

# **Optimal and Practical Real-Time Operation of Water Distribution Systems**

**Elad Salomons**

A THESIS SUBMITTED FOR THE DEGREE

"DOCTOR OF PHILOSOPHY"

Dissertation by Publications

University of Haifa

Faculty of Social Sciences

The Department of Natural Resources & Environmental Management

October 2021

# Optimal and Practical Real-Time Operation of Water Distribution Systems

**By: Elad Salomons**

**Supervised by: Dr. Mashor Housh**

A THESIS SUBMITTED FOR THE DEGREE

"DOCTOR OF PHILOSOPHY"

Dissertation by Publications

University of Haifa

Faculty of Social Sciences

The Department of Natural Resources & Environmental Management

October 2021

Recommended by:

*Housh M.*

Date: 01/10/2021

---

(Supervisor)

Approved by:

*Boris S. Asjw*

Date: 31/01/2022

---

(Chairperson of PhD Committee)

## **Acknowledgment**

This work could not have been done without the guidance of my supervisor, Dr. Mashor Housh. With his great teaching skills and patience, Dr. Housh has guided me throughout the past three years and enriched my knowledge in the theoretical and the practical aspects of my work. Only a handful of people have influenced the course of my professional career, and Dr. Housh is one of them.

Our discussions went way beyond the topic of my thesis. We spent endless hours talking about science, economics, life, children, education, and much more. I can only hope Dr. Housh has gained a fraction of what I did from our talks.

Mashor, thank you!

I also want to thank the Israeli Water Authority, the Bazan Group, and the Center for Cyber Law & Policy at the University of Haifa in conjunction with the Israel National Cyber Directorate in the Prime Minister's Office for their support of my work.

Haifa, January 2022.

## Table of Content

ABSTRACT .....	IV
Introduction .....	1
Paper I.....	9
Salomons, E.; Housh, M. "Practical Real-Time Optimization for Energy Efficient Water Distribution Systems Operation". Cleaner Production 2020, 275, 124148. <a href="https://doi.org/10.1016/j.jclepro.2020.124148">https://doi.org/10.1016/j.jclepro.2020.124148</a> .....	9
Statement of authorship - Paper I.....	10
Paper II .....	25
Salomons, E.; Housh, M. "A Practical Optimization Scheme for Real-Time Operation of Water Distribution Systems". Water Resources Planning and Management 2020, 146 (4), 04020016. <a href="https://doi.org/10.1061/(ASCE)WR.1943-5452.0001188">https://doi.org/10.1061/(ASCE)WR.1943-5452.0001188</a> .....	25
Statement of authorship - Paper II .....	26
Paper III .....	39
Salomons, E.; Shamir, U.; Housh, M. "Optimization methodology for estimating pump curves using SCADA data". Water 2021, 56 (9), e2020WR027917. <a href="https://doi.org/10.1029/2020WR027917">https://doi.org/10.1029/2020WR027917</a> .....	39
Statement of authorship - Paper III .....	40
Paper IV .....	55
Salomons, E.; Sela, L.; Housh, M. "Hedging for Privacy in Smart Water Meters". Water Resources Research 2020, 56 (9), e2020WR027917. <a href="https://doi.org/10.1029/2020WR027917">https://doi.org/10.1029/2020WR027917</a> .....	55
Statement of authorship - Paper IV .....	56
General Discussion .....	73
Appendix I.....	76
Sela, L., Salomons, E., and Housh, M. (2019). "Plugin prototyping for the EPANET software." Environmental Modelling & Software, Elsevier Ltd. <a href="https://doi.org/10.1016/j.envsoft.2019.05.010">https://doi.org/10.1016/j.envsoft.2019.05.010</a> .....	76
References .....	85

## List of Figures

Figure 1: Real-time control loop of WDS operation.....	2
--	---

# **Optimal and Practical Real-Time Operation of Water Distribution Systems**

**Elad Salomons**

## **ABSTRACT**

### **Background**

Despite the importance of smart Water Distribution Systems (WDSs) operation in obtaining energy-saving and environmentally friendly strategies, in many cases, the operation is still done according to expert opinion and rules of thumb which use local control schemes or some ad-hoc control rules. For example, in Israel, there is not even one water utility that has an operational real-time pump scheduling system that considers the changing demands, varying pumping electricity tariffs, and operational constraints. In general, the overall control loop of the real-time operations of WDS consists of reading the current state of the system, performing a demand forecast, optimizing the operations for a limited timeframe (e.g., 24 hours), and implementing the first time step of the optimized plan. Then this cycle is repeated for the next time step. During the past decades, many academic studies were conducted, and new methods were developed for the optimal control of WDSs. Nonetheless, most of the suggested control schemes are complex and impractical for real-world applications due to data availability and requirements, optimization difficulty, computational efficiency, and lack of central control system and technical staff.

### **The goal of the study**

This research aimed to explore and create new methodologies for optimal and practical real-time operation of WDSs while putting special emphasis on the practicality of the developed methods. Two control schemes are considered: a local scheme that is applied to a small operating zone (e.g., pump and tank) and a central scheme that is aimed to run at a control center and optimally operate a larger WDS.

## **Methods**

Two main strategies are adapted to seek practical tools: 1) Showing that the explicit hydraulic behavior may be excluded from the optimization formulation thus eliminating the non-linearity of the problem and, 2) Using wise binary coding of the discrete decision variables to obtain reduced size optimization problem. In addition, a demand forecasting algorithm is developed which requires a limited amount of historical demand data, thus making it practical for real-time applications.

## **Contribution**

To date, only a limited number of water utilities use a closed-loop optimal pump operations control scheme. This research contributes to the water sector by developing practical methods for the optimal operation of WDSs. From an academic perspective, previous work focused mainly on the open-loop operation while this research concentrates on a closed-loop control scheme with online feedback from the system in a rolling horizon mode. The tradeoff between the operation practicality and optimality has not been explored before in the literature since most of the previous work concentrated on the optimal solution for a given operation horizon without simulating the real-time behavior of the system in a close control loop.

## Introduction

In 2010, the United States' water-related energy use was 12.6% of the total energy consumption (Sanders and Webber 2012) of which one third (4% of the total energy consumption) is estimated to be consumed by pumping and treating water and wastewater (Goldstein and Smith 2002). There are about 52,000 community water systems in the US (Copeland and Carter 2017). Nearly 85% of the US population is supplied by about 5% of these systems, while the remaining 95% include many small systems serving 3,300 persons or fewer (Copeland and Carter 2017). About 80% of the energy consumed by these utilities is used by motors for pumping. Similar values are also reported in Israel with about 55 large water corporations and over 1000 small water suppliers.

Due to environmental regulations and increasing energy costs, energy conservation and efficiency are gaining importance in many water utilities. Different activities may help utilities with this goal, such as energy management, right sizing elements in the system, upgrading and replacing to more efficient equipment, self-generating energy, and optimizing operation. For the latter, Smart Water Distribution Systems (WDSs) can play a key role in achieving optimal operations which seek energy-saving and environmentally friendly strategies. Nevertheless, in many cases, the WDSs operation is still done according to expert opinion and rules of thumb which use local control schemes or some ad-hoc control rules. For example, in Israel, there are no local water utilities with an operational real-time pump scheduling system that considers the changing demands, varying electricity tariffs, and operational constraints.

During the past decades, many academic studies were conducted, and new methods were developed for optimal control of WDSs. In general, the overall control loop of the real-time operation of WDS is shown Figure 1. First, the current state of the system is read, usually, from a Supervisory Control and Data Acquisition (SCADA) system, a water demand forecast is performed for the next operation period, and the electricity tariffs are obtained. Then, an optimization problem is formulated and solved to obtain the operational settings (e.g., pumps and valve settings) which minimize the operational costs for the next operation period (e.g., 24 hours). These settings must fulfill the system's constraints, both physical constraints (e.g., power connection size) and operator's requirements (e.g., minimum tank levels for the reliability of the supply). Finally, the obtained operational settings are implemented for the current time and the process is repeated for the next time step. It is important to note that, at the beginning of each time step, the initial condition of the system is set (mainly the real tank

levels), thus "nullifying" any potential discrepancy between the design results and the real situation of the system before solving the next time step problem (Rao and Salomons 2007).

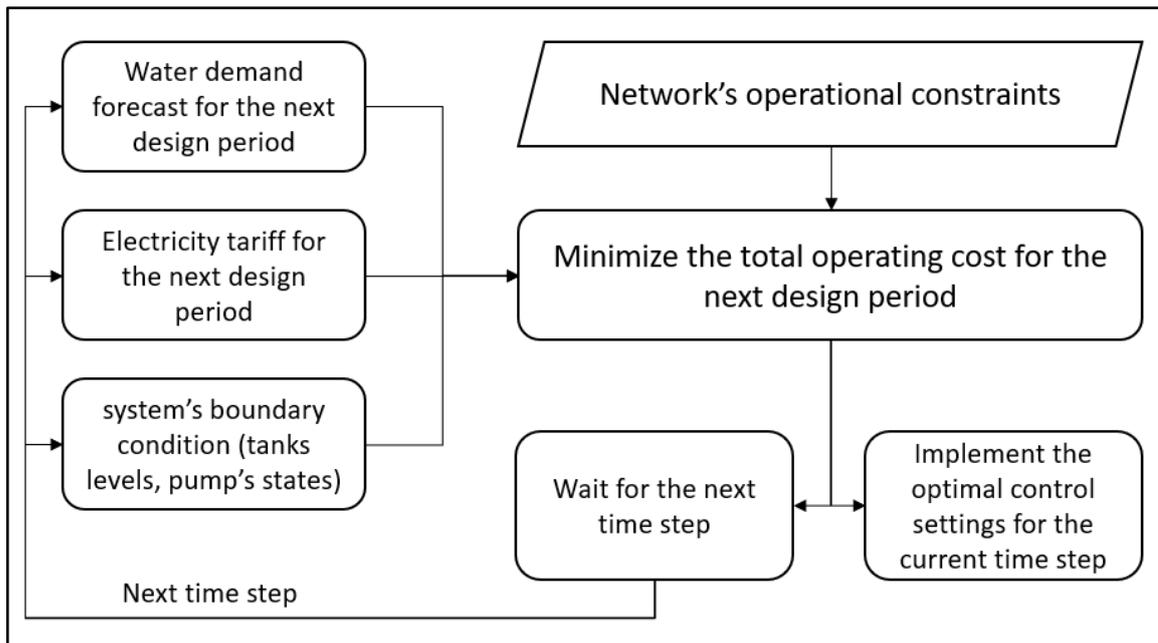


Figure 1: Real-time control loop of WDS operation

Several aspects of the abovementioned real-time control scheme make the task challenging and, in many cases, impractical for real-world applications. The main challenges include:

- Data availability and requirements.
- Optimization difficulty.
- Computational efficiency.
- Lack of central control system and technical staff.

Considering the above challenges, this research aimed to explore and create new methodologies for the optimal and practical real-time operation of WDS. These tools can help water utilities in utilizing advanced optimization methods to gain energy savings and more environmentally friendly operation strategies. Starting with a simple network (e.g., pumping station and a service tank) the aim was to develop a local control scheme that will be simple to implement at the local pumping station's Programmable Logic Controllers (PLC) and is robust enough for unknown demand scenarios. At a second stage, a larger and more complex network was used to develop a practical and central control scheme. From a practical point of view, considering real-world WDS, it can be shown that in many cases the explicit hydraulic behavior may be excluded from the optimization problem thus eliminating the non-linearity of the problem. Additionally, the size of the optimization problem may be reduced by using wise binary coding of the discrete decision variables in the optimization problem. This approach results in a Mixed

Integer Linear Program (MILP) with a relatively small number of integer variables which can be solved in reasonable time with state-of-the-art commercial solvers such as CPLEX (IBM Corp. 2009) or free and open-source solvers such as CBC (Lougee-Heimer 2003).

The real-time control loop of WDS operation (Figure 1) results in two major challenges: demand forecasting and energy cost minimization optimization. Although the real-time WDS operation problem is a holistic control loop, most of the research is concentrated on parts of the problem and less on the problem as a whole. An exception is the work of Coulbeck and Orr (1989) which considered the overall aspects of the control problem consisting of a demand predictor, an optimized pump scheduler, and a simulator. The architecture and activities of the control system were presented, these include a control computer, SCADA system, activity scheduler, data manager, and a performance monitor. In the POWADIMA research project (Potable Water Distribution Management) Jamieson et al. (2007), Rao and Salomons (2007) and Shamir and Salomons (2008) presented a similar platform coupling a short-term demand forecasting module with Genetic Algorithm (GA) optimization and Artificial Neural Networks (ANN) for hydraulic simulation. As opposed to the above research, most of the published studies have concentrated on limited aspects of the control problem as described below.

The main objective of a WDS is to supply water to the customers. As such, before any operational planning could be made, an estimation of the future water demands must be taken. Demand forecasting is the basis of all WDS operation problems. Yet, the different parameters of the forecasting algorithms differ according to the problem at hand (Donkor et al. 2014). For strategic decision making, such as system capacity expansion, a long-term forecasting horizon would be used (e.g., over 10 years with annual periodicity). For tactical planning, a shorter forecasting horizon would be used for revenue forecast or staging system improvement (e.g., 1-10 years with monthly periodicity). However, for operations purposes, a shorter forecasting horizon would be used which ranges from 24-168 hours up to one year with hourly, daily, or weekly periodicity. In the operation problem, considering the real-time nature of the problem, usually, the low end of the demand forecasting horizon is considered. That is a horizon of up to 48 hours with hourly periodicity. Zhou et al. (2002) developed a daily and 24 hours ahead demand forecasting module for the city of Melbourne (Australia) by dividing the demand into a base and seasonal consumption, thus characterizing it on a daily and monthly basis. Taking into account seasonal demand variations requires long history datasets, as such Zhou et al. (2002) used a dataset of six years. Similarly, Alvisi et al. (2007) constructed a daily and hourly water demand forecast for the real-time near-optimal operation of a WDS. The daily demand is modeled by a Fourier series which accounts for seasonal cycles. The hourly demand model is

fed from the daily model and composed of periodic and persistence components. Herrera et al. (2010) compared several hourly demand prediction models for a city in south-eastern Spain. These models used historic hourly and daily demands and explanatory climate variables (e.g., temperature, wind speed, and rain). The compared models include ANNs, projection pursuit regression (PPR), multivariate adaptive regression splines (MARS), support vector regression (SVR), and random forests. The common requirement for these models is an offline training stage using long history datasets. In a more recent study, Pacchin et al. (2017) suggested a simpler hourly demand forecast model which consists of two steps. First, the total demand for forecasting horizon is estimated and then the hourly pattern over this time window is predicted. The total demand estimation is based on the previous day with an adjustment coefficient, while the hourly pattern is based on the weighted patterns of the same type of days in the previous weeks. This kind of models is appealing for a real-time WDS operation problem since it requires only limited amount of historic data, and it is easy and fast to implement.

WDSs operations are a classical example of a cost versus reliability tradeoff. On the one hand, and with reliability in mind, the operators prefer to keep high water levels in the tanks to cope with possible emergencies, such as fire, pipe bursts, energy shortages, etc. On the other hand, and with efficiency in mind, water utilities seek to minimize operation costs and thus seek tanks' water level trajectories in accordance with the energy tariff during the day. In an attempt to balance these two conflicting objectives, minimum and maximum water levels are usually defined to guarantee the required reliability while allowing the operator to vary the tank water levels between this defined minimum and the maximum water levels to minimize operating costs. Operating inside these predefined bounds can utilize different control schemes. One of the properties of these control schemes is the controller's physical location. In this respect, it is possible to consider the following two extreme configurations: 1) A Local configuration in which the controller is installed locally (at the pumping station) or 2) A Central configuration in which the controller is installed remotely (at the main control center).

For a pumping station and a tank configuration, the simplest automatic local controller uses time set points to initiate the pump and pressure setpoint to stop it (Sanks and Tchobanoglous 1998). In this rarely used practice, a predefined time setting is used for starting the pump (which could be set to the start of the low electricity tariff period). This causes the water level in the tank to rise to the point where a water level valve is reached, and the closure of the tank inlet is invoked. The closure of the tank inlet causes the pump's suction pressure to increase up to a limit which invokes the pump's shutdown. The main benefit of this control scheme is that a telemetry system is not needed between the pumping station and the tank, which keeps the

control loop simple. On the other hand, time-based controls do not allow for any reaction when there are changes in the demand. The implementation of such a scheme may thus require frequent manual intervention (Sanks and Tchobanoglous 1998). Nowadays, telemetry and SCADA systems have become simple to configure and have relatively low installation and maintenance costs. This has decreased the popularity of earlier control setups and allows for new water level controls schemes which utilize telemetry. In these schemes, the tank's water level is constantly transmitted to a controller located in the pumping station which compiles these readings into pump ON/OFF decisions following a predefined control logic. The simplest level-based control uses a fixed ON level and a fixed OFF level for the pump (Paschke et al. 2001). However, these fixed levels do not allow for possible energy cost savings when variable electricity tariffs are implemented. To cope with the inefficiency of fixed level triggers, Blinco et al. (2016), Creaco et al. (2016), and Marchi et al. (2016) have suggested fixed ON-OFF triggers which vary according to the different tariff periods. This scheme would thus maintain low water levels during the peak period and high water levels during the off-peak period. This improvement of energy cost efficiency may come at the expense of adjacent pump switches, which is an undesired operation property. Unlike Blinco et al. (2016), Alvisi and Franchini (2017) have suggested variable level triggers which are considered as a linear or nonlinear function during each tariff period.

The above-mentioned controller configuration is relevant when the controller is installed locally (at the pumping station). Although such a control scheme has advantages, its main drawback is its lack of "system view" which could be achieved by the central control configuration. When the controller is installed at the main control center, it can utilize more sophisticated algorithms and computational resources to control large networks. An extensive review of pump scheduling algorithms that are suitable for central control schemes is available in Ormsbee and Lansley (1994) and Mala-Jetmarova et al. (2017).

## **Research problem**

Several aspects of the real-time control scheme make the task challenging and, in many cases, impractical for real-world applications:

- Data availability and requirements – many demand forecasting algorithms require long history datasets to obtain reliable predictions (Alvisi et al. 2007). Even if the data is available, the use of long-term historical data series is a limitation when there are changes in the distribution system and demands baselines over time.

- Optimization difficulty – due to the non-linearity, non-convexity, and the on/off operational states of pumps, the pump's scheduling problem is, in some cases, formulated as a mix-integer non-linear program (MINLP). That is an NP-hard problem that cannot be solved for global optimality, especially for large networks over a long-time horizon.
- Computational efficiency – by nature, real-time applications need to run fast to react to rapidly changing conditions which are expected in WDSs.

This research addresses the above three challenges by a range of methods that are specially designed with practicality and usability in mind.

## **Methods**

In this research, the emphasis is on the overall performance of the control process and rather than on each of its components as usually done in the literature. That is, simpler and practical methodologies that together yield good ("near-optimal") results are preferred over complicated components that might add small benefits for a large "price". The above challenges will be addressed with several main pillars:

1. Adoption of simple demand forecasting algorithms – an emphasis is made on-demand forecasting algorithms that do not involve long history datasets, are simple to calculate, and may adjust to near past demand changes.
2. Development of local and central control schemes - compared to the local control scheme which has a limited amount of information about the system as well as a limited computational capacity, the central control scheme may have the advantage of “looking” at the system as a whole and accounting for information from different WDS components. Nevertheless, the central control scheme will usually rely on controllers which use optimization methods to solve the pump scheduling problem, which requires intensive computational efforts to run the optimization algorithms in real-time. Although the energy costs could be minimized by these sophisticated pump scheduling algorithms, the complexity of these control schemes compared to a local control scheme makes them less popular in practice (Mala-Jetmarova et al. 2017). The local control scheme is popularly owed to its simplicity and its robustness because it does not rely on communication protocols and because most of the needed information to perform the operation task is available in situ. Moreover, the local scheme’s control logic is developed offline and the results are sent to the local PLC for implementation at the pumping station (Abdelmeguid and Ulanicki 2012). The control settings may also be used for a long period, or at least until

some major change in the system takes place (pump changes, major demand changes, etc.).

To that end, an optimal control process for the two control schemes will be developed:

- 2.1. Local control – first a simple local optimal control process is developed which can be utilized on a remote PLC with no real-time control from a central location. This is a control process for a small hydraulic zone (e.g., one or two pumping stations and a tank) which is very representative for many small water utilities around the world and in Israel. The main characteristics of such a control scheme are its limited data requirements and its mathematical simplicity which does not necessitate a dedicated optimization software.
- 2.2. Central control – at a central location (e.g., a control room) more powerful computational resources and more information from the WDS are available. This enables more sophisticated algorithms to run including dedicated optimization software, commercial or open-source, on a larger network. However, the NLP, MILP, and LP control schemes formulations described above all have drawbacks. Here, different MILP approximations (with varying accuracy) are explored and present that when tested in a real-time optimization framework the balance between approximation accuracy and solution efficiency is biased. That is, a simple low-accuracy approximation may yield to efficient and practical solution algorithm which results in a near-optimal solution when tested in real-time operating conditions.
3. Most of the research on WDS operations concentrated on the optimality problem for a specific time frame (e.g., 24 hours) and not on the closed control loop with the feedback from the system in a moving time window. As described earlier, only the first time step (or steps) of the operation plan are implemented and the optimization procedure is repeated thus the "investment" and the "efforts" put in the optimality search are not fully utilized. To that end, as opposed to many former studies, the focus in the new developments is on the overall process optimality and practicality.

## Structure

The above-described challenges are addressed in three published journal papers:

1. "***Practical Real-Time Optimization for Energy Efficient Water Distribution Systems Operation***": this paper address the local control problem by introducing an innovative flow allocation algorithm to control pumping stations in an operational zone. The

proposed algorithm can be implemented on a standard PLC and requires limited input data and computation power.

2. "***A Practical Optimization Scheme for Real-Time Operation of Water Distribution Systems***": in this paper, a practical MILP formulation is proposed to solve the central control problem. First, a new demand forecasting algorithm is presented which only requires a short history of demand data. Then, the traditional MILP formulation is reduced to include only a limited number of Integer decision variables, thus making the optimization problem solvable in a short computational time, making it practical for real-time applications.
3. "***Optimization methodology for estimating pump curves using SCADA data***": both the local and central control algorithms presented in the first two papers require operational parameters for the system's pumps. In the third paper, a new and practical procedure is presented to estimate pump curves from SCADA data even in cases where limited measurements are available.

While working with the SCADA data and considering automated meter reading (AMR) information, the question of end-user privacy arose. As smart water meters gaining popularity, the fine-grained information collected by smart meters raises growing concerns of privacy invasion due to personal behavior exposure (private activity, daily routine, etc.). In a fourth paper presented herein, "***Hedging for Privacy in Smart Water Meters***", the privacy concerns, related to smart meters are presented, and a hardware apparatus coupled with a software solution is suggested to hedge against the privacy risks. Noteworthy that this paper was selected as Editor's Highlight in the journal.

Additionally, during my Ph.D. study, I have co-authored a paper that focuses on software development for increasing the usability of the tools developed in the WDSs research community. In the paper presented in Appendix I, we developed a plug-in architecture for the popular EPANET hydraulic simulator and demonstrated the use of such architecture on different use-cases. I am a major contributor to this work, starting from conceptualizing the idea and solely developing the necessary code for the experiment.

## **Paper I**

*Salomons, E.; Housh, M. "Practical Real-Time Optimization for Energy Efficient Water Distribution Systems Operation". Cleaner Production 2020, 275, 124148.*

<https://doi.org/10.1016/j.jclepro.2020.124148>.

Journal Impact Factor: 7.246

5-Year Impact Factor: 7.491

Publication Status: Published (Accepted 8 September 2020)

## Statement of authorship - Paper I

### Principal Author

Title of Paper	Practical Real-Time Optimization for Energy Efficient Water Distribution Systems Operation		
Publication Status	<input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication		
Publication Details	Salomons, E.; Housh, M. "Practical Real-Time Optimization for Energy Efficient Water Distribution Systems Operation". Cleaner Production 2020, 275, 124148. <a href="https://doi.org/10.1016/j.jclepro.2020.124148">https://doi.org/10.1016/j.jclepro.2020.124148</a>		
Name of Principal Author (Candidate)	Elad Salomons		
Contribution to the Paper	Conceived the presented idea, developed the theory and performed the computations. Wrote the manuscript with support from the co-author.		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Name and Signature	Elad Salomons 	Date	September 8 <sup>th</sup> , 2020

### Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

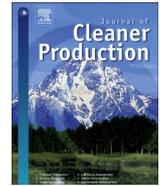
- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the dissertation

Name of Co-Author 1	Mashor Housh		
Contribution to the Paper	Verified the analytical methods, encouraged Elad Salomons to investigate the subject, and supervised the findings of this work.		
Name and Signature	Mashor Housh 	Date	September 8th, 2020



Contents lists available at ScienceDirect

## Journal of Cleaner Production

journal homepage: [www.elsevier.com/locate/jclepro](http://www.elsevier.com/locate/jclepro)

# Practical real-time optimization for energy efficient water distribution systems operation



Elad Salomons\*, Mashor Housh

Department of Natural Resources and Environmental Management, University of Haifa, Israel

## ARTICLE INFO

## Article history:

Received 11 June 2020  
 Received in revised form  
 27 August 2020  
 Accepted 8 September 2020  
 Available online 11 September 2020

Handling editor: Bin Chen

## Keywords:

Energy efficiency  
 Pump scheduling  
 Water distribution systems  
 Real-time operation

## ABSTRACT

The production, treatment and delivery systems of drinking water and wastewater is one of the largest energy consumers in the US with about 4% of the nation's power consumption. Roughly 80% of the water treatment and distribution costs are associated with electricity, mainly for pumping. Increasing the efficiency of drinking water pumping systems could benefit both the energy- and water-sectors. Despite the advancement of optimal pump scheduling technology, most water utilities are relatively small, and thus lack the funds, hardware and technical personal to support the use of sophisticated and computer intensive pump optimization programs. This study presents a simple and practical model predictive control methodology for real-time pump scheduling. This methodology can be deployed on a standard hardware (e.g., PLCs in pumping stations), which is currently in use by most water utilities. As such, it provides optimal pump scheduling benefits without necessitating large investment in new computational hardware (e.g., advanced controllers). The proposed methodology reduces both the energy consumption (by selecting the most efficient pumps' combinations) and the operation cost (by optimizing the pumps' operation according to electricity tariff periods). The results show that our practical methodology, which could be implemented in simple controllers, can provide near optimal decisions comparable with sophisticated optimization methods that require advanced hardware. To the best of our knowledge, there is no available methodology with such capabilities which is specifically designed for local control schemes. Thus, the novelty of this study is the utilization of this optimization methodology on a simple PLC hardware. Our results are of high importance for both academic and practical reasons, as it shows that the proposed methodology could be a kernel for a low-cost pumps' optimization technology.

© 2020 Elsevier Ltd. All rights reserved.

## 1. Introduction

The operation of Water Distribution Systems (WDSs) is energy intensive. In 2010, the energy consumed by treating and pumping of water and wastewater in the United States (US) is estimated as 4% of the total energy consumption (Copeland and Carter, 2017; Sanders and Webber, 2012). The US is not a special case, energy consumption of water systems is significant in other regions on the

globe. Lam et al. (2017) surveyed the energy use for water provision in 30 cities over 15 years. They report energy intensity, in most cities, of up to 1 kWh per m<sup>3</sup> of supplied water. This intensity is lumped to an annual energy use of 100 kWh per-capita. There are many factors that influence the energy use of water systems such as climate, topography, water use pattern, and operation efficiency (Lam et al., 2017). The latter is the focus of the current study. Optimal pumps' operation in WDSs has both economic and environmental benefits (Bunn and Reynolds, 2009). The economic benefits are achieved mainly by shifting the pumping times from periods of higher electricity cost to cheaper ones. Whilst the environmental benefits are achieved by choosing the most efficient pumps combination, which reduces the energy consumption. Thus, reducing the greenhouse gas (GHG) footprint of the water utility (Blinco et al., 2016; Torregrossa and Capitanescu, 2019). We note that shifting pumping times from periods of peak energy costs to non-peak hours require water storage facilities (water tanks),

*Abbreviations:* FAA Flow Allocation Algorithm, GHG Greenhouse Gas; HPZ Hydraulic Pressure Zone, LP Linear Programming; MAE Mean Absolute Error, MILP Mixed-Integer Linear Programming; MPC Model Predictive Control, NDF Naïve Demand Forecasting; PLC Programmable Logic Controller, SCADA Supervisory Control and Data Acquisition; SST Sorted States Table, TMC Theoretical Minimum Cost; US United States, WDS Water Distribution Systems.

\* Corresponding author.

E-mail address: [selad@optiwater.com](mailto:selad@optiwater.com) (E. Salomons).

<https://doi.org/10.1016/j.jclepro.2020.124148>

0959-6526/© 2020 Elsevier Ltd. All rights reserved.

which have capital cost and water quality implications. Several studies (Edwards and Maher, 2008; Farmani et al., 2006; Slavik et al., 2020) discuss the tradeoff between positive and negative implications of water storage in WDSs.

Optimal pumps' scheduling seeks the optimal pumps operation in space (which pumps combination?) and time (when to turn-On/Off the pumps?) that minimizes the energy cost subject to hydraulic and water supply reliability constraints. Because of hydraulic requirements, WDSs are usually comprised of one or more Hydraulic Pressure Zones (HPZs). An HPZ is a defined area of a WDS which receives water from a given hydraulic grade line. Thus, it is supplied by at least one pressure control device such as a water tank, pumping station or a pressure reducing/sustaining valve. The boundaries of an HPZ can include a closed pipe or valve (i.e., confined HPZ), single or multiple inlets and outlets. HPZs are widely used in practice due to their benefits in avoiding high pressure in the WDS, reducing leakage by pressure control, and the ability to isolate specific zones in the WDS during emergencies and more. Large networks may be divided to many HPZs, for example, the Barcelona (Spain) network is comprised of over 60 HPZs (Ocampo-Martinez et al., 2013) and the network of Haifa (Israel) have over 100 small HPZs due to its sloped topography. Nonetheless, many water utilities manage small networks. For example, 95% of the 52,000 community water systems in the US are small-scale systems serving 3300 persons or fewer (Copeland and Carter, 2017). For most of these small systems, it is impractical to be divided into many HPZs and thus usually consist of limited number of HPZs. Mei-Carmel (2020) states that the vast majority of WDSs in Israel consist of two or three HPZs. In terms of operation, the division of the network into zones helps to focus the control actions on limited number of devices and thus simplifying the optimal operation task. Generally, an HPZ can be operated without a storage tank by supplying water via a pressure reducing valve or a variable speed pumping station (Nowak et al., 2018). However, having a tank within the zone provides a more reliable water supply as well as the ability to reduce the energy cost of the pumped water by shifting pumping times between different electricity tariff periods. As such, including tanks within HPZs is a desirable property for better WDS operation. Herein, we consider HPZs that include storage tanks. Under this setting, there is a need for a control system to optimally operate the pumps and the trajectories in the tanks. For these purposes, Supervisory Control and Data Acquisition (SCADA) systems became popular in the past decades as their installation and maintenance costs decreased. Installing a SCADA system, allows for centralized control scheme that utilizes sophisticated and resource intensive operating methodologies (Predescu et al., 2020). These systems are often installed in a central location, such as a control room, that oversees the entire water network operation (Cembrano et al., 2000). A recent application of the centralized management system is the Digital Twin WDS presented by Conejos Fuertes et al. (2020). This Digital Twin resides in the control room and communicates with the SCADA system while aiming at providing a holistic overview of the system for improved operational decisions. Mala-Jetmarova et al. (2017) and previously Ormsbee and Lansey (1994), presented a detailed literature review of central operation control schemes. The centralized control schemes can be used to derive operational decisions that account for both water quantity and water quality aspects. For example, Abdallah and Kapelan (2019) suggested a pump scheduling method, based on an Evolutionary Algorithm, for optimum energy cost while accounting for water quality consideration in the WDS. Khatavkar and Mays (2018) used Genetic Algorithms for real-time control of WDSs while considering both water quality and limited electrical power availability. These centralized control schemes benefit from powerful computation resources which are installed in the control room, as

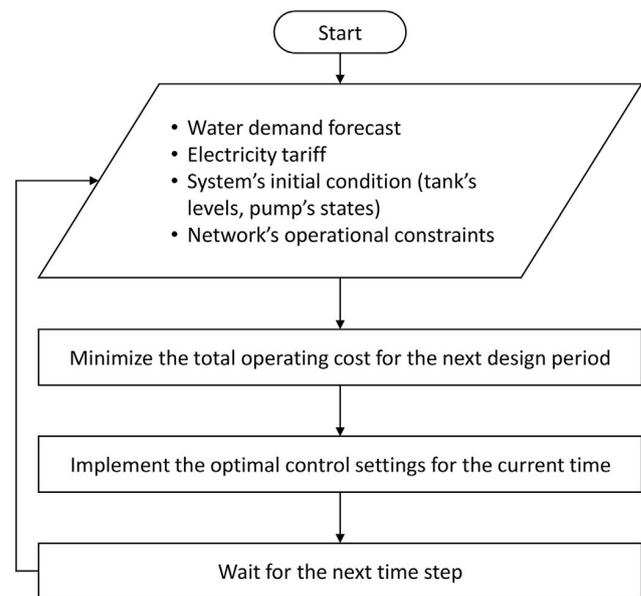


Fig. 1. Control loop for Water Distribution System (WDS) operation.

such this allows for using classical optimization methodologies such as Linear Programming (LP) (Jowitt and Germanopoulos, 1992), Non-Linear Programming (Candelieri et al., 2018; Yu et al., 1994), Mixed-Integer Linear Programming (MILP) (Salomons and Housh, 2020), Mixed-Integer Non-Linear Programming, Dynamic Programming (Carpentier and Cohen, 1993) and Evolutionary Algorithms (Cimorelli et al., 2020; Luna et al., 2019; Odan et al., 2015; Torregrossa and Capitanescu, 2019; Vieira et al., 2018). Typically, centralized control schemes that rely on classical optimization methods use dedicated software and hardware installed in the control room, thus requiring high level of technical personnel to operate and maintain them.

A simpler control, which is more widely used in practice, is the local control scheme in which the control logic is decentralized and embedded in local controllers installed in the pumping stations or water tanks. A comprehensive review of real-time local control schemes is presented in Creaco et al. (2019). Usually, local controllers are based on simple control rule. For example, in many practical applications, fixed speed pumping stations relay on tank trigger levels. These triggers invoke pump turn-on command when the tank level is low and pump turn-off command when the level is high. Optimized trigger levels can be computed off-line and then embedded in the local Programmable Logic Controller (PLC) for implementation (Alvisi and Franchini, 2017; Housh and Salomons, 2019; Paschke et al., 2001). The pumping station's PLC communicates with the PLC located at the water tank to receive its water level and determine the pump state based on the embedded logic.

