestudiet var det lave antal af relevante temperaturdata. Temperaturdata fra smart meters, som muligvis er tilgængelige fra hver husstand, kan bruges til at løse denne udfordring. I en anden forsynings målerdistrikt viste smart meter temperaturdata tilfredsstillende lighed med den simulerede temperatur (average root mean square error på 0,9 ºC). Dette understreger potentialet ved at bruge smart meter temperaturdata til mere avancerede undersøgelser, såsom lækagesporing og identifikation af ventilindstillinger.

Den typiske silo-tankegang i vandsektoren resulterer ofte i, at man overser værdien af drikkevandsdata uden for vandforsyningssektoren. I et casestudie blev det undersøgt om det kan betale sig at implementere varmepumper på drikkevandsvandledninger som et supplement til Københavns fjernvarmenet. Der blev fundet et varmepotentiale på 20,7 MW. Derudover steg andelen af drikkevandsforbrugere, hvor den øvre temperaturgrænse på 12 °C blev overholdet, fra 41% til 81% i august. I et andet casestudie blev smart meter vandforbrugsdata koblet til en afløbsmodel for at sammenligne den simulerede spildevandsstrøm med observationer fra afløbssystemet. Der blev identificeret fejl i målerne i afløbssystemet, og smart meter dataene blev derfor vurderet mere pålidelige til at estimere tørvejrsflowet end observationerne fra disse målere.

Generelt viste ph.d.-projektet at mange fejl i forsyningernes ledningsdatabaser og sensordata kun opdages ved at anvende dataene. Så længe forsyningerne ikke kan opretholde en høj datapålidelighed, er det tvivlsom om flere sensorer vil øge forsyningernes forståelse af deres system. Først når der kan sikres en højere pålidelighed af dataene, kan deres sande potentiale udnyttes. I de kommende år bør forsyningsselskaber, teknologileverandører og forskere identificere metoder til at reducere fejlene som florerer i forsyningernes lednings og sensordata, så forsyningerne kan nå et højere niveau af digital modenhed.

Table of contents

1 Introduction

"Water and wastewater utilities must embrace digital solutions. There is really no alternative" (Sarni et al., 2019).

This quotation originates from a report about digital water by the International Water Association and sets the framework for this thesis. Sarni et al. (2019) state that "digital water is already here", and that utilities are evolving from simple to complex and interconnected institutions. In general, digitalisation entails a long list of benefits for the water sector. But is digitalisation really a straightforward path to more success? The short answer is no, as each utility is unique in terms of levels of digital maturity and overall needs. However, to give a clearer and more elaborate answer to the above question it is important to understand the ongoing challenges of the water sector driving the transition towards digitalisation. Moreover, digitalisation requires that multiple obstacles be addressed properly before it is possible to refer to water systems as 'truly smart' (Moy de Vitry et al., 2019).

1.1 Challenges of water supply management

The United Nations Sustainability Goal 6 is about ensuring the availability and sustainable management of water and sanitation for all (United Nations, 2019). Around 10% of the global population lacks basic drinking water services, half of the global population lives in areas that experience water scarcity at least one month of the year, and rivers in Africa, Asia and South America are now more polluted than they were in the 1990s (United Nations, 2019). In addition, ever-increasing urbanisation, global population growth and climate change have put immense pressure on the drinking water systems as we know them today (for example, WWAP, 2019). These factors combined with socio-economic changes, will not only lead to changing consumption patterns but will also affect worldwide water consumption, which will continue to grow (about 1% each year since the 1980s), ultimately increasing global water stress (WWAP, 2019).

The summer of 2018 showed that temporal and regional water shortages occurred even in areas otherwise unfamiliar with water stress, such as Denmark and Germany. In the future, declining amounts and decreasing quality of water resources will lead to an increased competition between water users, demanding new ways of distributing water fairly between them (DANVA, 2018; German Environment Agency, 2019).

As a consequence of water stress, utilities need to pay increased attention to resource efficiency and the exploitation of alternative water resources (Rygaard et al., 2011). However, the introduction of alternative solutions, such as rainwater collection or wastewater reclamation, poses severe challenges that need to be addressed (Rygaard et al., 2011). Examples include thorough analysis of the solutions' energy requirements and effects on the quality of the delivered water (Rygaard et al., 2011). Thus, the overall complexity of drinking water supplies will continue to increase. This is also the case in Denmark, where, among other things, new and rising numbers of contaminants are being detected in the groundwater. The number of reported cases of pesticide levels exceeding guideline levels increased notably from 15 to 65 waterworks between 2013 and 2017 (Ministry of Environment and Food of Denmark, 2018a). This demands more advanced treatment processes than currently implemented.

Moreover, long-established utilities experience that a great proportion of their drinking water infrastructure is past its prime. Due to this unreliable infrastructure, not only the costs of operation and management (e.g. increased leakage) will grow, but also businesses and the standard of living will be affected negatively (ASCE, 2011). For example, investment in American water infrastructure does not keep up with need, reaching an estimated funding gap of up to $144 billion by 2040 (ASCE, 2011). Thus, there is an increasing incentive to cope with the failing infrastructure in an efficient manner due to lack of funds (Eggimann et al., 2017). Here, 'efficiency' comprises better rehabilitation planning (asset management) and extending the expected system lifetime, as well as an optimised operation and design of urban systems. For example, Nguyen et al. (2018) state that coarse and out-of-date information is often used during the design and planning of urban water infrastructure, leading to inefficient management. This is also the case for Denmark, where Kirstein (2016) and Kirstein et al. (2016), by analysing district metered area (DMA) data showed that there were significant differences between the actual consumption and apparently out-of-date demand patterns provided by the Danish environmental protection agency (Watertech A/S, 2005). For example, application of out-of-date demand patterns may lead to unintentional network augmentation during design (Gurung et al., 2016) and doubtful hydraulic model simulations.

1.2 Digitalisation

Digitalisation is often pointed out as the process capable of mitigating and solving parts of the previously mentioned challenges. Gartner's IT glossary (Gartner, 2019) defines the term as: "the use of digital technologies to change

a business model and provide new revenue and value-producing opportunities; it is the process of moving to a digital business."

In urban water management, digitalisation is envisioned to increase the effectiveness and flexibility of urban water systems and establish the opportunity for new services (Moy de Vitry et al., 2019). Such opportunities have boosted the interest of the research community on digitalisation in the water sector, notably since the start of the 2000s (Figure 1).

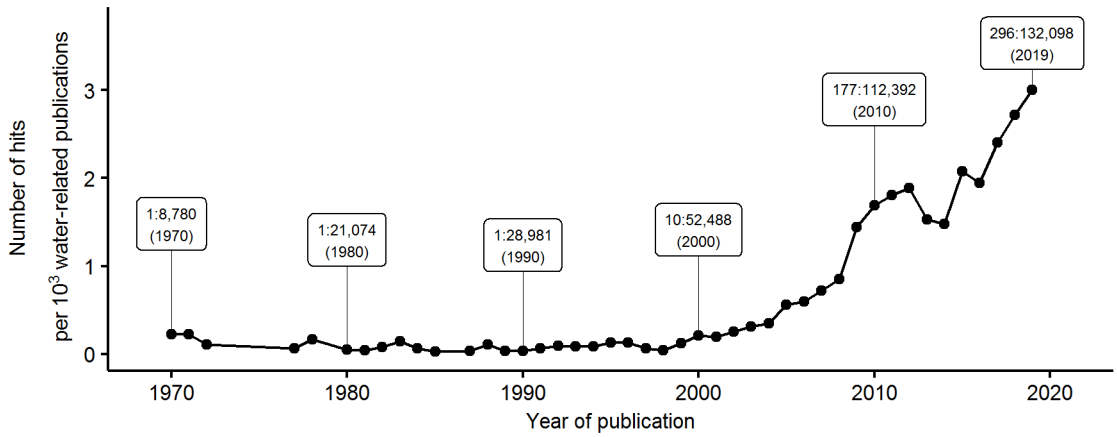


Figure 1. Normalised hits based on 1) publications concerning digitalisation in the water sector only and 2) publications related to water research in general. Number boxes display the actual hit ratio between these two categories. Scopus search (21/09/2019), Appendix A.

This boost in research is, among other things, driven by new technologies that are believed to benefit the water sector in multiple ways, providing more secure, resilient, reliable, efficient, cost-effective and innovative water solutions (Sarni et al., 2019). The literature has many promising examples justifying this claim (see, for example, Section 2). A comprehensive overview of the history and future of digitalisation in the urban water sector is given in Makropoulos and Savić (2019), indicating that the sector is far from being at the end of its digital transition journey. Sarni et al. (2019) outline a digital adaptation for utilities spanning steps of basic, opportunistic, systematic and transformational adaptation (Figure 2). Utilities around the world have already started the digital transformation, averaging an adoption level of 'opportunistic' according to a survey conducted in Sarni et al. (2019). Here, opportunistic accounts for, among other things, utilities with digital automation and control mechanisms and analytical tools during process optimisation.

The digital transformation is mainly enabled through recent advances in smart information and communication technology (ICT) and facilitated through applications and implementations of, for example, the Internet of Things (IoT)

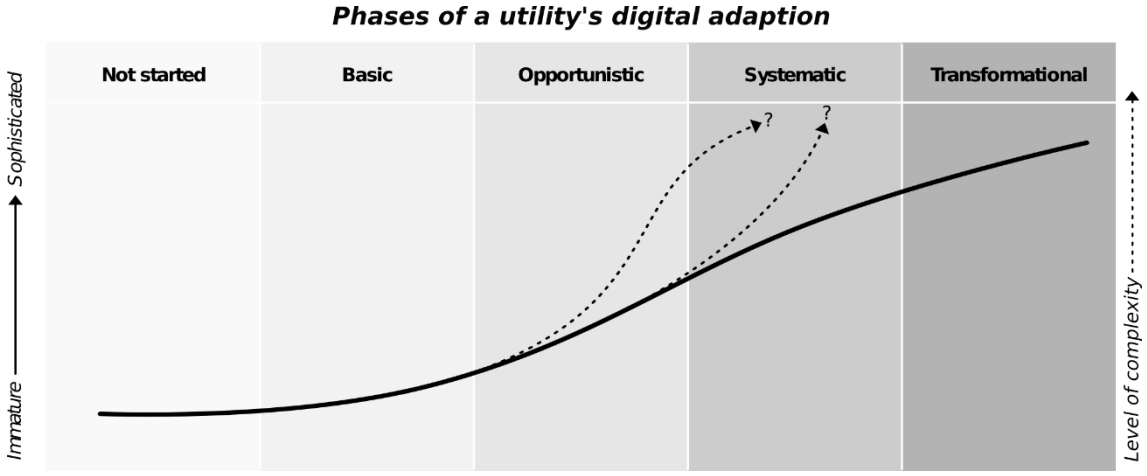


Figure 2. 'Digital Water Adoption Curve' adapted and modified from (Sarni et al., 2019).

and edge and cloud computing (Eggimann et al., 2017; Kulkarni and Farnham, 2016; Sun and Scanlon, 2019). In the water sector, these technologies have increased the speed of data collection and analytics notably over the last several decades in parallel with decreasing costs, ultimately opening up the world of 'Big Data' (e.g. Monks et al., 2019; Sun and Scanlon, 2019). An example of how fast the sector is evolving is highlighted in Cominola et al. (2015), where the installation of ultrasonic smart meters was described as too costly; now it is state-of-the-art, as shown, for example, by the deployment of 2,300 ultrasonic smart meters in Northern Copenhagen (LEAKman, 2018) or by the case study utilities in Papers IV and VI.

Digitalisation concerns the entire urban water sector, but this PhD study focuses only on data from water distribution systems. Whereas the collection and application of data from WDNs are not new to drinking water utilities, the increasing volumes of data introduce challenges and potentials within the field of water distribution system analysis. Here, the PhD study will to a large extent focus on the application of data from smart meters, being a prime example of digitalisation.

1.3 The digital challenges

Simply installing more devices and collecting more data do not result in a higher level of digital maturity for a utility. Due to the low usage of the collected data in many Danish utilities, I therefore speculate that reaching higher levels of digital maturity (e.g. systematic) is a more difficult (steeper) process than envisioned and may lead to unknown paths of rising complexity for utilities (dotted lines, Figure 2). Many important factors for becoming more *data-*

driven (i.e. relying on the collection and analysis of data) need to be considered, here divided into three steps:

1 **Data collection.** It should be considered what and how much data should be collected. For example, Kulkarni and Farnham (2016) and Mekki et al. (2019) showed that the implementation of ICT comes at various costs and depends on multiple factors such as the selected data transmission technologies. Eggimann et al. (2017) pointed out that data-driven urban water management should aim to reach a region of optimal data availability, meaning that there is a point where more data is not necessarily better. Moreover, there is still a lack of evidence about what types and frequencies of information best suit the needs of utilities and consumers (Boyle et al., 2013).

2 **Validation and reconstruction of data.** A reliability assessment of the collected data is necessary. I recognize 'data reliability' as a term covering the accuracy and completeness of collected data, which is secured only through forms of validation and reconstruction processes. With the increasing amount of data collected, it is of paramount importance that autonomous data collection and verification processes are in place to secure a high reliability of the data (Sun and Scanlon, 2019). In a survey of the urban water community about how ubiquitous sensing will shape the future of the sector, Blumensaat et al. (2019) showed that data validation and integrated management were ranked as the two most important topics among professionals. It is important to be able to trust the collected data with a very high reliability before it can be applied in decision-making processes (Blumensaat et al., 2019). Furthermore, integrated management will play an increasing role in future infrastructure design, but models used for such processes are limited by the availability and quality of the data (Eggimann et al., 2017).

3 **Application of data.** Simply assuring a high data reliability, however, does not release the biggest potential the sector is currently facing; the overall value created by the collected data from digital technologies is often too unclear (Sarni et al., 2019). In general, proof is still lacking as to whether digitalisation really leads to long-term savings and increased performance; this is often because it is unclear which exact challenges can and cannot be solved by mining the data (Blumensaat et al., 2019; Boyle et al., 2013; Cominola et al., 2015; Eggimann et al., 2017; Sarni et al., 2019).

I envision these three steps as an interlinked, cyclic process, whereby new insights from unsuccessful and successful data applications in step 3 may lead to a better understanding of data collection requirements in step 1. Furthermore,

Papers III, IV and VI highlighted that it is first the application of data that clarifies valid and invalid data, thus improving step 2.

It should be noted that many additional important challenges and topics related to digitalisation, such as privacy and cyber-security concerns or a changing customer–utility relationship raise important questions. These were, however, deemed outside the scope of this thesis.

1.4 Objectives

Currently, digitalisation is one of *the* main topics of the urban water sector. In this thesis, I will assess obstacles and opportunities as well as develop methods related to 1) data collection, 2) validation and reconstruction of data and 3) application of the data in water distribution systems and external water-related systems. This is done to guide utilities towards improving the operation and management of their systems. This thesis addresses topics 1–3 by answering the following questions.

1 Define adequate sampling resolution from smart meters.

- What is the effect of changed sampling resolutions from smart meters in hydraulic simulations?
- What is the effect of filling gaps between measurements with different methods?

2 Improve the reliability of collected data in urban water systems.

- What are the common rates and types of anomalies in water network data?
- How can we improve future data collection procedures?

3 Evaluate new ways to use collected data in water distribution systems.

- Which benefits can be obtained by analysing temperature data collected in the water distribution network and from smart meters?
- What is the impact on consumers and the distributed water quality when installing heat pumps in water distribution networks?
- How can smart meter data be used to estimate wastewater flows?

A major focus of this thesis is the application of ‘real-world’ data to answer the above stated questions. In short, ‘real-world’ case studies from five different utilities are used to outline the current state of digitalisation.

1.5 Thesis structure

The thesis structure follows the data flows in a water utility (Figure 3).