PLCs are industrial rugged computers, usually with tailored operating systems, which have limitations on the programs and control logic they can run. Thus, in practice, water utilities tend to have simple control logic in PLCs. Although modern and advanced PLCs may run almost any program, many water utilities are not upgrading their control systems due to high costs and broader effects such changes have on the organization (Water-and-Sewer, 2014; Water-Technology, 2013). In many cases, local control by PLCs, which is currently in use in many water utilities, is not suitable to run real-time pump scheduling programs that rely on the aforementioned classical optimization methods. Thus, the potential

of achieving efficient energy and cost savings are not fully utilized when local control scheme is adapted.

In control theory, the real-time control of WDSs can be addressed using a Model Predictive Control (MPC) framework. MPC is a framework for controlling dynamic processes under a set of constraints that utilizes three modules: (a) a simulation model to simulate the control of the dynamics of the system over a finite future time horizon; (b) a prediction model for predicting the unknown future conditions; and (c) an optimization procedure, that decides on the optimal decisions to optimize a pre-defined objective using the predicted conditions and the simulated dynamics. In MPC, the process (i.e., using the three modules a-c) is repeated for each time-step, where the state of the system is updated with a receding horizon strategy. The MPC framework is widely used in many applications, including centralized control of WDSs (Ocampo-Martinez et al., 2012; Wang et al., 2017) as well as traffic control (Jamshidnejad et al., 2016), energy management (Wytock et al., 2017) and many more applications. Fig. 1 shows the real-time MPC loop of a WDS operation. When started, and at every time-step, a water demand forecast is made for the next operation horizon and the electricity tariff for the same period is obtained. The current system conditions are read from the control system and are used as initial conditions for the next time-step. Finally, user defined operational constraints such as time-based minimum and maximum tanks levels, minimum and maximum pressures at the demand nodes, physical constraints on power limitations and water quality considerations (Darweesh, 2020; Khatavkar and Mays, 2019) are formulated. Next, an optimization procedure is carried out to generate the system's operation decisions for the next operation horizon (e.g., the next 24, 48 h) that yield minimum cost subject to the formulated constraints. Once the future operation decisions are obtained, the first time-step (i.e., the next hour) decision is implemented. At this stage, the procedure waits for the next time-step to repeat the same control loop again with a receding horizon strategy.

Within this control loop, the most computationally intensive tasks are the demand forecasting and the minimization of the energy cost. Demand forecasting is a basic element in all WDS design and operation problems, where different forecasting horizons are used according to the problem at hand. Typically, demand forecasting for operational purposes (unlike strategic and tactical planning) uses a short-term forecast of a few days with hourly periodicity (Donkor et al., 2014). Many methods have been proposed for short-term demand forecasts. Some methods use long data series (years) for seasonal demand variations (Alvisi et al., 2007; Zhou et al., 2002). Other methods use explanatory climate variables such as temperature and rain (Herrera et al., 2010). Herrera et al. (2010) considered demand forecasting using data driven models such as Projection Pursuit Regression, Support Vector Regression and Artificial Neural Networks which require both long demand datasets and computationally intensive tuning and learning stage. Thus, the aforementioned forecasting methods require large datasets and computational resources. While these requirements may be suitable for centralized control scheme (where dedicated software and hardware are available), using such demand forecasting methods in local control PLC may be impractical or impossible. To cope with this, Pacchin et al. (2017) suggested a simpler short-term demand forecast method in which the total daily demand is first estimated based on the previous day, and then, the hourly pattern is derived from a weighted average hourly patterns in previous weeks. In a recent study, Salomons and Housh (2020), further simplified this forecasting method and introduced the Naïve Demand Forecasting (NDF) method, which uses the arithmetic average of hourly demands in previous weeks as a prediction for future hourly demands. The NDF method requires

only a few weeks of historic demand data and uses simple mathematical expressions (e.g., summation and divisions) for deriving the prediction. These characteristics make the NDF demand forecasting method practical for implementation on PLCs. Owing to its simplicity and suitability to local control schemes, this method is adopted in this paper and will be further explained in Section 2.1.

The second challenge of optimizing the energy cost is even more computationally demanding. As detailed above, the central control scheme uses computation intensive solvers for optimization. Among these tools one can find commercial solvers such as CPLEX (IBM Corp, 2009), open-source tools such as CBC (Forrest and Lougee-Heimer, 2005), and tailored simulation-optimization software which uses hydraulic solvers (e.g., EPANET (Rossman, 2000)). Without heavy modifications, none of these tools can run on a typical PLC and thus are unsuitable for local control scheme. While these tools may appear essential for handling the nonlinear nature (e.g., nonlinear hydraulics) of the problem, in many situation the nonlinearity could be relaxed. Jowitt and Germanopoulos (1992) suggested that the explicit hydraulics (which is the source of nonlinearity) of the system may be relaxed and thus the optimal pump scheduling problem could be formulated as an LP problem. This relaxation assumes that any pump scheduling plan which satisfies the minimum and maximum water level constraints at the tanks will also satisfies the required nodal pressure constraints in the network. This assumption is valid when a well-designed WDS is considered. That is, when the water demand can be delivered in an appropriate pressure from the tanks even when pumps are not operating (Ormsbee and Lansey, 1994). Moreover, this assumption implies that flow and power consumption of the pumps are relatively not affected by other elements in the network. For example, this is satisfied when the magnitude of the static head is large compared to the dynamic head (Housh and Salomons, 2019), i.e. when a pump operates against relatively constant head that dictates the pump's operation point on the characteristic curve. This is a typical situation in HPZs, in which the pump stations work against relatively constant pressure in the zone. Salomons and Housh (2020) utilized the relaxation above to solve the real-time centralized control of WDSs by formulating the optimization problem as MILP problem. Nonetheless, the suggested approach is designed for centralized control scheme, since it requires off-the-shelf MILP optimization packages, which, as discussed previously, are incompatible with local PLCs.

This study presents a simple and practical MPC methodology which is specifically designed for local control schemes. This method could be deployed on local PLCs, which are currently in use by most water utilities. As such, it provides optimal pumps scheduling benefits without necessitating large investment in new computational hardware (e.g., advanced controllers or centralized control scheme). The core of the proposed framework is an efficient optimization procedure, the Flow Allocation Algorithm (FAA), which could be easily implemented in local PLCs, since it only requires a few basic operators that are available in standard PLCs, such as loops and if conditions.

With simplicity and practicality in mind, we designed an efficient local control framework, which achieves near optimal decisions comparable with sophisticated optimization methods that are exclusive to centralized control schemes. The proposed methodology reduces both the energy consumption (and as a result the GHG footprint of the water utility) and the utility's operation cost. The former is achieved by selecting the most efficient pumps' combinations, while the latter is achieved by optimizing the pumps' operation in accordance with electricity tariff periods.

The remainder of this paper is organized as follows: Section 2 details the proposed methodology. In Section 3 we present the test case and in Section 4 the results of the test case are discussed.

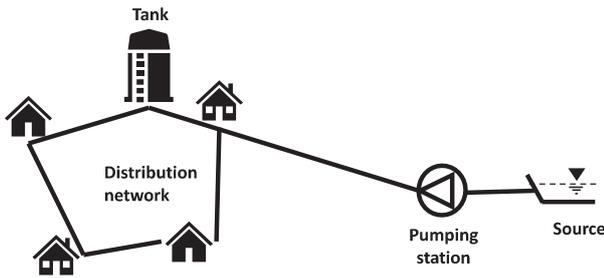


Fig. 2. Simple illustration of Hydraulic Pressure Zone (HPZ).

Table 1  
Sorted States Table (SST) of the pumping states in the illustrative network.

State	Q (m <sup>3</sup> /hr)	E (kWh)
1	40	4
2	75	11.25
3	100	22

Finally, in Section 5 we conclude the study results and propose future research directions.

## 2. Methodology

We consider an HPZ (Fig. 2) for which we need to develop an MPC framework, compatible with local control scheme deployment. The HPZ includes pump stations with multiple fixed speed

pumps and storage tanks. The task is to select the best combinations of the pumps and the best time-steps for operating them subject to water supply constraints and the system dynamics. For this MPC framework, Section 2.1 presents a practical demand forecasting methodology while Section 2.2 presents an efficient optimization procedure which is tailored for the system dynamics of HPZs in water distribution systems.

### 2.1. Demand forecast

Following the goal of utilizing simple and practical algorithms, we adopt the NDF method presented by Salomons and Housh (2020). For the completeness of this paper we briefly describe the NDF method. For the prediction of an hourly demand value, the NDF method averages the demand in the same hour in previous weeks. Denoting the current absolute time in hours passed from a predetermined time reference (e.g., the beginning of the year),  $h$ , the demand forecast for the next hour is given by Eq. (1).

$$\tilde{d}_h = \frac{1}{w} \sum_{i=1}^w d_{h-168 \cdot i} \quad (1)$$

where  $w$  is the number of previous weeks considered,  $d$  is the historic demand, 168 is the number of hours in a week and  $\tilde{d}$  is the forecasted demand. For implementing the MPC, there is a need to predict the demands for a future operation horizon  $T$  (e.g., optimization horizon of 48 h). That is, for each time instant  $t$  in the set  $\tau_h \equiv \left\{ h, h + \Delta t, \dots, h + \frac{T}{\Delta t} - 1 \right\}$  where  $\Delta t$  is the time-step (e.g., 1 h). Eq. (1) could be used for all elements in the set  $\tau_h$  to create an extended

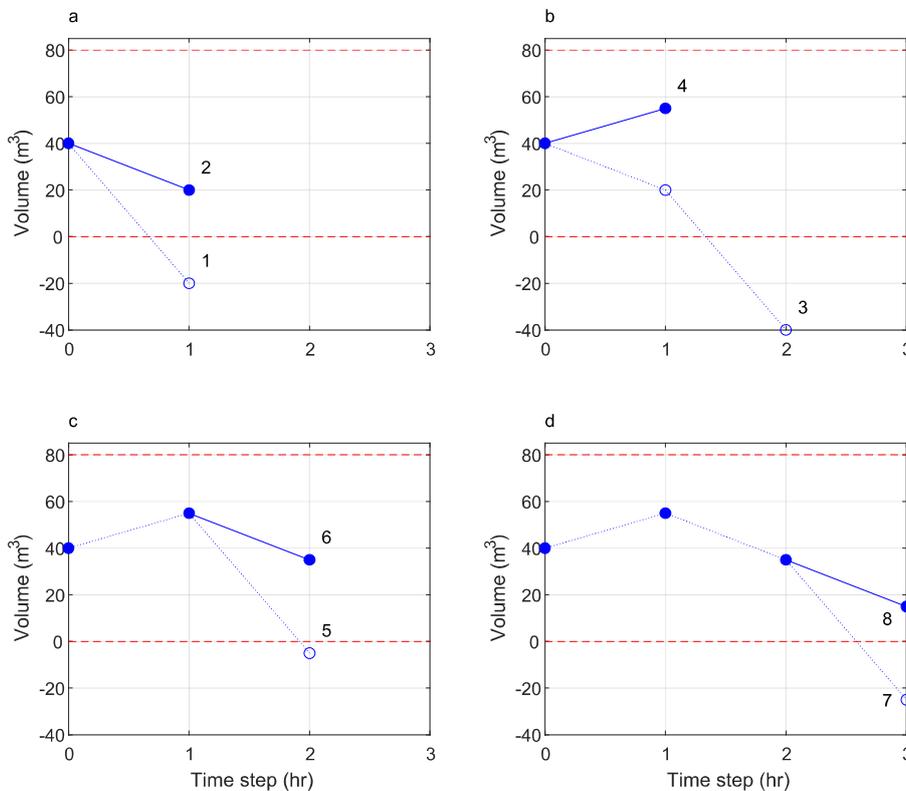


Fig. 3. Demonstration of the FAA steps, water trajectories: (a) first iteration, (b) second iteration, (c) third iteration and (d) fourth iteration. Dotted line: water trajectories before adding pumping in the current iteration. Solid line: water trajectories after adding pumping in the current iteration. Dashed lines: minimum and maximum tank's volume.

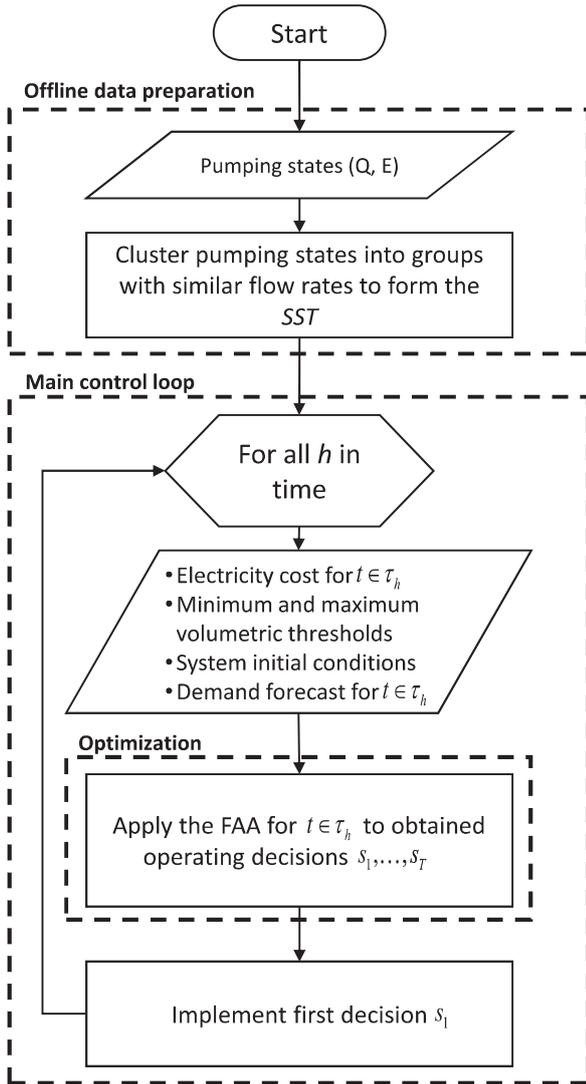


Fig. 4. Flowchart of the proposed methodology.

demand forecast as shown in Eq. (2).

$$\tilde{d}_t = \frac{1}{w} \sum_{i=1}^w d_{t-168 \cdot i} \quad \forall t \in \tau_h \quad (2)$$

## 2.2. Flow allocation algorithm

There are two main roles for any optimization procedure, optimizing the objective and satisfying the constraints. The FAA builds on a very basic property of the problem, which indicates that a decision of “do nothing” (i.e., no pumping at all) is the lowest possible cost, which obviously infeasible for the system dynamics. Then, the FAA iteratively looks for infeasibility in the system and tries to resolve it by incrementally allocating additional flow. Accounting for the system dynamics, the added flow must be allocated at time-steps before the infeasibility occurs in order to reduce it. While accounting for the minimum cost objective, indicates that we must add the flow with minimum additional cost.

To illustrate the FAA, we first consider a simple WDS in Fig. 2, which consists of one pumping station and one tank. There are three pumping states in the pumping station as detailed in Table 1. A pumping state represents a combination of pumps within the station. For example, the three pumping states in Table 1 represent three combinations of two pumps. The first pump working alone in the first state, the second pump working alone in the second state, and the two pumps working together in the third state. In addition to these three states, there is a fourth state representing the case when the two pumps are off. Each pumping state in the table is characterized by its flow ( $Q$ ) and its hourly energy consumption ( $E$ ). The states table is sorted according to flow and hence we denote it as Sorted States Table ( $SST$ ). For simplicity, we consider a tank with a minimum and maximum volumes of  $V_{MIN} = 0 \text{ m}^3$  and  $V_{MAX} = 100 \text{ m}^3$ , respectively and initial tank volume of  $V_0 = 40 \text{ m}^3$ . For demonstration purposes, we limit the operation horizon,  $T$ , to 3 h in which the electricity tariff,  $ET$ , is 1  $\$/kWh$  for the first and third hours and 2  $\$/kWh$  for the second hour. The water demand,  $d$ , is constant for the 3 h at the rate of  $60 \text{ m}^3/\text{hr}$ . The task is to select which pumping state,  $s$ , to operate at each time-step to minimize energy costs while maintaining the tank volume within the minimum and maximum constraints.

To solve this optimization problem using the FAA, we initialize the algorithm in the first iteration by assuming that the pumps are off during all time-steps, thus, starting with the initial volume of  $40 \text{ m}^3$  and with a demand of  $60 \text{ m}^3/\text{hr}$ , the tank volume is expected to reach a value of  $-20 \text{ m}^3$  at  $t = 1$  which is infeasible (point 1 in Fig. 3a). As such, in the next iteration some flow must be allocated in the first time-step (in which the infeasibility is encountered). Considering the available pumping states from Table 1, we allocate the first pumping state which is the most efficient one with the smallest energy consumption. With a flow rate of  $40 \text{ m}^3/\text{hr}$  for this pumping state, the tank volume is expected to reach a value of  $20 \text{ m}^3$  at  $t = 1$  (point 2 in Fig. 3a) which is within the feasible range of the tank's volume. However, the tank's volume at  $t = 2$  is expected to be  $-40 \text{ m}^3$  (point 3 in Fig. 3b) which is again not within the feasible range and thus additional flow must be allocated before  $t = 2$ . Now, we have two options: (1) increase the flow at  $t = 1$  to the second state, i.e. from  $40$  to  $75 \text{ m}^3/\text{hr}$  or (2) increase the flow at  $t = 2$  to the first state, i.e. from  $0$  to  $40 \text{ m}^3/\text{hr}$ . In the first option, the additional cost is  $(11.25kWh - 4kWh) \times 1\$/kWh = \$7.25$  (i.e., replacing the first state with the second) while the additional cost in the second option is  $4kWh/\text{m}^3 \times 2\$/kWh = \$8$ . Note that despite the lower energy consumption of the first state,  $4 \text{ kWh}$ , compared to the additional energy consumption of the second state  $11.25kWh - 4kWh = 7.25kWh$ , moving to the second state at  $t = 1$  is favorable due to the higher energy price at  $t = 2$  ( $2 \text{ \$/kWh}$  vs.  $1 \text{ \$/kWh}$ ). To this end, for the second iteration the algorithm allocates the second pumping state at  $t = 1$  with a flow rate of  $70 \text{ m}^3/\text{hr}$  bringing the tank volume at  $t = 1$  to  $55 \text{ m}^3$  (point 4 in Fig. 3b). The process is repeated for the third iteration, in which the tank's volume will be  $-5 \text{ m}^3$  at time  $t = 2$  (point 5 in Fig. 3c) which is outside the feasible range of the tank volume. As such, additional flow must be allocated in the first or second time-steps. Again, we have two options: (1) allocate the third pumping state in the first time-step or (2) allocate the first pumping state in the second time-step. Comparing the additional cost of the third pumping state in the first time-step,  $(22kWh - 11.25kWh) \times 1\$/kWh = \$10.75$ , to the first pumping state in the second time-step,  $4kWh/\text{m}^3 \times 2\$/kWh = \$8$ , reveals that the second option is preferable. Utilizing this selection, with a flow rate of  $40 \text{ m}^3/\text{hr}$ , will bring the tank volume at  $t = 2$  to  $35 \text{ m}^3$  which is within the feasible range (point 6 in Fig. 3c). In the fourth iteration, considering again the demand, the tank's volume in  $t = 3$  is expected to be  $-25 \text{ m}^3$  (point 7 in Fig. 3d) which is not feasible and requires additional flow to be allocated. In this

**Table 2**  
The flow allocation algorithm (FAA).

<b>Algorithm: Flow Allocation</b>	
1: <b>Input:</b>	$\tilde{d}, V_0, V_{MIN}, V_{MAX}, T, ET$ , Sorted States Table with columns $s_{ST} = \{S, G, E, Q\}$ and $n_{states}$ rows
2: <b>Output:</b>	Selected operation states for each time-step $s_t$
3: <b>Initialize:</b>	$s_t = 0 \forall t = 1 \dots T$
4:	Simulate $V_{i+1} = V_0 + \sum_{j=1}^i Q(s_j) - \sum_{j=1}^i d_j \quad \forall i = 1 \dots T$
5:	Check for infeasibility in $V_{MIN}$ , return the infeasibility time, $t_{viol}$ ; if none $t_{viol} = \infty$
6: <b>Execute:</b>	While $t_{viol} \leq T$ do:
7:	Set $\Delta C_{min} = \infty$
8:	For $t = t_{viol}$ to 1 do
9:	For $s_{imp} = s_t + 1$ to $n_{states} - 1$ do
10:	If $G(s_{imp}) > G(s_t)$ then
11:	Set $\Delta C = (E(s_{imp}) - E(s_t)) \cdot ET_t$
12:	If $\Delta C < \Delta C_{min}$ then
	Simulate
13:	$V_{i+1} = V_0 + \sum_{j=1}^i Q(s_j) + Q(s_{imp}) - \sum_{j=1}^i d_j \quad \forall i = 1 \dots T$
14:	If $V \leq V_{MAX}$ for all times, then
15:	Set $\Delta C_{min} = \Delta C$
16:	Set $t^* = t, s^* = s_{imp}$
17:	Break $s_{imp}$ loop
18:	End if
19:	End if
20:	End if
21:	End $s_{imp}$ loop
22:	End $t$ loop
23:	$s_t = s^*$
24:	Simulate $V_{i+1} = V_0 + \sum_{j=1}^i Q(s_j) - \sum_{j=1}^i d_j \quad \forall i = 1 \dots T$
25:	Check for infeasibility in $V_{MIN}$ , return the infeasibility time, $t_{viol}$ ; if none $t_{viol} = \infty$
26:	End while loop
27: <b>Return:</b>	$s_t \forall t = 1 \dots T$

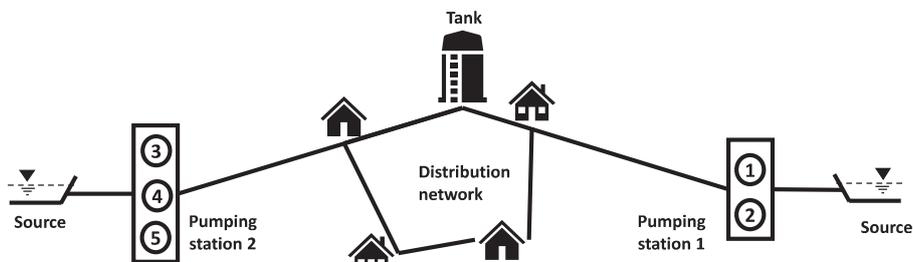


Fig. 5. Case study network layout.

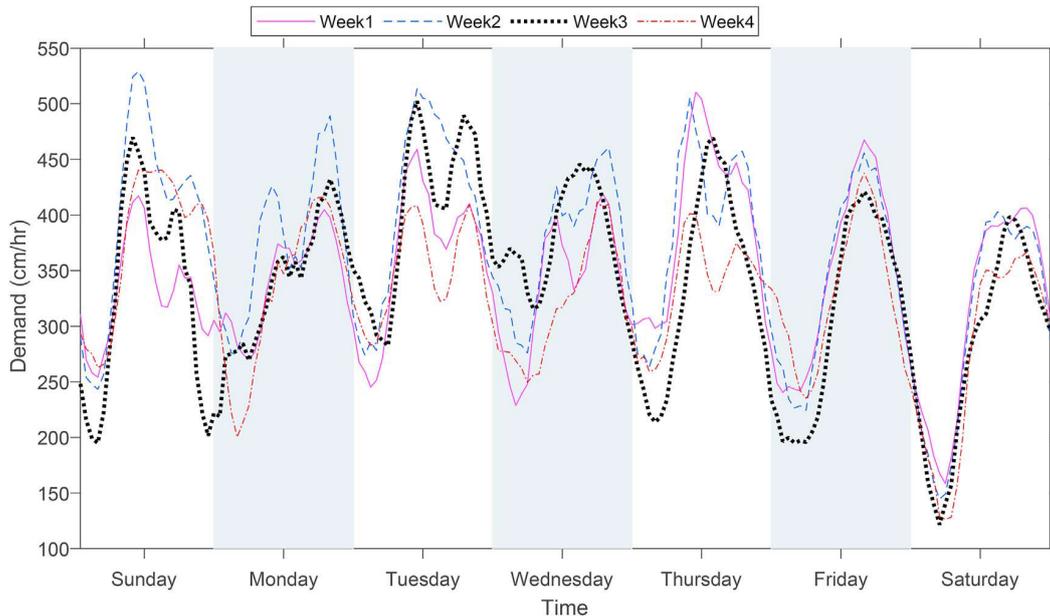


Fig. 6. Hourly demand profile during four weeks in the case study.

case we have three options: (1) the third pumping state in the first time-step; (2) the second pumping state in the second time-step or (3) the first pumping state in the third time-step. These options have additional costs of \$10.75, \$14.5 and \$4 respectively, making the third option the preferable one. Utilizing this option, with a flow rate of  $40\text{m}^3/\text{hr}$ , will bring the tank volume at  $t = 3$  to  $15\text{m}^3$  which is within the feasible range (point 8 in Fig. 3d). After the fourth iteration we end up with a feasible solution for all the time-steps and the optimization procedure is terminated.

Next, we present a detailed explanation of the methodology which consists of three main stages as outlined in Fig. 4. The first stage is an offline data preparation process in which the pumps' characteristics are analyzed. The pumps data may be obtained from SCADA system records, field measurements, pump tests or the original pump's manufacturer data sheets. The required data is a list of operational states with their flow and hourly energy consumption. If there are  $n$  pumps in the pumping station, the number of pumping states (i.e., pumps combinations),  $n_{states}$ , is  $2^n$  including the "do-nothing" state in which no pumps are working. In practice, not all theoretical states are feasible due to different constraints such as pressure restrictions in the system and/or limitation on the power connection to the pumping station. The final list of the states should include only the feasible states. Next, the list of feasible states is clustered into groups of states with similar flow values. This is done since the FAA is designed to add flow in each iteration to reduce the infeasibility of the system by selecting the best state from a higher flow group. For example, if there are two similar pumps in a pumping station, there are three distinct groups of states: (1) one state with no pump is working (2) two states in which only one pump is working, and (3) one state with the two pumps are working. In the illustrative example, presented in Table 1, which also includes two pumps, four groups could be identified because the two pumps are not similar and each one of them, when working alone, could be a separate group. Noteworthy that the groups' id ranges from 0 to  $N_G$  from the lowest flow group to the highest flow group.

To summarize, the off-line first stage, produces the states table which is sorted first by flow group ( $G$ ) and then by the energy ( $E$ ) to form the SST. This table also holds the states index ( $S$ ) and the states

flow ( $Q$ ).

The second stage is the main control loop which continuously runs in the PLC. The loop is repeated for every time-step (e.g., every hour) for any given time  $h$  with the following input: the electricity tariff ( $ET$ ), the minimum and maximum tank's volumetric constraints ( $V_{MIN}$  and  $V_{MAX}$  respectively), the demand forecast ( $\bar{d}$ ) for

Table 3  
Sorted States Table (SST) of the pumping states in the case study.

State	Group G	Flow Q ( $\text{m}^3/\text{hr}$ )	Energy E (kWh)
0	0	0	0
1	1	215	75.25
2		220	75.90
3		215	77.83
4		250	110.00
5		250	115.00
6	2	420	151.20
7		430	151.79
8		430	153.94
9		465	186.00
10		470	187.06
11		465	188.33
12		460	188.60
13		470	190.82
14		460	190.90
15		410	221.40
16	3	625	228.13
17		680	261.80
18		675	263.25
19		680	264.52
20		675	264.60
21		670	265.99
22		675	266.63
23		620	293.26
24		620	294.50
25		630	297.36
26	4	875	338.63
27		870	341.04
28		835	371.58
29		835	372.41
30		830	390.10
31	5	1030	448.05

the next operation horizon ( $T$ ), and the initial conditions of the system (e.g., the current tank level). Usually the pumping station's PLC can communicate directly with the tank's PLC without the need for a centralized system. With the availability of tanks' water level records and flow records in the pump station PLC, it can construct demand records through simple water mass balance. These demand records are used for demand prediction within the PLC using the NDF method. Next, the control loop invokes the optimization stage (i.e., FAA). The output of the optimization stage is the pumps schedule for the next operation horizon, that is, the list of pumping states to be operated at the next  $T$  time-steps. Once the list is obtained, only the decision of the current time-step is implemented and then the control loop waits for the next time-step and the process is repeated in a receding horizon manner. The FAA which is used as the optimization procedure in the third stage is the main novelty of the proposed framework, the details of the FAA are presented in Table 2.

The algorithm is invoked with the input parameters detailed in line L1 of Table 2. The output is the selected operation states for each time-step,  $s_t$  (L2). The decision variables are initialized with the do-nothing state (L3) and the tank's volume is simulated for all times (L4). Then, an initial feasibility test is conducted (L5) and the time of the violation ( $t_{viol}$ ) of the minimum volume is returned. The main loop of the FAA is initiated (L6) and will continue until no tank volume violation is observed. The FAA allocates flow with the minimum added cost so we initialize (L7) a local variable to hold the current minimum additional cost ( $\Delta C_{min}$ ). In order to reduce the infeasibility at time  $t_{viol}$  it is evident that the additional flow must be allocated at a time not exceeding  $t_{viol}$ . To this end, we

search to add flow between  $t_{viol}$  and the first time-step (L8). The reasoning for searching from  $t_{viol}$  backwards in time is that we prefer delaying our pumping decisions, such that we will have the option for recourse actions when time progresses. At each time  $t$  we loop through the SST, starting with states above the existing state of time  $t$  (L9), while limiting our search to states in a higher group (L10). During the search, we calculate (L11) the additional cost incurred ( $\Delta C$ ) which is the difference between the energy consumption of the examined state ( $s_{tmp}$ ) and the existing state, multiplied by the electricity tariff ( $ET_t$ ). If this additional cost is less than the current minimum additional cost ( $\Delta C_{min}$ ) (L12), we simulate the tank's expected volume with the proposed state (L13) and check for infeasibility in the tank's upper limit (L14). If this proposed state ( $s_{tmp}$ ) causes infeasibility in the upper limit, we move on to check the next state (L21). However, if there are no violation of the upper volume constraint, we update  $\Delta C_{min}$  with the new additional cost (L15) and record the selected state and time (L16). Now we break the states' loop (L16) and move to examine the possibility to add flow in earlier time-steps (L22). Once we cover all the optional time-steps we update the new selected state and time (L23) which completes one iteration of the FAA. Next, we simulate the tank's volume trajectories (L24) and check for infeasibility (L25). If there is an infeasibility,  $t_{viol}$  is updated and the loop continues (L6). If no infeasibility is found, the algorithm returns the set of the selected operational states (L27). An important observation is that the FAA described above, can be computed easily with the use of common operators "For\While" loops, "If" statements and basic mathematical operations. Hence, as described previously, it is compatible with local control scheme and it could be implemented

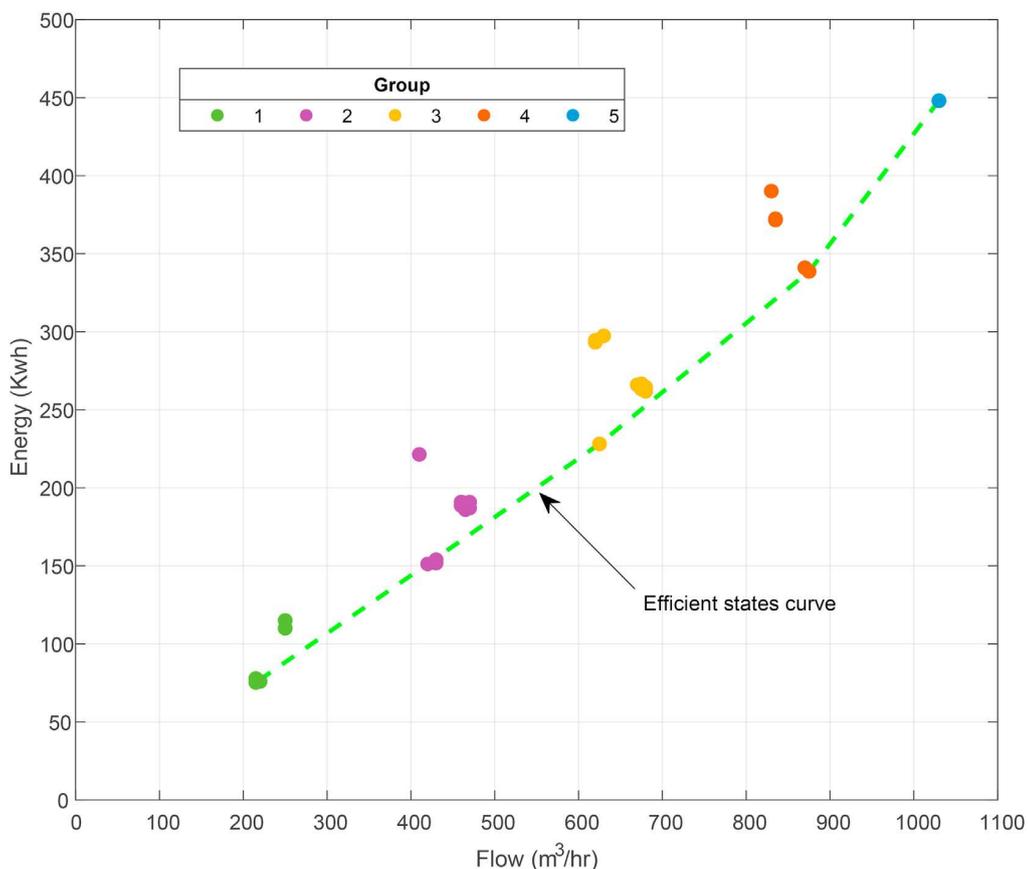


Fig. 7. Pumping states energy consumption vs. flow in the case study.

**Table 4**  
Electricity tariff structure.

Day	Off-peak hours	Mid-peak hours	Peak hours
Sun – Thu	00–05 22–23	20–21	06–19
Fri	00–05 20–23	06–19	–
Sat	00–16 20–23	17–20	–
Electricity price (\$/ kWh)	0.0842	0.1066	0.1339

on a standard PLC.