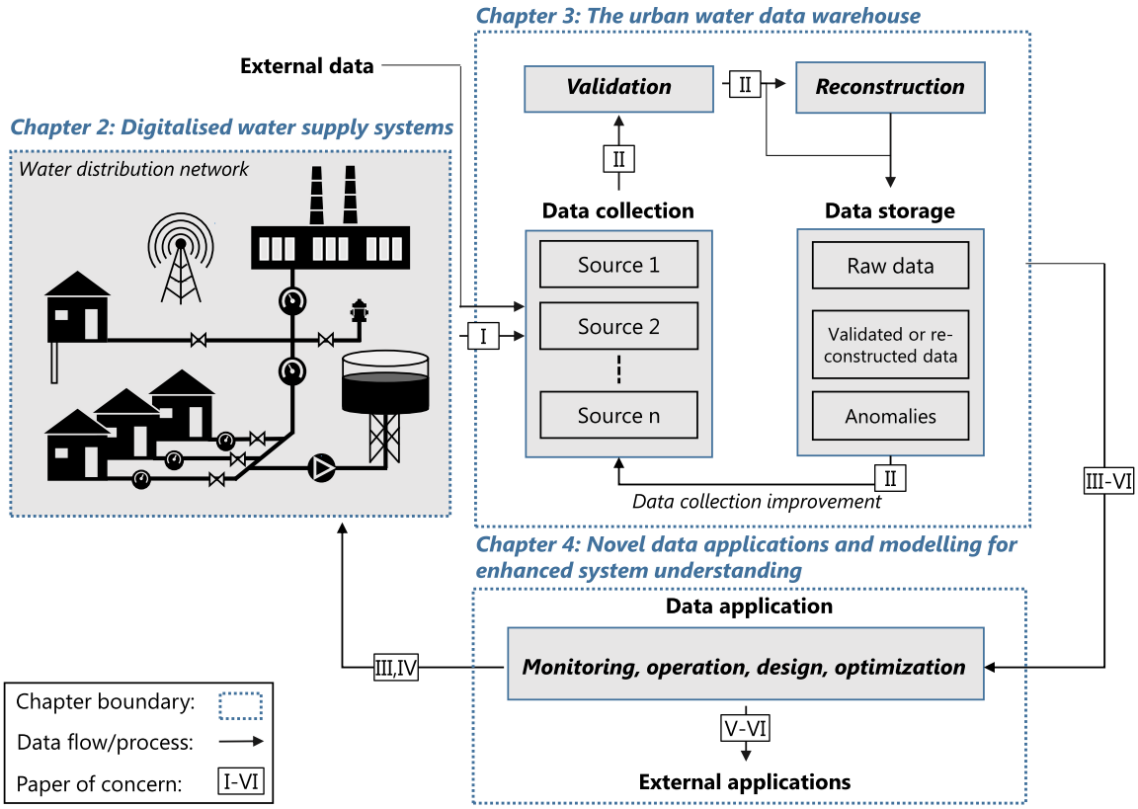


Figure 3. Conceptual overview of data flows in a water utility. The focus of each chapter is delineated by blue boxes.

The collected data needs to undergo various processes (indicated by boxes and arrows in Figure 3) prior to successful application, guiding the remainder of this thesis.

- Chapter 2 (Digitalised water supply systems) introduces commonly installed data-collection devices in drinking water distribution networks (WDNs), including the benefits and pitfalls of these technologies.
- Chapter 3 (The urban water data warehouse) highlights the need for utilities to rethink their data gathering (Step 1) and processing in forms of validation and reconstruction (Step 2) prior to successful application of the data (Papers I and II):
    - The first step includes data collection. Paper I highlights how sampling resolutions affect pressure and water age simulations as well as water loss assessments.

- Next, it is important to maintain a high level of reliability of the collected data. Paper II shows how the data can be validated and how errors in data collection can be identified. An example of reconstructing data is given in Section 3.2.4

- Chapter 4 (Novel data applications and modelling for enhanced system analysis) addresses new opportunities (Step 3) from data collected in WDNs for optimised water distribution system analysis (Papers III and IV) and external research fields (Papers V and VI).
  - WDN-related applications. Papers III and IV show how temperature data collected in the network can be used to improve the operation and management of WDNs.
  - External applications. Papers V and VI show that temperature and smart meter data are beneficial for management/operation of systems outside the 'WDN bubble', such as district heating and urban drainage.
- Chapter 5 (Conclusion) summarises the main results.
- Chapter 6 (Perspective) lists future research possibilities, risks and challenges that need to be addressed.

The applied real-world data across six case studies covers smart meters at the household level, meters at WDN level and manual sampling of flows, pressure and temperatures (Table 1).

Table 1. Overview of real flow (Q), pressure (P) and temperature (T) data collected by smart meters, sensors in the water distribution network and manual sampling.

Data source	**Smart meter (household)**		**Water distribution network**			**Manual sampling**	
Parameter	Q	T	Q	T	P	T	P
Paper I	X		X		X		
Paper II			X		X		
Paper III			X	X	X	X	X
Paper IV	X	X	X	X	X		
Paper V						X	
Paper VI	X		X				

2 Digitalised water supply systems

Currently, the amount of data collected from devices deployed in urban water systems is increasing notably. Examples include sensors monitoring various forms of parameters, such as flow and pressure meters, pumps and valves collecting data about their state (e.g. on or off) and acoustic loggers recording the level of noise at selected locations. The data collected from such devices is useful for monitoring, control, design and planning purposes. Kulkarni and Farnham (2016) gave an overview of (smart) ICT deployed in water distribution systems, subdividing the field of water monitoring into pressure and flow management, consumption monitoring, water loss management and water quality monitoring. I used these categories to classify commonly implemented data-collection devices according to their major objectives in WDNs (Table 2). Here, actuators (pumps and valves), as well as non-data-driven drivers (e.g. customer satisfaction) have been deemed outside the scope of this thesis. The following section addresses each of the devices listed in Table 2.

Table 2. Major data-collection devices deployed in water distribution networks and their application in water distribution system analysis, based on my categorisation.

Drivers / Devices	Pressure & flow management	Consumption monitoring	Water loss management	Water quality monitoring
Flow & pressure meters (distribution level)	X	X	X	
Water quality sensors				X
Noise loggers			X	
Smart meters (consumer level)	X	X	X	X

2.1 Flow and pressure meters (at distribution level)

Pressure & flow management. Flow and pressure sensors installed at DMA level or other critical locations (e.g. tanks or reservoirs) have been an integral part of many utilities for decades (even though Puust et al. (2010) described the installation of flow meters as “a recent trend”). For example, in combination with hydraulic models, the collected data may be used to simulate the flow and pressure throughout the network and highlight areas with pressure deficits or long retention times. Pressure monitoring and management are also directly connected to water loss management, as higher pressure may lead to higher leakage rates (see below). Furthermore, pressure and flow data play a major role in real-time control of WDNs, as the data is used to remotely control the states of pumps and valves (e.g. Creaco et al., 2019). Some of these pumps and valves are termed ‘intelligent’ as they are capable of taking decisions independent of remotely controlled set-points, such as in the case of blackouts or connectivity issues (e.g. AVK, 2018).

Consumption monitoring. In terms of consumption monitoring, the data can be used to gather information about the consumers in the area, e.g. based on DMA-level measurements useful for data reconstruction purposes, demand forecasting and design of networks (e.g. Kirstein, 2016; Kirstein et al., 2016). Application of forecast models that rely on demand data from flow meters, is expected to increase, as already seen for optimal pump control in Dutch water supply systems (Bakker et al., 2014).

Water loss management. DMA data can be used for leakage detection. One of the most common approaches in using flow data for leakage detection is minimum night flow (MNF) analysis (for example, Puust et al., 2010). In MNF analysis, the water inflow into an area is measured (e.g. between 02:00 and 05:00) and legitimate uses are subtracted. High deviations in MNF can then be used to detect leakages and bursts. Good knowledge of the downstream area is important when installing sensors, as otherwise it is possible that the installed sensors will not be sensitive enough to detect the MNF, as seen in the applied data in Paper II. In addition, more sophisticated approaches can be applied, such as data-driven methods for automatic burst detection in WDNs (Wu and Liu, 2017). Here, the quality of the data constitutes the main uncertainty, with flow measurements being more reliable (though also more expensive) than pressure measurements (Wu and Liu, 2017). Most sensors are deployed at DMA level, being capable of detecting bursts in the range of 1.5–50% of average DMA inflow, but the success of the reviewed methods were difficult to compare owing to the varying nature of the case studies (Wu and Liu, 2017). Wu and Liu (2017) also state that finer sampling rates and data communication frequencies (e.g. < 5 minutes) could reduce detection times.

2.2 Water quality sensors

Water quality monitoring. In general, the monitoring of water quality has been deemed to be one of the most difficult parameters to "monitor remotely and reliably" (Makropoulos and Savić, 2019) owing, among other things, to the high number of possible contaminants and entry points (Adedoja et al., 2018; Eggimann et al., 2017). Moreover, Makropoulos and Savić (2019) state that additional work is needed on novel water quality sensors. An overview of microbial sensors from 2015 can be found in Tatari et al. (2016), showing that the response time of many sensors varies (examples in Tatari et al. (2016) between 10 minutes and 18 hours), depending on which parameters are measured. Tang and Albrechtsen (2019) list commercially available technologies for real-

time or near-real-time monitoring of water quality (albeit from the food industry), of which the most common parameters included pH, turbidity and conductivity. Currently, technologies behind sensors measuring microbial water quality as well as physiochemical properties of the water are being rapidly developed (Tang and Albrechtsen, 2019). Thus, it is believed that the number of water quality sensors will increase notably in the coming years. Potentially, the number of water quality parameters monitored can also lead to the detection of leakages, among other things because turbidity can increase during pipe burst events (Puust et al., 2010).

Temperature measurements in the WDN can be used as a simple water quality indicator and guide utilities to locations of water quality audits, among other things because temperature affects the growth rates of biofilm-forming bacteria (Liu et al., 2016) and because temperature measurements are an indicator of the water's residence time in the WDN (Papers III and IV). Thus, temperature can be selected as a parameter useful for the identification of relevant monitoring sites (for example, Larsen et al., 2017).

2.3 Noise loggers

Water loss management. Typically, acoustic loggers attached to pipe fittings record the level and spread of noise in pipes to detect leakages by statistical analysis (Puust et al., 2010). Leakage correlators use the signal between two adjacent noise loggers to correlate and narrow down the area of leakages (see, for example, Li et al. (2015) for further information on acoustic detection methods).

2.4 The smart meter revolution

Here, smart meters are understood to be meters at the household level. Smart meters exemplify the digital transformation in the water sector more than any other device, as they affect both utility and consumers and can play a major role in integrated urban water management. The meters are termed 'smart' because water consumption and other parameters are measured on a less than daily basis and are collected remotely, opening up various possibilities within data analytics with benefits for the utility and consumers (Boyle et al., 2013; Cominola et al., 2015). Also, technological advances have made it possible to implement additional monitoring capabilities in smart meters, no longer limiting meters to only monitoring the demand.

Advances in sensing have made meters (e.g. ultrasonic smart meters) more accurate and less fragile compared to mechanical meters, and the implementation

of smart meters has increased notably over the last few decades (Boyle et al., 2013; Cominola et al., 2015; Kamstrup, 2019a). For example, based on a survey of 55–60 participating utilities, the number of remotely read meters at household level in Denmark increased from 15% to 46% between 2013 and 2017 (DANVA, 2018). However, the literature also points out that on a global scale, the majority of digital metering rollouts were conducted on smaller scale trials. Furthermore, the rollout was slower than expected, partly because many of the benefits are difficult to monetize or utilities struggle to see the benefits of the data (Monks et al., 2019; Stewart et al., 2018; Paper IV).

Based on a comprehensive literature search and survey among experts, Monks et al. (2019) revealed in total 75 benefits for utilities and customers when implementing digital metering. These included, among other things, enhanced interaction between consumers and the utility, greenhouse gas reductions due to reduced driving, and more efficient billing (Monks et al., 2019). Smart meter deployment reduces one of the major limiting factors of water distribution system analysis: the unknown demand. However, as long as utilities do not apply this data in analyses, such as in near-real-time water loss assessments, in hydraulic online models or in demand forecasting, the benefits of smart meter deployment are less obvious for utilities with a basic level of digital adaptation (Figure 2). In the following, data-driven benefits of smart meter deployment will be discussed concerning the categories listed in Table 2.

Pressure & flow management. The increased data quantity and quality from smart meters can be a major driver for improved hydraulic model accuracy in terms of flow and pressure simulations. Naturally, the resemblance of the DMA inflow with smart meter consumption data increases when using daily or less-than-daily consumption data, compared to coarser audit data (e.g. quarterly) (Paper I). Smart meters may also be used to remotely disconnect users from water, e.g. during maintenance or when bills have not been paid (e.g. Blokker, 2019).

Consumption monitoring. One major driver for installing smart meters is water demand management, as smart meters can help to increase awareness about water usage and thus lead to reduced consumption (Boyle et al., 2013; Nguyen et al., 2018). Here, disaggregation of the consumption data into specific usages can increase the awareness of users about their consumption (e.g. the specific water usage of certain appliances) and help the utility with managing demand peaks (Cole and Stewart, 2013; Cominola et al., 2015; Nguyen et al., 2018;

Stewart et al., 2018). Digital metering can also be used to improve infrastructure planning and reduce network augmentation, which is particular evident for consumers with little information prior to data collection (Gurung et al., 2016, 2014; Monks et al., 2019).

Water loss management. Smart meter data can help to establish temporary or permanent DMAs and enable detailed water balances and post-meter leakage detection, e.g. based on leak alarms (e.g. Monks et al., 2019). Furthermore, more accurate estimates of water loss components, such as background leakage, can be achieved (Loureiro et al., 2014). Newer smart meters may even be equipped with inbuilt noise loggers intended to improve post-meter and network leak detection (Kamstrup, 2019b). Moreover, Bragalli et al. (2019) analysed the impact of an increasing number of missing smart meters from apartment blocks and single houses on the error of the estimated water loss. The results of estimated water loss worsened notably for missing large-scale consumers (i.e. apartment blocks) compared to single houses. This is important for utilities, as the deployment of smart meters at large-scale consumers may be more troublesome, for example, because a different type of smart meter needs to be installed and a larger number of people/businesses are affected.

Water quality monitoring. Monks et al. (2019) list only few water quality related benefits of smart meters, but state that the implementation of water quality testing at customer meter level might help to reduce the number of required audits. Yet, according to Blokker (2019), no microbiological parameters (or parameters related to water quality, such as pH) are monitored with smart meters. Temperature is an often sampled parameter as it may be collected as spin-off from ultrasonic smart meters; however, utilities struggle to know how to use the data (Blokker, 2019; Paper IV). Additional potential benefits of (smart meter) temperature data based on Papers III and IV are listed in Sections 4.1 and 4.2.

Also, smart meter data can be beneficial as input to enhanced demand knowledge from a water-energy-nexus point of view (Stewart et al., 2018). Smart meter temperature and consumption data can help to identify certain pipe and soil characteristics and estimate the heat transfer potential of drinking water in WDNs, e.g. as a low-temperature source in district heating systems (Papers IV and V). Furthermore, smart meter data can be an integral part of estimating the wastewater flow and as a validation tool of in-sewer flow observations in urban drainage management (Paper VI).

3 The urban water data warehouse

Gartner's IT glossary (Gartner, 2019) defines a data warehouse as a "storage architecture designed to hold data extracted from transaction systems, operational data stores and external sources". I envision the *urban water data warehouse* as the location where data from multiple sources, such as sensors and asset data, is collected as well as made available for a variety of applications through extract, transform and load processes. As utilities become increasingly data-driven, it is important that they secure highly reliable data in the warehouse. However, prior to the rollout of new ICT, such as smart meters and flow or pressure sensors throughout their network, a utility should ask itself what its general goal is in installing the new technology. Is it to facilitate the billing of customers, to provide near-real time leakage monitoring and demand forecasts or to provide water quality estimates for use in the case of contamination?

If data-driven goals are of significance for a utility, the above-mentioned questions are important because the actual use of the data in the utility drives the maintenance of a high reliability of the collected data in WDNs (Papers I–IV and VI).

In other words, I speculate that data used proactively by utilities show fewer anomalies and these, if detected, are given higher priority. This is not the case for sparsely used data, where errors prevail for longer periods (Papers II and III). This may in some cases lead to a wrong understanding of the data quality and false trust in data.

The applicability of data depends, among other things, on where the data is collected. However, identifying optimal sensor locations – such as of water quality sensors for contaminant detection (e.g. Adedoja et al., 2018), pressure meters for burst detection (Wu and Liu, 2017), or which consumers should be favoured during smart meter enrolment (e.g. Bragalli et al., 2019) – were deemed outside the scope of this thesis.

As with the location of data collecting devices, the sampling resolution (i.e. the time step between measured data points) has a significant impact on the achievement of the 'digitalisation goals' of a utility, as described below.

3.1 Sampling resolution and gap filling

ICT deployments have a trade-off between transmission costs and applicability (amount of collected information). Among other things, transmission costs depend on the selected network, message sizes, and costs of data collection

(Kulkarni and Farnham, 2016; Mekki et al., 2019).

Cominola et al. (2018) and Nguyen et al. (2018) showed the benefits of having very fine sampling resolutions of smart meter demand observations (< 1 minute) resulting in successful end-use disaggregation. Furthermore, Cominola et al. (2018) showed that there can be up to 62% difference in the magnitude of peak demands based on a 10-second compared to a 24-hour sampling resolution. Gurung et al. (2014) showed that this is important because detailed knowledge about peak demands based on a fine sampling resolution can provide better demand patterns and subsequently reduce network augmentation. In terms of generating reliable water quality simulations, the spatial aggregation and sampling resolution of demands are of importance (Blokker et al., 2008). Depending on the set-up, a 1-hour time step can be sufficient for water quality simulations of large demand aggregations (e.g. transportation system models), but sampling resolutions below 5 minutes are required to generate reliable results for smaller networks (Blokker et al., 2010, 2008). Also, Creaco et al. (2017) showed that pressure simulations are largely affected by the sampling resolution and demand modelling approach. Whereas a top-down approach (applying a demand multiplier on coarse readings) first leads to acceptable simulations at sampling resolutions of demands and modelling time steps > 1 hour, a bottom-up stochastic approach (generating unique demand profiles for each consumer) can generate reliable pressure simulations at time steps > 2 minutes, which is important when running near-real-time simulations (Creaco et al., 2017).

Smart meter data represent the actual demand from each household. Incorporating such data into hydraulic models can enable the monitoring of WDNs in near real-time. Compared to stochastic bottom-up demand allocation approaches (e.g. Blokker et al. (2010) and Creaco et al. (2017)), Paper I applied the actual consumption from each individual household's smart meter as input to a hydraulic model of a case study DMA, to assess the effect of common smart meter sampling resolutions on pressure, water age and total consumption simulations. Accumulated volume readings from each smart meter were available with a sampling resolution averaging 30 minutes/sample. To align the random nature of each sample's timestamp, gaps between samples were filled by linear interpolation and a demand-pattern-based approach. Hereby, aligned and unique consumption data time series were generated for each smart meter, making it possible to aggregate the consumption from different consumers in the hydraulic model's nodes. The model was ran in 5-minute time steps to compare the model's results with the 5-minute sampling resolution of the DMA

inflow. This approach was repeated with gradually coarser sampling resolutions. Resolutions commonly sampled by Danish utilities were thus established, ranging between 30 minutes to 24 hours/sample (Paper I).

For pressure simulations, the results showed that there was little difference in the root mean square error (RMSE) between assessed sampling resolutions, mainly because of the overall low pressure head loss through the network. Even combining a representative demand pattern with very coarse demand readings (e.g. quarterly) turned out to be sufficient compared to finer sampling resolutions when simulating pressure (RMSE < 0.1 m) (Paper I).

In terms of water age, however, there is a clear benefit from collecting finer sampling resolutions. Thus, if utilities are concerned about retention times and interested in water quality simulations, finer sampling resolutions should be favoured (Paper I). This is also shown in Paper IV, where the sampling resolution of around 30 minutes/sample is the major restraining factor of more detailed temperature analyses (see also Section 4.2).

When very coarse readings (e.g. quarterly) are available in combination with reliable demand patterns, the RMSE of total DMA consumption was only around 0.5 m^3/hour higher (though approximately equal to the DMA's MNF) than when using a fine sampling resolution (Paper I). However, such a good demand pattern might only be available from smart meter data which has a fine resolution.

When having sampling resolutions finer or equal to 2 hours, there was little difference between the two gap-filling methods; at coarser resolutions, however, the demand-pattern-based gap-filling method outperformed linear interpolation.

Thus, the following could be concluded on sampling resolutions and gap filling from smart meters.

- Fine sampling resolutions (< 1 hour) should be used when utilities are interested in a high level of detail about the consumption in their systems. This knowledge improves the validity of water quality simulations relying on accurate simulation of the water's retention time in the WDN.
- Coarse sampling resolutions (1–24 hours) are sufficient when the overall goal includes daily water loss assessments and when assessing the pressure in areas with overall low pressure head loss.

- At coarse sampling resolutions (> 2 hours), a demand pattern based approach to gap filling between samples is preferred over linear interpolation.

3.2 Validating and reconstructing data

Data cleansing, including validation and reconstruction processes, is a major concern in almost all fields of research and applications related to environmental water management (Sun and Scanlon, 2019). Owing the increasing amount of data collected, the quality control, validation and easier access to data are "expected to be at the heart of the next steps in hydroinformatics" (Makropoulos and Savić, 2019). Without such steps, the value of the collected data decreases notably and might lead to false acts on incorrect data (e.g. Blumensaat et al., 2019). The first statement in the Introduction could thus be reformulated:

"Water and wastewater utilities embracing digital solutions have to include data validation and reconstruction processes. There is really no alternative."

Data validation has been part of the urban water sector for decades, covering fields such as urban hydrology (e.g. Branisavljević et al., 2011; Mourad and Bertrand-Krajewski, 2002) and water distribution systems (e.g. Cugueró-Escofet et al., 2016; García et al., 2017; Quevedo et al., 2010). Unfortunately, the implementation and focus on automatic data validation has not kept up with the interest in data collection and application. This needs to change as utilities become increasingly data-driven. The steps of validation and reconstruction may seem very obvious, but real-world examples (Papers II, III and VI) reveal that this is tricky and often only practised to a limited extent.

3.2.1 Need for automatic data cleansing

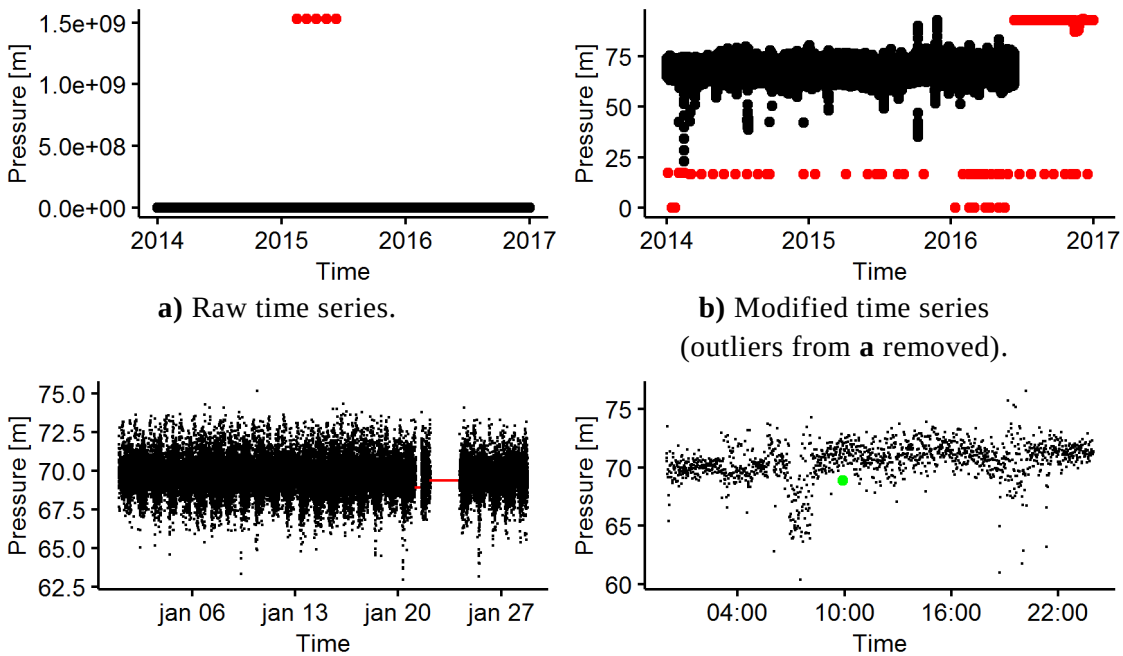
An example of why automatic validation and reconstruction of water network data is of such crucial importance is shown for a real-pressure time series collected at the exit of a DMA in Denmark (Figure 4). Visualization of the raw time series (Figure 4a) highlights data points that can be deemed unfeasible ($1.5 \cdot 10^9$ m). Removing these anomalies from further analysis results in Figure 4b, indicating that additional data points exist that are highly doubtful in the area (recurring pressure measurements). When omitting these anomalies, the time series looks more trustworthy, as shown in January 2014 (Figure 4c).

However, even here, anomalies are still visible in the form of two flatline segments. After a more exhaustive identification and potential reconstruction of invalid data points, the data quality may be considered as 'good'. For example, Figure 4d shows a day of plausible measurements in 2016. Nevertheless, a control measurement was conducted on the same day by the utility, highlighting a pressure difference of 3 m. Additional control measurements conducted at earlier and later stages confirmed similar deviations. This shows that, even though the data looks 'good', there might be essential flaws in it, as the pressure sensor might have incorrect time settings, drifted over many years, or wrong location coordinates in the utility's asset database.

Being one example out of thousands of sensors deployed means an obvious need for automatic validation and reconstruction of data. Even more important, this example shows that even though data looks 'good', users should retain some scepticism regarding the truth of the data, as this otherwise may lead to incorrect decisions, e.g. when this data is used in a real-time control setup.

3.2.2 Anomaly types

As seen in Figure 4, many types of anomalies exist in the data collected in WDNs, owing to transmission errors, meter malfunctions, system operations, etc. (for example, Loureiro et al., 2016; Quevedo et al., 2017). In Paper II,



a) Raw time series.

b) Modified time series (outliers from **a** removed).

c) Modified time series, zoom January 2014 (outliers from **a** and **b** removed).

d) Modified time series, zoom 7th June 2016 (outliers from **a** and **b** removed).

Figure 4. Example of pressure time series (● measurement ● anomaly ● control) collected at a district metered area outlet in Denmark.

anomaly types were subdivided into three different categories (Types 1–3), described in detail in Figure 5. However, Papers III, IV and VI revealed a more 'general' anomaly type that is difficult to quantify, but needs to be addressed: if the understanding of system attributes is misconceived (Type 0, Figure 5) the application of data will be troublesome.

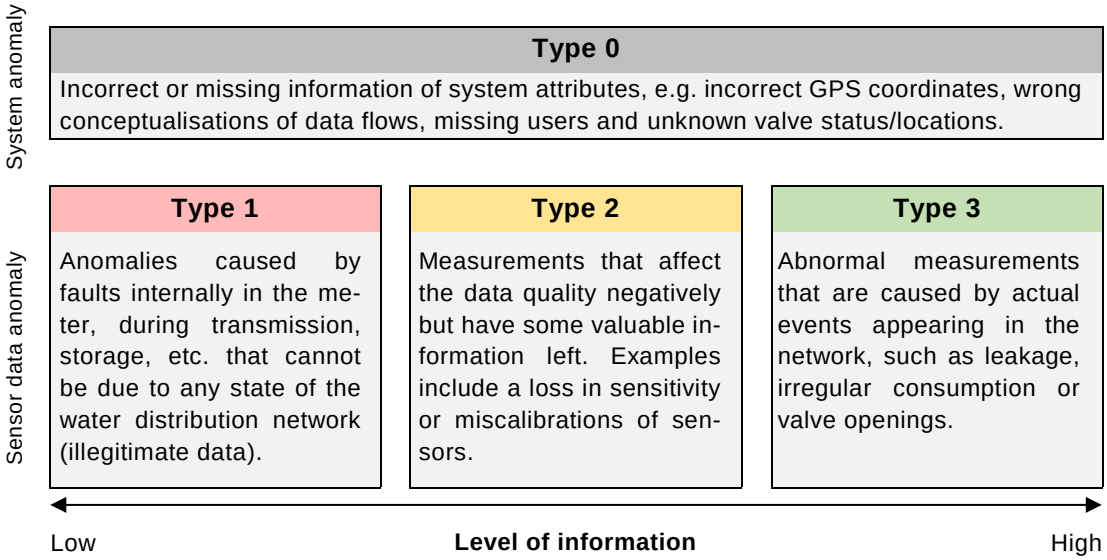


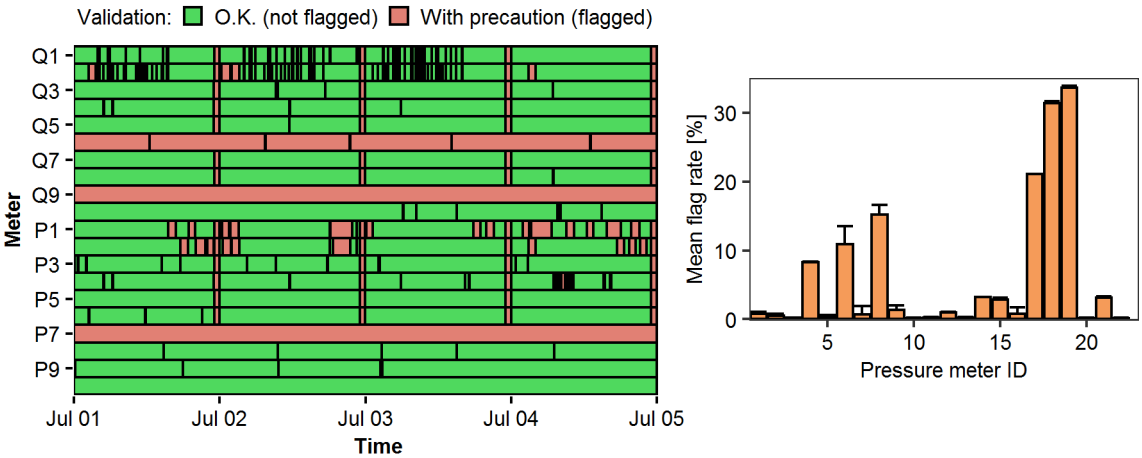
Figure 5. Classification of anomalous data collected in water distribution networks (Paper II) with additional system anomaly description of Type 0.

3.2.3 Identification of anomalies

Seven tests were proposed in Paper II to detect anomalies of Types 1 and 2, i.e. errors that do not reflect the true state of the water distribution systems: duplicate timestamp test; illegitimate format test; range test; rate of change test; flatline test; timestamp inconsistency test; and a timestamp drift test. An online tool was generated during the PhD research, whereby the seven tests can be run: https://leakagemanagement.net/meter-validate/.