**3. Case study**

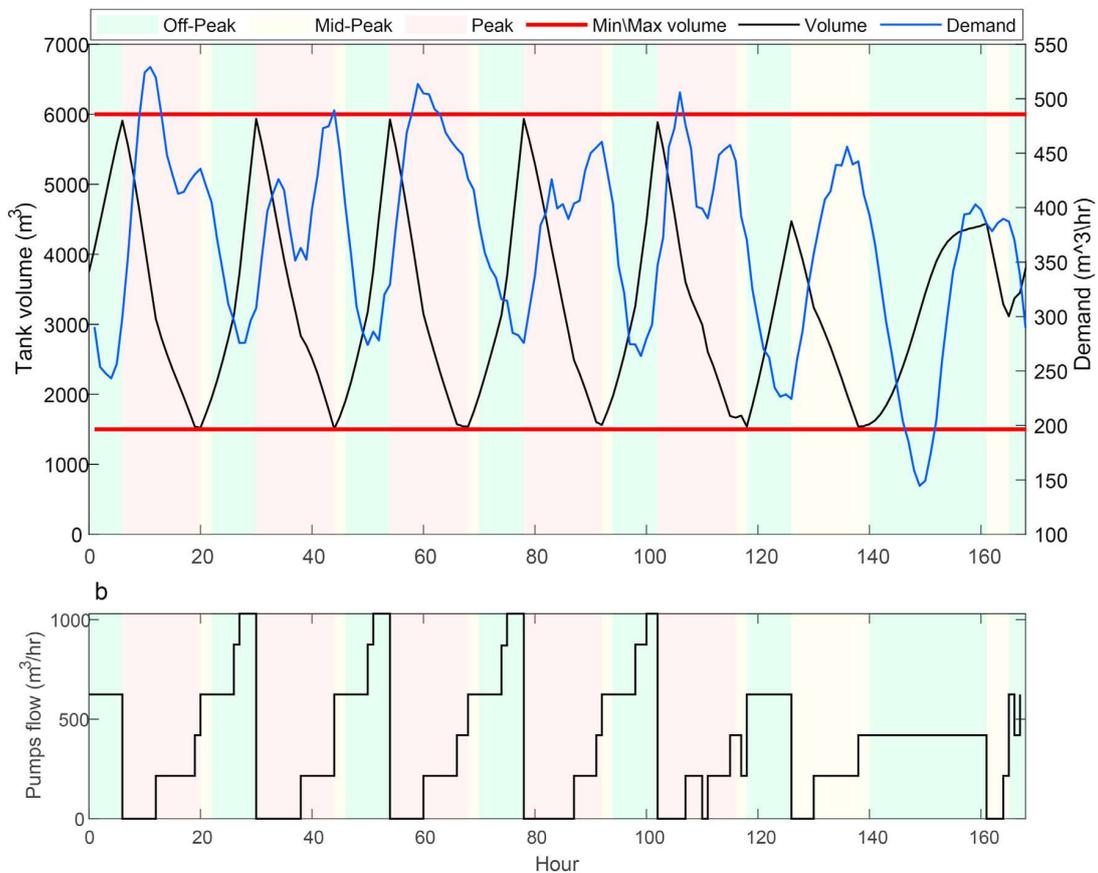
To demonstrate the suggested methodology we consider an HPZ (Fig. 5), which is supplied by two pumping stations, one with two pumps and the second with three pumps. The zone is served by one elevated tank with an operational minimum and maximum volume constraints of 1500 m<sup>3</sup> and 6000 m<sup>3</sup> respectively. The data of this case study are based on real-life measurements of pump flows, consumed power (i.e., pump states) and demand time series. Fig. 6 shows four weeks of demand data. Each series in Fig. 6 is an hourly demand profile for one week. The origin of this case study’s data is an Israeli city with a mixed Muslim and Jewish population; thus, Friday and Saturday are holydays. For the weekdays (i.e., Sunday to Thursday), a typical two demand peaks can be observed. The weekend days have a different pattern in which the demand

decreases in Friday afternoon.

In our case study, the operation of any pump in one pumping station would not significantly affect the operation of the pumps in the other, thus, pumping states in each station can work at the same time. This is because the two pumping stations work (almost) directly against an elevated tank. In this situation, the two pumping stations work against the topographic difference independently. Noteworthy that, in this real example, the change in the tank level is negligible compared to the high topographic difference. These conditions are also satisfied in other networks. Jowitt and Germanopoulos (1992) provide a thorough explanation on the validity of this assumption. They argue that in some practical networks, despite that the flow pattern in the network may change significantly as a result of pump switching, the magnitude of the nodal heads will not change significantly, and thus, the pumping stations will operate near the same operating point.

The states from the two stations could be combined to construct the SST. With five pumps feeding the HPZ, there are a total of  $n_{states} = 32$  pumping states, including the “do-nothing” state, as shown in the SST (Table 3). The pumping states in Table 3 are grouped by flow similarity which coincide with the number of pumps included in each state (e.g., in states 1–5 of group 1, only 1 pump is operated). The coefficient of variation (i.e., standard deviation divided by the average) of the flow in each group ranges between 0% (group 5) to 8% (group 1) as can be seen in Fig. 7.

Fig. 7 shows the hourly energy consumption of the pumping states as a function of their flow. The energy consumption raises with the flow between the pumping groups while the flow and



**Fig. 8.** Tank’s volume (a) and pumps flows (b) results for one week.

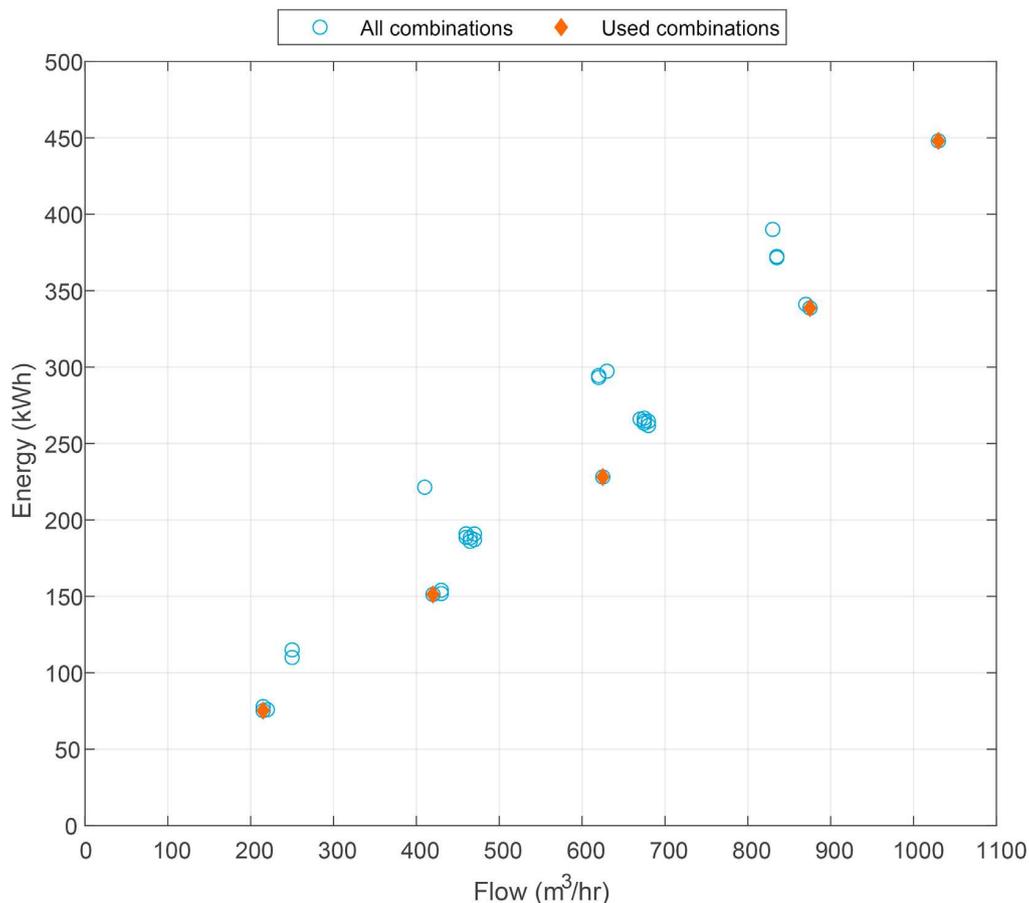


Fig. 9. Selected pumping states during one month operation of the test case.

energy vary within each group. The most efficient pumping state in each group is the one with the smallest energy consumption. This set of pumping states comprises the efficient states curve as marked in Fig. 7. These states correspond to the first state for each group in Table 3 (i.e., states 1, 6, 16, 26 and 31).

The electricity tariff is outlined in Table 4. The tariff has three cost levels: off-peak, mid-peak and peak which vary during the hours of the day and the days of the week. Sundays through Thursday have the same tariff structure while Fridays and Saturdays have a different structure each. These tariffs are based on the Israeli electricity tariff structure.

As the electricity tariff structure has a weekly cycle and the demand profile of this HPZ also vary in a weekly manner, the operation horizon is set to one week,  $T = 168$  hours, with a time-step of 1 h. The time step of 1 h is usually selected for practical reasons mainly to reduce frequent pump switches, which can cause mechanical damage, water hammer and water quality issues (Alvisi and Franchini, 2017; Housh and Salomons, 2019; Lansey and Awumah, 1994; Wood, 2005).

#### 4. Results

The abovementioned test case was run hour by hour in a receding horizon mode for a full month with a total of 720 runs according to the methodology outlined in Fig. 4. The process started with a tank volume of  $V_0 = 3750 \text{ m}^3$  which is the middle of the operational volume of the tank. The results of one-week, out of the

full month, are shown in Fig. 8. The tank volume for the week (Fig. 8a) fluctuates between the minimum and maximum allowed volumes, where, in general, the tank fills during the off-peak electricity tariff periods and empties during the peak periods. Fig. 8a also shows the hourly water demand for the week which exhibits decreasing demands during the weekend (Friday and Saturday). Fig. 8b shows the pumps flow over time. The results show that most of the pumping is done during the off-peak periods, where the high flows are always postponed to the end of the off-peak period. This property of delaying high flow, is due to the backward in time flow allocation strategy (L8 in Table 2) which we discussed earlier. During the weekend (the last 48 h) the tank does not totally fill since there are enough off-peak and mid-peak hours to allow more modest pumping rates (which are more energy efficient) over longer time as can be noticed in Fig. 8b. Additionally, the pumps operation is smooth and the changes in the flow rate is gradual. This is a desirable property in pumps operation, since frequent on/off operation of pumps may affect the WDS functionality as discussed in many studies (Lansey and Awumah, 1994). All the results in this paper were built using MATLAB version R2018b, the YALMIP toolbox (Lofberg, 2004) on a 64-bit Lenovo X1 ThinkPad with an Intel i7-7600U CPU @ 2.8 GHz and 16 GB of RAM.

It should be noted that the tank volume is not enough to supply the water during the entire peak period. Thus, some pumping must be made during the peak hours (e.g., hours 12–20 in Fig. 7b), but this is mostly done with the small flow rate of pumping state 1, i.e.  $215 \text{ m}^3/\text{hr}$  (Table 3). Another nice property of the FAA algorithm, is

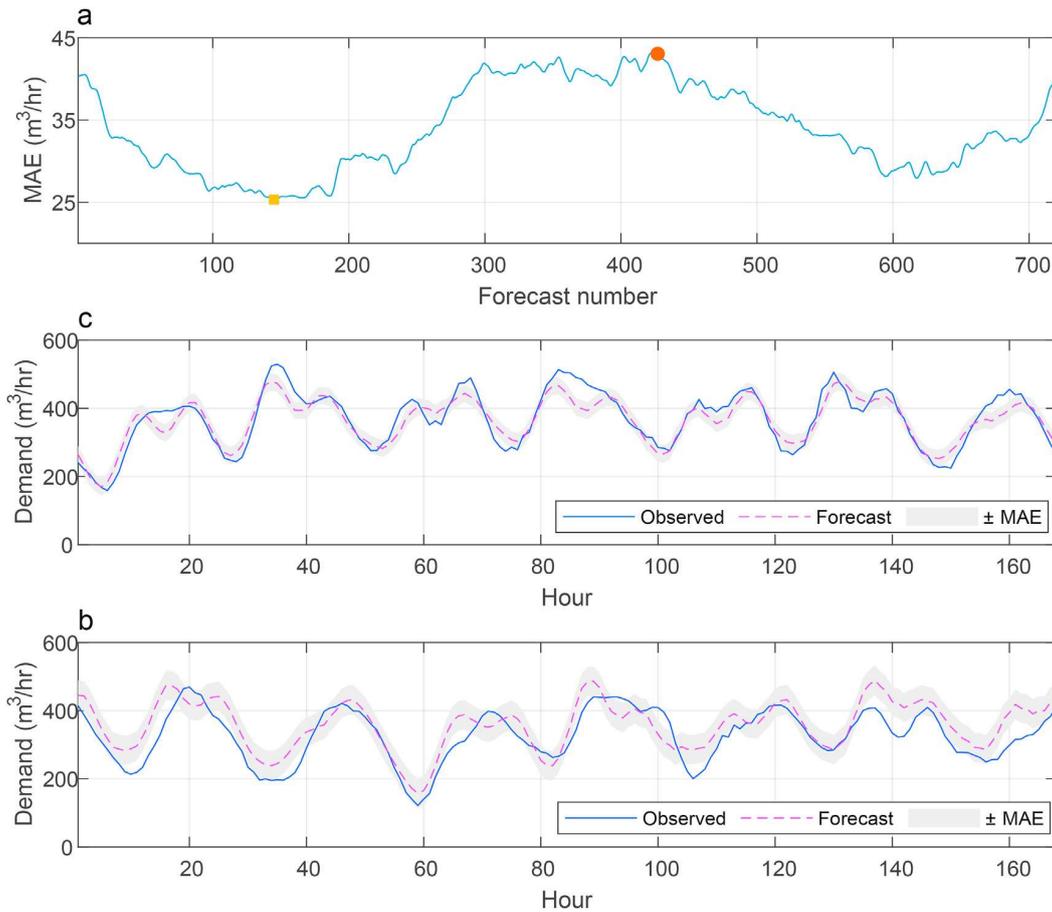


Fig. 10. NDF method performance: (a) MAE for all forecasts; and (b and c) observed and forecast demands for the minimum and maximum MAE, respectively.

Table 5  
Comparison of optimization methods.

Run	Optimization procedure	Strategy	Demand	Total cost (\$)
FAA(RH-FD)	FAA	Receding horizon	Forecasted	29,911
MILP(RH-FD)	MILP (CBC)	Receding horizon	Forecasted	29,248
$TMC_{FAA}$	FAA	Full month	True	29,753
$TMC_{MILP}$	MILP (CPLEX)	Full month	True	29,032

that all the selected pumping states during one month of operation belong to the efficient states set, as can be seen in Fig. 9. This indicates an efficient use of energy for pumping purposes, which also contributes to reducing the GHG footprint of the water utility, in addition to reducing the operation cost.

The results above are based on the framework presented in Fig. 4, which requires weekly demand forecasts for each of the 720 runs. To evaluate the accuracy of the NDF demand forecast, which is adapted in this study, we compare the forecasts with the real demand values by calculating the Mean Absolute Error (MAE) in Eq. (3).

$$MAE = \frac{1}{|\tau_h|} \sum_{t \in \tau_h} |d_{obs, t} - \tilde{d}_t| \quad (3)$$

where  $d_{obs, t}$  is the observed demand at time  $t$  and  $|\tau_h|$  is the number of time-steps in the forecasting horizon,  $T$ . The MAE for all 720 forecasts is shown in Fig. 10a with a minimum (best) and maximum (worst) values of  $25.3 \text{ m}^3/\text{hr}$  and  $43.1 \text{ m}^3/\text{hr}$  respectively. The observed and forecasted weekly demand pattern for the

best and the worst forecasts are shown in Fig. 10b and c, respectively. These results demonstrate the good performance of the NDF method despite its simplicity.

The total cost for the entire month of the Receding Horizon strategy with Forecasted Demands (RH-FD), as obtained by the FAA, is \$29,911. To evaluate the optimality of this solution, we compare it with other optimization methods (Table 5). First, using the receding horizon strategy and the NDF forecasted demands, we formulate the optimization problem as a MILP problem and solve it using the CBC solver (Forrest and Lougee-Heimer, 2005) and obtain a total cost of \$29,248 which is only 2.3% less than the FAA. The daily costs obtained by the FAA and the MILP solver are shown in Fig. 11a.

Next, we use the FAA and the MILP formulation and solve the full month with the true demands, here we solve the entire month as one problem without a receding horizon strategy. This is done to estimate the Theoretical Minimum Cost (TMC) of the WDS operation, when unrealistically assuming that the demand can be perfectly predicted for the entire month. Thus, this filters out the effect of the NDF method and focuses the comparison on the

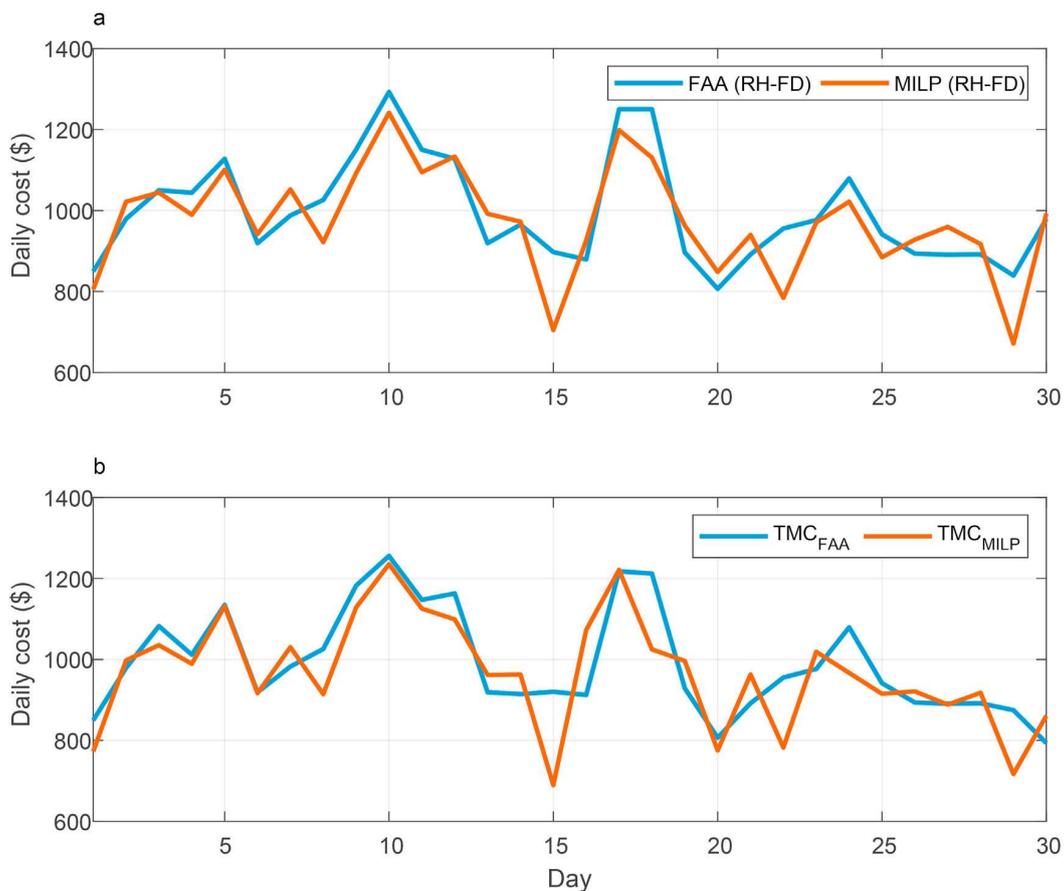


Fig. 11. Daily operation costs for FAA and MILP solvers.

optimality of the obtained schedule. When solving the TMC with the FAA we obtain a solution of \$29,753 while solving the TMC with the MILP solver we obtain a solution of \$29,032 which is only 2.5% cheaper than the FAA solution. Solving the large-scale MILP of one month required the use of the advanced commercial CPLEX solver (IBM Corp, 2009). Noteworthy that the difference between FAA(RH-FD) and  $TMC_{FAA}$  is only 0.5% indicating the good performance of the NDF despite its simplicity. The daily costs of the TMC as obtained by the FAA and CPLEX solver are shown in Fig. 11b. These results highlight the optimality of the FAA, as it can reach near optimal solutions in all days.

The near optimal solution of the FAA highlights its applicability to the real-time MPC framework in local control scheme. This is because, firstly, it solves the optimization problem three orders of magnitude faster than the MILP formulation. The cumulative probability distributions of the run time for the FAA and the MILP solvers are shown in Fig. 12 (note the horizontal log axis). As the optimization procedure is intended for real-time control scheme, a practical time limit for each optimization run is set at 5 min (Salomons and Housh, 2020). Despite that the time step is 1 h, a new operating plan for the next time step must be obtained much faster (e.g. 5 min). Allowing the optimization to run for 1 h, and only then deploying the plan, is not advisable since after 1 h the system state may be changed significantly. In such case the obtained results are not relevant anymore. The results show that the FAA solves the optimization problem in less than one tenth of a second in all cases! Whilst the MILP solver exceeds this threshold in 13% of the cases (in 4% of the cases the run time is above 1000 s which is the defined maximum

run time for the CBC MILP solver). Secondly, and this is the most important advantage of the FAA, the execution of the algorithm can be performed in a simple PLC with simple operators such as loops and conditional statements, as opposed to heavy software dependencies and heavy computational demand of the MILP solver.

## 5. Conclusions

Pump scheduling methods which utilize computation intensive optimization algorithms may be suitable for implementation as a centralized solution in dedicated control rooms where advanced software and hardware are available. However, these methods cannot be installed in local PLCs of pumping stations or water tanks sites. Typically, the control logic which runs on the local PLC should be simple and with less computational requirements. As such, in many water utilities which use local control schemes, the potential benefits of optimal pumps' scheduling are unexploited. This study presents a practical optimization methodology for real-time control which is compatible with local control schemes. Hence, it leverages optimization methods in simple PLCs without the need for large investments in centralized control infrastructure.

The core of the proposed framework is the practical optimization procedure depicted in the FAA. The algorithm iteratively allocates additional flow to the water network to reduce infeasibilities while balancing the optimality of the solution over the operation horizon. The pumping states are allocated according to the SST which ensures efficient pumps' states selection. The most reliable data for the pumps' combinations characteristics, which is used to

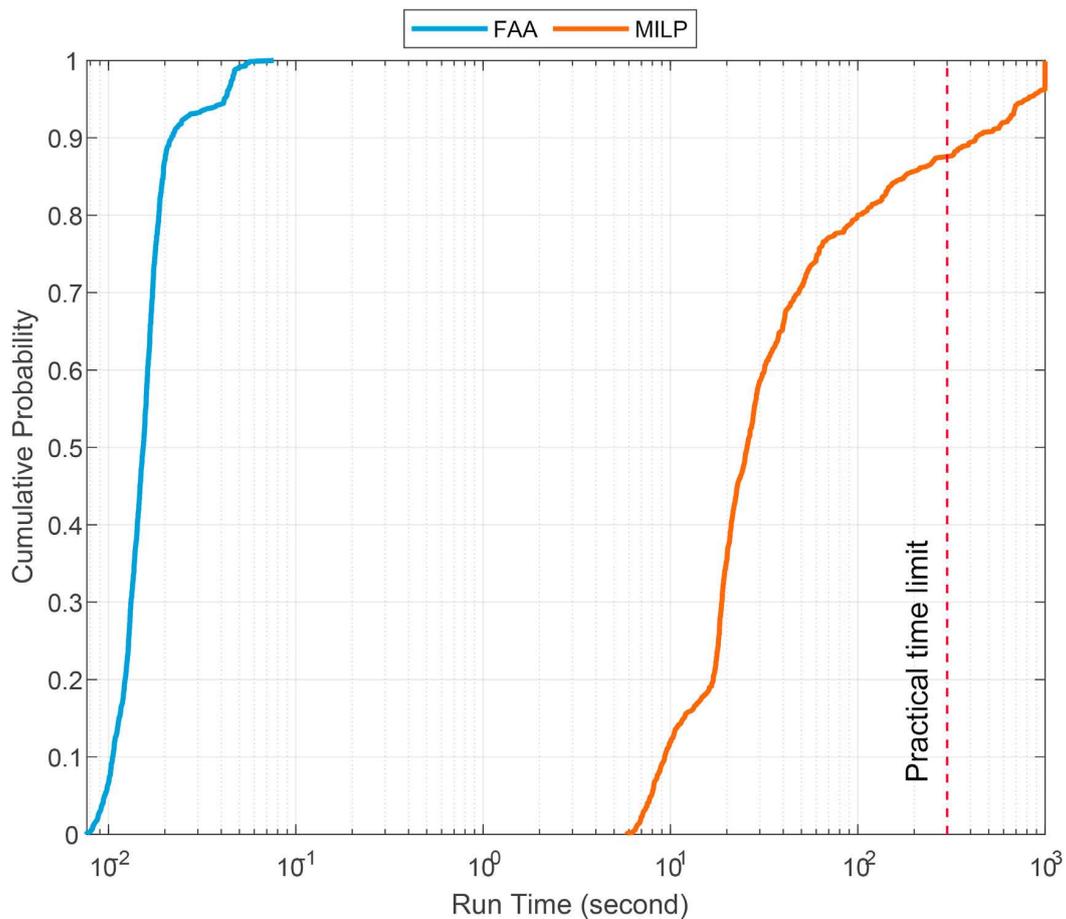


Fig. 12. Cumulative probability distributions of run time for FAA and MILP solver.

build the SST, can be obtained by analyzing SCADA data. This process is done offline and depends on both the specific pumps and the network characteristics. Our results indicate that the suggested framework not only reduces the operation costs, but also saves energy by utilizing the most efficient pumps. Thus, it can help the utility in reducing its GHG footprint. The proposed optimization procedure is compared to MILP solvers (which are typically used in centralized control schemes). The results show that it can provide near-optimal solutions, comparable with the obtained from MILP solvers, in a fraction of the time required by the solvers. In fact, the MILP implementation does not always run within a practical time limit suitable for real-time implementation, while the FAA runs in less than one tenth of a second.

The suggested framework is simple, practical, and applicable to local control schemes. This is achieved by its modest software computational requirements of loops, simple conditional statements, and basic mathematical operators. The proposed algorithm is designed for fixed speed pumps; however, variable speed pumps are also common. This is especially true in HPZs without storage in which the variable demand is supplied directly by the pumps and the pump's speed is dynamically adjusted to meet the demand. One way to incorporate variable speed pumps in the proposed method is to discretize the range of possible speeds into a set of states. Future work should examine the applicability of the proposed method for variable speed pumps. Other directions for future research may include better simulation of the water system with a hydraulic simulator acting as the real system, the inclusion of user

defined constraints such as mid-day tank volume targets and more complex electricity tariffs (e.g., power spot markets).

#### CRediT authorship contribution statement

**Elad Salomons:** Conceptualization, Methodology, Software, Data curation, Writing - original draft, preparation, Visualization, Investigation. **Mashor Housh:** Supervision, Software, Validation, Writing - review & editing.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

This study was supported by the Israeli Water Authority.

#### References

- Abdallah, M., Kapelan, Z., 2019. Fast pump scheduling method for optimum energy cost and water quality in water distribution networks with fixed and variable speed pumps. *J. Water Resour. Plann. Manag.* 145 [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001123](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001123).
- Alvisi, S., Franchini, M., 2017. A robust approach based on time variable trigger levels for pump control. *J. Hydroinf.* 1–12 <https://doi.org/10.2166/hydro.2017.141>.

- Alvisi, S., Franchini, M., Marinelli, A., 2007. A short-term, pattern-based model for water-demand forecasting. *J. Hydroinf.* 9, 39. <https://doi.org/10.2166/hydro.2006.016>.
- Blinco, L.J., Simpson, A.R., Lambert, M.F., Marchi, A., 2016. Comparison of pumping regimes for water distribution systems to minimize cost and greenhouse gases. *J. Water Resour. Plann. Manag.* [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000633](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000633).
- Bunn, S.M., Reynolds, L., 2009. The energy-efficiency benefits of pumpscheduling optimization for potable water supplies. *IBM J. Res. Dev.* 53 <https://doi.org/10.1147/JRD.2009.5429018>.
- Candelieri, A., Perego, R., Archetti, F., 2018. Bayesian optimization of pump operations in water distribution systems. *J. Global Optim.* 71, 213–235. <https://doi.org/10.1007/s10898-018-0641-2>.
- Carpentier, P., Cohen, G., 1993. Applied mathematics in water supply network management\*. *Automatica* 29, 1215–1250. [https://doi.org/10.1016/0005-1098\(93\)90048-x](https://doi.org/10.1016/0005-1098(93)90048-x).
- Cembrano, G., Wells, G., Quevedo, J., Pérez, R., Argelaguet, R., 2000. Optimal control of a water distribution network in a supervisory control system. *Contr. Eng. Pract.* 8, 1177–1188. [https://doi.org/10.1016/S0967-0661\(00\)00058-7](https://doi.org/10.1016/S0967-0661(00)00058-7).
- Cimorelli, L., D'Aniello, A., Cozzolino, L., 2020. Boosting genetic algorithm performance in pump scheduling problems with a novel decision-variable representation. *J. Water Resour. Plann. Manag.* 146, 04020023 [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001198](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001198).
- Conejos Fuertes, P., Martínez Alzamora, F., Hervás Carot, M., Alonso Campos, J.C., 2020. Building and exploiting a Digital Twin for the management of drinking water distribution networks. *Urban Water J.* 1–10 <https://doi.org/10.1080/1573062X.2020.1771382>.
- Copeland, C., Carter, N.T., 2017. Energy - Water Nexus: the Water Sector' S Energy Use R43200 13.
- 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.* <https://doi.org/10.1016/j.watres.2019.06.025>.
- Darweesh, M.S., 2020. Impact of optimized pump scheduling on water quality in distribution systems. *J. Pipeline Syst. Eng. Pract.* 11 [https://doi.org/10.1061/\(asce\)ps.1949-1204.0000486](https://doi.org/10.1061/(asce)ps.1949-1204.0000486), 05020004.
- Donkor, E.A., Mazzuchi, T.A., Soyer, R., Alan Roberson, J., 2014. Urban water demand forecasting: review of methods and models. *J. Water Resour. Plann. Manag.* 140, 146–159. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000314](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000314).
- Edwards, J., Maher, J., 2008. Water quality considerations for distribution system storage facilities. *J. Am. Water Work. Assoc.* <https://doi.org/10.1002/j.1551-8833.2008.tb09676.x>.
- Farmani, R., Walters, G.A., Savic, D., 2006. Evolutionary multi-objective optimization of the design and operation of water distribution network: total cost vs. reliability vs. water quality. *J. Hydroinf.* 8, 165–179. <https://doi.org/10.2166/hydro.2006.019b>.
- Forrest, J., Lougee-Heimer, R., 2005. CBC user guide. In: *Emerging Theory, Methods, and Applications*. <https://doi.org/10.1287/educ.1053.0020>.
- Herrera, M., Torgo, L., Izquierdo, J., Pérez-García, R., 2010. Predictive models for forecasting hourly urban water demand. *J. Hydrol.* <https://doi.org/10.1016/j.jhydrol.2010.04.005>.
- Housh, M., Salomons, E., 2019. Optimal dynamic pump triggers for cost saving and robust water distribution system operations, 145, pp. 1–9. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001028](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001028).
- IBM Corp., 2009. V12. 1: user's manual for CPLEX. *Int. Bus. Mach. Corp.* 12, 481.
- Jamshidnejad, A., Papamichail, I., Papageorgiou, M., De Schutter, B., 2016. A model-predictive urban traffic control approach with a modified flow model and endpoint penalties. *IFAC-PapersOnLine* 49, 147–152. <https://doi.org/10.1016/j.ifacol.2016.07.025>.
- Jowitt, P.W., Germanopoulos, G., 1992. Optimal pump scheduling in water-supply networks. *J. Water Resour. Plann. Manag.* 118, 406–422. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1992\)118:4\(406\)](https://doi.org/10.1061/(ASCE)0733-9496(1992)118:4(406)).
- Khatavkar, P., Mays, L.W., 2018. Model for real-time operations of water distribution systems under limited electrical power availability with consideration of water quality. *J. Water Resour. Plann. Manag.* 144 [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001000](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001000), 04018071.
- Khatavkar, P.N., Mays, L.W., 2019. Testing an optimization–simulation model for optimal pump and valve operations with required storage tank turnovers. *J. Water Manag. Model.* 2019 464. <https://doi.org/10.14796/jjwmm.c464>.
- Lam, K.L., Kenway, S.J., Lant, P.A., 2017. Energy use for water provision in cities. *J. Clean. Prod.* 143, 699–709. <https://doi.org/10.1016/j.jclepro.2016.12.056>.
- Lansley, K., Awumah, K., 1994. Optimal pump operations considering pump switches. *J. Water Resour. Plann. Manag.* 120, 17–35. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1994\)120:1\(17\)](https://doi.org/10.1061/(ASCE)0733-9496(1994)120:1(17)).
- Lofberg, J., 2004. YALMIP : a toolbox for modeling and optimization in MATLAB. In: 2004 IEEE Int. Conf. Robot. Autom. (IEEE Cat. No.04CH37508), pp. 284–289. <https://doi.org/10.1109/CACSD.2004.1393890>.
- Luna, T., Ribau, J., Figueiredo, D., Alves, R., 2019. Improving energy efficiency in water supply systems with pump scheduling optimization. *J. Clean. Prod.* 213, 342–356. <https://doi.org/10.1016/j.jclepro.2018.12.190>.
- Mala-Jetmarova, H., Sultanova, N., Savic, D., 2017. Lost in optimisation of water distribution systems? A literature review of system operation. *Environ. Model. Software* 93, 209–254. <https://doi.org/10.1016/j.envsoft.2017.02.009>.
- Mei-Carmel, n.d. Haifa's water sector - challenges, infrastructure and technology [WWW Document], 6.7.20. <https://www.mei-carmel.co.il/?CategoryID=190>.
- Nowak, D., Krieg, H., Bortz, M., Geil, C., Knapp, A., Roclawski, H., Böhle, M., 2018. Decision support for the design and operation of variable speed pumps in water supply systems. *Water (Switzerland)* 10, 734. <https://doi.org/10.3390/w10060734>.
- Ocampo-Martinez, C., Barcelli, D., Puig, V., Bemporad, A., 2012. Hierarchical and decentralised model predictive control of drinking water networks: application to Barcelona case study. *IET Control Theory & Appl.* 6, 62. <https://doi.org/10.1049/iet-cta.2010.0737>.
- Ocampo-Martinez, C., Puig, V., Cembrano, G., Quevedo, J., 2013. Application of predictive control strategies to the management of complex networks in the urban water cycle. *IEEE Contr. Syst. Mag.* 33 <https://doi.org/10.1109/MCS.2012.2225919>.
- Odan, F.K., Ribeiro Reis, L.F., Kapelan, Z., 2015. Real-time multiobjective optimization of operation of water supply systems. *J. Water Resour. Plann. Manag.* 141 [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000515](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000515).
- Ormsbee, L.E., Lansley, K.E., 1994. Optimal control of water supply pumping systems. *J. Water Resour. Plann. Manag.* 120, 237–252. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1994\)120:2\(237\)](https://doi.org/10.1061/(ASCE)0733-9496(1994)120:2(237)).
- Pacchin, E., Alvisi, S., Franchini, M., 2017. A short-term water demand forecasting model using a moving window on previously observed data. *Water* 9, 172. <https://doi.org/10.3390/w9030172>.
- Paschke, M., Spencer, K., Waniarcha, N., Simpson, A.R., Widdop, T., 2001. Genetic algorithms for optimising pumping operations. In: *19th Federal Convention. Australian Water Association, Canberra, Australia*.
- PreDESCU, A., Truică, C.-O., Apostol, E.-S., Mocanu, M., Lupu, C., 2020. An advanced learning-based multiple model control supervisor for pumping stations in a smart water distribution system. *Mathematics* 8, 887. <https://doi.org/10.3390/math8060887>.
- Rossman, L.A., 2000. EPANET manual. *Soc. Stud. Sci.* <https://doi.org/10.1177/0306312708089715>.
- Salomons, E., Housh, M., 2020. A practical optimization scheme for real-time operation of water distribution systems. *J. Water Resour. Plann. Manag.* 146 [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001188](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001188), 04020016.
- Sanders, K.T., Webber, M.E., 2012. Evaluating the energy consumed for water use in the United States. *Environ. Res. Lett.* 7 <https://doi.org/10.1088/1748-9326/7/3/034034>.
- Slavik, I., Oliveira, K.R., Cheung, P.B., Uhl, W., 2020. Water quality aspects related to domestic drinking water storage tanks and consideration in current standards and guidelines throughout the world – a review. *J. Water Health.* <https://doi.org/10.2166/wh.2020.052>.
- Torregrossa, D., Capitanescu, F., 2019. Optimization models to save energy and enlarge the operational life of water pumping systems. *J. Clean. Prod.* 213, 89–98. <https://doi.org/10.1016/j.jclepro.2018.12.124>.
- Vieira, T.P., Almeida, P.E.M., Meireles, M.R.G., Souza, M.J.F., 2018. Use of computational intelligence for scheduling of pumps in water distribution systems: a comparison between optimization algorithms. In: 2018 IEEE Congress on Evolutionary Computation, CEC 2018 - Proceedings. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/CEC.2018.8477833>.
- Wang, Y., Cembrano, G., Puig, V., Urrea, M., Romera, J., Saporta, D., Valero, J.G., Quevedo, J., 2017. Optimal management of Barcelona water distribution network using non-linear model predictive control. *IFAC-PapersOnLine.* <https://doi.org/10.1016/j.ifacol.2017.08.1070>.
- Water-and-Sewer, 2014. SCADA upgrade project [WWW document], 5.18.20. <http://waterandsewer.org/current-projects/scada-upgrade-at-wpfi>.
- Water-Technology, 2013. DEWA awards SCADA system project to monitor water network in Dubai [WWW Document], 5.18.20. <https://www.water-technology.net/uncategorised/newsdewa-awards-scada-system-project-monitor-water-network-dubai/>.
- Wood, D.J., 2005. Waterhammer analysis — essential and easy (and efficient). *Ascent* 131, 1123–1131. [https://doi.org/10.1061/\(ASCE\)0733-9372\(2005\)131:8\(1123\)](https://doi.org/10.1061/(ASCE)0733-9372(2005)131:8(1123)).
- Wytock, M., Moehle, N., Boyd, S., 2017. Dynamic energy management with scenario-based robust MPC. In: Proceedings of the American Control Conference. <https://doi.org/10.23919/ACC.2017.7963253>.
- Yu, G., Powell, R.S., Sterling, M.J.H., 1994. Optimized pump scheduling in water distribution systems. *J. Optim. Theor. Appl.* 83, 463–488. <https://doi.org/10.1007/BF02207638>.
- Zhou, S.L., McMahon, T.A., Walton, A., Lewis, J., 2002. Forecasting operational demand for an urban water supply zone. *J. Hydrol.* 259, 189–202. [https://doi.org/10.1016/S0022-1694\(01\)00582-0](https://doi.org/10.1016/S0022-1694(01)00582-0).

## Paper II

*Salomons, E.; Housh, M. "A Practical Optimization Scheme for Real-Time Operation of Water Distribution Systems". Water Resources Planning and Management 2020, 146 (4), 04020016. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001188](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001188).*

Journal Impact Factor: 2.406

5-Year Impact Factor: 2.864

Publication Status: Published (Accepted 2 October 2019)

## Statement of authorship - Paper II

### Principal Author

Title of Paper	A Practical Optimization Scheme for Real-Time Operation of Water Distribution Systems		
Publication Status	<input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication		
Publication Details	Salomons, E.; Housh, M. "A Practical Optimization Scheme for Real-Time Operation of Water Distribution Systems". Water Resources Planning and Management 2020, 146 (4), 04020016. <a href="https://doi.org/10.1061/(ASCE)WR.1943-5452.0001188">https://doi.org/10.1061/(ASCE)WR.1943-5452.0001188</a>		
Name of Principal Author (Candidate)	Elad Salomons		
Contribution to the Paper	Conceived the presented idea, developed the theory and performed the computations. Wrote the manuscript with support from the co-author.		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Name and Signature	Elad Salomons 	Date	October 2 <sup>nd</sup> , 2019

### Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- iii. the candidate's stated contribution to the publication is accurate (as detailed above);
- iv. permission is granted for the candidate to include the publication in the dissertation

Name of Co-Author 1	Mashor Housh		
Contribution to the Paper	Verified the analytical methods, encouraged Elad Salomons to investigate the subject, and supervised the findings of this work.		
Name and Signature	Mashor Housh 	Date	October 2 <sup>nd</sup> , 2019



# A Practical Optimization Scheme for Real-Time Operation of Water Distribution Systems

Elad Salomons<sup>1</sup> and Mashor Housh, M.ASCE<sup>2</sup>

**Abstract:** Pump scheduling is a key element in water distribution systems operation. Modeling this problem requires a mixed integer nonlinear program (MINLP) formulation. Even linearization schemes of mixed integer linear programs (MILPs) are typically beyond the capability of real-time optimization frameworks. In this study, we explore different levels of MILP approximations by reducing the number of binary decision variables (i.e., different binarization levels). In addition, we present a simple demand forecast model and evaluate the performance and approximation accuracy of the suggested approach in a real-time optimization framework under a receding horizon operation mode. The results show that the balance between approximation accuracy and solution efficiency is biased. That is, a simple low-accuracy approximation may yield an efficient and practical solution algorithm that results in a near-optimal solution. DOI: 10.1061/(ASCE)WR.1943-5452.0001188. © 2020 American Society of Civil Engineers.

**Author keywords:** Pump scheduling; Demand forecast; Water network operation; Real time; Model predictive control.

## Introduction

Because of environmental regulations and increasing energy costs, energy conservation and efficiency are gaining importance in many water utilities. There are different activities to help utilities achieve this goal, such as energy management, correct element sizing in the system, upgrading and replacing with more efficient equipment, self-generating energy, and optimizing the operation. For the latter, smart water distribution systems (WDSs) can play a key role in achieving optimal operation for energy saving and environmentally friendly strategies. In 2010, the United States' water-related energy use was 12.6% of the total energy consumption (Sanders and Webber 2012), of which one-third (4% of the total energy consumption) is estimated to be consumed by pumping and treating water and wastewater (Goldstein and Smith 2002). Nevertheless, in many cases, WDS operation is still done according to expert opinion and rules of thumb that use local control schemes or ad-hoc control rules, especially in small-scale systems. There are about 52,000 community water systems in the US. Nearly 85% of the US population is supplied by about 5% of these systems, whereas the remaining 95% are small-scale systems serving 3,300 persons or fewer (Copeland and Carter 2017). About 80% of the energy consumed by these utilities is used by motors for pumping. Similar asymmetry is also observed in Israel, with about 55 large water corporations and over 1,000 small water suppliers. The distributed nature of national WDSs with many small-scale systems, which typically lack smart systems and computational infrastructure, require practical and simple solutions that they can

afford. As such, these small-scale systems tend to use local control schemes or ad-hoc control rules for handling the system operation.

During the last decades, many academic studies have been conducted, and new methods have been developed for optimal real-time control of WDSs. Real-time WDS control has different implementations. Creaco et al. (2019) presented a comprehensive review of real-time control objectives, mainly local, that include ensuring minimum service pressure, controlling variable speed pumps in the presence of a pressure deficit, water level control in tanks, and flow control for maximizing energy production. Ormsbee and Lansey (1994) and lately Mala-Jetmarova et al. (2017) presented a literature review of systemwide central operation that includes pump operation in the context of real-time control. In general, the overall control loop of the real-time operation of WDS is shown Fig. 1. First, the current state of the system is read, usually from a supervisory control and data acquisition (SCADA) system; a water demand forecast is performed for the next operation period; and the electricity tariffs are obtained. Then, an optimization problem is formulated and solved in order to obtain the operational settings (e.g., pumps and valve settings) that minimize the operational costs for the next operation period (e.g., 24 h). These settings must fulfill the system's constraints, both physical constraints (e.g., power connection size) and operator requirements (e.g., minimum tank levels for reliability of the supply). Finally, the obtained operational settings are implemented for the current time step and the process is repeated for the next time step. At the beginning of each time step, the initial condition of the system is set (mainly the real tank levels), thus nullifying any potential discrepancy between the design results and the real situation of the system before solving the next time step problem.

In control theory, the previously mentioned model (Fig. 1) is known as a model predictive control (MPC) model. MPC is a method for controlling dynamic processes under a set of constraints. The MPC uses a model of the dynamic system to simulate its behavior and optimize control settings while satisfying the constraints. This is done for a future finite time horizon in which some parameters are unknown; thus, it is predictive. At each time step, the process is repeated and the system state is updated within a receding horizon strategy. MPC is used in different fields, such as economics (Ellis et al. 2017) and traffic control (Jamshidnejad et al. 2016),

<sup>1</sup>Ph.D. Candidate, Dept. of Natural Resources and Environmental Management, Univ. of Haifa, Haifa 3498838, Israel. ORCID: <https://orcid.org/0000-0002-1522-1805>

<sup>2</sup>Senior Lecturer, Dept. of Natural Resources and Environmental Management, Univ. of Haifa, Haifa 3498838, Israel (corresponding author). Email: mhoush@univ.haifa.ac.il

Note. This manuscript was submitted on May 9, 2019; approved on October 2, 2019; published online on February 13, 2020. Discussion period open until July 13, 2020; separate discussions must be submitted for individual papers. This paper is part of the *Journal of Water Resources Planning and Management*, © ASCE, ISSN 0733-9496.

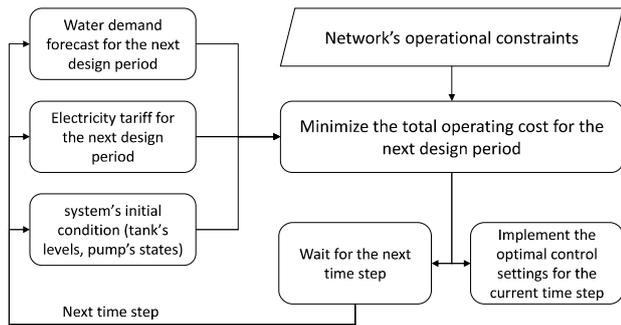


Fig. 1. Real-time control loop of WDS operations.

as well as for water resources systems (Galuppini et al. 2019; Grosso et al. 2014; Ocampo-Martinez et al. 2012). Even though the real-time WDS operation problem is a holistic control loop, as shown in Fig. 1, most of the research has concentrated on parts of the problem and less on the problem as a whole (i.e., it has focused on certain blocks such as demand forecasting and energy cost optimization). An exception is the work of Coulbeck and Orr (1989), which considered the overall aspects of the control problem consisting of a demand predictor, optimized pump scheduler, and simulator. The architecture of the control system included a control computer machine, SCADA system, activity scheduler, data manager, and performance monitor. In the potable water distribution management research project (POWADIMA), Jamieson et al. (2007), Rao and Salomons (2007), and Shamir and Salomons (2008) presented a similar platform coupling a short-term demand forecasting module with genetic algorithm (GA) optimization and artificial neural networks (ANNs) for hydraulic simulation. As opposed to the previous studies, most of the published work has concentrated on limited aspects of the control problem. For example, most of the research on optimal WDS operation has concentrated on the optimality problem for a specific time frame (e.g., 24 h) and not on the closed control loop with feedback from the system in a receding horizon mode. As described earlier, only the first time step (or steps) of the operation plan is implemented and the optimization procedure is repeated; thus, the investment and efforts put into the optimality search are not fully utilized. To that end, as opposed to many former studies, the focus of this paper is on the overall process optimality and practicality.

Several aspects of the real-time control scheme make the task challenging and, in many cases, impractical for real-world applications: (1) data availability and requirements—many demand forecasting algorithms require long history data sets in order to obtain reliable predictions (Alvisi et al. 2007). Even if the data are available, the use of long-term history data series is a limitation when there are changes in the WDS and demand baseline over time; (2) optimization difficulty—due to the on-and-off operational states of pumps, the pump scheduling problem is, in some cases, formulated as a mixed-integer linear program (MILP). That is, it is a nondeterministic polynomial time (NP)-hard problem that is difficult to solve for global optimality for large networks over a long-time horizon; and (3) computational efficiency—by nature, real-time applications need to run quickly in order to react to the rapidly changing conditions that are expected in WDSs.

Considering these challenges, in this study, the emphasis is on the overall performance of the control process rather than on each of its components, as usually done in the literature. That is, we would prefer simpler practical methodologies that together yield good (near-optimal) results over complicated components that

might add small benefits for a large price. The previous challenges will be addressed with two main pillars: (1) adoption of simple demand forecasting algorithms—an emphasis will be made on demand forecasting algorithms that do not involve long history data sets, are simple to calculate, and may adjust to near past demand changes, and (2) reducing the size of the optimization problem by using wise binary coding of the discrete decision variables. This approach should result in an optimization problem with a relatively small number of binary variables that can be solved in reasonable time with state-of-the-art commercial solvers such as CPLEX (IBM 2009) or free and open-source solvers such as CBC (Lougee-Heimer 2003). This study develops a practical method for optimal operation of WDSs and explores the tradeoff between the operation efficiency and optimality within a receding horizon operation mode as opposed to concentrating on the optimal solution for a given operation horizon, as usually done in the literature. The rest of the paper is structured as follows: first, a demand forecasting algorithm is presented. Next, we introduce the reduced MILP formulation, which is used to derive the control variables in an efficient manner, and then a realistic case study is presented while comparing the different MILP approximations. The last section summarizes the main conclusions of the study.

### Demand Forecast

The main objective of a WDS is to supply water to customers. However, before any operation planning can be done, an estimation of the future water demands must be given. Demand forecasting is the basic element in all WDS design and operation problems, where different forecasting settings are used according to the problem at hand (Donkor et al. 2014). For strategic decision-making, such as system capacity expansion, a long-term forecasting horizon should be used (e.g., over 10 years with annual periodicity). For tactical planning, such as revenue forecast or staging system improvement, a shorter forecasting horizon should be used (e.g., 1–10 years with monthly periodicity). However, for operations purposes, a short-term forecasting horizon should be used, which typically ranges from 24–48 h up to 1 year with hourly, daily, or weekly periodicity.

In the real-time operation problem, which is the focus of this study, we consider the low end of the demand forecasting horizon, that is, 24–48 h with hourly periodicity. Zhou et al. (2002) developed daily total and 24-h-ahead demand forecasting models for the city of Melbourne (Australia) by dividing the demand into base and seasonal consumption, thus characterizing it on a daily and monthly basis. However, taking into account seasonal demand variations requires long history data sets; for example, Zhou et al. (2002) used a historical data set of 6 years. Similarly, Alvisi et al. (2007) constructed a daily total and hourly water demand forecast for real-time near-optimal operation of a WDS. The daily demand is modeled by a Fourier series, which accounts for seasonal cycles. The hourly demand model is fed from the daily model and composed of periodic and persistence components. Herrera et al. (2010) compared various hourly demand prediction models for a city in southeastern Spain. The data set included historic hourly and daily demands and explanatory climate variables (e.g., temperature, wind speed, and rain). The compared prediction models include artificial neural networks, projection pursuit regression (PPR), multivariate adaptive regression splines (MARS), support vector regression (SVR), and random forests. The common requirement for these models is an offline tuning stage using long history data sets. The previous demand forecast models all require a long demand data history, which is not available in some cases, and mostly do not account for recent changes in demand patterns. In a more recent study, Pacchin et al. (2017) suggested a simpler hourly demand forecast

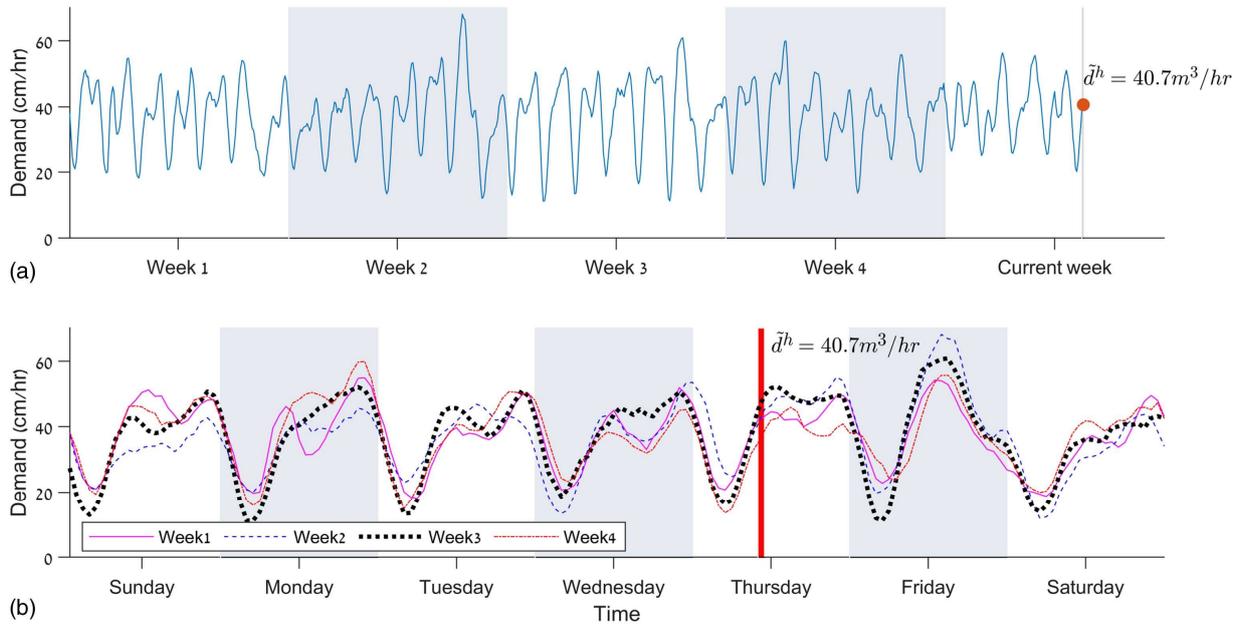


Fig. 2. Historic demand data.

model that consists of two steps. First, the total demand for a predefined forecasting horizon is estimated, and then the hourly pattern over this time window is predicted. The total demand estimation is based on the previous day with an adjustment coefficient, whereas the hourly pattern is based on the weighted patterns of the same type of days in the previous weeks. This kind of models is appealing as a practical prediction model for real-time WDS operation because it requires only a limited amount of historic data and has ease of implementation and low computation complexity. In addition, from a practical point of view, such a simple forecasting model is easier to implement and maintain compared to different machine learning algorithms such as ANNs. Inspired by the work of Pacchin et al. (2017), for the purpose of this research, we propose the naïve demand forecasting (NDF) model, which also requires limited historic demand data (e.g., 4 weeks) and is fast and simple to calculate. The NDF model is solely based on the average demand in the same hour in previous weeks. Let  $h$  be the current absolute time denoted in hours passed from a given base reference (e.g., hours from the beginning of the year, decade, or any other given time). Given  $h$ , the demand prediction for the next hour is given by Eq. (1)

$$\tilde{d}^h = \frac{1}{w} \sum_{i=1}^w d^{h-168i} \quad (1)$$

where  $w$  = number of historic weeks considered;  $d$  = historic demand;  $\tilde{d}$  = forecast demand; and 168 = number of hours in a week. To demonstrate the NDF prediction model, a real-world historic demand time series is given in Fig. 2(a). It includes 4 full historic weeks, 4 days of the current week, and 9 h of the current day. The end of this time series is the current time, denoted as  $h$ . Fig. 2(b) presents the same 4 historic weeks of hourly demand data over a weekly cycle; that is, each series presents 1 historic full week. The daily pattern can easily be observed with a morning and afternoon peak. A distinctive different demand pattern can easily be observed for Friday and Saturday. These different patterns can be explained by the different holidays of the mixed Jewish and

Muslim population of the city at hand. The predicted demand using the NDF model for time  $h$ ,  $\tilde{d}^h$ , is shown in Fig. 2(a). This prediction is merely the average of the same hour demand in the previous 4 weeks, as shown in Fig. 2(b) by the vertical line on Thursday.

Similarly, the NDF model may be used to forecast the demands for a longer time period as required by the WDS optimization model. Considering the operation window length  $T$  (e.g., optimization horizon of 48 h) and a time step  $\Delta t$  (e.g., 1 h), at each time  $t \in \tau_h \equiv \{h, h + \Delta t, \dots, h + \Delta t[(T/\Delta t) - 1]\}$  a demand forecast is required; thus, Eq. (1) takes a time series form as described in Eq. (2)

$$\tilde{d}^t = \frac{1}{w} \sum_{i=1}^w d^{t-168i} \quad \forall t \in \tau_h \quad (2)$$

In consistency with Pacchin et al. (2017), which inspired the development of the NDF model, we evaluate the performance of the NDF model by using the mean absolute error (MAE%) as given in Eq. (3)

$$\text{MAE}\% = \frac{100\%}{n} \sum_{t \in \tau_h} \left| \frac{d_{\text{obs}}^t - \tilde{d}^t}{d_{\text{obs}}^t} \right| \quad (3)$$

where  $d_{\text{obs}}^t$  = observed demand at time  $t$ ;  $\bar{d}_{\text{obs}}$  = average of the observed demands; and  $n$  = number of time steps in the forecasting horizon,  $T$ .

### Optimization Problem Formulation

The most general formulation of the WDS operation problem, which fully represents the hydraulics of the system, is mixed-integer nonlinear programming (MINLP) such as the considered in Biscos et al. (2003) and Vigerske et al. (2012), where the total pumping cost is minimized. The integer variables are associated with the pumps and valves in the systems that have an on and off state (open and close for valves). The nonlinearity is a result of energy balance equations such as the Hazen–Williams equation

and the pump and valve curves. However, despite their generality, these formulations are difficult to solve for global optimality in large networks. In many cases, the optimization problem may be formulated as linear programming (LP) or mixed-integer linear programming problems by removing the explicit hydraulics formulation. This optimization scheme is relevant when the pumps are working against a point with relatively constant head [e.g., when the static head is significantly larger than the dynamic head, as Housh and Salomons (2019) demonstrated]. The nodal pressure requirements in the network are assumed to be satisfied if the water level in the tanks is within its operational levels (Ormsbee and Lansley 1994), given that the network is properly designed (in terms of pipe sizing). Jowitt and Germanopoulos (1992) also noted that the LP and the MILP formulation may be applied when the flow and power in the pumping stations are not affected by pump and valve controls in the network. The real-world test case demonstrated in this study satisfies the previous requirements. Other real-world networks, such as C-town (Ostfeld et al. 2012), hold the same properties. However, the LP formulation, such as in Jowitt and Germanopoulos (1992), may result in short pump operation time, which is undesirable, mainly because of mechanical reasons. With the previous assumption, we first formulate a full MILP problem.

The WDS considered in this study includes sources such as groundwater wells, fixed and variable speed pumps, tanks, and junctions. Each junction may be an aggregation of a group of real network junctions. At time  $h$ , the optimization problem aims at determining how to operate the WDS for the next time step  $\Delta t$  (e.g., optimization time step of 1 h). Nevertheless, to prevent myopia in the solution process, the first decision must account for the conditions (e.g., energy costs and demands) in a predefined future operation horizon. As such, the optimization problem should be solved for an operation horizon  $T$  (e.g., optimization horizon of 48 h), although only the first time step decision will be executed in practice. Specifically, at each time  $t \in \tau_h \equiv \{h, h + \Delta t, \dots, h + \Delta t[(T/\Delta t) - 1]\}$  of the operation horizon  $T$ , the state (i.e., on or off) of each fixed speed pumping station (FPS) and the flow for each variable speed pumping station (VPS) should be determined. The sets of pumps are defined by  $\text{FPS} = \{\text{FPS}_1, \dots, \text{FPS}_{n_{\text{FPS}}}\}$  and  $\text{VPS} = \{\text{VPS}_1, \dots, \text{VPS}_{n_{\text{VPS}}}\}$ , where  $n_{\text{FPS}}$  and  $n_{\text{VPS}}$  are the number of fixed and variable speed pumping stations in the network, respectively.

For each fixed speed pumping station  $p \in \text{FPS}$ , there is a set  $S_p$  of operation states that represent any combination of pump units, including the off state in which no pumping units are working. Each of these states  $s \in S_p$  is accompanied by its flow  $Q_s$  and power  $P_s$ . The values  $Q_s$  and  $P_s$  are not decision variables but rather given parameters that could be determined by analyzing the operation points of different pump unit combinations in the fixed speed pumping stations. As opposed to the fixed speed stations, in the variable speed pumping stations  $p \in \text{VPS}$ , the flow  $Q_p$  is a continuous decision variable, and the power is a function of the station's total flow  $P_p(Q_p)$ .

Given the above, the objective of the optimization problem is defined in Eq. (4), in which the energy cost is minimized over the operation horizon  $T$

$$\min \text{obj} = \sum_{t \in \tau_h} \text{EC}^t \cdot \Delta t \left( \sum_{p \in \text{FPS}} \sum_{s \in S_p} P_s \cdot I_s^t + \sum_{p \in \text{VPS}} P_p(Q_p^t) \right) \quad (4)$$

$$I_s^t \in \{0, 1\} \quad \forall t \in \tau_h, \forall s \in S_p, \forall p \in \text{FPS} \quad (5)$$

where  $\text{EC}^t$  = electricity cost at time  $t$ ;  $I_s^t$  = binary decision variable determining if a specific operation state  $s$  of a fixed speed pumping station is operating at time  $t$  or not; and  $Q_p^t$  = continuous decision

variable determining the flow rate of the variable speed pumping station  $p$  at time  $t$ .

The previous objective function is subject to a set of physical and operational constraints. At each time step, for each FPS, only one operating state could be selected; see Eq. (6). That is, logically, only one combination of pumping units could be chosen to switch on during the time step. For example, for a pumping station with two fixed pumping units, four states are defined: State #1 for which no pump is on, State #2 in which only pump #1 is on, State #3 in which only pump #2 is on, and State #4 in which both pumps are on. The condition in Eq. (6) specifies that only one of these possible realizations can exist at the same time

$$\sum_{s \in S_p} I_s^t = 1 \quad \forall t \in \tau_h, \forall p \in \text{FPS} \quad (6)$$

The mass balance conservation law must be satisfied for each junction  $j \in J$  of the network junctions set  $J$ . For each junction  $j$ , we define sets of ingoing and outgoing flows from fixed speed pumps ( $\text{FPS}_j^{\text{in}}$ ,  $\text{FPS}_j^{\text{out}}$ ), variable speed pumps ( $\text{VPS}_j^{\text{in}}$ ,  $\text{VPS}_j^{\text{out}}$ ), flows to and from operational tanks ( $R_j$ ), and a set of supplied demand zones ( $D_j$ )

$$\begin{aligned} \sum_{p \in \text{FPS}_j^{\text{in}}} \sum_{s \in S_p} Q_s \cdot I_s^t + \sum_{p \in \text{VPS}_j^{\text{in}}} Q_p^t + \sum_{r \in R_j} Q_r^t - \sum_{p \in \text{FPS}_j^{\text{out}}} \sum_{s \in S_p} Q_s \cdot I_s^t \\ - \sum_{p \in \text{VPS}_j^{\text{out}}} Q_p^t = \sum_{i \in D_j} d_i^t \quad \forall t \in \tau_h, \forall j \in J \end{aligned} \quad (7)$$

where  $Q_r^t$  = decision variable for the flow from (positive) or to (negative) tank  $r$  at time  $t$ ; and  $d_i^t$  = demand in demand zone  $i$  at time  $t$ .

In addition to the spatial mass balance at each time step as given in Eq. (7), the mass balance for the tanks must be ensured over time, as specified in Eq. (8). Moreover, the water volume in each tank should be kept within a predefined operational range, as detailed in Eq. (9). Usually, the maximum volume denotes the physical capacity of the tank, whereas the minimum volume requirement is to account for emergency capacity such as fire prevention regulations and other reliability considerations

$$V_r^{t+\Delta t} = V_r^t - Q_r^t \cdot \Delta t \quad \forall r \in R, \forall t \in \tau_h \quad (8)$$

$$\underline{V}_r^t \leq V_r^t \leq \bar{V}_r^t \quad \forall r \in R, \forall t \in \tau_h \quad (9)$$

where  $V_r^t$  = water volume in tank  $r$  at time  $t$ ;  $R$  = set of all tanks in the system; and  $\underline{V}_r^t$  and  $\bar{V}_r^t$  = minimum and maximum volumes at time  $t$ , respectively.

Typically, variable speed pumps are designed to operate in wide ranges of flow; still, this flow is bounded by minimum and maximum flow rates specified by the installed equipment. These bounds are defined in Eq. (10)

$$\underline{Q}_p^t \leq Q_p^t \leq \bar{Q}_p^t \quad \forall t \in \tau_h, \forall p \in \text{VPS} \quad (10)$$

where  $\underline{Q}_p^t$  and  $\bar{Q}_p^t$  = minimum and maximum flow rates, respectively.

In some practical cases, pumping stations have limits on the amount of water they are allowed to pump in a specific period of time. For example, such a scenario may arise when the daily capacity of the water treatment plant is below the daily capacity of the pumping station. Another practical example is when water wells have restrictions on water volume production because of hydrological withdrawing constraints. Eqs. (11) and (12) constrain the total amount of water that FPS and VPS can pump in pre-specified time periods, respectively

$$\underline{Q}_{p,i}^{\text{total}} \leq \sum_{t \in K_i} \sum_{s \in S_p} Q_s \cdot I_s^t \cdot \Delta t \leq \bar{Q}_{p,i}^{\text{total}} \quad \forall i \in \text{FTS}_{h,p}^{\text{total}}, \forall p \in \text{FPS} \quad (11)$$

$$\underline{Q}_{p,i}^{\text{total}} \leq \sum_{t \in K_i} Q_p^t \cdot \Delta t \leq \bar{Q}_{p,i}^{\text{total}} \quad \forall i \in \text{VTS}_{h,p}^{\text{total}}, \forall p \in \text{VPS} \quad (12)$$

where  $\text{FTS}_{h,p}^{\text{total}}$  = time segments (TSs) defined at time  $h$  for the total volume constraints;  $K_i$  = times in the  $i$ th TS; and  $\underline{Q}_{p,i}^{\text{total}}, \bar{Q}_{p,i}^{\text{total}}$  = minimum and maximum volume constraints for pumping station  $p$  in the  $i$ th TS, respectively. For example, if, for a pumping station, there are two time segments with volume restriction, say, hours 2–6 and hours 10–14, then we define a set  $K_{2-6}$  that includes all hours between 2 and 6 and a set  $K_{10-14}$  that includes all hours between 10 and 14.

Frequent flow rate changes between consecutive time steps are in some cases restricted to avoid mechanical and water quality issues (Housh and Salomons 2019). Eqs. (13) and (14) constrain the change of the flow rate that FPS and VPS can have between consecutive time steps, respectively

$$\Delta \underline{Q}_{p,i} \leq \sum_{s \in S_p} Q_s^t [I_s^t - I_s^{t-\Delta t}] \leq \Delta \bar{Q}_{p,i} \quad (13)$$

$$\forall t \in K_i, \forall i \in \text{TS}_{h,p}^{\text{diff}}, \forall p \in \text{FPS}$$

$$\Delta \underline{Q}_{p,i} \leq Q_p^t - Q_p^{t-\Delta t} \leq \Delta \bar{Q}_{p,i} \quad \forall t \in K_i, \forall i \in \text{TS}_{h,p}^{\text{diff}}, \forall p \in \text{VPS} \quad (14)$$

where  $\text{TS}_{h,p}^{\text{diff}}$  = time segments defined at time  $h$  for the change in the flow rate constraints; and  $\Delta \underline{Q}_{p,i}$  and  $\Delta \bar{Q}_{p,i}$  = minimum and maximum flow rate change constraints for pumping station  $p$  in the  $i$ th TS, respectively.

Pumping stations consume energy for their operation that is supplied by power stations. Power stations can be a connection to the electricity grid, a power generator, or any other type of energy supply. The power supplied to the pump stations is limited by power station capacity or prespecified contracts with water utilities. Such constraints related to power supply are imposed in Eq. (15), which limits the simultaneous operation of a predefined set of pumping states. Several pumping stations may be connected to one power station, and the constraint applies to their total power consumption

$$\sum_{s \in S_m} I_s^t \leq 1 \quad \forall t \in K_i, \quad (15)$$

$$\forall i \in \text{TS}_{h,m}^{\text{power}}, \forall m \in \text{POS}$$

where POS = set of all power stations in the network;  $S_m$  = set of pumping states that are not allowed to operate simultaneously; and  $\text{TS}_{h,m}^{\text{power}}$  = time segments defined at time  $h$  for power station constraints.

The previous detailed optimization problem [Eqs. (4)–(15)] is formulated as a MILP. The number of integer variables in the problem is determined by the number of states for each pumping station multiplied by the number of the time steps within the optimization horizon  $T$ . Thus, for large networks with a large number of time steps in the optimization horizon, solving the MILP to global optimality is not guaranteed given the limited central processing unit (CPU) time in real-time operation of the system (even in the mid-size case study herein, 50% of the cases will exceed a runtime of 5 min). To cope with this challenge, we propose a reduced approximated formulation of the original MILP. We hypothesize that for real-time receding horizon control loop implementation (where efficiency is important), such a reduced MILP model will yield to an

efficient and practical solution algorithm that results in a near-optimal solution. As such, we emphasize the overall performance of the control process rather than the accuracy of the optimization problem as usually done in the literature. That is, we would prefer simpler and practical methodologies that together yield good (near-optimal) results over complicated components that might add small benefits for a large price.

The reasoning behind this hypothesis is that, as described earlier, only the first time step of the operation plan is implemented; thus, the investment and efforts put into the optimality search are not fully utilized. The real decision we are looking for is the first time step decision; we only account for a longer operation horizon to ensure that the optimization problem is not myopic, so that the obtained first time step decision takes into account the future conditions. As such, when adapting a receding horizon control loop framework, any MILP approximation that yields a first time step decision equal to the optimal one in the original MILP should be equally good when implemented in a real-life system.