The seven tests were run on flow and pressure data sets from three Danish utilities, covering on average 32 months. The results showed that a varying proportion of anomalies were found in all utilities' data sets, averaging between 3% and 35%. Even though only a small proportion of data points was identified by the timestamp inconsistency test, the analysis showed that these anomalies covered around 10% of the time in the data sets of the three utilities (i.e., on average, 10% of data was missing). This clearly highlights the strong need for validation and reconstruction of the collected data in Danish utilities. Furthermore, the tool's practicability for reducing the amount of anomalies prevailing for long periods is exemplified in a monitoring or operational setup (Figure 6). Figure 6a shows the incoming data for ten flow (Q) and ten pressure

(P) sensors over time, with invalidated data coloured in red from the seven tests. The figure shows 1) that some anomalies occur at the same time, and 2) that there are specific sensors with either no data or a higher number of anomalies. Such a visualisation can help the utility to discover overall issues within data collection and react quickly to such data quality and collection issues. Figure 6b shows the percentage of flatline anomalies (being the major anomaly contributor) for all 22 pressure meters in one of the utilities from Paper II. Such a visualisation can be used to prioritise which meters should be calibrated first and to rank the reliability of the meters.



(a) Daily data verification. (b) Flatline anomaly test analysis.

Figure 6. Example of anomaly visualisation for operational use. (a) Raw meter data validation from ten flow (Q) and ten pressure (P) meters in July 2015. (b) Mean flag rate based on all raw data points for the flatline test in pressure meters. Whiskers display the total flag rate based on all anomaly tests in the individual pressure meter, highlighting that flatline anomalies are the major anomaly source. From supporting information, Paper II.

3.2.4 The value of anomalies

Analysis of anomalies can help to identify general errors in data collection and transmission. For example, Figure 6a shows that data from multiple sensors were repeatedly invalidated for one hour around midnight. Whereas these meters were not physically attached to the same part of the WDN, the data was collected in the same database and the error is therefore likely to have originated from the utility's database setup. Therefore, anomalies were stored as an amendment to the operational database in a 'malfunction indicator database' in Paper II. For example, the Jaccard Index (e.g. Tan et al., 2006) can then be applied to identify similarities between anomalies. Application of this similarity measure revealed multiple issues, potentially related to transmission and connectivity problems (Paper II).

3.2.5 Higher-level validation of anomalies and reconstruction

As stated in Paper II, various methods exist for reconstructing invalidated data. These include time-series analysis, physically based models or machine learning approaches such as artificial neural networks (ANNs) (e.g. Branisavljević et al., 2011; García et al., 2017; Mounce et al., 2010; Quevedo et al., 2010). Figure 7 shows an example of training multiple ANNs to reconstruct missing or invalidated pressure and flow data, respectively, and to validate Type 3 anomalies. The ANN training was solely based on validated data from other sensors with similar dynamics deployed in the utilities' networks (see Appendix B for further information).

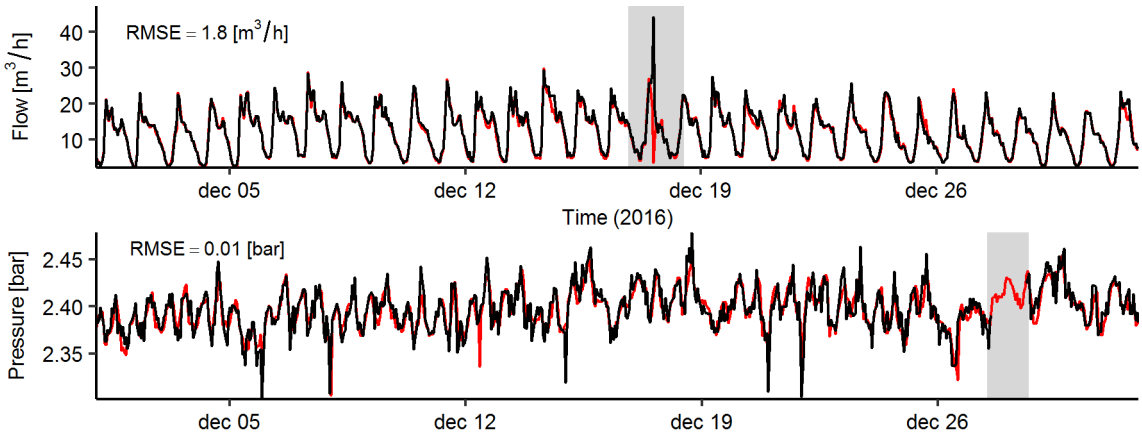


Figure 7. Measured (—) and reconstructed (—) pressure and flow time series from two Danish utilities. See Appendix B for more information. RMSE = root mean square error.

In both time series, the reconstructed data follow the real measurements adequately (Figure 7). The upper panel shows the potential of reconstructing data for validation purposes and identification of Type 3 anomalies. On 18 December a much higher flow rate was measured than predicted. As the reconstructed data is solely based on other sensors in the network, this might indicate that other sensors in the network were affected simultaneously, indicating that this event was a Type 3 anomaly caused by a legitimate measured network operation affecting larger parts of the network. The lower panel indicates that missing data around 28 December can be filled by ANNs to generate uniform data streams.

Thus, the following can be concluded based on the work from Paper II.

- Pressure and flow data sets from three Danish utilities revealed a high need for validation, as a large proportion of anomalies prevailed for long periods in the utilities' data sets, with one data set having 35% invalid data (Types 1 and 2).
- The proposed methodology helps utilities to monitor their data collection and can be used as an operational tool to quickly act on and reduce Type 1 and Type 2 anomalies.
- On average, 10% of the period covered by the utilities' data sets was missing. This highlights the need for reconstruction processes, as data applications require uniform and reliable data streams. Reconstruction can also be used to further validate Type 1, 2 and 3 anomalies.

4 Novel data applications and modelling for enhanced system analysis

After collection and quality control, the data can be applied for the monitoring, operation, control and design of WDNs. Research has begun into digital 'multi-utilities', whereby data from other sectors, such as gas and electricity, are coupled with water data to produce enhanced demand management strategies and the opportunity for new services and businesses (Stewart et al., 2018). However, silo thinking may be a limiting factor for sharing of data from different sources both within and between institutions, because utilities are complex organisations with multiple departments and differing objectives (Kulkarni and Farnham, 2016; Sarni et al., 2019). Thus, the first stage of data collection and processing (Figure 3) should not be seen as a process covering only data from the WDN, but also include data from other sectors and sources (that needs to be easily accessible (e.g. Makropoulos and Savić, 2019; Sarni et al., 2019)). Likewise, the application of data is not restricted solely to the field of WDN analysis (Figure 3).

Even if data sharing between institutions is successful, one major problem of the increased data collection is the missing expertise, experience and examples of possible applications of the collected data streams. In the following, I will demonstrate four examples of 'surplus value' from the collected data, with a major focus on temperature data, as utilities often do not know what to do with this data. The first two examples concern mainly optimised water distribution system analysis, whereas the latter two display the external fields of district heating and urban drainage.

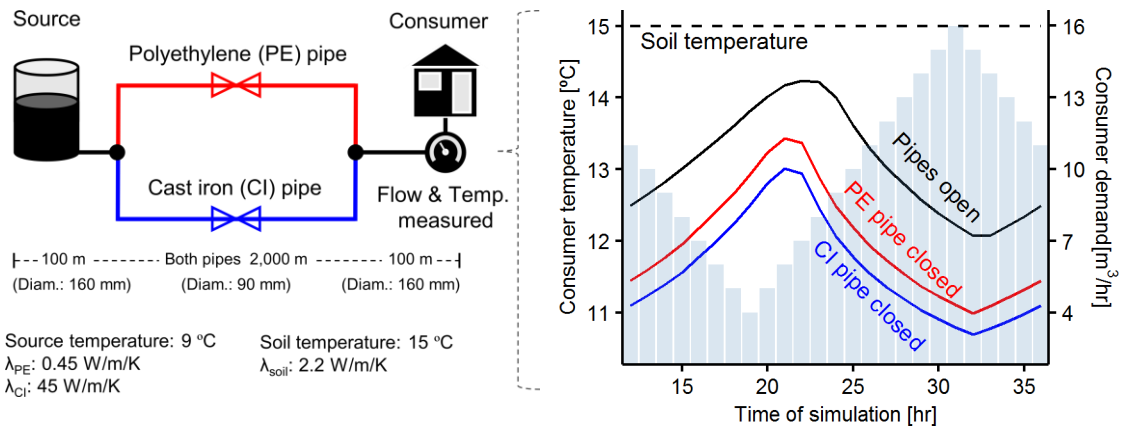
Temperature simulations in water distribution networks

Besides the simple water quality monitoring benefits discussed in Chapter 2, I believe that temperature measurements from both smart meters and the WDN have an important function that is not yet fully exploited: As the water is heated up or cooled down throughout the network, temperature measurements store to some extent the history about the path of the water. It is this effect that is explored in Papers III–V.

Heat transfer models describe the change in water temperature over time in a WDN (Blokker and Pieterse-Quirijns, 2013; De Pasquale et al., 2017; Paper V), and include four particularly important parameters that need to be considered: 1) the undisturbed soil temperature, determining whether the drinking

water is heated up or cooled down throughout the network; 2) the inlet (initial) water temperature; 3) the time that the water has spent in the network; and 4) the heat transfer coefficient, taking into account thermal resistances of soil and pipe materials, among other things. The latter is subject to variations in the literature, and Papers III–V apply the Hubeck-Graudal implementation (Paper V).

Figure 8 illustrates the concept and effect of these four parameters on a simple network setup (Figure 8a) with two parallel running pipes with different insulating materials. Varying demand and closing of either pipe will change the time spent by the water in the WDN. Figure 8b shows the effect of varying demand on the retention time of the water in the WDN and thus the simulated end temperature, depending on whether the unlined cast iron or polyethylene pipe are closed or both pipes are kept open.



(a) Network setup. (b) Temperature simulation.

Figure 8. Modelled temperature variation when closing a polyethylene (PE) or unlined cast iron (CI) pipe, or having both pipes open. Based on the temperature model described in Paper IV.

In the following, two examples are described showing that temperature data from the distribution system, as well as from smart meters, can be used to update utilities' understanding of their systems.

4.1 Improved system understanding through temperature modelling

As shown in Figure 8, the state of valves in particular can have a major effect on the temperature fluctuations in the WDN. To show whether the application of temperature data can help in identifying major system anomalies (Type 0, Figure 5) otherwise overseen in WDNs, the effect of open and closed valves on temperature simulations was analysed in Paper III.

Knowing the location and correct status of valves (i.e. open or closed) is of paramount importance for utilities. The list of possible negative consequences is long and includes: areas incorrectly narrowed down during contamination events; improperly working fire hydrants; and a greater proportion of affected customers without service during construction works (Deb et al., 2012; Delgado and Lansey, 2009; Wilson, 2011). Moreover, anecdotal evidence from Denmark indicates that DMA water balances that do not add up are often a result of unknown valve status. To mitigate these problems, utilities often maintain databases on the location and status of valves (Walski et al., 2003). Even if utilities are unaware about the actual status of a given number of valves, the state of the valves is often assumed known in WDN model analyses (e.g. Sophocleous et al., 2017; Wu et al., 2012). Including such possible incorrect valve settings in a hydraulic model generates unreliable results, particularly when simulating water quality (Savic et al., 2009).

In the literature, examples exist where the correct location and status of valves were identified by combining a hydraulic model with pressure and flow measurements (e.g. Delgado and Lansey, 2009; Do et al., 2018; Sophocleous et al., 2017; Walski et al., 2014; Wu et al., 2012). Yet none of these studies has incorporated temperature measurements as an additional parameter. One major issue of combining models with flow and pressure data to determine the valve status is the lack of significant head loss throughout the WDN. For example, there was as little as an average of 0.2 m head loss throughout the DMA in Paper I. The head loss can be increased, e.g. by opening hydrants, but it is a rather labour-intensive approach. Thus, it was investigated whether temperature data could simplify this approach and increase chances of correct valve status identification and location.

In Paper III, a semi-synthetic and a real case study were presented and a genetic algorithm (GA) was applied to identify the location and status of valves, including temperature data. This type of optimisation algorithm was run to identify the best combination of open and closed valves in the WDN by minimizing a fitness function. As the search space of open and closed valve combinations can be large, the number of assessed valves was reduced where possible. For example, valves that disconnected consumers entirely from the network when closed were not considered in the search space, since this would be reported by consumers otherwise. The aim of the semi-synthetic case study was to show whether temperature measurements alone, e.g. available from smart meters, could be utilised to identify correct valve status. In the second case study, real

temperature, flow and pressure measurements from a transportation network were applied and the valve settings were manually tested in the field.

4.1.1 Semi-synthetic case study

The semi-synthetic case study included real temperature measurements at the inlet of a DMA. First, the synthetic 'true' nodal temperatures were generated by closing five out of the 379 valves included in the search space. The remaining valves were left open. Next, starting with the assumption of all valves being open, the GA was applied to identify the five closed valves based solely on the 'true' nodal temperatures. These would in a real case system stem from smart meters; however, Paper IV indicated that in reality the collection of this 'true' data set is difficult (see also Section 4.2).

A total of 48 different GA setups were run, each identifying a best set of closed valves termed 'best fit'. In these runs, the perfect solution was identified 16 times. In total, the 'best fits' identified 183 out of 240 possible 'truly' closed valves and 21 incorrectly closed valves. However, these incorrectly closed valves were often in close proximity to the actual closed valves, and thus the exercise could potentially still guide utilities to areas of concern. In some runs, the GA was stuck in local minima, among other things, because of small temperature differences occurring when closing selected valves, the relative large search space and the specified GA parameters.

The GA runs showed that filling the initial GA population with a weighted approach (incorporating some prior knowledge based on simulations) and a larger initial population had the highest impact on the outcome. This is of importance, as it can improve future modelling approaches. Where possible, the search space should be reduced by incorporating the knowledge of preliminary model runs and operators: "This valve should be closed in the model, and be excluded from optimisation, as it was tested last week", increasing the overall success rate of the GA.

4.1.2 Real transportation network case study

The second case study analysed valve settings, a transportation WDN model and a week's measured flow and pressure measurements at ten DMA inlets, one tank and two waterworks. As in the synthetic case, a GA was used to identify valves status. Even though some of these were marked as 'closed' by the utility, this information was not used in the modelling procedure owing to the high likelihood that it was incorrect. In the analysis, temperature measurements

at two DMA inlets in combination with all pressure measurements were incorporated into a fitness function. The GA was then run with no prior knowledge (i.e. all valves open) about the valves' actual status. Among other things, different weighting scenarios between temperature and pressure measurements were applied in the fitness function to highlight the individual and combined effect of the two parameters on valve status identification.

Identification of 'system anomalies' (Type 0) prior to application
The interest in temperature data revealed various Type 0 anomalies (Figure 5). First, two temperature meters had to be excluded from analysis as they measured doubtful values, ranging between < -5 and > 40 ºC. Moreover, one meter with reliable measurements was thought to be located at the inlet of a tank, but preliminary temperature simulations resulted in a contradiction between the meter's temperature measurements and another meter's data set. Discussions with the utility revealed two different SCADA setup diagrams, where the meter on the first diagram was located at the tank inlet (incorrect) and at the outlet (correct) on the second diagram.

All pressure measurement sites showed reliable dynamics, but test samples taken at the sites revealed pressure offsets between -1.3 and 3 m. Consequently, the fitness function had to be modified by subtracting the simulated and measured median pressure to account for unknown drifts but still incorporating the dynamics in the data.

Identification of 'system anomalies' (Type 0) after application
The best fit of 30 GA runs resulted in the identification of 41 different out of a total of 106 possible closed valves. This number was reduced to nine valves by only considering those reappearing in at least 50% of the weighting scenarios. Two of these valves were identified only by the weighting scenario runs that solely included temperature measurements. Going to the field to check the actual settings of the valves revealed the following about the nine valves:

- Two valves were locked in and reported faulty (thus, these valves' status were unknown).
- Three valves were inaccessible, e.g. owing to heavy vegetation.
- One valve, marked as closed by the utility and scoring the highest score in all GA runs (closed in 24 out of 30 runs) was possibly open (unfortunately, the technician was in doubt).