To derive the MILP approximation, we propose a reduced-size MILP in which only part of the binary decisions  $I_s^t$  is required to be binary. Specifically, we require that only the binary decision variables in the near future should be restricted to the set  $\{0, 1\}$ , whereas the far-future variables are relaxed to continuous variables between 0 and 1. By doing so, we still include the far future in the operation horizon to prevent myopia of the solution, but far-future decisions are modeled in an imperfect way (i.e., relaxed to continuous) relying on the reasoning described previously. In such a framework, the relaxed binary variables could be interpreted as the fraction of time (out of the time step length) in which a pumping station state is selected. In light of the previous, the MILP formulation in the previous section should be modified by replacing Eq. (5) with Eqs. (16) and (17). Eq. (16) denotes the binary decision variables for the first time steps, whereas Eq. (17) sets the continuous variables for the rest of the optimization horizon

$$I_s^t \in \{0, 1\} \quad \forall t \in \tau_h^{\text{binary}}, \forall s \in S_p, \forall p \in \text{FPS} \quad (16)$$

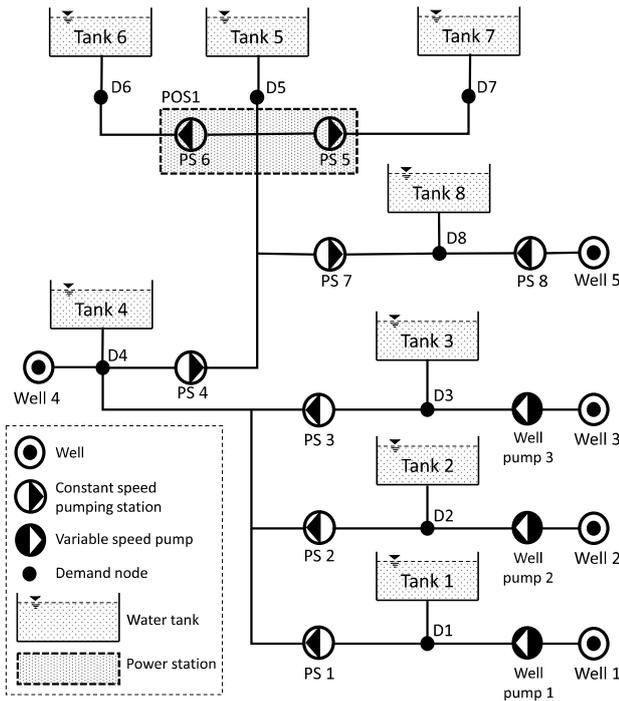
$$0 \leq I_s^t \leq 1 \quad \forall t \in \tau_h - \tau_h^{\text{binary}}, \forall s \in S_p, \forall p \in \text{FPS} \quad (17)$$

where  $\tau_h^{\text{binary}}$  = set of times representing the near future in which the decision variables are restricted to be binary. The size of the set  $\tau_h^{\text{binary}}$  defines the binarization level in the reduced MILP, such that binarization level 1 is when  $\tau_h^{\text{binary}}$  includes the first time step only, whereas binarization level 2 is when  $\tau_h^{\text{binary}}$  includes the first two time steps, and so on.

## Case Study

The test case is based on a real-world network introduced in Seleky et al. (2012). The network is a regional water network of Sopron, a city in Hungary near the Austrian border with a population of about 62,000 inhabitants (2017). The layout of the network and its connectivity are shown in Fig. 3. The network consists of five wells, three variable speed pumping stations, five constant speed pumping stations, eight tanks, and eight junctions, of which five are demand nodes. The aim is to minimize the operation costs of the system while satisfying all the constraints discussed previously. The operational cost is a function of power consumption of the pumps and the electricity tariff. The electricity tariff consists of two periods: low tariff with a cost of 1 €/kW · h for hours 0–7, 13–17, and 20–24 and peak tariff with a cost of 1.25 €/kW · h for hours 7–13 and 17–20.

Wells 1–3 are equipped with variable speed pumps that are limited in their maximum hourly flow. In addition, these wells have daily minimum and maximum production volumes, as shown in



**Fig. 3.** Case study network layout.

**Table 1.** Parameters of the wells in the case study

Well	$\bar{Q}$ (m <sup>3</sup> /h)	$\bar{Q}_{total}$ (m <sup>3</sup> /day)	$\bar{Q}_{total}^{diff}$ (m <sup>3</sup> /day)	$Q_{initial}$ (m <sup>3</sup> /h)
1	270	3,000	6,000	253
2	250	1,000	3,000	72
3	460	5,000	11,000	179
4	66	1,584	1,584	66
5	$\infty$	0	$\infty$	—

Note:  $Q_{initial}$  is the initial well flow at the beginning of the 720-h time horizon.

Table 1. In a receding horizon operation mode, such as in our case, daily constraints are handled by adjusting the parameters of the constraint in Eq. (12) according to the pumped water up to the current time. Specifically, in Eq. (12),  $VTS_{h,p}^{total}$  changes for every  $h$ , and  $\bar{Q}_{p,i}$  and  $\bar{Q}_{total}$  are adjusted according to the pumped water during the part of the day before the current time step. Because of operational and hydrological constraints, the flow of Wells 1–3 may be changed only a limited number of times during the day. As such, the operators of the network decided that these changes in the wells' flow will occur only when the electricity tariff changes (i.e., hours 7, 13, 17, and 20 during each day). Using Eq. (14), this requirement could be easily adapted in our formulation by setting  $TS_{h,p}^{diff}$  to include all times except the tariff change times and setting  $\Delta\bar{Q}_{p,i} = \Delta\bar{Q}_{p,i} = 0$  during these times for the variable speed pump associated with Wells 1–3. Well 4 has a constant flow rate of 66 m<sup>3</sup>/h; this could be adapted by including the entire operation horizon (e.g., 48 h) in  $TS_{h,p}^{diff}$  and setting the allowed change to zero  $\Delta\bar{Q}_{p,i} = \Delta\bar{Q}_{p,i} = 0$ . Well 5 has no direct supply constraints and can supply any flow Pumping Station 8 requires. As such, it could be defined by setting  $TS_{h,p}^{diff}$  to an empty set.

There are eight pumping stations (PS) in the network (PS1–PS8). At each time step, the pumping stations may be operated

**Table 2.** Parameters of the pumping stations in the case study

Pumping station	State	$Q_s$ (m <sup>3</sup> /h)	$P_s$ (kW)
1	1	0	0
	2	150	50
	3	360	125
2	1	0	0
	2	110	5.5
	3	270	30
3	1	0	0
	2	500	60
	3	550	210
4	1	0	0
	2	66	18.7
	3	116	37.5
5	1	0	0
	2	66	3.8
	3	118	7.4
	4	148	10.6
6	1	0	0
	2	90	9.2
	3	114	9.2
7	1	0	0
	2	72	24
	3	130	46
8	1	0	0
	2	72	24
	3	130	46

**Table 3.** Parameters of the tanks in the case study

Tank	$V$ (m <sup>3</sup> )	$\bar{V}$ (m <sup>3</sup> )	$V^{init}$ (m <sup>3</sup> )
1	200	500	391
2	0	1,000	498
3	1,008	2,000	1,226
4	983.6	1,901.6	1,370
5	1,784	3,766	2,210
6	2,500	4,950	3,783
7	930	2,132	1,345
8	620.7	1,179	989

according to one pumping state from a set of possible states that represent a certain combination of pump unit operation. This set also includes a state in which all pumping units are off. For each pumping station, the possible pumping states are detailed in Table 2. Because of power limitations, Pumping Stations 5 and 6 together, which are connected to the same power station, may not consume more than 35 kW during high-tariff periods. This constraint means that State 3 of PS5 and State 4 of PS6 may not operate together during high-tariff periods because their combined power consumption is 37.5 + 10.6 = 48.1 kW, which exceeds the 35 kW limit. This restriction could be implemented using Eq. (15), where  $S_{POS1}$  includes these two pumping states and  $TS_{h,POS1}^{power}$  includes the high-tariff period at every  $h$ . All other state combinations are allowed in terms of power constraints. Table 3 summarizes the parameters of the eight operation tanks in the system.

### Methodology Implementation

The optimization horizon is set to  $T = 48$  h with a time step of  $\Delta t = 1$  h. This network has a daily demand pattern; to account for the daily pattern, many previous studies typically use a 24-h horizon with cyclic constraints on the water volumes for the tanks, aimed at preventing a myopic solution that does not account for an

extended operation horizon. However, there is no guarantee that a cyclic constraint will result in the optimal solution. Furthermore, in the case of multiple tanks, it is not clear why each tank should return to its initial volume. This becomes even more evident in the case of a fault in the system where the day starts with nonoptimal volumes in the tanks. In this case, it is not preferable to set these volume levels to the target volumes at the end of the optimization horizon. In this study, owing to the efficiency of the reduced MILP model, we allow for a longer horizon (e.g., 48 h) instead of the cyclic constraint as typically done in previous studies.

According to the process described in Fig. 1, at each time step, the water volumes from the previous time step are set as the tanks' initial volumes, and the demand prediction for the five demand nodes is made for the next 48 h. The reduced MILP is built along with the variable electricity tariff data and constraints as detailed previously. As described previously, this network has a daily operation cycle and demand pattern. As such, the reduced MILP formulation is tested with different binarization levels (i.e., different sizes of  $\tau_h^{\text{binary}}$ ), which range from Binarization Level 1, in which  $\tau_h^{\text{binary}}$  includes the first time step, up to Binarization Level 24, in which  $\tau_h^{\text{binary}}$  includes the next 24 time steps. Because the optimization solver may converge to multi-optimal solutions (i.e., different solutions with the same operation cost), and in order to minimize the number of switches between pumping states from the current time to the next time step, a discount of 1% in the electricity tariff was given to the current working states at the first time step in every optimization problem. For example, if one of the pumping states is working at the starting time of the optimization, and from the cost perspective it is indifferent whether the pump is on at the first time step or the second one, we introduce a small cost reduction of 1% in current states of the first time step, thus giving preference toward keeping the states unchanged. This procedure helps in avoiding undesired frequent pump switches. A discount of 1% is small enough not to interfere with the optimal solution; it is just to distinguish between multi-optimal solutions that result in the same objective value. This is because the difference in power between the different states is significantly above 1% (as shown in Table 3), and the difference in electricity tariffs is more than 25%. Once the optimization process is ended, the first time step of the obtained solution is implemented in the real network conditions by simulating its operation using the true demands (true demand is assumed unknown to the optimization model, which only uses the demand forecast) to obtain the simulated tanks levels that will be used as initial conditions for the next optimization problem. In this study, we consider a simulation of the real system with only water demand as uncertain. As such, the process dynamics are assumed perfectly known. However, in many applications, other uncertainties exist in the model, such as pump and pipe characteristics, that influence the physical behavior of the system. A more detailed model that takes these additional uncertainties into account could be developed (e.g., a full hydraulic model). Our future work will be extended to cover more uncertainties.

## Results

The proposed receding horizon control loop was tested for a period of 30 days (i.e., 720 h). The optimization program was built using MATLAB version R2018b, the YALMIP toolbox (Lofberg 2004), and the open-source CBC MILP solver (Forrest and Lougee-Heimer 2005) on a 64-bit Lenovo X1 ThinkPad with an Intel i7-7600U CPU @ 2.8 GHz and 16 GB of RAM. A series of runs with different sizes of  $\tau_h^{\text{binary}}$  were performed, varying from Binarization Level 1 (with binary variables for the first hour and continuous variables for the next 47 h) up to Binarization Level 24.

**Table 4.** Results for 1-month optimal control with different binarization levels

Binarization level	Total cost (Euro)	Cost error from TMC (%)
1	199,760	11.21
4	197,360	9.88
8	196,480	9.39
12	196,910	9.63
16	197,600	10.01
20	197,650	10.04
24	196,970	9.66

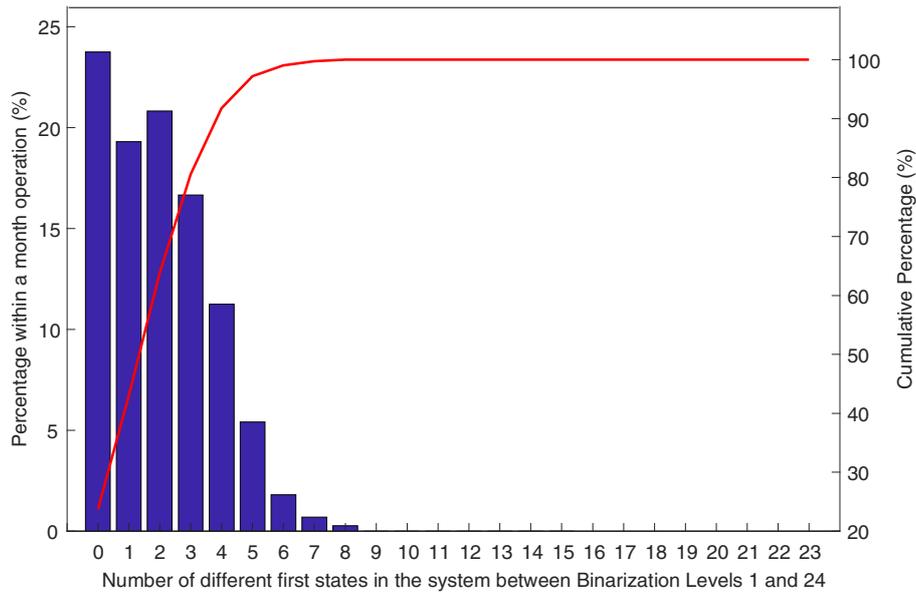
The total cost for Binarization Level 1 was 199,760 Euros, with a solution runtime of up to 1.6 s for each time step, whereas the total cost for Binarization Level 24 was 196,970 Euros, with a solution runtime of up to 1,000 s for each time step, which is the time limit for the optimization solver. A summary of the results is given in Table 4, in which the total cost is calculated according to the objective function of the optimization problem, Eq. (4).

Although we allowed the solver to run for 1,000 s (for comparison reasons), in practice, the solver must run in a shorter time (e.g., 5 min). Allowing the solver to run for 1 h and only then deploying the operational plan is not advisable because after 1 h, the system state will be changed significantly.

To evaluate the performance of the previous solutions, we compared them to the theoretical minimum cost (TMC). The TMC was obtained by solving the full MILP problem for 30 days with the true demands, which of course are unknown in the real case, thus making it a theoretical best benchmark, which we only used for comparison purposes. To obtain the TMC solution, a large-scale MILP problem of 16,560 binary variables was solved using the commercial CPLEX solver (IBM 2009). The solution was obtained in 4.5 h using the same hardware described previously and resulted in a total cost of 179,620 Euros.

As shown in Table 4, the errors obtained were between 9.4% and 11.2%, with negative correlation between binarization level and error. An error of order 10% when compared to the best benchmark solution (i.e., TMC), in which we solved a deterministic optimization problem for the entire horizon (i.e., 1 month) with perfect demand forecast, is not a high error. When comparing Binarization Level 1 with Binarization Level 24, which is typically adapted in the literature, we found that the model with Binarization Level 1 produced an error of 1.44% compared to the model with Binarization Level 24, whereas Binarization Level 4 produced an error as low as 0.2% compared to the model with Binarization Level 24.

Intuitively, one might expect that increasing the binarization level always improves the total cost. Nevertheless, our results show that this is not necessarily the case. The results show that, for the case study conditions, Binarization Level 8 produces the closest result to the best benchmark. To interpret this result, one should remember that any MILP with binarization level less than 720 (i.e., that does not include the entire operation horizon of 1 month with binary decisions) is in fact an approximation of the "true" optimization problem. At each time step, this approximation is solved for 48 h, but only the first time step's decision is implemented. As such, no matter how gross the approximation, if it produces first time step decisions (when implemented in a receding horizon mode) closer to TMC decisions, it will produce better results. For example, let us consider a reduced MILP with Binarization Level 1 versus Binarization Level 24. Intuitively, one may think that a model with Binarization Level of 24 that includes more



**Fig. 4.** Statistics for the number of different states in the first time step decisions as obtained from binarization levels of 1 and 24.

binary decisions is a better approximation. This is because it is closer to the true optimization problem in which all pumping decisions should be binary, as opposed to the case of Binarization level 1, in which only the first time step is binary and the rest of the time horizon is considered with continuous decisions. Nevertheless, recalling that only the first decision will be implemented from both approximations, the model with Binarization Level 1 might produce a first time step decision better than the one obtained from the model with Binarization Level 24. Indeed, each optimization problem in Binarization Level 1 is a reduced MILP; thus, it does not even produce feasible decisions after the first time step (infeasible in the sense that binary variables are treated as continuous ones), but what really matters when operating in receding horizon mode is the optimality of the first time step decisions.

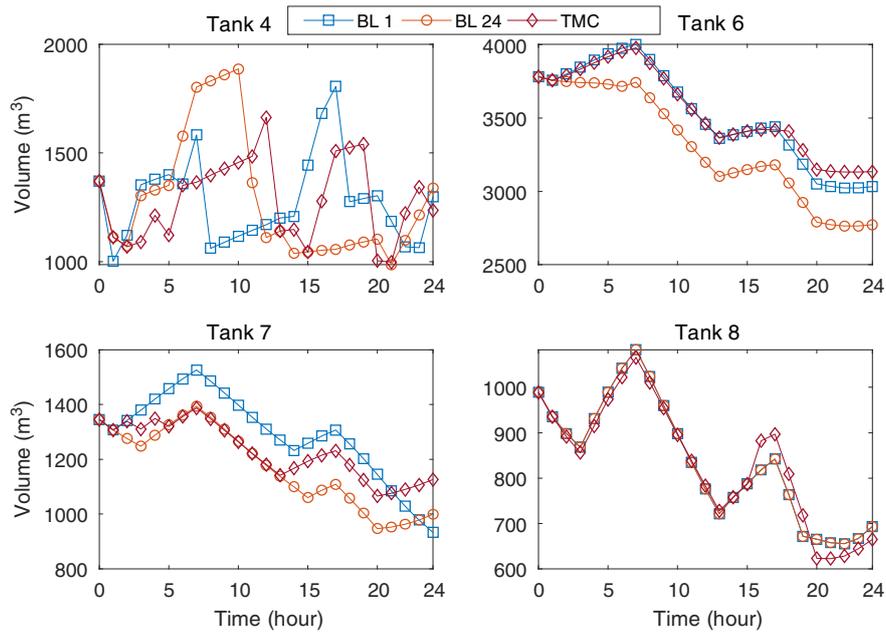
Acknowledging the importance of the first time step decision, next we compared the first time step decisions from Binarization Levels 1 and 24. Fig. 4 presents the percentage and cumulative percentage of the number of different states in the first time step decisions between the two binarization levels. The results show that in 23.75% out of the 720 optimization problems (i.e., solved for 720 h in a receding horizon model), both binarization levels produced the same pumping states, whereas in 19.31% of the cases, the two binarization levels only differed in one pumping state. The cumulative percentage graph in Fig. 4 shows that in 92% of the cases, there were only up to four different states between the two binarization levels. This similarity of the first time step decisions obtained from the very distinct binarization levels, explains how Binarization Levels 1 and 24 obtained very close total costs (i.e., Binarization Level 24 was closer to the TMC by only 1.5%).

Different binarization levels could have different impacts on different parts of the network. To illustrate these impacts, Fig. 5 presents tank volumes of four tanks in the system during the first 24 h for Binarization Levels 1 and 24 as compared to the TMC solution. In Tank 8, the two binarization levels and the TMC solution resulted in similar volume trajectories; as such, the two binarization levels had the same level of optimality in terms of Tank 8 operation. On the other hand, in Tank 6, the volume trajectory had the same trend in all models, but the model with Binarization Level

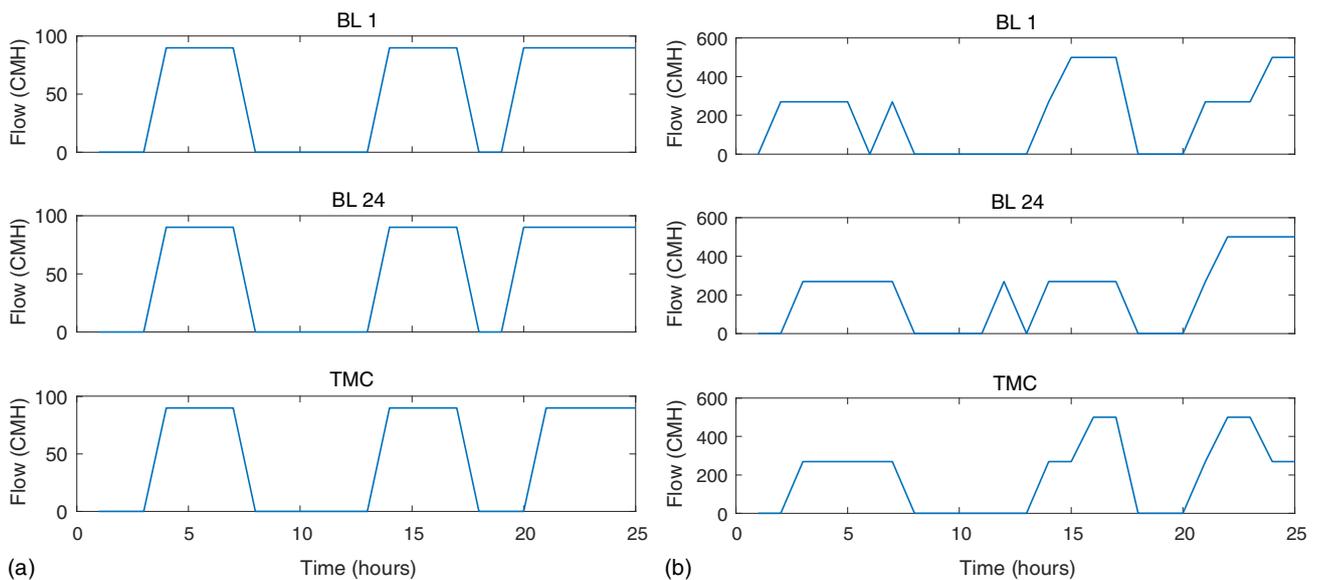
24 deviated from the TMC trajectory. As such, this demonstrates that in terms of Tank 6 trajectory, the higher binarization level is in fact inferior. In Tank 7, the model with Binarization Level 24 produced a closer trajectory to the TMC trajectory, whereas in Tank 4 both binarization levels produced trajectories that were significantly different from the TMC trajectory. These results in Fig. 5 emphasize that there is no general rule on how the binarization level might impact different parts of the network and fortifies the fact that increasing the binarization level does not necessarily improve the optimality of the operation. The same behavior can be observed for pumping station flow, which is a function of the pumping state control variables. Fig. 6 shows pumping station flow trajectories for Binarization Levels 1 and 24 compared to the flows trajectories in the TMC solution for PS3 and PS7. For PS3 [Fig. 6(a)], the hourly flows were exactly the same for the two binarization levels, and both had a small change at Hour 19 compared to the TMC solution. On the contrary, for PS7 [Fig. 6(b)], there were noticeable differences between the two binarization levels and the TMC solution.

An important aspect of receding horizon operation mode is the optimization runtime. As the number of binary variables increases, the optimization solver runtime is expected to increase. However, the state of the system changes during the 1-h time step (e.g., consumers change their demands and the tanks' levels change); as such, it is important to reach the solution as quickly as possible. We set a reasonable limit of 5 min CPU time. The cumulative probability of the different solution runtimes is shown in Fig. 7. The results show that for Binarization Levels 1 and 4, over 99% of the runs end in less than 3.4 s. As the binarization level increases, the probability of reaching a solution within the practical time limit of 5 min decreases: 99%, 93%, 90%, 73%, and 52% for Binarization Levels 8, 12, 16, 20, and 24, respectively. That is, for this test case, using binarization level of more than 8 is not a practical methodology for real-time operations.

Clearly, there is a tradeoff between the binarization level and error from the best benchmark TMC solution. The tradeoff between the error and solution time is shown in Fig. 8. Every point in Fig. 8 presents the solution error and the 95th runtime percentile of a



**Fig. 5.** Tank volumes for Binarization Levels 1 and 4 compared to the TMC trajectories.

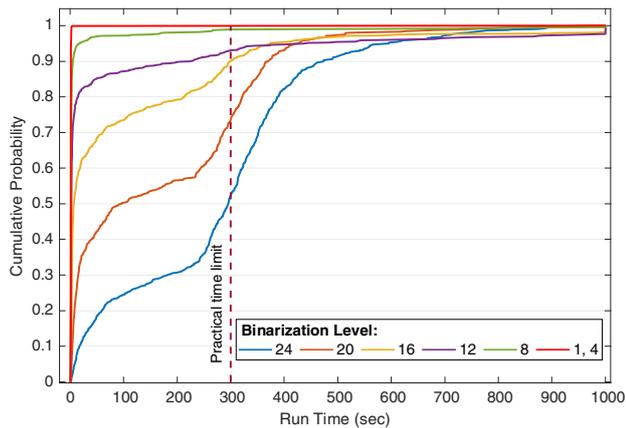


**Fig. 6.** Pumping station flows for Binarization Levels 1 and 24 compared to the TMC flows: (a) PS3; and (b) PS7.

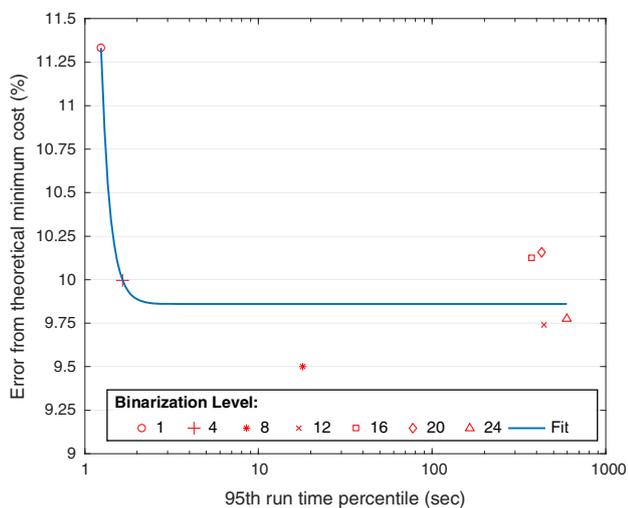
given binarization level. For example, for Binarization Level 1, the error from TMC is 11.21%, whereas 95% of the 720 runs performed in the receding horizon mode finished within 1.28 s. Fig. 8 also presents the best fit ( $y = 8.628 \cdot 10^{-8.274x} + 9.86$ ) between error and runtime, which shows a constant asymptotic relationship above Binarization Level 4. Indeed, with a binarization level of only 4, the obtained error compared to the best benchmark solution is almost equal to the obtained from Binarization Level 24 (9.88% versus 9.66%). The runtime for the model with Binarization Level 24, which is typically adapted in the literature, will exceed 595 s in

5% of the cases (Fig. 7) and will exceed the 5-min limit in 52% of the cases (Fig. 7). The model with Binarization Level 4 produces an error as low as 0.2% compared to the model with Binarization Level 24 and solves 95% of the cases under 1.65 s, with 100% of the cases within the 5-min limit (Fig. 7). This makes the model with Binarization Level 4 a good compromise for obtaining a practical model for the WDS at hand.

To evaluate the proposed naïve demand forecasting model performance, we considered 720 forecasts of 48 h each over a period of 32 days using a sliding window of length 48 h. To avoid seasonal



**Fig. 7.** Runtime cumulative probability for different binarization levels. (Binarization Levels 1 and 4 produce very close cumulative probability; as such, they are plotted as one line in this figure.)



**Fig. 8.** Solution errors from theoretical minimum cost versus runtime tradeoff.

change influences, we used a period in midwinter. The mean MAE% over all forecasts was 14.7%, with a minimum of 4.8% and a maximum of 26.3%. Fig. 9(a) presents the MAE% for all 720 forecasts with the minimum and maximum values highlighted. Figs. 9(b and c) show the observed and forecast demands for the minimum and maximum MAE%, respectively. Although these error values may seem high, we claim that, unlike previous studies, the impact of the demand forecast accuracy should be evaluated in view of the robustness of the applied control approach with respect to model uncertainties (i.e., the accuracy provided by NDF is sufficient for the proposed application). As such, we favor the simplicity of the prediction model over its accuracy as a separate block in the overall optimal control scheme. For our purpose, the NDF model yields acceptable forecasts with low computational burden and requires a small set of demand history data, making it a practical prediction model.

To examine the sensitivity of the obtained results to the accuracy of the demand forecast, we solved the reduced model with

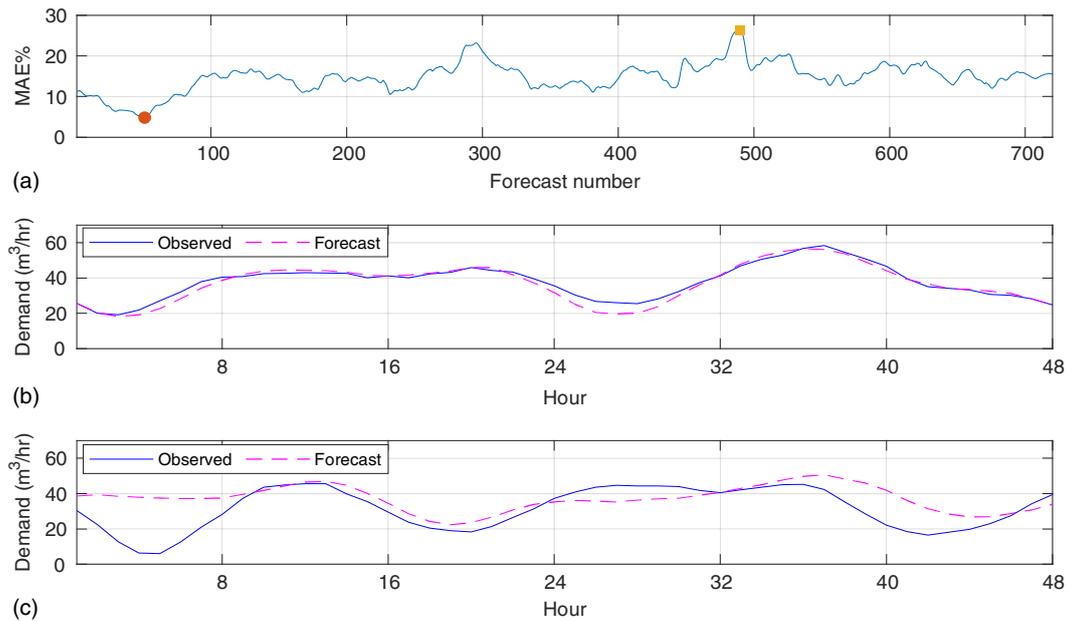
Binarization Level 1 using the true demand for every optimization problem in the receding horizon optimization (i.e., perfect demand forecasting). The total cost under this setting was 199,990 Euros, which was, surprisingly, 0.1% worse than the results obtained with the forecast demands. This is a counterintuitive result in which a perfect forecaster results in lower performance than a naïve forecaster when implemented in a feedback control loop. Nevertheless, this result fortifies our claim that the focus should be put on the overall performance of the system rather than a component of the control system (e.g., a forecaster in this case). This result could be explained by the fact that the optimization model does not solve the entire operation horizon of 30 days, but it solves pieces of 48 h each. Thus, in terms of the 30-day problem optimality, it is not guaranteed that solving the 48-h subproblems with perfect demand prediction will result in a better performance than solving the 48-h subproblems with imperfect demand prediction. Indeed, our results indicate that when solving with reduced MILP, imperfect demand prediction results in slightly better performance in terms of the 30-day cost. Of course, this is not a general conclusion, but it highlights the fact that emphasis on the demand prediction accuracy and optimization problem optimality, which have been the focus of many studies, is overrated if not tested within a complete control loop framework that resembles the real-life implementation of such systems.

## Conclusions

Although real-time WDS operation problem is a holistic control loop, many previous studies focused on different parts of the operation problem, such as demand forecasting accuracy and optimization problem optimality, rather than on the holistic control loop. Herein, we considered the entire problem in a closed control loop scheme and showed that emphasis on the demand prediction accuracy and optimization problem optimality is overrated if not tested within a complete control loop framework that resembles the real-life implementation of such systems.

The general real-time WDS operation problem requires a formulation of MINLP that is, for many real-world networks, beyond the capability of a real-time optimization framework, making it an impractical formulation. Nevertheless, in many cases, explicit hydraulic formulation can be neglected, and MILP approximation can be formulated. This study presented practical reduced MILP formulation to solve the optimization problem. Relying on the fact that in a receding horizon mode, only the first time step decisions will be implemented, our reduced MILP is configured by the binarization level, which is the number of first time steps that are modeled with binary decision variables. That is, we allow for binary decision relaxation in the far future (i.e., beyond the defined binarization level). We show that such reduced MILPs with small binarization levels can provide a practical solution strategy that balances solution accuracy and efficacy. The results show that different binarization levels do not significantly differ in cost. In contrast, increasing the binarization level increases the computational burden up to a point that exceeds a practical runtime limit imposed in real-time operation.

The obtained results are counterintuitive in two aspects. First, as the binarization level increases, that is, when the formulation is closer to the real MILP problem, better overall cost results are expected. However, the test case results show that this expectation is not always realized. In some cases, lower binarization levels yield better results than higher ones. This highlights the importance of first time step decision optimality, which is the only decision implemented when solving in receding horizon mode as expected in



**Fig. 9.** NDF model performance estimation: (a) MAE% for all forecasts; and (b and c) observed and forecast demands for the minimum and maximum MAE%, respectively.

real-life settings. This also indicates that optimization problem optimality, which has been the focus of many studies, is overvalued if not tested within a receding horizon mode. The results also counter-intuitively show that when testing this control scheme with perfect demand forecasting, using the true demands, the imperfect forecast slightly outperforms the perfect one. This lends support to the fact that demand forecast accuracy is not important by itself if not tested in the context of the entire control loop.

### Data Availability Statement

Some or all data, models, or code generated or used during the study are available from the corresponding author by request.

### Acknowledgments

This study was supported by the Israeli Water Authority (Contract number 4501284516).

### References

Alvisi, S., M. Franchini, and A. Marinelli. 2007. "A short-term, pattern-based model for water-demand forecasting." *J. Hydroinf.* 9 (1): 39. <https://doi.org/10.2166/hydro.2006.016>.

Biscos, C., M. Mulholland, M. V. Le Lann, C. A. Buckley, and C. J. Brouckaert. 2003. "Optimal operation of water distribution networks by predictive control using MINLP." *Water SA* 29 (4): 393–404. <https://doi.org/10.4314/wsa.v29i4.5044>.

Copeland, C., and N. T. Carter. 2017. *Energy–water nexus: The water sector's energy use R43200*. Washington, DC: Congressional Research Service.

Coulbeck, B., and C. H. Orr. 1989. "Real-time optimized control of a water distribution system." *IFAC Proc.* 22 (6): 441–446. <https://doi.org/10.1016/B978-0-08-037034-7.50080-8>.

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

Donkor, E. A., T. A. Mazzuchi, R. Soyer, and J. Alan Roberson. 2014. "Urban water demand forecasting: review of methods and models." *J. Water Resour. Plann. Manage.* 140 (2): 146–159. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000314](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000314).

Ellis, M., J. Liu, and P. D. Christofides. 2017. "Economic model predictive control." In Vol. 2 of *Encyclopedia of systems and control*, 1420–1430. Oxford, UK: Pergamon Press.

Forrest, J., and R. Lougee-Heimer. 2005. "CBC user guide." In *Emerging theory, methods, and applications*, 257–277. Aurora, CO: INFORMS.

Galupini, G., E. Creaco, C. Toffanin, and L. Magni. 2019. "Service pressure regulation in water distribution networks." *Control Eng. Pract.* 86 (May): 70–84. <https://doi.org/10.1016/j.conengprac.2019.03.007>.

Goldstein, R., and W. Smith. 2002. *Water and sustainability (volume 4): US electricity consumption for water supply and treatment—The next half century*. Technical Rep. 1006787. Concord, CA: Electric Power Research Institute (EPRI).

Grosso, J. M., C. Ocampo-Martínez, V. Puig, and B. Joseph. 2014. "Chance-constrained model predictive control for drinking water networks." *J. Process Control* 24 (5): 504–516. <https://doi.org/10.1016/j.jprocont.2014.01.010>.

Herrera, M., L. Torgo, J. Izquierdo, and R. Pérez-García. 2010. "Predictive models for forecasting hourly urban water demand." *J. Hydrol.* 387 (1–2): 141–150. <https://doi.org/10.1016/j.jhydrol.2010.04.005>.

Housh, M., and E. Salomons. 2019. "Optimal dynamic pump triggers for cost saving and robust water distribution system operations." *J. Water Resour. Plann. Manage.* 145 (2): 1–9. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001028](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001028).

IBM (International Business Machines). 2009. "V12. 1: User's manual for CPLEX." *Int. Bus. Mach. Corporation* 46 (53): 157.

Jamieson, D. G., U. Shamir, F. Martinez, and M. Franchini. 2007. "Conceptual design of a generic, real-time, near-optimal control system for water-distribution networks." *J. Hydroinf.* 9 (1): 3. <https://doi.org/10.2166/hydro.2006.013>.

Jamshidnejad, A., I. Papamichail, M. Papageorgiou, and B. De Schutter. 2016. "A model-predictive urban traffic control approach with a

- modified flow model and endpoint penalties." *IFAC-PapersOnLine* 49 (3): 147–152. <https://doi.org/10.1016/j.ifacol.2016.07.025>.
- Jowitt, P. W., and G. Germanopoulos. 1992. "Optimal pump scheduling in water-supply networks." *J. Water Resour. Plann. Manage.* 118 (4): 406–422. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1992\)118:4\(406\)](https://doi.org/10.1061/(ASCE)0733-9496(1992)118:4(406)).
- Lofberg, J. 2004. "YALMIP : A toolbox for modeling and optimization in MATLAB." In *Proc., 2004 IEEE Int. Conf. on Robotics and Automation (IEEE Cat. No.04CH37508)*, 284–289. Piscataway, NJ: IEEE. <https://ieeexplore.ieee.org/document/1393890>. <https://doi.org/10.1109/CACSD.2004.1393890>.
- Lougee-Heimer, R. 2003. "The common optimization interface for operations research: Promoting open-source software in the operations research community." *IBM J. Res. Dev.* 47 (1): 57–66. <https://doi.org/10.1147/rd.471.0057>.
- Mala-Jetmarova, H., N. Sultanova, and D. Savic. 2017. "Lost in optimisation of water distribution systems? A literature review of system operation." *Environ. Modell. Software* 93 (Jul): 209–254. <https://doi.org/10.1016/j.envsoft.2017.02.009>.
- Ocampo-Martinez, C., D. Barcelli, V. Puig, and A. Bemporad. 2012. "Hierarchical and decentralised model predictive control of drinking water networks: Application to Barcelona case study." *IET Control Theory Appl.* 6 (1): 62. <https://doi.org/10.1049/iet-cta.2010.0737>.
- Ormsbee, L. E., and K. E. Lansley. 1994. "Optimal control of water supply pumping systems." *J. Water Resour. Plann. Manage.* 120 (2): 237–252. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1994\)120:2\(237\)](https://doi.org/10.1061/(ASCE)0733-9496(1994)120:2(237)).
- Ostfeld, A., et al. 2012. "Battle of the water calibration networks." *J. Water Resour. Plann. Manage.* 138 (5): 523–532. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000191](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000191).
- Pacchin, E., S. Alvisi, and M. Franchini. 2017. "A short-term water demand forecasting model using a moving window on previously observed data." *Water* 9 (3): 172. <https://doi.org/10.3390/w9030172>.
- Rao, Z., and E. Salomons. 2007. "Development of a real-time, near-optimal control process for water-distribution networks." *J. Hydroinf.* 9 (1): 25. <https://doi.org/10.2166/hydro.2006.015>.
- Sanders, K. T., and M. E. Webber. 2012. "Evaluating the energy consumed for water use in the United States." *Environ. Res. Lett.* 7 (3): 034034.
- Selek, I., J. G. Bene, and C. Hs. 2012. "Optimal (short-term) pump schedule detection for water distribution systems by neutral evolutionary search." *Appl. Soft Comput. J.* 12 (8): 2336–2351. <https://doi.org/10.1016/j.asoc.2012.03.045>.
- Shamir, U., and E. Salomons. 2008. "Optimal real-time operation of urban water distribution systems using reduced models." *J. Water Resour. Plann. Manage.* 134 (2): 181–185. [https://doi.org/10.1061/\(ASCE\)0733-9496\(2008\)134:2\(181\)](https://doi.org/10.1061/(ASCE)0733-9496(2008)134:2(181)).
- Vigerske, S., W. Huang, H. Held, and A. Gleixner. 2012. "Towards globally optimal operation of water supply networks." *Numer. Algebra, Control Optim.* 2 (4): 695–711. <https://doi.org/10.3934/naco.2012.2.695>.
- Zhou, S. L., T. A. McMahon, A. Walton, and J. Lewis. 2002. "Forecasting operational demand for an urban water supply zone." *J. Hydrol.* 259 (1–4): 189–202. [https://doi.org/10.1016/S0022-1694\(01\)00582-0](https://doi.org/10.1016/S0022-1694(01)00582-0).

## **Paper III**

*Salomons, E.; Shamir, U.; Housh, M. "Optimization methodology for estimating pump curves using SCADA data". Water 2021, 56 (9), e2020WR027917.*

<https://doi.org/10.1029/2020WR027917>.

Journal Impact Factor: 2.544

5-Year Impact Factor: 2.709

Publication Status: Published (Accepted 19 February 2021)

## Statement of authorship - Paper III

### Principal Author

Title of Paper	Optimization methodology for estimating pump curves using SCADA data		
Publication Status	<input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication		
Publication Details	Salomons, E., Shamir, U., & Housh, M. (2021). Optimization Methodology for Estimating Pump Curves Using SCADA Data. Water 2021, Vol. 13, Page 586, 13(5), 586. <a href="https://doi.org/10.3390/w13050586">https://doi.org/10.3390/w13050586</a>		
Name of Principal Author (Candidate)	Elad Salomons		
Contribution to the Paper	Conceived the presented idea, developed the theory and performed the computations. Wrote the manuscript with support from the co-authors.		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Name and Signature	Elad Salomons 	Date	February 25 <sup>th</sup> , 2021

### Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- v. the candidate's stated contribution to the publication is accurate (as detailed above);
- vi. permission is granted for the candidate to include the publication in the dissertation

Name of Co-Author 1	Uri Shamir		
Contribution to the Paper	Verified the analytical methods, encouraged Elad Salomons to investigate the subject, and supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.		
Name and Signature	Uri Shamir 	Date	February 25 <sup>th</sup> , 2021
Name of Co-Author 2	Mashor Housh		
Contribution to the Paper	Verified the analytical methods, encouraged Elad Salomons to investigate the subject, and supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.		
Name and Signature	Mashor Housh 	Date	February 25 <sup>th</sup> , 2021

## Article

# Optimization Methodology for Estimating Pump Curves Using SCADA Data

Elad Salomons <sup>1</sup>, Uri Shamir <sup>2</sup> and Mashor Housh <sup>1,\*</sup>

<sup>1</sup> Department of Natural Resources and Environmental Management, University of Haifa, Haifa 3498838, Israel; selad@optiwater.com

<sup>2</sup> Civil and Environmental Engineering, Technion, Haifa 3200003, Israel; shamir@technion.ac.il

\* Correspondence: mhoush@univ.haifa.ac.il

**Abstract:** Water distribution systems (WDSs) deliver water from sources to consumers. These systems are made of hydraulic elements such as reservoirs, tanks, pipes, valves, and pumps. A pump is characterized by curves which define the relationship of the pump's head gain and efficiency with its flow. For a new pump, the curves are provided by the manufacturer. However, due to its operating history, the performance of a pump deteriorates, and its curves decline at an estimated rate of about 1% per year. Pump curves are key elements for planning and management of WDSs and for monitoring system efficiency, to determine when a pump should be rehabilitated or replaced. In practice, determining pump curves is done by field tests, which are conducted every few years. This leaves the pump's performance unmonitored for long time periods. Moreover, these tests often cover only a small range of the curves. This study demonstrates that in the era of IoT and big data, the data collected by Supervisory Control And Data Acquisition (SCADA) systems can be used to continuously monitor pumps' performance and derive updated pump characteristic curves. We present and demonstrate a practical methodology to estimate fixed and variable speed pump curves in pumping stations. The proposed method can estimate individual pump curves even when the measurements are given only for the pumping station as a whole (i.e., total flow, pumping station head gain). The methodology is demonstrated in a real-world case study of a pumping station in southern Israel.

**Keywords:** pump curves; water distribution systems modeling; pump monitoring by SCADA



**Citation:** Salomons, E.; Shamir, U.; Housh, M. Optimization Methodology for Estimating Pump Curves Using SCADA Data. *Water* **2021**, *13*, 586. <https://doi.org/10.3390/w13050586>

Academic Editor: Joaquim Sousa

Received: 25 January 2021

Accepted: 19 February 2021

Published: 24 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The flow of water requires energy. In nature, rivers flow from the mountains to the sea due to gravitational energy. Similarly, in some water distribution systems (WDSs), the flow of water in pipes is also governed by gravity. However, to supply water to higher areas and to overcome the energy losses in the water delivery system, additional energy must be added. This is the role of pumps, which convert electrical power to mechanical power, and then into hydraulic energy (head). Energy consumption for pumping water in distribution systems constitutes an important element of the overall energy budget. For example, the energy consumed for treating and pumping of water and wastewater in the United States in 2010 was estimated to be 4% of the total energy consumption [1,2]. In a survey conducted by Lam et al. [3], it was found that the energy used for water provision in 30 cities over 15 years, is up to 1 kWh per m<sup>3</sup> which amounts to an average annual per-capita energy use of 100 kWh. Pump performance and efficiency deteriorate over time [4]. It has been estimated that the energy use of a potable water pump will increase by about one percent per year [5]. The reasons and rates of performance degradation over time have been analyzed by Eaton et al. [6]. They conclude that the head at constant flow declines in a non-linear fashion, reaching 1% after two years of service, and growing to 10% after 9 years of service. They recommend that pump performance should be tracked over time but also point to the practical difficulties of doing this under field conditions.

Thus, monitoring the performance and condition of pumps is important. The operation of a pump is described by its characteristic curves that show the pump's head gain, power, and efficiency over a range of flow rates. For new pumps, these curves are supplied, together with the pump, by the pump's manufacturer. For pumps already installed and in operation, it is customary to perform field tests to evaluate their current performance curves [7]. For example, Israeli regulations require that a certified pump test should be performed once every 30 months or after 7500 h of operation (whichever comes later) [8]. According to these regulations, it is not permitted to operate a pump with an efficiency of less than 65%, or a well pump with an efficiency of less than 55%. The pump curves are used for monitoring its performance as well as a central component in hydraulic models of WDSs. These models are used for design, operation, and monitoring tasks in WDSs, including: pipes' sizing and expansion [9,10], network calibration [11], detecting cyber-attacks [12], optimizing pumps operation [13,14], water quality modelling [15,16], sensor placement [17], minimizing greenhouse gas emissions [18], and analysis of water hammer effects [19].

Obtaining the pump curve by a pump test is a simple procedure for a fixed speed pump (FSP) as the pump can be tested at a few operational points (flows and head), with different suction and/or discharge pressures. By utilizing flow, pressure and power measurements, the curves can be constructed. On the other hand, variable speed pumps (VSP) are used to maintain a desired flow or pressure. Therefore, it is not feasible to test the pump across its range of speeds, so they are usually tested at a single speed, which is its nominal (i.e., maximum) speed. VSPs are common because they provide a number of advantages [20] including: (a) the pump flow can change gradually and give the upstream process (e.g., treatment plant) time to adjust; (b) no water storage is required on the demand side, as the pump can adjust to changing demands while maintaining the required pressure in the demand zone; (c) the flow can be changed gradually to reduce water hammer, and (d) motor life can be extended since fewer starts and stops are needed [21].

On the other hand, Gottliebson et al. [20] argue that VSPs also have disadvantages compared to FSPs: (a) they are more expensive in both installation and maintenance; (b) they may be less efficient; (c) controlling a VSP is more complex, and (d) a VSP may not be suitable for flat H-Q system curves as high efficiency is difficult to maintain over the entire flow range. In spite of these disadvantages, VSPs are most popular in systems in which no water storage is available, and where there is a need to regulate the flow using a demand-following mechanism.

With the VSPs gaining popularity in practice, they have been modelled in most simulation software, such as EPANET [22], and their modelling and simulation continue to be an active research topic [23–26]. Many studies of VSPs address the operation of WDSs and optimization of pumps scheduling [27–31]. In a recent review, Wu et al. [32] report improved system efficiency due to the use of VSPs, as well as other benefits of increased levels of flexibility in controlling WDSs in real time. Lima et al. [33] suggested the use of VSPs as a tool to recover energy and reduce leakage in WDSs. Wu et al. [34] incorporated VSPs in the design stage of water networks and transmission lines. Huo et al. [35] explored the option of using VSPs in deep injection well systems. All these studies assumed a fully known pump curves for their VSPs. Thus, as the effort in VSPs modelling and simulation grows, the need for accurate representation of the VSP curves increases.

Although not explicitly mentioned in most published research, the performance of pumps, including VSPs, may be evaluated from the abundance of historical records which is often available in WDSs that have Supervisory Control And Data Acquisition (SCADA) systems [5,36]. To produce the FSP's curves, one needs the flow, suction and discharge pressures (in fact only the difference is needed), and power readings for the pump. For a VSP the speed readings (or the motor frequency) are also required. Due to budget restrictions, many water utilities do not install all the required sensors in their pumping stations. In some cases, flow meters are not installed for each individual pump and only a single flow measurement is available for the entire pumping station. In other cases, the

individual pump's speed is not recorded. Power measurements are missing in many cases for the individual pumps and, if at all, are available only for the entire station. Facing these problems in real systems motivates the development of a methodology for deriving pump curves under limited data availability.

The methodology presented in this study is designed to produce the pump's curves when some of the data are missing. The implementation is demonstrated on a pressure zone in the water supply system of Mey-Sheva, which is the largest water utility in southern Israel. The entire system contains 6 pumping stations, 11 water tanks, and serves a population of 143,000 in Be'er-Sheva and the adjacent town Ofakim. The system has 670 km of water pipelines, of which about 100 km are part of the pressure zone.

## 2. Methodology

The methodology for pump curve calculation is designed to estimate them using SCADA data. Following the problem statement, the procedures are developed for single and combination of FSPs, and for single and combination of VSPs.

### 2.1. Problem Statement

The objective is to determine the curves for individual pumps in a pumping station that contains pumps in parallel, using SCADA data. The pumps may operate alone or in various combinations with other pumps. Ideally, a pumping station that consists of  $n$  pumps,  $p \in Pumps$ , where  $Pumps$  is the set of pumps in the station, will have real-time measurements of each pump's flow, power, and on/off state, as well as the suction and discharge pressures of the station. However, due to budget limitations, individual pump flow and power are often unavailable and only the total station's flow is measured. In this study, we consider that the following data are available at multiple times: each pump's state,  $I_p$  (a binary variable where a value of 0 denotes that the pump is off and 1 when it is running), the total station's flow,  $Q_{obs}$ , and the suction and discharge pressures. The difference between the discharge and suction pressures is the station's head gain,  $H_{obs}$ . For each variable speed pump, its speed,  $n_p$ , is also recorded. All these values change over time and are available for each time step,  $t \in T$ , where  $T$  is a set of given historical time steps. We aim to estimate the pumps' curves in a quadratic form, which is a common form used to approximate the concave nonlinear association between head gain and pump flow [22,37].

$$H_p = a_p - b_p \cdot Q_p^2 \quad p \in Pumps \quad (1)$$

where,  $H_p$  and  $Q_p$  are the pump's head gain and flow, respectively.  $a_p$  (the shutoff head) and  $b_p$  are the function parameters.

### 2.2. Fixed Speed Pumps

We start with the simple case of a single fixed speed pump (Figure 1) operating alone. Using Equation (1) and the SCADA data, we calculate, in Equation (3), the head errors,  $e_H$ , between the estimated pump head,  $H_{est}$  (i.e., the head resulted from the H-Q curve, Equation (2)), and the observed head,  $H_{obs}$ . Since we are dealing with a single pump, the index  $p$  is omitted.

$$H_{est,t} = a - b \cdot (Q_{obs,t})^2 \quad \forall t \in T \quad (2)$$

$$e_{H,t} = H_{est,t} - H_{obs,t} \quad \forall t \in T \quad (3)$$

To estimate the parameters of Equation (1) the sum of the squares of  $e_H$  could be minimized, namely using least square regression. However, this method is sensitive to outliers in the SCADA data, and large deviations may bias unduly the fitting procedure. To avoid this deficiency of a least square regression, we use the least absolute errors (LAE) method in which the sum of absolute errors,  $E_H$ , as defined in Equation (4), is minimized.

$$E_H = \sum_{t \in T} |e_{H,t}| \quad (4)$$

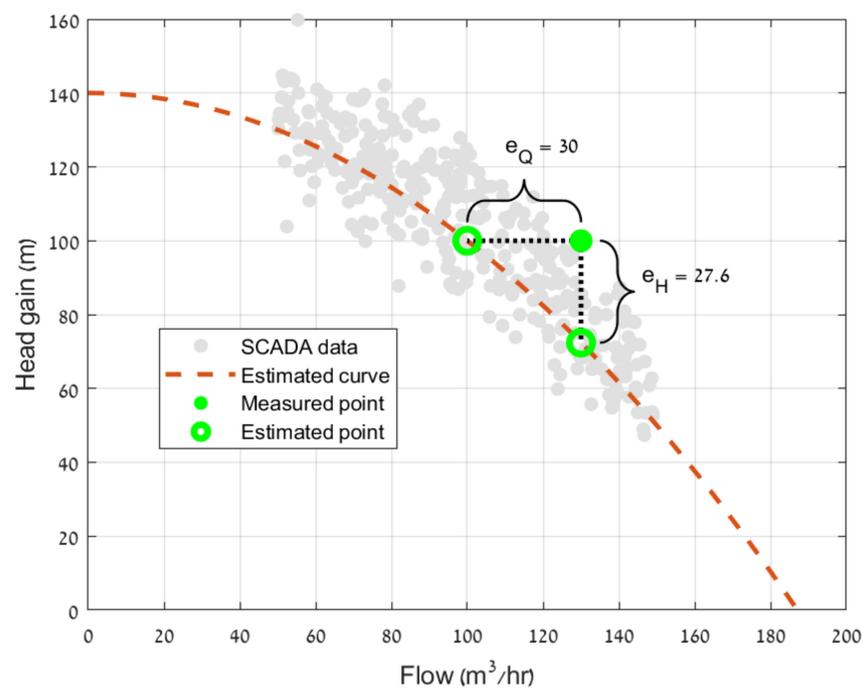
In the ordinary case of curve fitting, it is customary to assign the errors to vertical distances (i.e., the head in our context) between the observations and the assumed curve. However, in general, one can assign the errors to horizontal distances (i.e., the flow in our context) between the observations and the curve (Figure 1). Thus, the curve fitting process can also be defined on the flows, by minimizing  $E_Q$  as defined in Equations (5)–(7).

$$Q_{est,t} = \sqrt{\frac{a - H_{obs,t}}{b}} \quad \forall t \in T \quad (5)$$

$$e_{Q,t} = Q_{est,t} - Q_{obs,t} \quad \forall t \in T \quad (6)$$

$$E_Q = \sum_{t \in T} |e_{Q,t}| \quad (7)$$

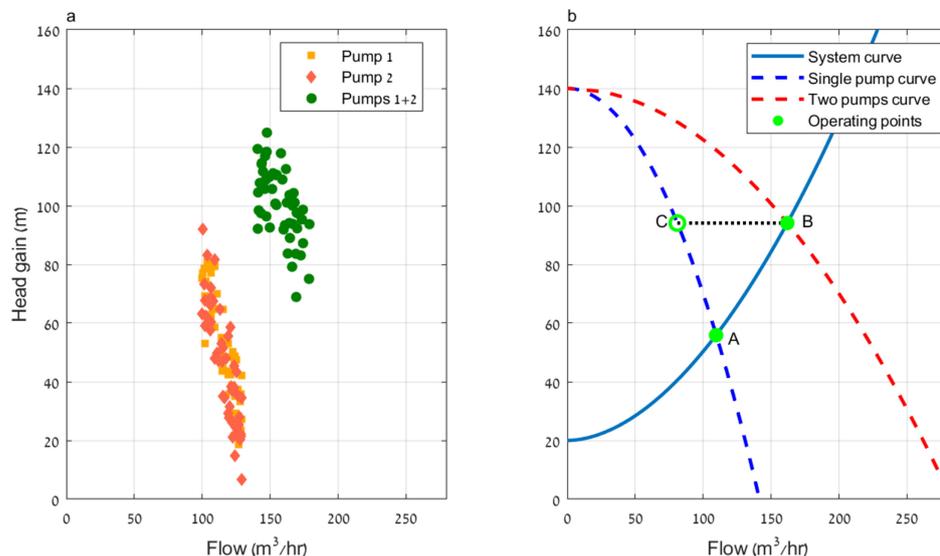
where,  $e_Q$  is the flow error,  $Q_{est}$  is the estimated flow,  $Q_{obs}$  is the observed flow, and  $E_Q$  is the sum of absolute flow errors. Figure 1 shows the vertical and horizontal distances of a measured point from the estimated pump curve. Generally, the estimated curve will not be the same for  $H$  and  $Q$  error minimization. While for a single fixed speed pump there is no clear advantage of one method over the other, and both vertical and horizontal errors can be easily calculated, the advantages of the horizontal calculations will become evident for more complicated cases that will be discussed next.



**Figure 1.** Vertical and horizontal errors between observations and assumed curve.

Pumping stations may have more than one pump operating in different combinations at different times. Consider, for example, a station with two identical pumps, Figure 2a shows the measured points of each pump alone and the two pumps working together. The case in which the two pumps operate together must be included in the procedure for deriving the curves for the individual pumps, since if only data for the pumps operating individually are considered, the curve's parameters may be biased because the entire operating region of the pump is not covered. To see this, consider the system curve in Figure 2b, when a pump operates alone it will be on the "right side" of its curve with high flow rates, as can be seen in Point A in Figure 2b. For two pumps operating together in parallel, the combined curve is obtained by summing the flows of the two pumps for each value of the head. The operating point in this case is Point B which is at the intersection of

the combined pump curve with the system curve. This operating point corresponds to a higher head compared to Point A, and thus each pump provides less flow at a higher head (Point C) than when it operates alone (Point A).



**Figure 2.** Two pumps operating in parallel (a) SCADA data (b) pumps and system curve.

This example demonstrates how individual pumps can work in different regions on the curve depending on the active pump combination. This behavior will be even more pronounced when there are several non-identical pumps in the pumping station with many combinations of operations. If we only use the data of a single pump operating alone (e.g., data near Point A) the estimated pump curve will tend to fit points in that region without considering the entire range of possible pump flows. To estimate the parameters of the individual curves based on the entire set of SCADA points (i.e., all combinations) there are two options:

Method 1

The combined curve for any specific combination can be derived as a function of the individual pumps’ parameters. For example, given two pumps with curves, Equations (8) and (9), the combined function must be written explicitly, as Equation (10), and curve fitting technique with vertical error,  $e_H$ , can be used to estimate the parameters for each of the two pumps curves.

$$H_1 = a_1 - b_1 Q_1^2 \tag{8}$$

$$H_2 = a_2 - b_2 Q_2^2 \tag{9}$$

$$H_{est} = f(a_1, b_1, a_2, b_2, Q_{obs}) \tag{10}$$

While this procedure is valid in theory, deriving an explicit analytical function, even for only two pumps and certainly for more pumps operating together, is not practical due to the nonlinearity of the curves. To derive the function  $f(\cdot)$ , Equations (8) and (9) are written for  $Q$  as given in Equations (11) and (12).

$$Q_1 = \sqrt{\frac{a_1 - H_1}{b_1}} \tag{11}$$

$$Q_2 = \sqrt{\frac{a_2 - H_2}{b_2}} \tag{12}$$

It should be noted that for any number of pumps operating in parallel, their head gain is equal, thus  $H_1 = H_2 = H_{est}$ . Summing the two flows yields the total flow of the pumps station as given in Equation (13)

$$Q_{obs} = \underbrace{\sqrt{\frac{a_1 - H_{est}}{b_1}} + \sqrt{\frac{a_2 - H_{est}}{b_2}}}_{g(a_1, b_1, a_2, b_2, H_{est})} \tag{13}$$

The function  $f(\cdot)$  is the inverse of the function  $g(\cdot)$  defined in Equation (13). Thus, even for the simple case of two pumps, one cannot derive an explicit form for the function  $f(\cdot)$ . At best, the procedure is tractable for two pumps, by a numerical solution of Equation (10) or (13) for given values of  $a_1, b_1, a_2, b_2$ . For a larger number of pumps this is not practical, since one needs to prepare in advance the combined functions for all possible combinations of pumps and then numerically solve all combination for each time step. Method 2 addresses this challenge by considering the horizontal error.

Method 2

Figure 3 demonstrates the calculation of the horizontal error for two pumps operating together. The observed flow,  $Q_{obs}$ , of the station is the combined flow of the two pumps, and  $H_{obs}$  is the measured head gain of the pumping station, which is equal for all pumps running in parallel. To calculate the horizontal error,  $e_Q$  in Equation (6), the estimated total flow,  $Q_{est}$ , is calculated based on the individual pumps' curves as given in Equation (14). That is, as shown in Figure 3, the observed head is used to estimate the flows from each pump,  $Q_{est,1}$  and  $Q_{est,2}$ , which are then summed as an estimate for the total flow of the station. Given this total flow estimate the horizontal error,  $e_Q$ , can be calculated.

$$Q_{est} = \sqrt{\frac{a_1 - H_{obs}}{b_1}} + \sqrt{\frac{a_2 - H_{obs}}{b_2}} \tag{14}$$

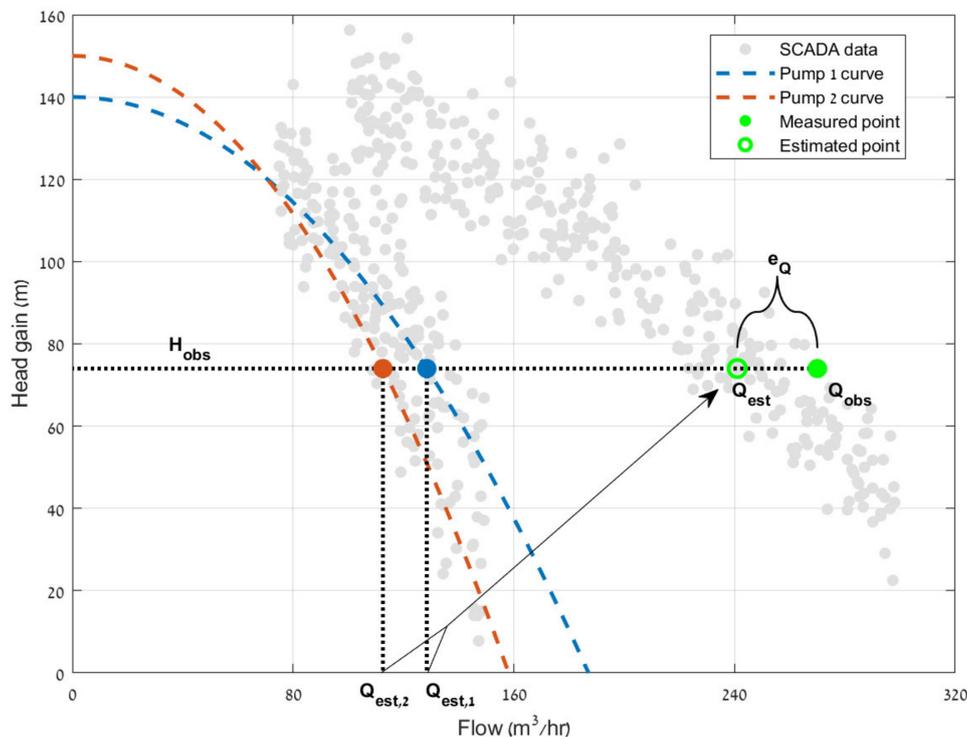


Figure 3. The horizontal error for two fixed speed pumps.

The advantage of this method is that it does not require an explicit derivation of the combined curve based on the individual curves' parameters (i.e., it does not require knowing the explicit function  $f(\cdot)$ ). In fact, unlike Method 1, which requires preprocessing the curves of all combinations, using Method 2 leads to a generic optimization problem, which can consider any arbitrary number of pumps, Equations (15)–(20).

$$\min_{a_p, b_p} \sum_{t \in T} |e_{Q,t}| \quad (15)$$

Subject to:

$$Q_{est,p,t} = \sqrt{\left| \frac{a_p - H_{obs,t}}{b_p} \right|} \quad \forall p \in Pumps, \forall t \in T \quad (16)$$

$$Q_{est,t} = \sum_{p \in Pumps} I_{p,t} \cdot Q_{est,p,t} \quad \forall t \in T \quad (17)$$

$$e_{Q,t} = Q_{est,t} - Q_{obs,t} \quad \forall t \in T \quad (18)$$

$$a_p \geq 0 \quad \forall p \in Pumps \quad (19)$$

$$b_p \geq 0 \quad \forall p \in Pumps \quad (20)$$

Equation (15) is the objective function which minimizes the summed absolute horizontal errors by deciding on the pumps' curves parameters  $a_p$  and  $b_p$ . In Equation (16) the estimated individual pumps flows,  $Q_{est,p,t}$  are calculated, and the total estimated flow for the pumping station is defined in Equation (17). It should be noted that the estimated flow is summed only for the pumps which are operating at each time by multiplying the individual pump's flow by its given binary state,  $I_{p,t} \in \{0, 1\}$ . In Equation (18) the horizontal error is calculated, and Equations (19) and (20) maintain the non-negativity of the parameters. While the absolute value in the objective function might be converted to set of linear constraints, the nonlinearity in the constraint in Equation (16) cannot be eliminated. Therefore, the obtained optimization problem is nonlinear. Readily available solvers such as `fmincon` shipped within Matlab [38] and the IPOPT [39] open source solver can handle this type of nonlinearity in the constraints. Moreover, since the number of decision variables is not expected to be high (e.g., a pumping station with 10 pumps will only result in 20 decision variables) and since this problem will be solved offline, one can utilize more computationally demanding global solvers such as Baron [40] to solve the optimization problem.

### 2.3. Variable Speed Pumps

A change in the rotational speed of a VSP changes its curves and operating point, according to the affinity laws:

$$\frac{Q_{p,t}}{\bar{Q}_{p,t}} = \frac{n_{p,t}}{\bar{n}_p} \quad \forall p \in Pumps, \forall t \in T \quad (21)$$

$$\frac{H_{p,t}}{\bar{H}_{p,t}} = \left( \frac{n_{p,t}}{\bar{n}_p} \right)^2 \quad \forall p \in Pumps, \forall t \in T \quad (22)$$

$$\frac{P_{p,t}}{\bar{P}_{p,t}} = \left( \frac{n_{p,t}}{\bar{n}_p} \right)^3 \quad \forall p \in Pumps, \forall t \in T \quad (23)$$

where,  $\bar{n}_p$  is the nominal speed of pump  $p$ , and  $n_{p,t}$ ,  $\bar{Q}_{p,t}$ ,  $\bar{H}_{p,t}$  and  $\bar{P}_{p,t}$  are the speed, nominal flow, nominal head gain, and nominal power for time  $t$  respectively. We examine first the case in which a single pump is operating. Its flow and head are that of the pumping station,  $Q_{p,t} = Q_{obs,t}$  and  $H_{p,t} = H_{obs,t}$ . Since  $\bar{n}_p$  is a known constant value (usually equivalent to the pump's speed at 50 Hz),  $\bar{Q}_{p,t}$  and  $\bar{H}_{p,t}$  can be calculated from Equations (21) and (22) for each time  $t$  according to the speed of the pump,  $n_{p,t}$ . Once these

values are calculated for different points in time, a curve fitting method can be employed to derive the pump’s curve (i.e.,  $\bar{H}_p = a - b \cdot \bar{Q}_p^2$ ) in the same way as for a fixed speed single pump (Section 2.2).

Thus, for times of a single operating pump we can easily map observed points to the nominal pump curve. However, as discussed previously, because the time instances of single pump operation are only a subset (sometimes only a small subset) of the entire operation, we may end up with a limited number of points on the curve, which causes a bias in the pump curve estimation process. This is especially true, when the points for the single pump operation fall on a narrow range of the pump curve. To overcome this bias, the pump’s operation with other pumps must be included.

Next, we derive the equations for fitting all the pumps’ curves simultaneously without restriction to specific time instances in which an individual pump is operating alone. That is, we also consider the time instances with joint operation of different pump combinations. From Equations (21) and (22) we obtain Equations (24) and (25).

$$\bar{Q}_{p,t} = Q_{p,t} \left( \frac{\bar{n}_p}{n_{p,t}} \right) \quad \forall p \in Pumps, \forall t \in T \tag{24}$$

$$\bar{H}_{p,t} = H_{p,t} \left( \frac{\bar{n}_p}{n_{p,t}} \right)^2 \quad \forall p \in Pumps, \forall t \in T \tag{25}$$

For VSPs we aim at fitting the nominal pumps’ curves as shown in Equation (26):

$$\bar{H}_{p,t} = a_p - b_p \cdot \bar{Q}_{p,t}^2 \quad \forall t \in T, p \in Pumps \tag{26}$$

Substituting Equations (24) and (25) in Equation (26) and extracting  $Q_{p,t}$  will yield a relationship between the individual pump heads and flows at any speed.

$$Q_{p,t} = \sqrt{\frac{\left( \frac{n_{p,t}}{\bar{n}_p} \right)^2 a_p - H_{p,t}}{b_p}} \quad \forall t \in T, p \in Pumps \tag{27}$$

Since the measured head is the same for all pumps,  $H_{p,t} = H_{obs,t} \forall p \in Pumps \forall t \in T$ , by substituting  $H_{obs,t}$  in Equation (27) we can obtain an estimate for the flows of the individual pumps.