- Three valves no longer existed, and these and adjacent parallel pipes should be removed from the utility's asset and hydraulic WDN model. The WDN model did thus not reflect the real system.

Overall, application of the temperature model and field tests casts doubt on the validity of the hydraulic model. Moreover, at one critical location, temperature measurements indicated clearly that the water was distributed and mixed differently than anticipated by the utility. Pressure measurements and simulations indicated that the water originated from a nearby waterworks, whereas temperature measurements indicated that the water originated from a nearby tank. The latter could partially be validated during a period where the waterworks was running on low capacity. Only the weighting scenarios considering temperature pointed to this critical location and a 'valve station' was discovered with locked-in and rusty valves and bypasses that were not part of the asset database.

The application of the methodology in Paper III revealed that poor data quantity and quality are major restricting factors when using hydraulic and temperature models for determining valve status. Additional validation tests of the temperature measurements and a more thorough analysis of the real status of the valves are required. At present, the method can be used to identify errors in the utility's asset database (i.e. Type 0 anomalies, Figure 5) and of Types 1 and 2, but it is first possible to pinpoint valves that really are closed if a high data reliability is also assured. Thus, whereas the valves might be closed as indicated by the models, the high uncertainty in the input data and unclear valve tests make it difficult to trust the modelling results at present.

The following can be concluded on the applied methodology.

- The evaluation of temperature data in combination with hydraulic data, prior to application of the methodology, revealed various errors otherwise overseen by the utility: invalid temperature and pressure observations and incorrect conceptualizations of where data is collected.
- The applied methodology resulted in the discovery of forgotten and faulty valves and other errors originating from the utility's asset database. However, as long as the high uncertainty in the input data is not reduced it is not possible to apply advanced methods such as the one presented for valve status detection.
- At one location, temperature data showed that the water was mixed differently than anticipated by the utility which, among other things, based its knowledge on (perhaps faulty) pressure measurements. This highlighted

the value of using temperature data for an improved understanding of the WDN.

4.2 Smart meter temperature data

One of the restraining factors of the applied methodology in Paper III was the reduced number of available temperature measurements as well as their limited spatial distribution. Thus, Paper IV looked into the question of whether smart meter temperature data can be used to overcome this limit. Moreover, Paper IV was applied to identify whether smart meter temperature data can be used to validate the temperature model also applied in Paper III (Section 4.1).

In the main, smart meter data is collected for improved billing purposes or enhanced water loss assessments; the temperature data, however, is collected as a 'spin-off' product and not optimised for usage. Whereas the temperature data applied in Paper III could be compared directly with temperature observations from the WDN, temperature data from smart meters does not directly represent the WDN temperature. This is because the water may spend a substantial amount of time in service pipes and (in-) house connections before being metered. Based on varying residence times in such connections, the smart meter temperature data can represent nearly the WDN temperature, the soil temperature, a mixture in between the two, or a temperature affected by other external heat sources such as the indoor air temperature (Figure 9). The applied real smart meter data in Paper IV thus required some classification and filtration prior to application.

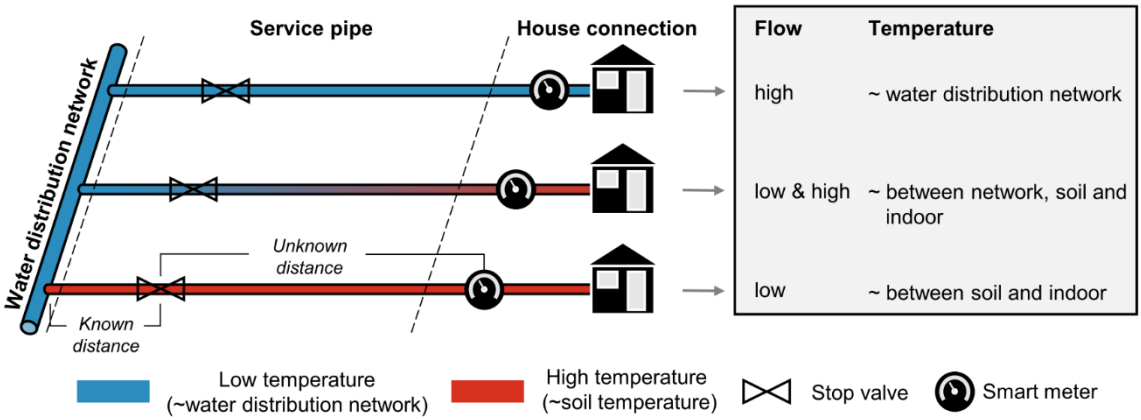


Figure 9. Variation in demand affects the residence time of the water in service pipes and house connections, leading to differing smart meter temperature measurements at the consumers' homes. Example from summer, with higher soil than water source temperatures.

4.2.1 Filtration and classification of smart meter temperature data

In the case study DMA, two weeks of temperature and consumed volume smart

meter data were available at all homes. The data covered a period during August, with higher soil than DMA inlet temperatures. For most consumers, information about service pipes from the asset database were available, only covering the distance between WDN mains and stop valves on the consumers' properties (Figure 9). However, no information about the pipe stretch from stop valves to the smart meter locations existed and the smart meter location was unknown, as the obtained GPS coordinates only represented the coordinates of the consumers' properties. Based on this scarce information, the service pipes were prolonged linearly from each stop valve to the GPS coordinates of each property.

Identification of 'system anomalies' (Type 0) through soil temperature estimates

The soil temperature was used as a boundary condition in the heat transfer model, but the actual soil temperature in the DMA was unknown. To mitigate this, the soil temperature was estimated using the smart meter temperature data. Low or no consumption lead to stagnant or slow-flowing water in service pipes, with the water temperature gradually going into equilibrium with the soil temperature (Figure 9). Sudden demand pushing this 'stagnant volume' into consumers' homes can result in temperature samples which, to some extent, represent the soil temperature. Owing the uncertainty of service pipe lengths (and their diameters) between stop valves and smart meter locations and the time of actual consumption, a conservative filtering was applied to identify samples representing the soil temperature. This was done by only accepting samples representing: 1) retention times > 3 hours in the service pipe; 2) a consumption equal to 20–80% of the service pipe volume; and 3) temperature samples that were at most 15 minutes old (otherwise they may potentially have been affected by, and represented, indoor temperatures).

Samples passing this conservative filtration revealed a slowly decreasing soil temperature over two weeks with some clear outliers (> ±3 ºC from mean). Manual inspection of all smart meter locations in a geographic information system revealed that these outlying temperatures were found in the only location with two-storey houses (each storey had its own smart meter). Thus, the actual service pipe and house connections were much longer than anticipated. Furthermore, outliers of soil temperature, being colder than the mean temperature, revealed locations where no stop valve information was available; thus, the model-building process connected service pipes to the nearest main. The too-cold soil temperature estimates indicated that the service pipe locations,

diameter or length at these locations were incorrect. Thus, analysing the temperature data based on knowledge of service pipes revealed incorrect and missing information about consumers in the utility's asset database (i.e. Type 0 anomalies, Figure 5). Real soil temperatures are currently being sampled to validate the ones estimated from smart meters.

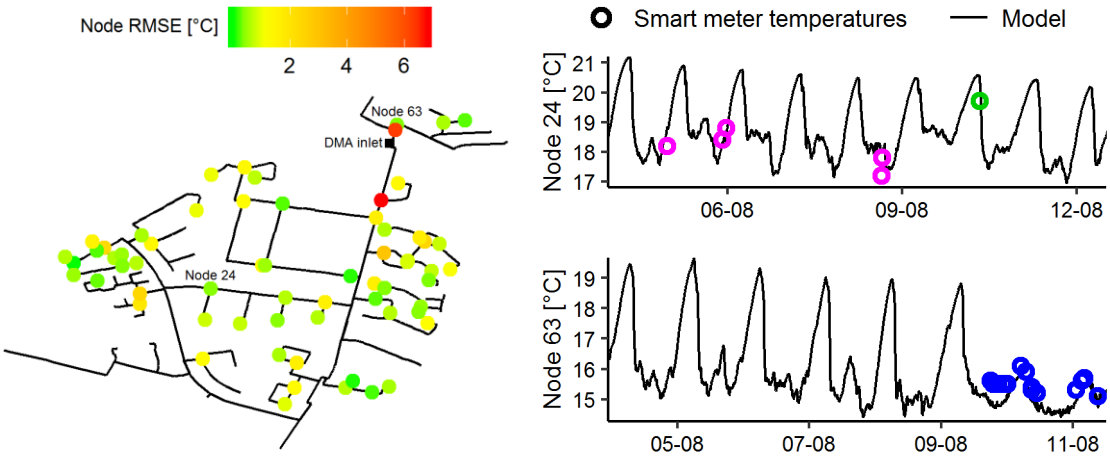
Water distribution network temperatures

The heat transfer model is validated by comparing the simulated network temperature with 'measured' smart meter temperatures. Except for the DMA inlet, however, the actual temperature throughout the DMA was unknown, as no temperature sensors were installed directly on the WDN mains. Like the soil temperature, the 'measured' WDN temperature was therefore estimated using smart meter temperature data. WDN temperatures were estimated using smart meter samples where at least 400% of the service pipe volume was consumed within 15 minutes. This conservative filtering was applied to overcome issues based on too coarse network skeletonisation, and to reduce uncertainties owing to unknown service pipe and home connection lengths and diameters. Some samples passing the filtering were above the soil temperature, indicating uncertainty in the applied filtering process. This uncertainty is expected to stem from external heating sources, incorrect pipe characteristics estimations or higher local soil temperatures than expected. The remaining temperature samples showed expected variations and were used to assess the temperature model's validity.

4.2.2 Temperature model analysis with smart meter data

The temperature model, with the soil temperature estimated from smart meter measurements and the measured inlet DMA temperature used as boundary conditions, was run and compared to the estimated 'measured' network temperatures for validation purposes. Samples representing the WDN temperature showed good match with the simulated temperature in all but two nodes having a mean RMSE of 1 ºC (Figure 10). Two nodes were removed from the analysis, as their RMSE stood out (> 6 ºC, Figure 10a), improving the mean RMSE to 0.9 ºC. For one node, the reason was a too coarse skeletonisation during the building process of the hydraulic model. In the hydraulic model, smart meter temperatures were used to estimate the network temperature in the nearest node; in the real WDN, however, the temperature samples represented the temperature in a side-branch, explaining the high temperature deviation between model and samples. In the other case, discussions with the utility revealed that a consumer had incorrect GPS coordinates and was not part of the DMA. Thus, running a temperature model showed that the model could validate data and

identify larger flaws such as incorrect consumers that would otherwise not be detected (i.e. Type 0 anomalies, Figure 5).



(a) Overview of nodal RMSE in case study DMA. (b) Smart meter and modelled temperatures.

Figure 10. Examples of temperature variation throughout the case study water distribution network (Paper IV). (a) Nodal root mean square error variation (RMSE) based on temperature model and smart meter temperatures. (b) Example of two nodes with an RMSE of 0.36 ºC (upper panel) and 0.32 ºC (lower panel) where distinct colours represent individual homes' smart meters.

I believe that this small analysis revealed only a small fraction of smart meter temperature potentials. Future applications include, among other things, the following (see Paper IV for additional potentials).

- Leakage detection, as higher flow rates lead to different temperature gradients than anticipated. Reliable temperature simulations can then be used to pinpoint locations with large differences between model output and temperature observations.
- Valve status identification (similar to Paper III), as the water is distributed differently than anticipated, changing the temperature profile throughout the WDN.
- Consumer alerts, e.g. when temperatures are too high or too low over a certain period.

Hydraulic network setup and sampling resolution

In particular, the way smart meter temperature data was sampled posed a challenge to the application of the data. Smart meters in the case study DMA sampled a large proportion of data, where the temperatures represented a mixture between service pipe and household connection temperatures. In Paper IV, less than 0.2% of all smart meter temperature data passed the conservative filtration

for estimates of soil and network temperatures. A more intelligent metering than sampling every hour, or quasi-random as in the case study, would be preferable. For example, sampling each time a certain volume has passed the meter could increase the percentage of applicable data. The average sampling resolution in the case study of around 30 minutes was insufficient. Thus, a finer sampling resolution (at least during periods of consumption) would improve the applicability, as the uncertainty of when the water actually has been used is then reduced. The second challenge includes the level of skeletonisation of the hydraulic model, which revealed a high impact on temperature simulations. A higher level of detail of the hydraulic model is required in future applications (for example, not bundling consumers in nodes), making it possible to represent temperatures at the actual location in the WDN, and also improving water quality simulations.

Currently the following can be concluded on smart meter temperature data.

- The temperatures throughout the DMA were simulated to a satisfying degree (average RMSE of 0.9 ºC). This made it possible to use the temperature model and compare the output with smart meter temperature data to highlight incorrect information about service pipes and consumers in the asset database, not found easily without these temperature data.
- Based on the applied filtering, only a small proportion (< 0.2%) of the collected data could be used to represent WDN and soil temperatures, revealing that smart meter temperature data is not sampled in a preferable manner. An improvement would include finer sampling resolutions (< 30 minutes) or additional sampling during times of high consumption, which would increase the possibility of representing WDN temperatures.
- The list of potential valuable applications of smart meter temperature data needing further research is long, including improved leakage detection and valve status identification.

4.3 Drinking water as a low-temperature source in district heating systems

When electricity prices are low, the operation of electrically driven heat pumps, extracting energy from WDNs and transferring it to district heating systems, seems to be a perfect match. This is firstly because the heat provided from the WDN to the district heating system is potentially replacing less sustainable sources of energy, such as fuels used in combined heat and power

plants. Secondly, it is because WDNs have stable temperatures and are available in many places around the world. Finally, low drinking water temperatures may reduce the risk of biofilm growth (Liu et al., 2016) and increase the share of water complying with the recommended upper temperature limit of 12 ºC at the tap in Denmark (Ministry of Environment and Food of Denmark, 2018b). Real-case examples of heat pumps installed in WDNs already exist, as exemplified by a smaller utility in Northern Jutland, Denmark, supplying around 15% of the annual required energy for district heating (Cronborg A/S, 2014). However, there is a potential socioeconomic downside: does the lower water temperature lead to more energy being required to heat the water at consumers' homes?

To analyse this effect, Paper V simulated the temperature throughout Copenhagen's WDN and included the effect of eight heat pumps, deployed at favourable locations in the model on the WDN mains. Figure 11 exemplifies the linkage between the WDN and a district heating system with a heat pump. The drinking water is cooled significantly at the heat pump location and, in this example, due to the soil being warmer than the water, some of the extracted energy is regained from the ground over time. Hereby, the water temperature slowly re-approaches the soil temperature downstream of the heat pump (Figure 11).

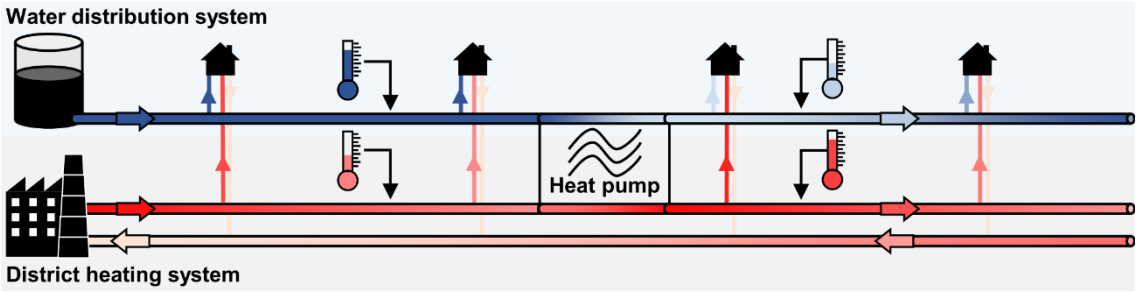


Figure 11. Linkage between a water distribution system and district heating system by a heat pump cooling the drinking water below soil temperature and transferring heat to the district heating system.

Case studies exist in the literature that assess the benefits and consequences of heat pump installations in WDNs (Blokker et al., 2013; De Pasquale et al., 2017; Paper V). However, owing to differing WDN characteristics (such as residence times) and differences in the temperature model setup (for example, whether a term describing the thermal resistance of the soil is included and whether soil temperatures were estimated or measured), results are not easily comparable and depend on the system setup.

4.3.1 Case study: Greater Copenhagen Utility, Denmark

The Hubeck-Graudal heat transfer model (Paper V) was validated against weekly sampled temperature measurements covering two years at 15 sampling locations. The simulated results showed a good match with the measurements.

The case study assessed the effect of reducing the drinking water temperature by 5 ºC at eight favourable heat pump locations. The analysis revealed a total heat extraction potential of 29.2 MW for Copenhagen's WDN. When including the work supplied by the heat pumps (35.9 MW), this amounts to around 2% of Copenhagen's peak heat demand during January and is 50% higher than previously estimated by Bach et al. (2016). This large difference was mainly explained by not deploying heat pumps to reservoir locations only. The case study simulations revealed that 38% (11.1 MW) of the extracted heat will be returned by the soil (i.e. water that has been cooled through heat pumps will regain heat downstream). Moreover, 33% (9.6 MW) of the energy has, based on water demand assumptions (Rygaard et al., 2013), no effect on the district heating demand of consumers, resulting in a net heat potential of 20.7 MW from the source. In other words, a heat utilisation degree of 71% was computed for Copenhagen's WDN.

The coefficient of performance (COP) is a term describing the ratio between the amounts of energy provided and the work (i.e. electricity) required to run heat pumps. Usually, this value exceeds unity, showing that more heat is extracted than work required to run the heat pump. The COP ratio can be applied to heat pumps individually and to the entire system. When applied as system COP, the additional electricity and district heating demand required for heating at consumers' homes, because of cooler water temperatures downstream of heat pumps, are taken into consideration. The application of a conservative heat pump COP of 2.9 resulted in a system COP of 1.7. This value described a relatively low ratio of heat provided over the total work required by the system, when compared to more generic integrations of heat pumps in district heating systems (having a system COP of up to 5 (e.g. Ommen et al., 2014)). Even a less conservative heat pump COP of 4 would only increase the system COP by up to 1.9. Thus, a more thorough (economic) analysis is required to assess whether heat pumps extracting energy from WDNs are more competitive when other solutions, including wastewater, seawater and fresh water sources, are not available (e.g. Elías-Maxil et al., 2014).

The heat utilisation analysis of the WDN revealed that the thermal resistance

from pipe materials in general was insignificant compared to the thermal resistance from the soil material. As the exact thermal conductivity of the soil was unknown, a sensitivity analysis was conducted modifying the parameter by ±30%. Whereas the analysis showed that the degree of heat utilisation of the soil changed by 11–15%, the overall degree of heat utilisation in this specific case study changed by only 2–4 percentage points.

Even though the energy benefits were debatable, the application of heat pumps had a significant positive impact on the end-use temperature, as the share of water delivered to users complying with the upper limit of 12 ºC increased from 41% to 81% during August.

In Paper V, the degree of heat utilisation of the soil increased from 34% to 38% when including the heat recovery occurring in service lines. This additional heat recovery was, however, based on scarce information about service lines, such as average service pipe lengths and diameters and limited demand information stored in the utility's hydraulic model. The application of smart meter temperature data in Paper IV showed that the water in service pipes can be stagnant or slow-flowing for long periods, indicating that the heat transfer in service lines may be underestimated significantly. Moreover, the applied WDN model in Paper V did not contain up-to-date demand patterns, nor did the model reflect seasonal demand variations. Smart meter data, as available in Paper IV, could be useful to improve the demand dynamics in the system and thus be used to validate the heat transfer model even further, such as during periods where the WDN model's demand pattern very likely does not resemble the real demand in Copenhagen's WDN.

In Paper V, the heat transfer model was a valuable tool for assessing the effect of installing heat pumps and cooling the water in a WDN.

- Whereas a net potential of 20.7 MW can be extracted from Copenhagen's WDN, the system COP was as low as 1.7, indicating that other sources of energy should be considered prior to heat pump deployment in Copenhagen's WDN.
- The share of water at consumers' nodes complying with a statutory recommendation (< 12 ºC) increased from 41% to 81% during August when installing heat pumps, thus improving the quality of the distributed water. This indicated a clear benefit of heat pump deployment.

- Future investigations should incorporate additional information, such as smart meter temperature data, to further validate the temperature model and assess the effect of heat recovery occurring in service lines.

4.4 Smart meter data as an add-on to urban drainage management

In Copenhagen, at least 83% of the consumed drinking water is expected to end up in the urban drainage system (Rygaard et al., 2013). This highlights the potential of using and linking smart meter consumption data with urban drainage models to estimate, for example, wastewater flows and constituents, as mentioned in the literature (Cole and Stewart, 2013; Monks et al., 2019; Nguyen et al., 2018). Comparing these estimates with in-sewer flow observations may furthermore be used to estimate infiltration/exfiltration in the urban drainage system. In Paper VI, this potential was tested on a real case study. Smart meter consumption data was bundled according to the meters' wastewater catchments (the 'DMA-equivalent' in urban drainage management) and compared to in-sewer observations from a combined sewer system. Hereby, the potential of using smart meter data to simulate the wastewater component of the dry weather flow in an urban drainage system was established, and the methodology could furthermore be used as a data anomaly and system validation tool.

4.4.1 Case study: Utility Elsinore, Denmark

In Paper VI, data from five in-sewer sensors covering three monitoring periods without rain was available for testing. Figure 12 displays the concept behind the applied method in Paper VI, where in-sewer observations were available from five catchments.

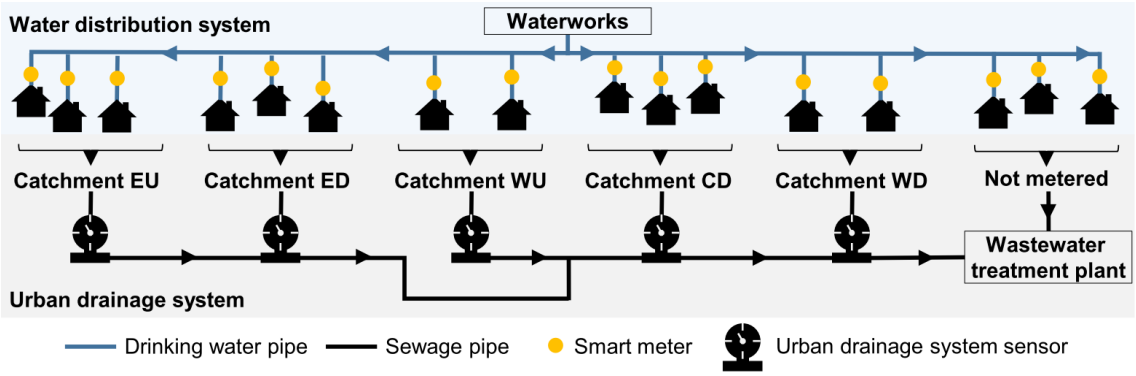


Figure 12. Linkage between a water distribution system with smart meters deployed and an urban drainage system consisting of five catchments. In-sewer sensors were installed downstream of these five catchments. Utility Elsinore, Denmark (Paper VI).

Moreover, data on the outflow from the waterworks and the inflow to wastewater treatment plants was available. Using this data, two approaches were tested:

1 Summing the smart meter data, based on upstream catchments.

2 Linking the smart meter data to an urban drainage model.

The first approach, being a more simple process, did not include the routing time through the urban drainage system.

Prior to the application, two typical problems related to digitalisation were identified. First, the evaluation of smart meter data from Utility Elsinore revealed Type 0 anomalies (Figure 5). These included a missing sub-catchment belonging to catchment 'EU' (Figure 12) in the wastewater plans of the utility as well as consumers without smart meters (only accounting for < 2% of the demand in the analysed area). Second, a substantial effort was put into gaining access to data from the utility and its partners. The smart meter and in-sewer sensor data, the WDN model and the urban drainage system model were managed by four individual contractors, and the wastewater treatment plant and waterworks data was obtained from various utility employees.

The summed smart meter data was compared with outflow measurements from the utility's waterworks. The results showed a waterworks outflow around 15% higher than the summed data in all three periods. This deficit might have been caused by, among other things, leakage, unknown consumers or sensor errors. Next, the wastewater treatment plant inflow was compared to the simulated inflow based on smart meters upstream of the five in-sewer sensors and additional catchments (Figure 12). Here, the simulated inflow was 50% higher than the observed one during two monitoring periods, but showed a high match during the third monitoring period. The smart meter data, however, showed a high match with audit data prior to smart meter deployment and did not markedly change its dynamics over the three monitoring periods. Thus, the smart meter data was deemed more valid than wastewater treatment plant observations. Most likely, deviations were caused by erroneous sensors at the wastewater treatment plant or other factors, such as exfiltration, occurring in the urban drainage system.

Looking at the five in-sewer observations only (Figure 12), the wastewater flow did not increase as expected in all downstream catchments, indicating possible exfiltration and infiltration of water into the urban drainage system, or in-sewer sensor errors. Moreover, residuals between observed and simulated

wastewater flow, varied greatly in all measuring periods and catchments, ranging from around -20 L/s (observation < simulation) to 70 L/s (observation > simulation). Different sources of error were discussed extensively in Paper VI, including consumed water not discharged to the sewer system or construction dewatering unintentionally released to the urban drainage system; however, no consistent positive or negative trend was seen in the mass balances of the catchments. It was thus concluded that the most likely reason for the observed discrepancies was erroneous in-sewer sensors. For example, the measurements in the most downstream catchment were at times 25 L/s higher than simulated, being an unrealistic additional volume of water ending up in the wastewater treatment plant (where differences between smart meter simulation and observed values were already too high). It is therefore believed that the degree of uncertainty of the in-sewer observations even exceeded expected uncertainties of up to 20%.

For each monitoring period, summed and simulated smart meter data flows upstream of each in-sewer sensor increased the further downstream the in-sewer sensors were placed (Figure 12). When compared with in-sewer observations (whose time dynamics are assumed trustworthy despite the uncertain flow magnitude), the included routing time by the simulated smart meter data showed a better timing of peaks and low points. This could be potentially useful for estimating infiltration of water into the urban drainage system and when used in a real-time control setup.

As the in-sewer observations were deemed the least trustworthy, the smart meter data posed a more reliable source for estimating the dry weather flow in the urban drainage system, even though it only represents the wastewater component. The case study showed that the application of smart meter data can also be used to highlight system anomalies, such as identifying unreliable data collected in the urban drainage system. Owing to the high uncertainty in the applied datasets, independent data sources are needed to verify that the in-sewer observations are indeed erroneous and to further estimate the other dry weather flow components, such as exfiltration and infiltration.

Thus, the application of smart meter data for comparison with in-sewer observations showed the following.

- Coupling smart meter data with urban drainage models increases the understanding of urban drainage systems and can be used as a tool to identify anomalies, including erroneous in-sewer observations.

- Owing to the large deviations between simulated wastewater flow and in-sewer observations, and because of various possible sources of error in the in-sewer data, smart meter data was deemed a more valid source for estimating the dry weather flow in urban drainage systems.
- Whereas linking the data from WDNs and urban drainage data was relatively simple, the access to data and information was complicated by being distributed between four contractors and several employees internally in the utility, which is an obstacle for further digitalisation of the water sector.

5 Conclusions

The water sector conceives digitalisation as a way to solve many of the challenges currently faced. Using six real-world case studies and data from five utilities, this thesis developed a range of novel methods to identify and address challenges and potentials for utilities to become increasingly data-driven.

Being data-driven requires action within at least three fields: 1) data collection, 2) data validation and reconstruction, and 3) application of the data. The investigated case studies revealed that these actions are highly interlinked and will enhance and complement each other.

Data collection. One typical example of digitalisation is the enrolment of smart meters, a process familiar to all five case study utilities. The sampling resolution of smart meters can have a major impact on reaching different data-driven goals. In terms of water loss assessment, commonly implemented sampling resolutions of consumption data ranging between 1 and 24 hours proved to be sufficient if representative demand patterns are available and used to interpolate in between adjacent data points. It is especially relevant to use such weighted demand-pattern-based approaches over linear interpolation for sampling resolutions > 2 hours. When it comes to water age simulations, however, sampling resolutions finer than 1 hour are needed to increase the validity of the simulations significantly. Moreover, application of smart meter temperature data showed that this data was not sampled optimally in regards to using the data for analysing the temperatures in the WDN. Only a small fraction of the smart meter temperature data (< 0.2% over 2 weeks) represented WDN and soil temperatures, which were essential inputs to simulate the temperatures in WDNs. Thus, additional sampling points or a finer sampling resolution than available in the case study utility (< 30 minutes) is likely needed.

Data validation and reconstruction. The case studies highlighted the need for an increased focus on data validation. This PhD project developed a systematic approach by categorizing anomaly types into four groups. Type 0 anomalies described ‘static’ data, representing missing and incorrect system attributes (e.g. wrong pipe diameters or consumer affiliations). Type 1 and 2 anomalies described sensor data that did not reflect the real state (e.g. illegitimate data) and misrepresented the actual state (e.g. drifted data from a miscalibrated sensor) of the WDN, respectively. Type 3 anomalies were used to classify abnormal measurements representing actual events (e.g. bursts) in the WDN.

The project implemented seven validation tests to identify Type 1 and 2 anomalies. Application of these tests revealed that many errors prevailed for long periods, with 3–35% invalidated and 10% missing data in the data sets from three utilities. Running such tests in an operational setting as well as analysing similarities in the occurrence of anomalies may help identify major flaws in the data collection schemes and help utilities to react in a fast manner, reducing the overall number of anomalies in the utilities' data sets.

High numbers of missing and invalidated data points also lead to a need for reconstruction processes. The thesis demonstrated data reconstruction based on artificial neural networks. Reconstructed consistent data streams can serve as an important input in, for example, online models. The reconstructed data can also help to identify Type 3 anomalies.

Application. In the literature, numerous studies highlight many potential data-driven applications of WDN analyses, but their success depends on the quality of the data. In this PhD thesis, the utilisation of real-world sensor data and application of advanced methods were particularly useful in highlighting Type 0 anomalies, i.e. flawed registrations in the utilities' asset databases. A large number of Type 0 anomalies limits the success of more advanced analyses.

A novel approach combined WDN temperature data, temperature simulations and a hydraulic WDN model, with the intention of identifying valves with unknown status. However, the analysis led to identification of errors in the utility's asset and sensor data instead, ultimately casting doubt on the validity of the hydraulic model. The combined temperature and hydraulic modelling thus turned out to have another benefit than originally intended. Moreover, WDN temperature data proved to have a clear value for the case study utility, as it indicated the water's origin and path taken in the WDN, which is otherwise not easily understood from pressure and flow data alone.

A low number of WDN temperature measurements can be overcome by using smart meter temperature data, which is potentially available from each household. In another case study, comparing smart meter temperatures with simulated temperature values throughout a DMA highlighted the applicability of the data to represent WDN temperatures (mean RMSE of 0.9 ºC). Well-sampled data and detailed WDN models having distinct nodes for all consumers may provide temperature simulations suitable for leakage localisation or valve status identification.