Owing to the advantages of the horizontal error formulation, the optimization problem is formulated in Equations (28)–(33). The objective function, Equation (28) is minimized by deciding on the pump’s curve parameters  $a_p$  and  $b_p$ . The individual pumps’ flows are estimated in Equation (29). To obtain the total estimated flow of the pumping station, the individual pumps’ flows are summed in Equation (30) and the horizontal error can be calculated in Equation (31), while maintaining the non-negativity of the parameters in Equations (32) and (33).

$$\min_{a_p, b_p} \sum_{t \in T} |e_{Q,t}| \tag{28}$$

Subject to:

$$Q_{est,p,t} = \sqrt{\frac{\left( \frac{n_{p,t}}{\bar{n}_p} \right)^2 a_p - H_{obs,t}}{b_p}} \quad \forall t \in T, p \in Pumps \tag{29}$$

$$Q_{est,t} = \sum_{p \in Pumps} I_{p,t} \cdot Q_{est,p,t} \quad \forall t \in T \tag{30}$$

$$e_{Q,t} = Q_{est,t} - Q_{obs,t} \quad \forall t \in T \tag{31}$$

$$a_p \geq 0 \quad \forall p \in Pumps \tag{32}$$

$$b_p \geq 0 \quad \forall p \in Pumps \quad (33)$$

The process for calculating the horizontal error is shown in Figure 4. First, the normalized heads for the two pumps are calculated (at points A and B) using their operating speed by Equation (25). Then, using the normalized heads the normalized flows can be estimated by inverting Equation (26). Next, using each pump's speed, their estimated individual flows ( $Q_{est,1}$ ,  $Q_{est,2}$ ) are estimated (at Points C and D) by inverting Equation (24). In the optimization model, these operations are lumped together in Equation (29). Finally, the estimated flows are summed in Equation (30) to obtain the total estimated flow  $Q_{est}$ , which is then used in calculating the horizontal error in Equation (31).

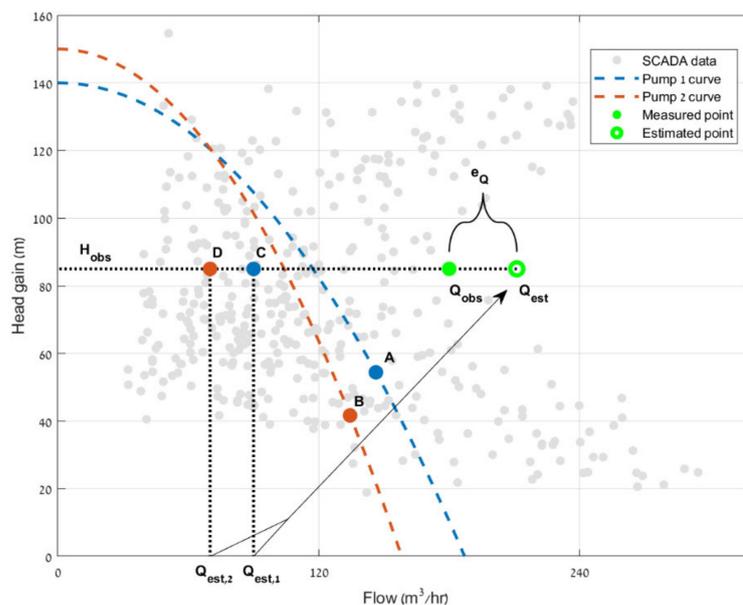


Figure 4. Calculation of the horizontal error for two variable speed pumps.

One should note the similarity between the formulation of the fixed speed pumps in Equations (15)–(20) and the variable speed pumps formulation in Equations (28)–(33). In fact, the former is a particular case of the latter when  $n_{p,t} = \bar{n}_p \forall t$ . Furthermore, in this way, the hybrid case of a pumping station featuring both fixed and variable speed pumps could also be addressed. As such, to use this methodology it is sufficient to implement the formulation in Equations (28)–(33) for an arbitrary combination of fixed and variable speed pumps.

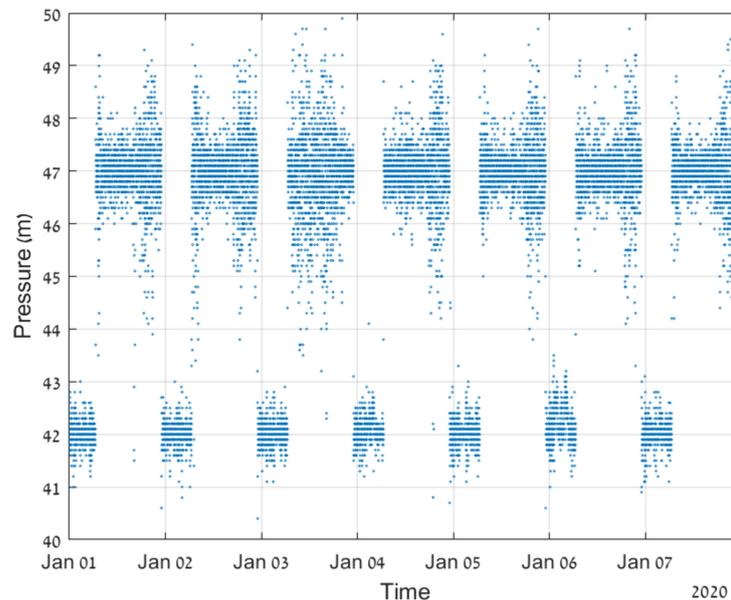
### 3. Test Case and Results

As a test case, we consider a large pressure zone (PZ) in the Southern Israeli city of Be'er-Sheva which is operated by the Mey-Sheva (<https://mey7.co.il/en> accessed on 25 January 2021) water cooperation.

The PZ is supplied by a single pumping station and has no storage tanks. The station has four variable speed pumps, and each can operate over a range of frequency settings which are correlated with their speeds. SCADA measurements (31 December 2019–03 May 2020) are available at 30 s intervals: suction and discharge pressures, total station flow, and individual pump frequencies (SCADA data is available in Data Set S1 in the Supplementary Materials). The frequencies are recorded as percentage [0, 100] for a range of 35–50 Hz. When the value is 0% it indicates that the pump is turned off or operating at its minimum frequency. There are no individual pump flow data and no power data. Power data are not available even for the entire station.

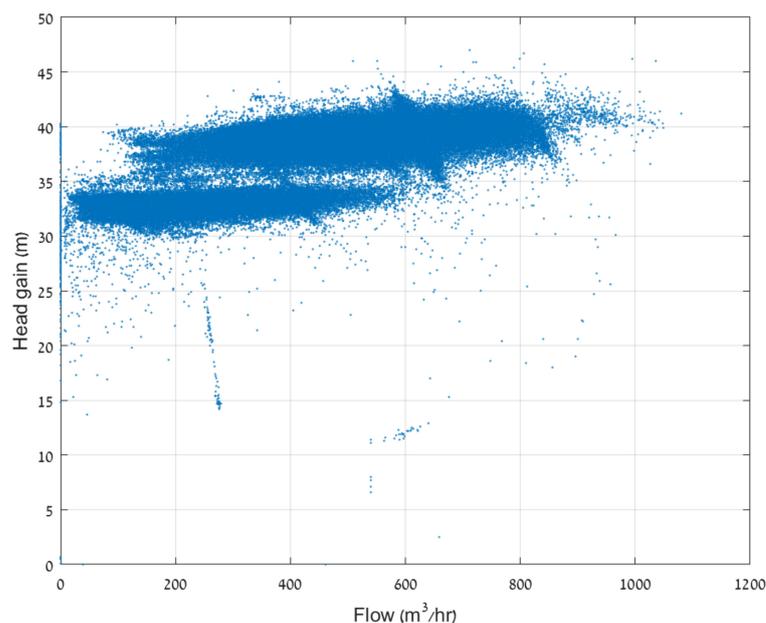
The pumps are operated to maintain the discharge pressure of the station within pre-specified range. As demand in the PZ increases, the pressure in the network decreases, and

the speed of the operating pump is raised to meet the required pressure. The speed is raised to the maximum speed and then another pump is switched on at its lowest speed, which is increased if the pressure continues to drop. The controlled pressure is set to be around 47 m during daytime (06:30–23:00) and to about 42 m during night hours (23:00–06:30) as seen in Figure 5.



**Figure 5.** Pumping station discharge pressure.

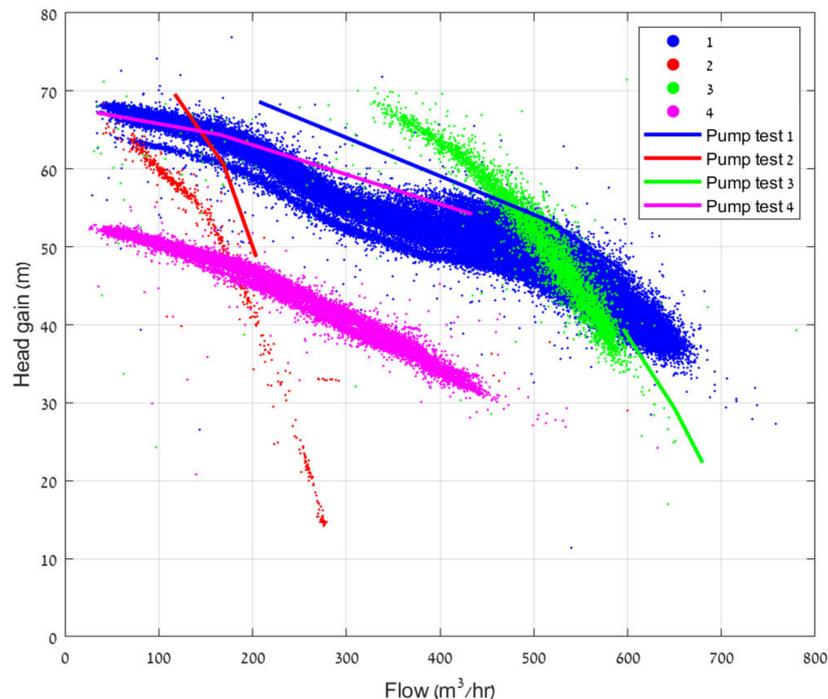
Observing the station's overall Flow-Head Gain plot in Figure 6, it can be seen that the station operates to produce a relatively constant discharge head for a wide range of flows, owing to the control strategy setting stated above. The two regions ("data clouds") in Figure 6 correspond to the day and nighttime discharge-head points.



**Figure 6.** Pumping station Flow-Head Gain points.

We begin by scanning the SCADA data for points that correspond to a single pump in operation, to determine the curves for each pump alone. This was done by searching for

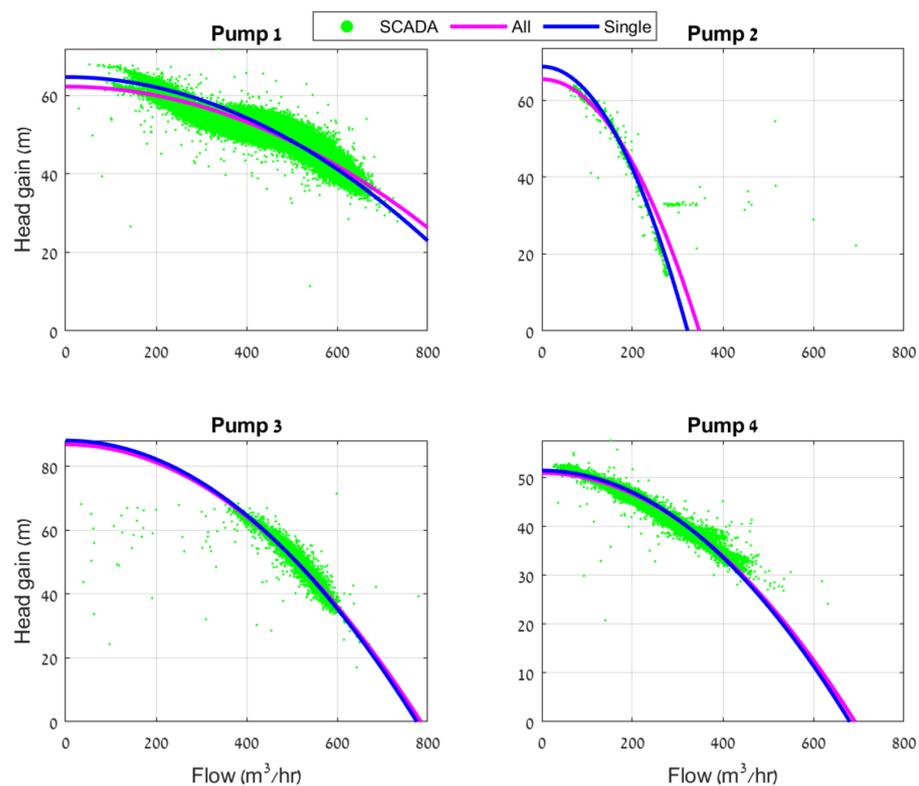
times where only one pump has a frequency greater than 35 Hz ( $>0\%$  in the data) while all other pumps are at a frequency of 35 Hz ( $=0\%$ ). Figure 7 shows the pump curves at the nominal speed (as per Equations (21) and (22)) for the individual pumps as recorded by the SCADA system. Figure 7 also shows the most recent available pump tests, which were performed in September 2017 by the Israeli Water Works Association (IWWA).



**Figure 7.** Individual pump curves for the test case pumping station.

The following observations can be made from Figure 7: (a) the pump test curves are parallel to the SCADA data; (b) the SCADA data are below the pump tests, as expected for deterioration of pump performance over time; (c) the SCADA data for pump 4 is about 20 m below the test curve, which seems excessive and requires further investigation; and (d) for pumps 2 and 3 the test covers only a small part of the actual flows experienced in operation. In fact, for pump 3, the test only covers a small region which was rarely used in the system.

Since the pumps operate alone only part of the time, it is necessary to evaluate their curves while they operate in different combinations. Utilizing the optimization methodology outlined in Section 2.3, (Equations (28)–(33)) we calculated the updated curves (Figure 8). The curve parameters appear in Table 1. For the case studied herein, the updated curve coefficients are not very different from the ones obtained when the pumps operate individually, except for pump 1 where the difference is more pronounced. Still, our generic methodology accounts for all available data for deriving the pump curves, thus it can deal with situations where the pumps operate in different regions under different conditions as was discussed in Figure 2.



**Figure 8.** Estimated pumps curves for the test case pumping station.

**Table 1.** Pumps' curve coefficients.

Pump	Single		All	
	a	b ( $\times 10^{-4}$ )	a	b ( $\times 10^{-4}$ )
1	64.00	0.621	66.29	0.701
2	66.63	6.163	65.78	5.826
3	85.62	1.364	83.93	1.309
4	51.05	1.072	51.07	1.073

#### 4. Conclusions

In this study we present a practical pump curve constructing methodology for obtaining individual pumps' curves from partially available SCADA data. Using it, each pump's performance can be monitored continuously between physical on-premises inspections. We considered the case in which the analyst aims at estimating the individual pump curves whilst the measurements are only performed on the pump station level (i.e., only total flow is available). We show that, unlike ordinary curve fitting techniques that minimize the vertical error between observations and estimating curve, it is advantageous to use the notion of horizontal errors (i.e., the deviation in terms of flow) since it does not require deriving explicit functions for all pump combinations in the pumping station.

The proposed methodology is formulated as a non-linear optimization problem with a small number of decision variables (two for each pump) which can be solved with open source or commercial global solvers. The proposed optimization formulation is generic for any number and type of pumps. Thus, it can be utilized for a single or multiple pumps, and for fixed or variable speed pumps (and a combination of these types). The method can estimate the pumps' curves for any given amount of historical SCADA dataset, without increasing the size of the optimization problem.

The methodology developed herein constitutes a contribution to an increasing trend to "back-figure" system information from SCADA data. Relying on SCADA data for estimat-

ing pump curves has significant practical implications. Field pump tests are expensive and may even be impractical. Indeed, our results on a real test case show that the individual pumps' SCADA data and their corresponding field test results can be significantly different. Furthermore, the field tests, for some of the pumps, cover only part of operating range. This emphasizes the importance of the proposed methodology that derive the pump's curve over the entire operating range under different operating conditions. That is, the method utilizes measurements not only from times when the pump operates alone, but also when it operates together with other pumps. It is worth noting that these situations, where different pump combinations operate at different time slots, are more common when pumps are connected in parallel. Our methodology is tailored for pumps connected in parallel, but still, a dedicated methodology for pumps connected in series is warranted and will be considered for future work.

Unfortunately, in our test case the station power consumption data were not available, and, therefore, the power and individual efficiency curves could not be calculated. Still, if the station power consumption data were available, the methodology could be used to estimate the individual power consumption curves.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/2073-4441/13/5/586/s1>, Dataset S1: The SCADA data used in this study.

**Author Contributions:** Conceptualization, E.S., M.H. and U.S.; methodology, E.S. and M.H.; software, E.S.; validation, E.S. and M.H.; writing—original draft preparation, E.S.; writing—review and editing, M.H. and U.S.; supervision, M.H. and U.S.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data used in this study are available in the Supplementary Materials.

**Acknowledgments:** This study was supported by the Israeli Water Authority. The help and support of Mey-Sheva management and staff is highly appreciated.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Copeland, C.; Carter, N.T. Energy—Water Nexus: The Water Sector's Energy Use R43200. 2017. Available online: <https://crsreports.congress.gov/product/pdf/R/R43200> (accessed on 25 January 2021).
2. Sanders, K.T.; Webber, M.E. Evaluating the energy consumed for water use in the United States. *Environ. Res. Lett.* **2012**, *7*, 034034. [CrossRef]
3. Lam, K.L.; Kenway, S.J.; Lant, P.A. Energy use for water provision in cities. *J. Clean. Prod.* **2017**, *143*, 699–709. [CrossRef]
4. ETSU; CETIM; Reeves, D.T.; NESA; Technical University Darmstadt. Study on Improving the Energy Efficiency of Pumps. European Commission: UK, 2001. Available online: <http://www.jakob-albertsen.dk/komposit/Darmstadtrapport.pdf> (accessed on 25 January 2021).
5. Bunn, S.M.; Reynolds, L. The energy-efficiency benefits of pumpscheduling optimization for potable water supplies. *IBM J. Res. Dev.* **2009**, *53*, 5:1–5:13. [CrossRef]
6. Eaton, A.; D'Alessandro, F.; Ahmed, W.; Hassan, M. On the performance degradation of centrifugal pumps. In Proceedings of the International Conference on Fluid Flow, Heat and Mass Transfer, Niagara Falls, Canada, 7–9 June 2018; Avestia Publishing: Orléans, ON, Canada, 2018; p. 158.
7. Sanks, R.L.; Tchobanoglous, G.; Bosserman Ii, B.E.; Jones, G.M. *Pumping Station Design*, 2nd ed.; Butterworth-Heinemann: Oxford, UK, 1998.
8. Israeli Regulations: Energy Efficiency Test in Pumping Facilities. Available online: [https://www.nevo.co.il/law\\_html/law01/99\\_9\\_315.htm](https://www.nevo.co.il/law_html/law01/99_9_315.htm) (accessed on 10 December 2020).
9. Costa, A.L.H.; De Medeiros, J.L.; Pessoa, F.L.P. Optimization of pipe networks including pumps by simulated annealing. *Braz. J. Chem. Eng.* **2000**, *17*, 887–895. [CrossRef]
10. Stokes, C.S.; Simpson, A.R.; Maier, H.R. A computational software tool for the minimization of costs and greenhouse gas emissions associated with water distribution systems. *Environ. Model. Softw.* **2015**, *69*, 452–467. [CrossRef]

11. Ostfeld, A.; Salomons, E.; Ormsbee, L.; Uber, J.G.J.G.; Bros, C.M.C.M.; Kalungi, P.; Burd, R.; Zazula-Coetzee, B.; Belrain, T.; Kang, D.; et al. Battle of the Water Calibration Networks. *J. Water Resour. Plan. Manag.* **2012**, *138*, 523–532. [[CrossRef](#)]
12. Housh, M.; Ohar, Z. Model-based approach for cyber-physical attack detection in water distribution systems. *Water Res.* **2018**, *139*, 132–143. [[CrossRef](#)]
13. Marchi, A.; Simpson, A.R.; Lambert, M.F. Optimization of Pump Operation Using Rule-Based Controls in EPANET2: New ETTAR Toolkit and Correction of Energy Computation. *J. Water Resour. Plan. Manag.* **2016**, *142*, 04016012. [[CrossRef](#)]
14. Khataavkar, P.; Mays, L.W. Model for real-time operations of water distribution systems under limited electrical power availability with consideration of water quality. *J. Water Resour. Plan. Manag.* **2018**, *144*, 04018071. [[CrossRef](#)]
15. Shang, F.; Uber, J.G.; Rossman, L.A. Modeling reaction and transport of multiple species in water distribution systems. *Environ. Sci. Technol.* **2008**, *42*, 808–814. [[CrossRef](#)]
16. Rossman, L. Modeling of chlorine residual in drinking water distribution system. *J. Environ. Eng. ASCE* **1995**, *120*, 803–820, ISSN 0733-9372/94/0004-0803. [[CrossRef](#)]
17. Ostfeld, A.; Salomons, E. Optimal layout of early warning detection stations for water distribution systems security. *J. Water Resour. Plan. Manag.* **2004**, *130*, 377–385. [[CrossRef](#)]
18. Blinco, L.J.; Simpson, A.R.; Lambert, M.F.; Marchi, A. Comparison of Pumping Regimes for Water Distribution Systems to Minimize Cost and Greenhouse Gases. *J. Water Resour. Plan. Manag.* **2016**, *142*, 04016010. [[CrossRef](#)]
19. Wood, D.J. Waterhammer Analysis—Essential and Easy (and Efficient). *ASCE* **2005**, *131*, 1123–1131. [[CrossRef](#)]
20. Gottliebson, M.; Sanks, R.L.; Vause, A.; Abelin, S.M.; Benfell, R.S.; Bicknell, H.C.; Densmore, B.; Duncan, R.W.; Hong, S.S.; Jones, C.; et al. Variable-Speed Pumping. In *Pumping Station Design*; Elsevier Ltd.: Amsterdam, The Netherlands, 2008; ISBN 9781856175135.
21. Lansey, K.; Awumah, K. Optimal Pump Operations Considering Pump Switches. *J. Water Resour. Plan. Manag.* **1994**, *120*, 17–35. [[CrossRef](#)]
22. Rossman, L.A. *EPANET2 User's Manual*; U.S. Environmental Protection Agency: Cincinnati, OH, USA, 2000.
23. Marchi, A.; Simpson, A.R. Correction of the EPANET Inaccuracy in Computing the Efficiency of Variable Speed Pumps. *J. Water Resour. Plan. Manag.* **2013**, *139*, 456–459. [[CrossRef](#)]
24. Ulanicki, B.; Kahler, J.; Coulbeck, B. Modeling the efficiency and power characteristics of a pump group. *J. Water Resour. Plan. Manag.* **2008**, *134*, 88–93. [[CrossRef](#)]
25. Marchi, A.; Simpson, A.R.; Ertugrul, N. Assessing variable speed pump efficiency in water distribution systems. *Drink. Water Eng. Sci.* **2012**, *5*, 15–21. [[CrossRef](#)]
26. Todini, E.; Tryby, M.E.; Wu, Z.Y.; Walski, T.M. Direct computation of Variable Speed Pumps for water distribution system analysis. In *Proceedings of the Combined International Conference of Computing and Control for the Water Industry, CCWI2007 and Sustainable Urban Water Management, SUWM2007*, Leicester, UK, 3–5 September 2007; pp. 411–418.
27. AbdelMeguid, H.; Ulanicki, B. Feedback Rules for Operation of Pumps in a Water Supply System Considering Electricity Tariffs. *Water Distrib. Syst. Anal.* **2010**. [[CrossRef](#)]
28. Coulbeck, B.; Orr, C.H. Real-Time Optimized Control of a Water Distribution System. *IFAC Proc. Vol.* **1989**, *22*, 441–446. [[CrossRef](#)]
29. Abdallah, M.; Kapelan, Z. Fast Pump Scheduling Method for Optimum Energy Cost and Water Quality in Water Distribution Networks with Fixed and Variable Speed Pumps. *J. Water Resour. Plan. Manag.* **2019**, *145*, 04019055. [[CrossRef](#)]
30. Hashemi, S.S.; Tabesh, M.; Ataekia, B. Ant-colony optimization of pumping schedule to minimize the energy cost using variable-speed pumps in water distribution networks. *Urban Water J.* **2014**, *11*, 335–347. [[CrossRef](#)]
31. Bene, J.G.; Hos, C.J. Finding least-cost pump schedules for reservoir filling with a variable speed pump. *J. Water Resour. Plan. Manag.* **2012**, *138*, 682–686. [[CrossRef](#)]
32. Wu, W.; Maier, H.R.; Dandy, G.C.; Arora, M.; Castelletti, A. The changing nature of the water–energy nexus in urban water supply systems: A critical review of changes and responses. *J. Water Clim. Chang.* **2020**, *11*, 1095–1122. [[CrossRef](#)]
33. Lima, G.M.; Luvizotto, E.; Brentan, B.M.; Ramos, H.M. Leakage control and energy recovery using variable speed pumps as turbines. *J. Water Resour. Plan. Manag.* **2018**, *144*, 04017077. [[CrossRef](#)]
34. Wu, W.; Simpson, A.R.; Maier, H.R.; Marchi, A. Incorporation of variable-speed pumping in multiobjective genetic algorithm optimization of the design of water transmission systems. *J. Water Resour. Plan. Manag.* **2012**, *138*, 543–552. [[CrossRef](#)]
35. Huo, J.; Eckmann, D.H.; Hoskins, K.P. Hydraulic simulation and variable speed pump selection of a dual-function pumping station: Pumping treated water to a reclaimed water system or deep injection well system. In *World Environmental and Water Resources Congress*; American Society of Civil Engineers: Reston, VA, USA, 2007; pp. 1–7.
36. Sunela, M.I.; Puust, R. A visual tool to calculate optimal control strategy for non-identical pumps working in parallel, taking motor and VSD efficiencies into account. *Water Sci. Technol. Water Supply* **2015**, *15*, 1115–1122. [[CrossRef](#)]
37. Gostoli, C.; Spadoni, G. Linearization of the head-capacity curve in the analysis of pipe networks including pumps. *Comput. Chem. Eng.* **1985**, *9*, 89–92. [[CrossRef](#)]
38. *MATLAB 9.5.0.1033004 (R2018b)*; The MathWorks Inc.: Natick, MA, USA, 2018.
39. Wächter, A.; Biegler, L.T. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program.* **2006**, *106*, 25–57. [[CrossRef](#)]
40. Tawarmalani, M.; Sahinidis, N.V. A polyhedral branch-and-cut approach to global optimization. In *Proceedings of the Mathematical Programming*; Springer: Berlin/Heidelberg, Germany, 2005; Volume 103, pp. 225–249.

## **Paper IV**

*Salomons, E.; Sela, L.; Housh, M. "Hedging for Privacy in Smart Water Meters".  
Water Resources Research 2020, 56 (9), e2020WR027917.  
<https://doi.org/10.1029/2020WR027917>.*

Journal Impact Factor: 4.309

5-Year Impact Factor: 5.036

Publication Status: Published (Accepted 10 August 2020)

*Statement of authorship - Paper IV*

**Principal Author**

Title of Paper	Hedging for Privacy in Smart Water Meters		
Publication Status	<input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication		
Publication Details	Salomons, E., Sela, L., & Housh, M. (2020). Hedging for privacy in smart water meters. <i>Water Resources Research</i> , 56, e2020WR027917. <a href="https://doi.org/10.1029/2020WR027917">https://doi.org/10.1029/2020WR027917</a>		
Name of Principal Author (Candidate)	Elad Salomons		
Contribution to the Paper	Conceived the presented idea, developed the theory and performed the computations. Wrote the manuscript with support from the co-authors.		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Name and Signature	Elad Salomons 	Date	September 17th, 2020

**Co-Author Contributions**

By signing the Statement of Authorship, each author certifies that:

- vii. the candidate's stated contribution to the publication is accurate (as detailed above);
- viii. permission is granted for the candidate to include the publication in the dissertation

Name of Co-Author 1	Lina Sela		
Contribution to the Paper	Verified the analytical methods, encouraged Elad Salomons to investigate the subject, and supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.		
Name and Signature	Lina Sela 	Date	September 17th, 2020
Name of Co-Author 2	Mashor Housh		
Contribution to the Paper	Verified the analytical methods, encouraged Elad Salomons to investigate the subject, and supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.		
Name and Signature	Mashor Housh 	Date	September 17th, 2020

# Water Resources Research

## RESEARCH ARTICLE

10.1029/2020WR027917

## Hedging for Privacy in Smart Water Meters

Elad Salomons<sup>1</sup> , Lina Sela<sup>2</sup> , and Mashor Housh<sup>1</sup>

<sup>1</sup>Department of Natural Resources and Environmental Management, University of Haifa, Haifa, Israel, <sup>2</sup>Department of Civil, Architectural and Environmental Engineering, University of Texas at Austin, Austin, TX, USA

### Key Points:

- Smart meters gain popularity with utilities providers; consumers' concerns of possible privacy infringement are growing as well
- Practical technology coupling hardware and software for hedging personal privacy in the presence of a smart water meter is proposed
- The trade-off of the level of privacy protection and the size of the system is presented

### Supporting Information:

- Supporting Information S1

### Correspondence to:

E. Salomons,  
selad@optiwater.com

### Citation:

Salomons, E., Sela, L., & Housh, M. (2020). Hedging for privacy in smart water meters. *Water Resources Research*, 56, e2020WR027917. <https://doi.org/10.1029/2020WR027917>

Received 13 MAY 2020

Accepted 10 AUG 2020

Accepted article online 13 AUG 2020

**Abstract** Smart water meters at household connections are being installed in large numbers throughout the world due to the benefits they are expected to bring to the water utilities and water consumers. Smart metering provides high-resolution readings and promises benefits to the water utilities, such as demand forecasting, regulating time-of-use watering, and making intelligent operation and planning decisions. For the consumers, smart metering promises improved billing and demand reduction by providing detailed and timely information about their water use and early notification of possible water leaks in their premises. However, the fine-grained information collected by smart meters raises growing concerns of privacy invasion due to personal behavior exposure (private activity, daily routine, etc.). Nevertheless, there is no readily available technology for protecting water consumers from revealing their in-home private activities. Thus, a viable argument in favor of smart metering technologies will not be possible without proactively accounting for the associated privacy challenges. Here, we present a practical technology coupling a dedicated apparatus with a control model for increasing personal privacy. We quantify the level of privacy achieved using information-theoretic criterion and an empirically based occupancy detection method between the smart meter readings and actual water use. Furthermore, we evaluate and compare privacy protection using the best effort approach previously developed for masking activities revealed from smart electricity meters. The main results reveal that simple control actions can disguise personal behavior patterns and, thus, hedge against privacy breach in smart water meters. Furthermore, we quantify the trade-off between the size of the apparatus and the level of privacy protection it provides. Our results demonstrate how “privacy friendly” smart water metering technology could be implemented in real-life systems and reduce the privacy concerns of water consumers.

## 1. Introduction

Smart water meters, continuously collecting and transmitting water usage, are changing the paradigm for managing residential water use. Installation of smart water meters is expected to surpass 100 million units by 2030, with water-stressed developed cities leading the rollout of smart water meters to address growing concerns of water shortages (Smart Energy International, 2018). Smart water meters promise benefits to the water utilities battling with inaccurate, labor-intensive meter readings, consumer billing, nonrevenue water losses, water scarcity, and growing demands (Gurung et al., 2014, 2015; Stewart et al., 2018). The increased spatial and temporal water consumption data enables improved demand modeling, prediction, and demand management strategies as well as making optimized operation and planning decisions (Cominola et al., 2015; Gurung et al., 2017; Nguyen et al., 2018; Zhuang & Sela, 2020). Several recent works have demonstrated the successful application of disaggregating smart water meter data to infer specific usage associated with human activities at the individual household level (e.g., shower and sink activities). Although most rely on high-temporal resolution of 1–60 s (Clifford et al., 2018; Cominola et al., 2019; Nguyen et al., 2013; Vasak et al., 2015), others have demonstrated the applicability of identifying personal activities with coarser temporal resolution of 15 min sampling rate (Chen et al., 2011).

Despite the fact that the aforementioned benefits advocate for smart water meters, the fine-grained information collected by smart meters raises growing concerns of privacy invasion due to personal behavior exposure (private activity, daily routine, etc.) and potential data misuse (McKenna et al., 2012; Véliz & Grunewald, 2018). Currently, privacy is a sweeping concept, encompassing, in the context of smart water meters, control over personal information (Jamieson, 2009; Solove, 2010; Weaver, 2014; Zipper et al., 2019). Multiple ongoing and unresolved debates regarding the definition of the illusive concept of privacy arise beyond the revealing personal household activities, including ownership of information and,

moreover, how this information may be used for discrimination, search, and surveillance (Ząbkowski & Gajowniczek, 2013). Interestingly, the privacy paradox reveals the discrepancy between users' attitude toward privacy concerns and their actual behavior in limited undertaking of countermeasures to protect their privacy (Barth & de Jong, 2017). In the smart water meter setting, this paradox maybe partially attributed to the lack of understanding or unawareness of the consumers of the embedded privacy in their water consumption data. At the very least, through simple data mining techniques, smart meter data could be easily used to detect home occupancy, and through more sophisticated data mining techniques, detection of specific devices (e.g., dish and clothes washing machines) can be discerned and utilized for targeted advertisements. Analyzing private water use activity can be used to evaluate personal hygiene habits to help reduce the spread of infectious disease, use to track hygiene compliance in retail food establishments and hospitals (Allwood et al., 2004; Pittet, 2001), and understand gender and ethnic differences in hygiene practices (Anderson et al., 2008; Horsburgh et al., 2017). A striking example of embedded privacy in water metered data occurred when the information revealed from smart water meters was used as evidence against a murder suspect in Arkansas, USA, who was accused partially based on smart water meter readings obtained without a warrant from the water department. The water meter data were used to claim that the suspect had used a large amount of water in the middle of the night to clean up the crime scene (Jerome, 2017). Information from smart energy meters about detected electricity theft has been associated with the detection of cannabis plantations (Depuru et al., 2011). Furthermore, Quinn (2008) argued that the privacy threat of smart meter goes beyond exposing private information to a large-scale spy.

The protection of privacy is supported in different legal frameworks including laws, policies, and regulations (Jawurek et al., 2012). In Europe, data privacy is covered under the General Data Protection Regulation (European Union, 2016), which requires that personal data should “be collected for a specified purpose and not be further processed for other purposes”. In the United States, individual privacy is protected, from the state, by the Fourth Amendment law, which was recently employed against smart meters in the U.S. Court of Appeals (United States Court of Appeals, 2018). Privacy tools, relying on institutional rules and mechanisms, can help the different stakeholders (e.g., service providers, organizations, utilities, and tech companies) to implement privacy tools without the need to change the technical system. In addition to privacy tools, smart metering systems are expected to uphold the highest security requirements. While security and privacy are often used interchangeably, security requirements of smart meters involve providing reliable delivery of data in terms of data integrity, for example, prevention of malicious unauthorized modification of the data, which might lead to incorrect billing or operation of the system; and confidentiality, in terms of unauthorized access to information by third parties (Souri et al., 2014). The security requirements, similarly to privacy tools, are addressed at the organization and the service provider level and have been the preferred approach for addressing privacy issues associated with smart meters. Nevertheless, fundamental weakness of privacy and security tools stems from the inherent reliance on the trustworthiness of all parties involved (e.g., trust between end user and water utility) in the process (Jawurek et al., 2012). In light of the above, there is a need for a privacy preserving system to *prevent privacy breach a priori through privacy technology* rather than policy. That is, a technological solution that does not depend on trust relationships and enables the individual to conceal their personal activity.