Another use of a WDN temperature model was identified by analysing the effect of eight heat pumps deployed in Copenhagen between the drinking water and district heating systems. The analysis revealed a positive net heat extraction potential of 21 MW; this may be an important supplement, especially during peak heat demands. In addition, the analysis revealed that the installation of heat pumps may improve summer water temperatures markedly by increasing the share of consumers receiving water under 12 ºC from 41% to 81% during the month of August.

In another case study, smart meter data from the WDN was coupled with an urban drainage model. Access to the data was complicated by being distributed between different contractors and staff members, representing a typical obstacle of digitalisation. Large deviations were observed between simulated wastewater flows and the in-sewer observations, and the smart meter data was deemed as a more valid source for estimating the dry weather flow in urban drainage systems.

This thesis revealed that there is a great potential hidden in the often unused data from water distribution systems. For example, it was demonstrated that smart meter data, and in particular temperature data (which is often overseen in optimisation and modelling tasks), add great value to the understanding and management of water distribution systems and beyond. Another important outcome of this thesis is the realization that it may be more challenging for utilities than expected to become increasingly data-driven. This is, among other things, because too many errors and uncertainties still exist in the utilities' asset databases and sensor data. In the years to come, utilities, technology providers and researchers need to collaborate to identify and reduce these uncertainties; thereby allowing the water sector to reach higher levels of digital maturity.

6 Perspective

The application of real-world data revealed various potentials, but also challenges that need to be addressed by utilities, researchers and technology providers, before utilities may become more data-driven.

Digitalised utilities. Even though the case study utilities have invested in monitoring capabilities in the form of SCADA systems and have started full smart meter rollouts, the use of the collected data is very limited. Many reasons may exist for this, including limited resources for analyses, missing analytical tools, or lack of funding for managing data and devices. This limited usage also means that utilities do not know the actual quality of their data. However, this thesis revealed that a generally low data reliability was a major issue in all case study utilities. Until reliability is improved, this will limit the potential of data-driven applications. I suspect that if utilities want to climb from 'basic' or 'opportunistic' digital adaptation to 'systematic' and later become 'transformational' (Figure 2), the following three steps are of importance.

1 High-quality asset data and a thorough system understanding need to be established. For example, are pipes, valves and consumers located as stated in the utility's database? To the best of my knowledge, fully automated procedures for identifying such discrepancies are missing, but the application and model-aided evaluation of data is a first step to identifying and reducing such errors.

2 It is of paramount importance to secure a high data reliability. Otherwise, there might be no reason for deploying additional sensors. Applications depending on the data may run flawlessly in the beginning, but the number of errors normally increases over time owing to limited resources for re-calibrating sensors and keeping system changes up-to-date in databases. Thus, utilities need to allocate resources for a proper validation to maintain a high reliability of the data, and the use of data validation tools should be part of the daily tasks in a utility.

3 When the reliability of the data is assured and the number of major system anomalies is reduced, utilities need to increase their usage of the data. This usage will likely highlight additional minor 'system anomalies' and improve the understanding of the system. This process is expected to result in further improvements in data collection and reliability. This will also require involvement from the research community and technology providers, who

should help utilities to identify and provide optimal data collection possibilities and analytical tools as well as demonstrate the application of these in real-world case studies.

Research opportunities. Use of temperature data and simulation of the temperature throughout the WDN revealed various benefits. To further highlight the true potential of the temperature simulations for valve status identification, additional studies need to be conducted. These include in-depth sensitivity analyses discussing when the method can be applied successfully and when not (for example, identifying which temperature gradients are needed throughout the WDN). Moreover, the applied genetic algorithm and temperature modelling software (EPANET-MSX) turned out to have limitations in terms of speed and convergence success at the applied fine sampling resolutions and are therefore currently not suitable for daily operations; thus, there is a demand for additional research in computational optimisation.

The true value of using smart meter temperature data was inhibited by how the data was sampled and categorized. Whereas the sampling resolution could be increased by the technology provider, additional procedures for categorizing the temperature data into soil and WDN temperatures will be useful. Owing to the uncertain nature of service line data, approaches that are independent of the physical properties of service lines should be investigated. These include, for example, unsupervised clustering algorithms, identifying smart meter data that represent soil and WDN temperatures. Next, it should be investigated to what extent the data can be used for more advanced methods such as valve status detection or leakage localisation. Smart meter temperature data could also be used to improve the knowledge of utilities about heat recovery in service lines. If there is indeed a higher recovery potential, the benefit of heat pumps deployed in WDNs may be greater than estimated here.

7 References

Adedoja, O.S., Hamam, Y., Khalaf, B., Sadiku, R., 2018. A state-of-the-art review of an optimal sensor placement for contaminant warning system in a water distribution network. Urban Water J. 15, 985–1000. doi:10.1080/1573062X.2019.1597378

ASCE, 2011. Failure to Act: The Economic Impact of Current Investment Trends in Water and Wastewater Treatment Infrastructure. American Society of Civil Engineers, VA.

AVK, 2018. Smart pressure management. Available at https://www.avkvalves.com/-/media/websites/avk-global/files/downloads/avk-pressure-management-brochure (accessed November 12, 2019).

Bach, B., Werling, J., Ommen, T., Münster, M., Morales, J.M., Elmegaard, B., 2016. Integration of large-scale heat pumps in the district heating systems of Greater Copenhagen. Energy 107, 321–334. doi:10.1016/j.energy.2016.04.029

Bakker, M., van Duist, H., van Schagen, K., Vreeburg, J., Rietveld, L., 2014. Improving the Performance of Water Demand Forecasting Models by Using Weather Input. Procedia Eng. 70, 93–102. doi:10.1016/j.proeng.2014.02.012

Blokker, E.J.M., Pieterse-Quirijns, E.J., 2013. Modeling temperature in the drinking water distribution system. J. Am. Water Work. Assoc. 105, E19–E28. doi:10.5942/jawwa.2013.105.0011. Erratum (2018) in J. Am. Water Work. Assoc. 110, 98-98. doi:10.1002/awwa.1178

Blokker, E.J.M., van Osch, A.M., Hogeveen, R., Mudde, C., 2013. Thermal energy from drinking water and cost benefit analysis for an entire city. J. Water Clim. Chang. 4, 11–16. doi:10.2166/wcc.2013.010

Blokker, E.J.M., Vreeburg, J.H.G., Beverloo, H., Klein Arfman, M., van Dijk, J.C., 2010. A bottom-up approach of stochastic demand allocation in water quality modelling. Drink. Water Eng. Sci. 3, 43–51. doi:10.5194/dwes-3-43-2010

Blokker, E.J.M., Vreeburg, J.H.G., Buchberger, S.G., van Dijk, J.C., 2008. Importance of demand modelling in network water quality models: a review. Drink. Water Eng. Sci. 1, 27–38. doi:10.5194/dwes-1-27-2008

Blokker, M., 2019. Making water meters truly smart by adding additional features. Poster presented at the 17th International Computing & Control for the Water Industry Conference (CCWI). University of Exeter, UK.

Blumensaat, F., Leitão, J.P., Ort, C., Rieckermann, J., Scheidegger, A., Vanrolleghem, P.A., 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, 8488–8498. doi:10.1021/acs.est.8b06481

Boyle, T., Giurco, D., Mukheibir, P., Liu, A., Moy, C., White, S., Stewart, R., 2013. Intelligent Metering for Urban Water: A Review. Water 5, 1052–1081. doi:10.3390/w5031052

Bragalli, C., Neri, M., Toth, E., 2019. Effectiveness of smart meter-based urban water loss assessment in a real network with synchronous and incomplete readings. Environ. Model. Softw. 112, 128–142. doi:10.1016/j.envsoft.2018.10.010

Branisavljević, N., Kapelan, Z., Prodanović, D., 2011. Improved real-time data anomaly

detection using context classification. J. Hydroinformatics 13, 307–323. doi:10.2166/hydro.2011.042

Cole, G., Stewart, R.A., 2013. Smart meter enabled disaggregation of urban peak water demand: precursor to effective urban water planning. Urban Water J. 10, 174–194. doi:10.1080/1573062X.2012.716446

Cominola, A., Giuliani, M., Castelletti, A., Rosenberg, D.E., Abdallah, A.M., 2018. Implications of data sampling resolution on water use simulation, end-use disaggregation, and demand management. Environ. Model. Softw. 102, 199–212. doi:10.1016/j.envsoft.2017.11.022

Cominola, A., Giuliani, M., Piga, D., Castelletti, A., Rizzoli, A.E., 2015. Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review. Environ. Model. Softw. 72, 198–214. doi:10.1016/j.envsoft.2015.07.012

Creaco, E., Campisano, A., Fontana, N., Marini, G., Page, P.R., Walski, T., 2019. Real time control of water distribution networks: A state-of-the-art review. Water Res. 161, 517–530. doi:10.1016/j.watres.2019.06.025

Creaco, E., Pezzinga, G., Savic, D., 2017. On the choice of the demand and hydraulic modeling approach to WDN real-time simulation. Water Resour. Res. 53, 6159–6177. doi:10.1002/2016WR020104

Cronborg A/S, 2014. Energi fra drikkevand bruges til fjernvarme. Available at https://csr.dk/energi-fra-drikkevand-bruges-til-fjernvarme (accessed November 12, 2019).

Cugueró-Escofet, M.À., García, D., Quevedo, J., Puig, V., Espin, S., Roquet, J., 2016. A methodology and a software tool for sensor data validation/reconstruction: Application to the Catalonia regional water network. Control Eng. Pract. 49, 159–172. doi:10.1016/j.conengprac.2015.11.005

DANVA, 2018. Vand i tal 2018 (in Danish). Danish Water and Wastewater Association, Skanderborg. Available at https://www.danva.dk/media/5002/2018_vand-i-tal.pdf (accessed November 12, 2019).

De Pasquale, A.M., Giostri, A., Romano, M.C., Chiesa, P., Demeco, T., Tani, S., 2017. District heating by drinking water heat pump: Modelling and energy analysis of a case study in the city of Milan. Energy 118, 246–263. doi:10.1016/j.energy.2016.12.014

Deb, A.K., Snyder, J.K., Grayman, W.M., 2012. Management of valves to improve performance reliability of distribution systems. J. Am. Water Works Assoc. 104, E370–E382. doi:10.5942/jawwa.2012.104.0087

Delgado, D.M., Lansey, K.E., 2009. Detection of Closed Valves in Water Distribution Systems, in: Water Distribution Systems Analysis 2008. American Society of Civil Engineers, Reston, VA, pp. 1–7. doi:10.1061/41024(340)82

Do, N.C., Simpson, A.R., Deuerlein, J.W., Piller, O., 2018. Locating Inadvertently Partially Closed Valves in Water Distribution Systems. J. Water Resour. Plan. Manag. 144, 04018039. doi:10.1061/(ASCE)WR.1943-5452.0000958

Eggimann, S., Mutzner, L., Wani, O., Schneider, M.Y., Spuhler, D., Moy de Vitry, M., Beutler, P., Maurer, M., 2017. The Potential of Knowing More: A Review of Data-Driven Urban Water Management. Environ. Sci. Technol. 51, 2538–2553.

doi:10.1021/acs.est.6b04267

Elías-Maxil, J.A., van der Hoek, J.P., Hofman, J., Rietveld, L., 2014. Energy in the urban water cycle: Actions to reduce the total expenditure of fossil fuels with emphasis on heat reclamation from urban water. Renew. Sustain. Energy Rev. 30, 808–820. doi:10.1016/j.rser.2013.10.007

García, D., Creus, R., Minoves, M., Pardo, X., Quevedo, J., Puig, V., 2017. Data analytics methodology for monitoring quality sensors and events in the Barcelona drinking water network. J. Hydroinformatics 19, 123–137. doi:10.2166/hydro.2016.048

Gartner, 2019. IT Glossary. Available at https://www.gartner.com/it-glossary/ (accessed November 12, 2019).

German Environment Agency, 2019. Trockenheit in Deutschland - Fragen und Antworten (in German). Available at https://www.umweltbundesamt.de/themen/trockenheit-in-deutschland-fragen-antworten (accessed November 12, 2019).

Gurung, T.R., Stewart, R.A., Beal, C.D., Sharma, A.K., 2016. Smart meter enabled informatics for economically efficient diversified water supply infrastructure planning. J. Clean. Prod. 135, 1023–1033. doi:10.1016/j.jclepro.2016.07.017

Gurung, T.R., Stewart, R.A., Sharma, A.K., Beal, C.D., 2014. Smart meters for enhanced water supply network modelling and infrastructure planning. Resour. Conserv. Recycl. 90, 34–50. doi:10.1016/j.resconrec.2014.06.005

Kamstrup, 2019a. Mechanical vs. Ultrasonic. Available at https://www.kamstrup.com/en-en/water-solutions/themes/mechanical-vs-ultrasonic (accessed November 12, 2019).

Kamstrup, 2019b. flowIQ 2200 - Advanced residential water meter with acoustic leak detection. Available at https://www.kamstrup.com/en-en/water-solutions/smart-water-meters/flowiq-2200 (accessed November 12, 2019).

Kirstein, J.K., 2016. Using high-frequency data to optimize water supply operation - Case Study: Halsnæs Water Supply, Denmark. MSc Thesis. Technical University of Denmark, DTU Environment, Kgs. Lyngby.

Kirstein, J.K., Borup, M., Rygaard, M., Høgh, K., 2016. Målerdata kan gemme på gratis informationer for forsyningerne (in Danish). Danskvand 84, 50–51.

Kulkarni, P., Farnham, T., 2016. Smart City Wireless Connectivity Considerations and Cost Analysis: Lessons Learnt From Smart Water Case Studies. IEEE Access 4, 660–672. doi:10.1109/ACCESS.2016.2525041

Larsen, S.L., Christensen, S.C.B., Albrechtsen, H.-J., Rygaard, M., 2017. GISMOWA: Geospatial Risk-Based Analysis Identifying Water Quality Monitoring Sites in Distribution Systems. J. Water Resour. Plan. Manag. 143, 04017018. doi:10.1061/(ASCE)WR.1943-5452.0000754

LEAKman, 2018. 2,300 smart meters installed in demonstration areas. Available at https://leakagemanagement.net/demo-facilities/2300-smart-meters-installed-in-demonstration-areas/ (accessed November 12, 2019).

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, 429–441. doi:10.2166/ws.2014.131

Liu, S., Gunawan, C., Barraud, N., Rice, S.A., Harry, E.J., Amal, R., 2016. Understanding,

Monitoring, and Controlling Biofilm Growth in Drinking Water Distribution Systems. Environ. Sci. Technol. 50, 8954–8976. doi:10.1021/acs.est.6b00835

Loureiro, D., Alegre, H., Coelho, S.T., Martins, A., Mamade, A., 2014. A new approach to improve water loss control using smart metering data. Water Sci. Technol. Water Supply 14, 618–625. doi:10.2166/ws.2014.016

Loureiro, D., Amado, C., Martins, A., Vitorino, D., Mamade, A., Coelho, S.T., 2016. Water distribution systems flow monitoring and anomalous event detection: A practical approach. Urban Water J. 13, 242–252. doi:10.1080/1573062X.2014.988733

Makropoulos, Savić, 2019. Urban Hydroinformatics: Past, Present and Future. Water 11, 1959. doi:10.3390/w11101959

Mekki, K., Bajic, E., Chaxel, F., Meyer, F., 2019. A comparative study of LPWAN technologies for large-scale IoT deployment. ICT Express 5, 1–7. doi:10.1016/j.icte.2017.12.005

Ministry of Environment and Food of Denmark, 2018a. Miljø- og fødevareministerens endelige besvarelse af spørgsmål nr. 775 (MOF alm. del) stillet 8. juni efter ønske fra Julie Skovsby (S) (in Danish). Available at https://www.ft.dk/samling/20171/almdel/mof/spm/775/svar/1510482/1938569.pdf (accessed November 12, 2019).

Ministry of Environment and Food of Denmark, 2018b. Bekendtgørelse om vandkvalitet og tilsyn med vandforsyningsanlæg (statutory order no. 1068, in Danish). Available at https://www.retsinformation.dk/Forms/R0710.aspx?id=202768 (accessed November 12, 2019).

Monks, Stewart, Sahin, Keller, 2019. Revealing Unreported Benefits of Digital Water Metering: Literature Review and Expert Opinions. Water 11, 838. doi:10.3390/w11040838

Mounce, S.R., Boxall, J.B., 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, 309–318. doi:10.1061/(ASCE)WR.1943-5452.0000030

Mourad, M., Bertrand-Krajewski, J.-L., 2002. A method for automatic validation of long time series of data in urban hydrology. Water Sci. Technol. 45, 263–270. doi:10.2166/wst.2002.0601

Moy de Vitry, M., Schneider, M.Y., Wani, O., Manny, L., Leitão, J.P., Eggimann, S., 2019. Smart urban water systems: what could possibly go wrong? Environ. Res. Lett. 14, 081001. doi:10.1088/1748-9326/ab3761

Nguyen, K.A., Stewart, R.A., Zhang, H., Sahin, O., Siriwardene, N., 2018. Re-engineering traditional urban water management practices with smart metering and informatics. Environ. Model. Softw. 101, 256–267. doi:10.1016/j.envsoft.2017.12.015

Ommen, T., Markussen, W.B., Elmegaard, B., 2014. Heat pumps in combined heat and power systems. Energy 76, 989–1000. doi:10.1016/j.energy.2014.09.016

Puust, R., Kapelan, Z., Savic, D.A., Koppel, T., 2010. A review of methods for leakage management in pipe networks. Urban Water J. 7, 25–45. doi:10.1080/15730621003610878

Quevedo, J., Garcia, D., Puig, V., Saludes, J., Cugueró, M.A., Espin, S., Roquet, J., Valero,

F., 2017. Real-time Monitoring and Operational Control of Drinking-Water Systems, Real-Time Monitoring and Operational Control of Drinking- Water Systems, Advances in Industrial Control. Springer International Publishing, Cham. doi:10.1007/978-3-319-50751-4

Quevedo, J., Puig, V., Cembrano, G., Blanch, J., Aguilar, J., Saporta, D., Benito, G., Hedo, M., Molina, A., 2010. Validation and reconstruction of flow meter data in the Barcelona water distribution network. Control Eng. Pract. 18, 640–651. doi:10.1016/j.conengprac.2010.03.003

Rygaard, M., Alsbjørn, L., Ejsing, M., Godskesen, B., Hansen, R., Hoffmann, B., Jørgensen, C., Ledgaard, K., Friis Møller, H.-M., Poulsen, M.-B.B., Schmidt, M., Tarp-Johansen, L., Zambrano, S.V.K., 2013. Sekundavand i Nordhavn: En forundersøgelse til strategi for alternativ vandleverance. Technical University of Denmark, DTU Environment, Kgs. Lyngby.

Rygaard, M., Binning, P.J., Albrechtsen, H.-J., 2011. Increasing urban water self-sufficiency: New era, new challenges. J. Environ. Manage. 92, 185–194. doi:10.1016/j.jenvman.2010.09.009

Sarni, W., White, C., Webb, R., Cross, K., Glotzbach, R., 2019. Digital Water - Industry leaders chart the transformation journey. International Water Association. London. Available at https://iwa-network.org/wp-content/uploads/2019/06/IWA_2019_Digital_Water_Report.pdf (accessed November 12, 2019).

Savic, D.A., Kapelan, Z.S., Jonkergouw, P.M.R., 2009. Quo vadis water distribution model calibration? Urban Water J. 6, 3–22. doi:10.1080/15730620802613380

Sophocleous, S., Savić, D.A., Kapelan, Z., Giustolisi, O., 2017. A Two-stage Calibration for Detection of Leakage Hotspots in a Real Water Distribution Network. Procedia Eng. 186, 168–176. doi:10.1016/j.proeng.2017.03.223

Stewart, R.A., Nguyen, K., Beal, C., Zhang, H., Sahin, O., Bertone, E., Vieira, A.S., Castelletti, A., Cominola, A., Giuliani, M., Giurco, D., Blumenstein, M., Turner, A., Liu, A., Kenway, S., Savić, D.A., Makropoulos, C., Kossieris, P., 2018. Integrated intelligent water-energy metering systems and informatics: Visioning a digital multi-utility service provider. Environ. Model. Softw. 105, 94–117. doi:10.1016/j.envsoft.2018.03.006

Sun, A.Y., Scanlon, B.R., 2019. How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. Environ. Res. Lett. 14, 073001. doi:10.1088/1748-9326/ab1b7d

Tan, P.N., Steinbach, M., Kumar, V., 2006. Introduction to Data Mining, Pearson International Edition. Pearson Addison Wesley, Boston.

Tang, C., Albrechtsen, H.-J., 2019. Water quality monitoring in the food and beverage industry: Sensors for monitoring physical, chemical and microbial water quality. Technical University of Denmark, DTU Environment, Kgs. Lyngby. [in press].

Tatari, K., Corfitzen, C.B., Albrechtsen, H.-J., Christensen, S.C.B., 2016. Sensors for microbial drinking water quality. Technical University of Denmark, DTU Environment, Kgs. Lyngby.

United Nations, 2019. The Sustainable Development Goals Report 2019. United Nations Publications, New York. Available at https://unstats.un.org/sdgs/report/2019/The-

Sustainable-Development-Goals-Report-2019.pdf (accessed November 12, 2019).

Walski, T., Sage, P., Wu, Z., 2014. What Does it Take to Make Automated Calibration Find Closed Valves and Leaks?, in: World Environmental and Water Resources Congress 2014. American Society of Civil Engineers, Reston, VA, pp. 555–565. doi:10.1061/9780784413548.059

Walski, T.M., Chase, D. V, Savic, D.A., Grayman, W., Beckwith, S., Koelle, E., 2003. Advanced water distribution modeling and management. Haestad press, Waterbury, CT.

Watertech A/S, 2005. Vandforbrug og forbrugsvariationer (in Danish). Miljøstyrelsen (Danish EPA). Available at https://www2.mst.dk/Udgiv/publikationer/2005/87-7614-592-1/html/default.htm (accessed November 12, 2019).

Wilson, C., 2011. Sustainable distribution system asset management: Risk mitigation through control point management. J. Am. Water Works Assoc. 103, 100–103. doi:10.1002/j.1551-8833.2011.tb11456.x

Wu, Y., Liu, S., 2017. A review of data-driven approaches for burst detection in water distribution systems. Urban Water J. 14, 972–983. doi:10.1080/1573062X.2017.1279191

Wu, Z.Y., Song, Y., Syed, J.L., 2012. Optimization Model for Identifying Unknown Valve Statuses and Settings, in: WDSA 2012: 14th Water Distribution Systems Analysis Conference, 24-27 September 2012 in Adelaide, South Australia. pp. 975–983.

WWAP, 2019. The United Nations World Water Development Report 2019: Leaving No One Behind. UNESCO World Water Assessment Programme, Paris. Available at https://unesdoc.unesco.org/ark:/48223/pf0000367306 (accessed November 12, 2019).

8 Appendix

A Scopus search query

For hits related to publications that cover digitalisation in the water sector, the following query was used on https://www.scopus.com:

TITLE-ABS-KEY(("digitalisation" OR "digitalization" OR "digital transformation" OR "data-driven" OR "digital water" OR "water 4.0" OR "internet of water" OR "smart water" OR "intelligent meter" OR "smart meter") AND "water") AND (EXCLUDE(SUBJAREA,"MEDI") OR EXCLUDE (SUBJAREA,"IMMU") OR EXCLUDE(SUBJAREA,"HEAL") OR EXCLUDE(SUBJAREA,"PSYC") OR EXCLUDE(SUBJAREA, "PHAR") OR EXCLUDE(SUBJAREA, "NURS") OR EXCLUDE(SUBJAREA, "VETE") OR EXCLUDE(SUBJAREA, "DENT") OR EXCLUDE(SUBJAREA, "ARTS") OR EXCLUDE(SUBJAREA, "NEUR"))

To account for general water-related publications, ("digitalisation" OR "digitalization" OR "digital transformation" OR "data-driven" OR "digital water" OR "water 4.0" OR "internet of water" OR "smart water" OR "intelligent meter" OR "smart meter") was removed from the upper query. The hit ratio was then computed between both outputs.

B Data reconstruction

During anomaly testing of the raw data in Paper II, missing data periods were identified and dubious observations flagged; however, most post-applications of the collected meter data will benefit from a continuous and flawless input data stream. Missing and flagged data points (as well as valid data points for visualization purposes) were reconstructed based on estimates from feed-forward artificial neural networks (ANNs) (Tan et al., 2006). The ANN models were trained based on the resilient back-propagation algorithm with weight backtracking (Riedmiller and Braun, 1994)[1]. To reconstruct meter observations, it was decided to construct simple ANNs that consisted of a maximum of 10 input neurons, 1 hidden layer including a maximum of 10 hidden neurons, and 1 output neuron. The hourly observations of a meter were predicted based on an ANN model that was trained on the hourly or specific lagged hourly

[1] Riedmiller, M., Braun, H., 1994. A direct adaptive method for faster backpropagation learning: the RPROP algorithm, in: IEEE International Conference on Neural Networks. IEEE, pp. 586–591. doi:10.1109/ICNN.1993.298623

values from other meters in a utilities network. The selection of other meters favoured highly correlated data sets. To ensure uniform time intervals, the data was aggregated to hourly data similar to test VII (see Paper II). Only pre-validated data of tests I–VII were used for creating the hourly averages. This improved the reliability of the model results, as neural networks are sensitive to the presence of noise in the training data (Tan et al., 2006).

Applied training of artificial neural networks

ANNs comprise three types of layer: one input, multiple hidden and one output layer. Neurons connect the different layers by specific weights and an activation function, which in our case is the hyperbolic tangent function (tanh):

$$\tanh(z) = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}}, \tag{1}$$

where z represents a linear combination of weighted inputs from neurons. The complete selection process of input and hidden neurons and training of ANN models for a meter data set is summarized in the following four steps (Figure B-1):

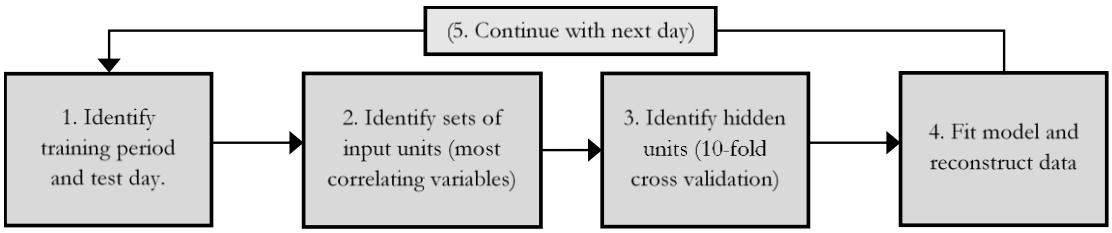


Figure B-1. Daily meter data reconstruction process based on artificial neural network models.

1 A meter data set was selected for reconstruction. As utilities may change their network set-up frequently, significant changes may be observed in the statistical properties of pressure and flow time series. Therefore, each daily model is only trained on the available data sets of the previous 60 days. Thus, a test day and the previous 60 days were used as test and training periods, respectively. If more than one third of the data was missing from the test meter in the training period, the test day was skipped for data reconstruction.

2 The correlation between the meter to be reconstructed and all available data sets was computed based on the training period. The five meter data sets with the highest absolute correlation were selected as ANN model input units. Also, the hourly lags 1, 2, 6, 12 and 24 of the most correlating meter

data set were included as input neurons. The response variable itself was not used as an input variable; this was done to overcome issues that may arise when only the analytical redundancy of the meter is used for model building. For example, time series models based on one meter might fail to predict measurements correctly that include sudden network operations (e.g. Quevedo et al., 2010).

3 For each set of input units, the number of neurons in the hidden layer was determined by 10-fold cross validation (CV) (Tan et al., 2006) applied on the training data set to avoid model overfitting. In short, the training set was divided into 10 partitions, where one partition was equal to the 'CV test set' in each fold. An ANN model was trained on the remaining nine parts and the sum of squared error between the model prediction and each 'CV test set' was computed. This procedure was repeated ten times. The number of neurons having the lowest mean squared sum of errors was selected for model training in the next step.

4 Having determined a final number of input units and number of neurons in the hidden layer, the model setup was trained on the entire training set available. The test period was then applied to the model(s) and the daily data was reconstructed.

Several conditions could occur that require the training of additional models or the discarding of model training and thus data reconstruction. Relating to step 2 of the ANN construction process illustrated in Figure B-1, the following conditions had to be met regarding the training and testing of data sets with missing data:

- If any input unit in combination with the reference meter had more than one third missing data in the training period, the input unit was excluded and another, based on its correlation with the reference meter, was selected.
- If the set of input units had more than one third missing timestamps in the training period, the input unit with the most missing timestamps was removed and a new input unit, based on its correlation with the reference meter, was selected.
- If an input unit had one or multiple missing timestamps in the training and/or test data set, additional input-unit sets and thus models covering these timestamps had to be constructed. The input units not covering the period were excluded from the new models.

- If no input-unit set could be determined that met all of the above-stated criteria for certain timestamps, the number of most correlating variables and hence input units was reduced by one. In this case, the entire procedure of step two was repeated until it is possible to reconstruct all timestamps. If it was not possible to find a set of input units, selected timestamps were skipped for reconstruction.

9 Papers

I **Kirstein, J.K.,** Høgh, K., Rygaard, M. & Borup, M. (2019). Effect of data sampling resolution of smart meter readings in water distribution network simulations. *Manuscript in preparation.*

II **Kirstein, J.K.,** Høgh, K., Rygaard, M. & Borup, M. (2019). A semi-automated approach to validation and error diagnostics of water network data. *Urban Water Journal*, **16**(1)**,** 1–10, doi:10.1080/1573062X.2019.1611884

III **Kirstein, J.K.,** Liu, S., Høgh, K., Borup, M. & Rygaard, M. (2019). Valve status identification by temperature modelling in water distribution networks. *Manuscript in preparation.*

IV **Kirstein, J.K.,** Høgh, K., Rygaard, M. & Borup, M. (2019). Using smart meter temperature and consumption data for water distribution system analysis. *Manuscript in preparation.*

V Hubeck-Graudal, H., **Kirstein, J.K.,** Ommen, T., Rygaard, M. & Elmegaard, B. (2019). Drinking water supply as low-temperature source in the district heating system: a case study for the city of Copenhagen. *Submitted*.

VI Lund, N.S.V., **Kirstein, J.K.,** Mikkelsen, P.S., Madsen, H., Mark, O. & Borup, M. (2019). Using smart meter water consumption data and in-sewer flow observations for model based sewer system analysis. *Submitted.*

In this online version of the thesis, Papers **I–IV** are not included but can be obtained from electronic article databases e.g. via www.orbit.dtu.dk or on request from:

DTU Environment
Technical University of Denmark
Miljoevej, Building 113
2800 Kgs. Lyngby
Denmark
info@env.dtu.dk

The Department of Environmental Engineering (DTU Environment) conducts science based engineering research within five sections: Air, Land & Water Resources, Urban Water Systems, Water Technology, Residual Resource Engineering, Environmental Fate and Effect of Chemicals. The department dates back to 1865, when Ludvig August Colding gave the first lecture on sanitary engineering.

Department of Environmental Engineering
Technical University of Denmark

DTU Environment
Bygningstorvet, Building 115
2800 Kgs. Lyngby
Tlf. +45 4525 1600
Fax +45 4593 2850

www.env.dtu.dk