Naturally, the aforementioned privacy concerns arise in other metered utilities at the household level including electricity and gas (Véliz & Grunewald, 2018). Privacy enhancing technologies have been previously proposed in the smart electricity grid literature (Jawurek et al., 2012; Kalogridis et al., 2010; McKenna et al., 2012). Several approaches have been proposed predominantly utilizing battery-based load hiding (BLH) paradigm. The main idea of BLH approaches is to utilize the battery to conceal the real energy used for household activities by charging and discharging the battery and, thus, offsetting the actual timing and volumes of the true demands. Given the size of the battery, the control actions prescribe the timing and rate of battery charging/discharging. In Kalogridis et al. (2010), the Best Effort (BE) algorithm was proposed, which minimized changes in meter readings in consecutive time steps. Privacy levels were measured using relative entropy for quantifying the distance between two signals (i.e., real and observed demands), cluster classification, and regression coefficients of the signals with and without privacy preserving system. In McLaughlin et al. (2011), the nonintrusive load leveling control scheme was proposed, which offset spikes and dips in usage by charging or discharging the battery, targeting instantaneous energy usages that expose human behavior. Several evaluation metrics were proposed including relative mass, residual features, and

entropy, which count the number of features in the signal that can disclose activity and quantify the amount of information encapsulated in the signal. A randomized approach for generating noises was proposed in Zhao et al. (2014) achieving differential privacy by controlling noise selection. An interesting exception to BLH approaches was proposed in Chen et al. (2015) using electric water heaters to modulate the energy embedded in the hot water. Matthews correlation coefficient was used to evaluate the performance of occupancy detection (OD) of four different detection methods, including thresholds,  $k$ -nearest neighbors, support vector machines, and hidden Markov models. The use of water heater for storing energy showed to outperform BLH approaches.

Despite the abovementioned technologies for privacy protection in the energy sector, the water sector is lagging behind in addressing privacy concerns. It follows that addressing privacy concerns stemming from smart electricity meters without smart water meters will inhibit achieving desired privacy goals. To address this gap, this paper aims to develop a *privacy preserving system (PPS) to hedge for privacy in smart water meters*. The approach proposed in this study is motivated by related work for increasing privacy in smart electricity meters (Asghar et al., 2017; Jawurek et al., 2012; Kalogridis et al., 2010; Rial & Danezis, 2011). Specifically, we propose a practical technology coupling a flow control valve and precharged water tank with a control model for increasing personal privacy in a household equipped with a smart water meter. Noteworthy that while privacy hedging can be achieved by simply controlling the sampling rate of the smart meter, for example, a reduced sampling rate implies greater privacy protection (Chen et al., 2011; Cominola et al., 2018), the sampling and transmission rate are controlled by the water utility or the service provider and are not controlled by the water user. In this paper the proposed approach is controlled and implemented by users that seek increased privacy. This approach is tested using real demand data collected from multiple houses of different characteristics across North America (DeOreo et al., 2016), and the results show a significant increase in privacy protection even with a small-size water tank.

This paper is structured as follows. In section 2.1, we present the proposed design of the PPS, and in section 2.2, we describe the flat-flow (FF) privacy moderation algorithm. In section 2.3, we propose an information-theoretic and empirically based approaches to evaluate the level of privacy in smart meter data. Ideally (from a privacy viewpoint), all water usage events should be concealed; however, in practice, perfect privacy is costly, which introduces a trade-off between cost and privacy protection level. Hence, in section 3, we demonstrate the approach and explore the trade-off between the level of privacy protection and the size of the PPS and compare the system performance under different privacy moderation algorithms, that is, the FF algorithm proposed in this work and the BE algorithm proposed in Kalogridis et al. (2010) for achieving privacy in electricity data. In section 4, we discuss the implications of wide adaption of the PPS for water utilities and the end users.

The contribution of this paper is as follows. First, it reveals some of the discordance between the benefits and the associated privacy concerns of smart water meters. Second, it proposes a new and practical technology coupling a hardware apparatus with a software control model for hedging personal privacy in the presence of a smart water meter. Third, it presents quantifiable privacy measures that enable the comparison of different privacy control actions. Lastly, it promotes the research toward privacy-sensitive design and implementation of smart metering systems.

## 2. Methodology

The proposed PPS is a combined hardware-software solution, installed by the water consumer within their premises, where the designed physical component can support different control strategies. The proposed PPS is shown in Figure 1. The main objective of the PPS is to conceal the true end use water consumption data by modulating the inflow and, thus, the readings of the smart water meter. Next, we describe the physical setup, the control actions (i.e., privacy moderation algorithm), and the evaluation metric.

### 2.1. System Overview

In the proposed setup, water is supplied from the distribution network through the service line to the premises. The smart water meter is located at the connection between the service and the premise lines, measuring the flow rates and transmitting the data to a centralized database. The collected measurements provide detailed end use water consumption data, and in turn, end user private information. The PPS,

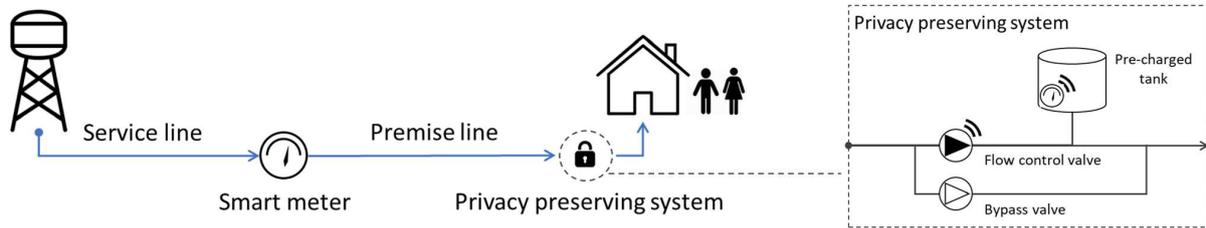


Figure 1. Privacy preserving system schematics.

located downstream of the smart water meter, consists of two main physical components: a flow control valve (FCV) and a precharged water tank. The privacy moderation algorithm regulates the flow in the FCV according to the specified control rules and in accordance with the state of the precharged water tank.

A precharged, usually a metal, water tank includes a flexible water bladder, while the rest of the tank is filled with compressed air at a precharged pressure (see illustration in Figure 2). The tank is refilled and drained based on pressure set points and acts like a buffer between the water meter and premises. Tank operation includes four states: (a) the water bladder is empty and the air expands to fill the tank's volume outside the bladder (Figure 2a); (b) water is filling the bladder while the air is compressed (Figure 2b); (c) when a preset air pressure is reached, the water flow into the bladder stops and the tank is considered full (Figure 2c); and (d) water flows out of the bladder by the compressed air pressure (Figure 2d). Note that precharged tanks are readily available hydraulic devices, which have common uses in domestic plumbing (Pentair, 2020). In addition to the FCV and the precharged tank, a bypass valve is included in case the system should be overridden, for example, during emergency situations (Figure 1). The proposed PPS is suitable for single-family residential users and can be conveniently placed in the premises by using a pipeline extension downstream of the water meter. Additionally, as we will show in the results, the size of the precharged water tank is comparable with the typical residential water heater, thus practically feasible for installation in single-family residential homes. We note that in some single-family residential buildings and apartment buildings, it may be difficult to find a place for such a system. Moreover, feasibility of installation of the PPS in multifamily residential buildings will require further investigation.

Given the physical setting in Figure 1, the problem is how to maximize the privacy protection by designing a privacy moderation algorithm to control the PPS given any set of end user demand,  $Q_D$ , and predetermined tank capacity  $V_{MAX}$ , while recognizing the uncertain nature of the demand. That is, we seek  $T: Q_D \rightarrow Q_{FCV}$ , a transformation  $T$  from the real end user demand,  $Q_D$ , to the measurable FCV flow,  $Q_{FCV}$ .

## 2.2. Privacy Moderation Algorithm

In the proposed setup, at any given time  $t$ , the smart water meter records the flow upstream of the water tank,  $Q_{FCV}(t)$ , rather than the actual end use demand,  $Q_D(t)$ , downstream of the tank. Intuitively, the control model minimizes the variation in the flow, as recorded by the smart water meter, by controlling the FCV to maintain constant flow when possible. The control actions and the resulting privacy hedging depend on the volume of water in the tank, as follows. For a given time  $t$ , the mass balance in the tank is given by Equation 1.

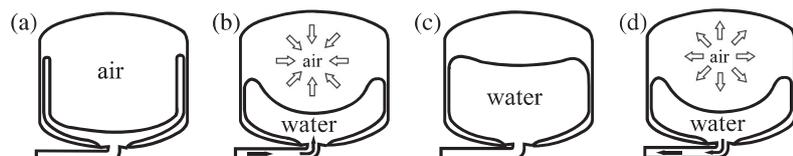


Figure 2. (a–d) Precharged water tank schematics.

**Table 1**  
Privacy Moderation Algorithm

Algorithm 1: Flat-flow algorithm		
1:	<b>Input:</b>	Water demand $Q_D$ , size of the precharged tank, $V_{MAX}$ and $V_{MIN}$
2:	<b>Output:</b>	Water meter readings $Q_{FCV}$
3:	<b>Initialize:</b>	$t = t_0, V(t) = V_0, Q_{FCV}(t), Q_D(t)$
4:	<b>Set:</b>	Flow through the control valve $Q_{FCV}(t+\Delta t)$ : if tank is full $V(t) = V_{MAX}$ set flow to 0 <b>elseif</b> tank is not full $V_{MIN} < V(t) < V_{MAX}$ set flow to $Q_{REF}$ <b>else</b> tank is empty $V(t) = V_{MIN}$ set flow to $Q_{MAX}$
5:	<b>Update:</b>	Water volume in tank $V(t+\Delta t) = (Q_{FCV}(t) - Q_D(t))\Delta t + V(t)$
6:	<b>Repeat:</b>	$t = t + \Delta t$

$$\frac{dV(t)}{dt} = Q_{FCV}(t) - Q_D(t) \quad (1)$$

where  $V(t)$  is the water volume in the tank,  $Q_{FCV}(t)$  is the flow through the control valve, and  $Q_D(t)$  is consumer's demand. The demand is determined by the consumer, and we assume that the water utility can supply the maximum historical demand,  $Q_{MAX}$  at any given time. The controlled variable, in this setup, is the flow through the FCV, that is,  $Q_{FCV}(t)$  at every time step, which is also the flow measured by the water meter. Clearly, when the volume of the tank does not change, that is, when  $\frac{dV}{dt} = 0$ , the water meter will measure the demand,  $Q_{FCV} = Q_D$ . The proposed control scheme, aiming at keeping a constant inflow (i.e., flat flow), evaluates the water level in the tank and determines the setting of the FCV in the next time step: When the volume in the tank is within the acceptable range, the flow through the FCV is set to constant reference flow rate  $Q_{REF}$ ; when the tank empties, FCV fully opens and the flow is set to  $Q_{MAX}$ , to supply the demand and recharge the tank; when the tank is full, the FCV closes and no flow is supplied from the service line. The control actions for the FF algorithm are summarized in Table 1. The reference and maximum flow rates are determined based on the historical demand of the specific end user, and the time step is determined based on the temporal sampling resolution of the smart meter.

### 2.3. Evaluation of Privacy Protection

To judge the effectiveness of a proposed transformation  $T: Q_D \rightarrow Q_{FCV}$ , it is necessary to define a privacy metric that quantifies the level of privacy protection that the transformation can offer. We propose two complementary approaches to evaluate the level of privacy in smart meter data: information-theoretic and empirically based approaches. While both measure the potential amount of information revealed about the end user behavior from the smart meter data, the former provides a more abstract measure of privacy and the latter specifically targets the ability of a privacy intruder to detect occupancy. The details of the proposed evaluation methods are as follows.

#### 2.3.1. Normalized Mutual Information

First and foremost, we seek a privacy metric that takes the true end user demand,  $Q_D$ , and the smart meter readings,  $Q_{FCV}$ , as inputs and returns a numerical measure of privacy level as an output. As we do not know how a privacy intruder might exploit the collected data, it is necessary to develop a metric that quantifies the amount of leaked information regardless of the data mining algorithm or the computational capabilities of the intruder. Intuitively, privacy is maximized when the recorded data are completely independent of the true end user demand data. Under these conditions, the information leakage is minimized, and it will not be possible to infer any private information from the smart meter readings. Information-theoretic metrics can be used to measure the amount of inherent information available for exploitation by a privacy intruder. Evidently, if the amount of private information available for learning from the measurable data,  $Q_{FCV}$ , is small, then the privacy loss is bounded by this small amount regardless of how the intruder operates. Thus, if the PPS satisfies privacy in an information-theoretic sense, user's privacy is protected, and privacy breach is reduced. An abundance of previous studies explored privacy issues from a fundamental information-theoretic perspective (e.g., Ma & Yau, 2015; Wagner & Eckhoff, 2018). A widely used privacy metric to measure the leaked information from smart meter readings is the Mutual Information (MI),

which measures the dependence between the true demand and smart meter readings while considering them as two random variables (Giaconi et al., 2015; Rajagopalan et al., 2011; Tan et al., 2013; Varodayan & Khisti, 2011; Wu et al., 2006). In this study, we adopt the MI metric to quantify the amount of leaked information from the transformation  $T$  between end user demand,  $Q_D$ , and the smart meter readings,  $Q_{FCV}$ , that is,  $MI(Q_D, Q_{FCV})$ .

The MI between two discrete random variables  $X, Y$  is defined in Equation 2.

$$MI(X, Y) = H(X) - H(X|Y) \quad (2)$$

where  $H(X)$  is Shannon entropy and  $H(X|Y)$  is conditional entropy of the random variable  $X$  given the random variable  $Y$ . The two random variables have realizations,  $x, y$ , which take values from finite support sets,  $\mathcal{X}, \mathcal{Y}$ , according to a joint probability mass function  $P_{XY}$  and marginal probability mass functions  $P_X, P_Y$  for  $X$  and  $Y$ , respectively.

As seen in Equation 2, MI is defined relying on Shannon entropy (Shannon, 1948),  $H(X)$ , which is the basis of many information-theoretic metrics. Shannon entropy measures the amount of randomness in the random variable based on the probabilities of the possible realizations as defined in Equation 3. The larger the amount of randomness in the random variable the larger the entropy.  $H(X|Y)$  is the conditional entropy of the random variable  $X$  given the random variable  $Y$ . The conditional entropy measures the amount of remaining randomness in  $X$  after observing  $Y$  (i.e., the unexplained randomness) and is calculated based on the conditional probability as shown in Equation 4.

$$H(X) = - \sum_{x \in \mathcal{X}} P_X(x) \cdot \log_2(P_X(x)) \quad (3)$$

$$H(X|Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{XY}(x, y) \cdot \log_2 \left( \frac{P_{XY}(x, y)}{P_Y(y)} \right) \quad (4)$$

Based on the above,  $MI(X, Y)$  measures the reduction in the amount of randomness of  $X$  after observing  $Y$ , that is, the explained randomness in  $X$  after observing  $Y$ , which could be interpreted as the MI between  $X$  and  $Y$ .

In our context,  $X$  represents the true end user demand,  $Y$  represents the smart meter readings, and MI quantifies how much information is leaked from the true demand by revealing the smart meter data. Namely, if the intruder were to apply a disaggregation algorithm to detect usages based on the smart meter data, then MI quantifies how much information about the true demand will be revealed (i.e., the amount of leaked information from the PPS). Low values of MI imply small information leakage and, thus, higher privacy protection. For example, in case of constant smart meter readings, the amount of randomness in the end user data given the meter reading (i.e., conditional entropy) will be equal to the amount of randomness in the end user data itself (i.e., its entropy), as such the MI will be 0. At the other extreme, in case the smart meter measures the actual end user demand, the conditional entropy will be 0 and the MI will be equal to the entropy of the end user data. As such, MI ranges between 0 and the entropy of the end user demand,  $H(X)$ . To allow for comparison between different scenarios (e.g., different end users), the normalized MI metric,  $NMI(X, Y)$ , can be defined as given in Equation 5, which takes values between 0 and 1.

$$NMI(X, Y) = 1 - \frac{H(X|Y)}{H(X)} \quad (5)$$

The NMI can be interpreted as the proportion of explained randomness in  $X$  when the variable  $Y$  is observed. In our context, NMI is the fraction of information about the end user data that is exposed from the smart meter readings (Koo et al., 2012).

### 2.3.2. OD

The NMI measure quantifies the level of privacy regardless of the specific detection algorithm that a privacy intruder may use. In addition, we propose a simple OD algorithm that can be used to evaluate and compare the performance of the PPS for different tank sizes and control algorithms. We consider a simple duration-based threshold algorithm that attempts to reveal the time instances in which the household is occupied (or not). Specifically, if zero flow is recorded continuously by the water meter for a specified

**Table 2**  
Summary of the Data Sets Used

Data set	City	Number of people	Data duration (days)	Total volume consumed (m <sup>3</sup> )	Average flow rate $Q_{REF}(10^{-6} \text{ m}^3/\text{s})$	Maximum flow rate $Q_{MAX}(10^{-4} \text{ m}^3/\text{s})$
1	Denver	1	11	1.63	1.72	2.84
2	Mississauga	2	12	6.04	5.37	3.98
3	Kissimmee	3	12	10.09	9.57	2.82
4	Denver	4	12	7.14	6.40	3.49

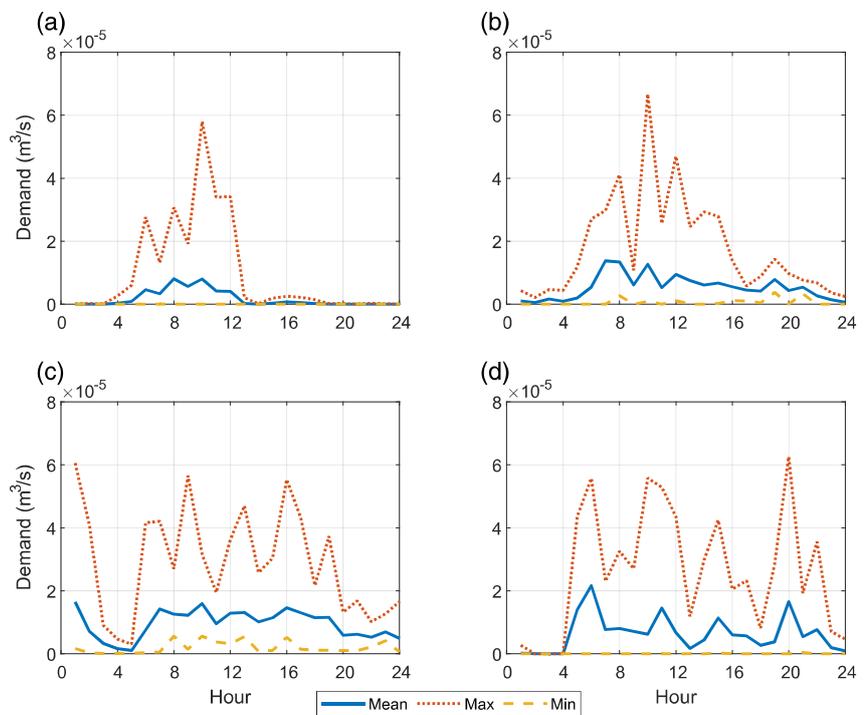
epoch time,  $t_e$ , it is assumed that the household is not occupied. Then a comparison can be made between the inferred occupancy without and with different PPS. The characteristic epoch time,  $t_e$ , can be estimated and calibrated for different households and stratified for different times of day based on the historical water usage. While more sophisticated demand disaggregation algorithms targeting specific water usages exist, occupancy detection provides a single best measure of privacy intrusion, as well as it does not require detailed and fully labeled water usage data (Nguyen et al., 2015).

### 3. Results

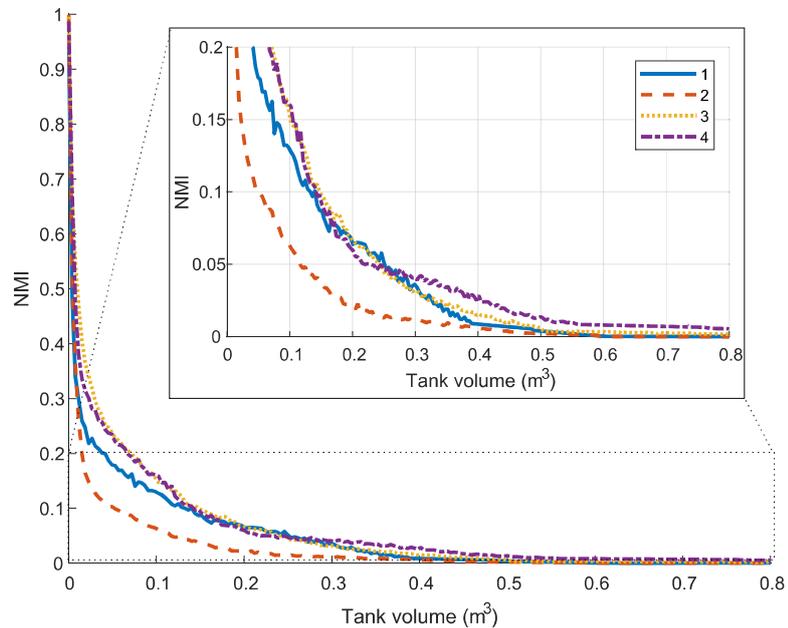
#### 3.1. Data Set Description

We test the performance of the proposed algorithm using real demand data sets obtained from the Residential End Uses of Water study (DeOreo et al., 2016). The results are demonstrated using four households (with one to four residents) in North America, which differ in the number of residents (one to four), their geographic location (Central, South and North regions of North America) and their daily demand patterns.

Table 2 summarizes the four data sets, and Figure 3 shows the average, minimum and maximum hourly demands profiles of the four households. As can be observed, the demand patterns for the different data sets vary in the hourly distribution and the peaks. For example, Data Set 1 (Figure 3a), is a one-person household,



**Figure 3.** Mean, maximum, and minimum hourly demand for Data Sets 1–4 (a–d, respectively).



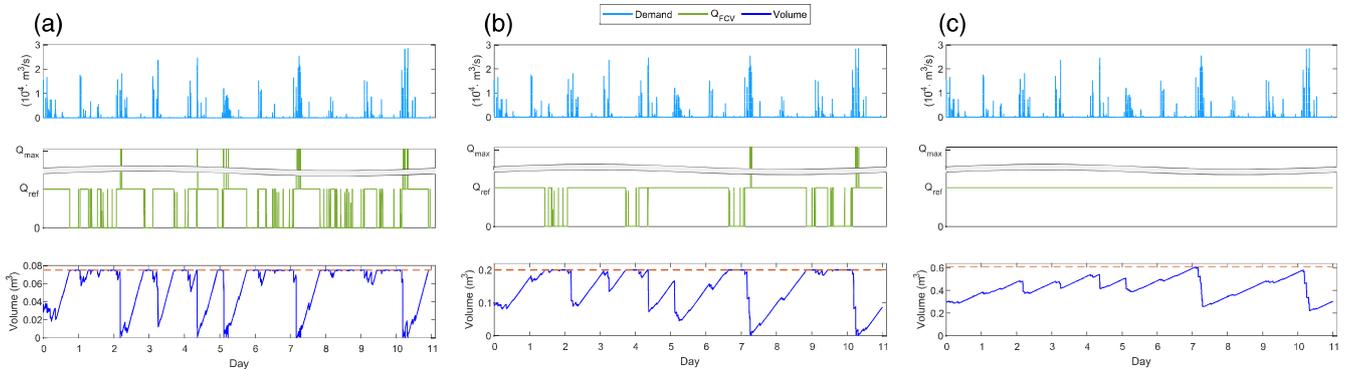
**Figure 4.** NMI as function of the tank volume for the four data sets.

which has most of its demand in the morning hours and no demand in the rest of the time as probably no one is present in the house at these times. For Data Set 3 (Figure 3c), which is a three-person household, there is a more “classical” demand pattern with small peaks in the morning and afternoon and less demand during the night hours.

### 3.2. Privacy Moderation Algorithm

For each data set, we tested the proposed FF algorithm (Table 1), where the average and maximum flow rates were calculated based on the historical demand patterns of each household (as listed in Table 2),  $V_{MIN}$  and  $V_{MAX}$  were set to 0 and the size of the water tank, respectively. To explore the trade-off between the privacy level and the size of the water tank, we evaluated the NMI metric by incrementally increasing the size of the water tank between 0 and 0.8 m<sup>3</sup>, where the largest water tank represents the average daily consumption of the highest water consuming household (Data Set 3). Figure 4 illustrates the resulting NMI scores for all households as a function of the size of the storage tank. For all data sets, the maximum value of 1 is achieved for a zero-volume tank, that is, without PPS. In this case, the demand is directly supplied by the water utility, thus entirely exposing the water use pattern. As the tank’s volume is increased, the NMI value decreases. In the tested range (0–0.8 m<sup>3</sup>), the NMI for Data Sets 1 and 2 decreases to 0 at 0.61 and 0.59 m<sup>3</sup>, respectively. In these cases, maximum privacy is obtained, since the flow measured by the water utility at the water meter is constant and does not reveal any pattern of the water use. Interestingly, Household 2 exhibits a sharper decrease in NMI scores compared with Household 1, despite having a higher total demand. This is attributed to fact that the temporal variation of water usage throughout the day contains valuable information regarding the end user’s behavior.

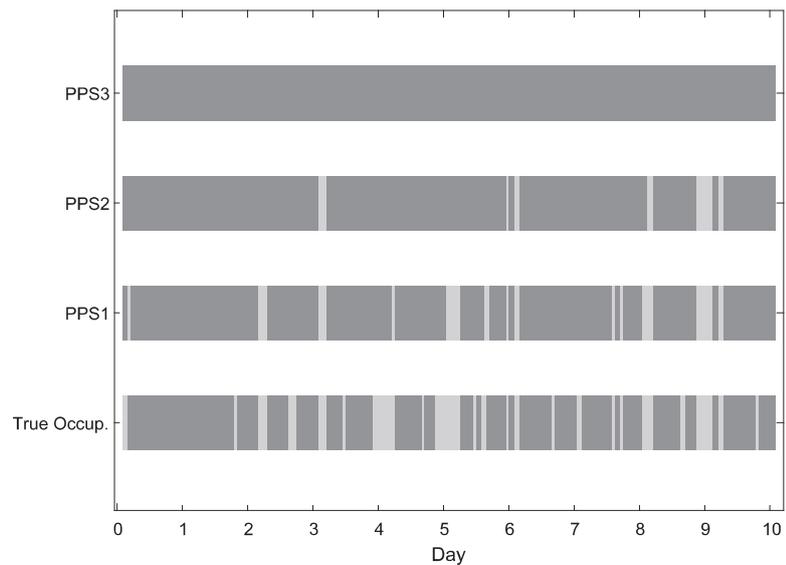
Next, in Figure 5 we show the variation in the inflow,  $Q_{FCV}$  (which is also the flow through the smart water meter), the water volume in the tank,  $V(t)$ , and the true water demand for Data Set 1 for three different tank sizes: 0.075, 0.2, and 0.61 m<sup>3</sup>. For a tank size of 0.075 m<sup>3</sup> (Figure 5a) the value of NMI is 0.15. The results show that the metered inflow observed by the smart meter is significantly different from the actual end user water demand, where during the majority of the time steps, the inflow,  $Q_{FCV}$ , is equal to the average flow,  $Q_{REF}$ . Occasionally, when the demand is high and the tank is empty, a high inflow from the valve,  $Q_{MAX}$ , is engaged to refill the tank and satisfy demand (as can be seen at the beginning of Day 3 in Figure 5a). For the cases where the tank is full and the demand is less than  $Q_{REF}$ , a valve closure is invoked, such



**Figure 5.** Changes in demand, smart meter readings, and water volume with time in Data Set 1 for different tank sizes: (a) 0.075, (b) 0.2, and (c) 0.61 m<sup>3</sup>.

that  $Q_{FCV}$  is equal to 0 (as can be seen at the end of Day 1 in Figure 5a). Similar behavior is observed for a larger tank size of 0.2 m<sup>3</sup> (Figure 5b); however, a smoother behavior is observed for the water volume in the tank and the inflow through the smart meter. As expected, a lower NMI value of 0.0638 is achieved. Note that the size of a typical residential water heater for three to four family size ranges between 0.19 and 0.3 m<sup>3</sup>, which is comparable in size to the proposed precharged water tank. In the last case, with 0.61 m<sup>3</sup> tank (Figure 5c) NMI reduces to 0. We observe that the tank is never fully empty or full, and the inflow is maintained at constant value, that is,  $Q_{FCV}$  is always equal to  $Q_{REF}$ . Similar plots are provided for the rest of the households in Figures S1–S3 in the supporting information (SI).

In addition to NMI, we evaluate the level of privacy by deploying the occupancy detection algorithm and evaluating the detection for different tank sizes. We start by detecting the occupancy using the original meter data without the PPS, setting the epoch time to 2 hr during the day and 5 hr during the night. In other words, if zero flow was recorded by the water meter for over 2 hr during the day or 5 hr during the night, it was assumed that the residence is not occupied. Figure 6 shows the occupancy detection for Data Set 1 for the three different tank sizes: 0.075 (PPS1), 0.2 (PPS2), and 0.61 (PPS3) m<sup>3</sup> with the FF moderation algorithm.



**Figure 6.** Occupancy detection in Data Set 1 for different tank sizes: (PPS1) 0.075, (PPS2) 0.2, and (PPS3) 0.61 m<sup>3</sup> with the FF moderation algorithm.

**Table 3**  
Best Effort Moderation Algorithm

**Algorithm 2:** BE algorithm

<b>1: Input:</b>	Water demand $Q_D$ , size of the precharged tank, $V_{MAX}$ and $V_{MIN}$
<b>2: Output:</b>	Water meter readings $Q_{FCV}$
<b>3: Initialize:</b>	$t = t_0, V(t) = V_0, Q_{FCV}(t), Q_D(t)$
<b>4: Set:</b>	Flow through the control valve $Q_{FCV}(t+\Delta t)$ : <b>if</b> tank is full $V(t) = V_{MAX}$ set flow to 0 <b>elseif</b> tank is not full $V_{MIN} < V(t) < V_{MAX}$ set flow to $Q_{FCV}(t)$ <b>else</b> tank is empty $V(t) = V_{MIN}$ set flow to $Q_D(t)$
<b>5: Update:</b>	Water volume in tank $V(t+\Delta t) = (Q_{FCV}(t) - Q_D(t))\Delta t + V(t)$
<b>6: Repeat:</b>	$t = t+\Delta t$

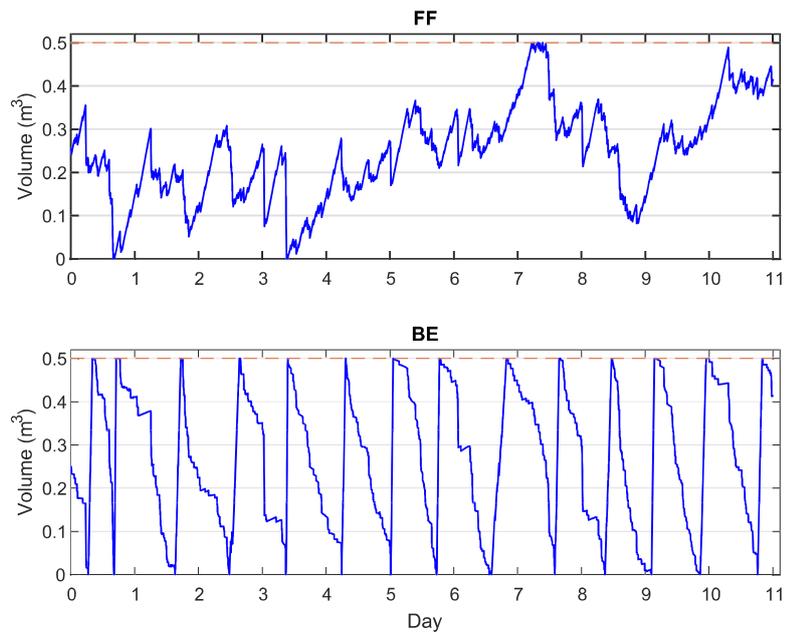
The true occupancy is shown at the bottom, where the dark and light colors represent time periods with and without occupancy, respectively. It is evident that the PPS can hedge against occupancy information leakage even for a small-size tank (i.e., PPS1). As expected, for a larger size tank (i.e., PPS3) the occupancy is almost entirely hidden.

**3.3. Comparison With the Best Effort Algorithm**

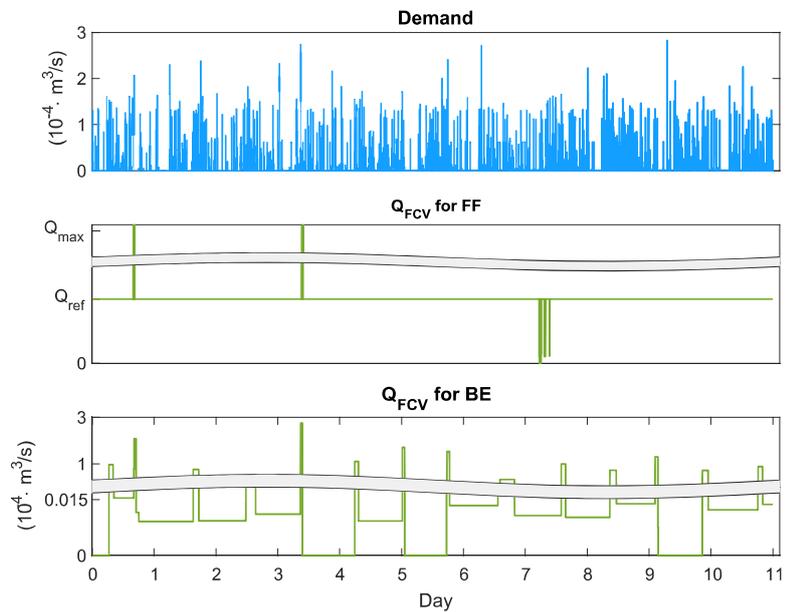
Next, we compare the performance of the proposed FF privacy moderation algorithm, with the best effort algorithm (BE) previously proposed by Kalogridis et al. (2010) for moderating home electricity load signature to hide the home appliance usage information. The BE algorithm aims at minimizing the change in the metered value exposed to the utility between two successive time steps, that is, minimize

$\Delta Q_{FCV}(t) = Q_{FCV}(t) - Q_{FCV}(t - \Delta t)$ . In this case,  $Q_{FCV}$  only changes when the tank is empty or full and then it is assigned the value of the exact demand,  $Q_D$ . Unlike, the proposed FF algorithm, which tries to maintain a constant inflow, the BE algorithm tries to adjust the current inflow to the inflow of the previous time step (i.e., in a sense BE algorithm uses dynamic reference flow as opposed to a fixed reference as in the FF algorithm). The BE control actions are summarized in Table 3.

Figure 7 shows the variation in the tank water volumes applying the FF and BE algorithms given a  $0.5 \text{ m}^3$  size tank for Data Set 3. Under the FF control, the tank empties and fills only several times during the simulation period, whereas using the BE algorithm, the tank empties and fills frequently on a (almost) daily cycle. Thus, the FF algorithm achieves a greater privacy hedging compared to the BE algorithm, NMI of 0.0059 and 0.0838 for the FF and the BE algorithms, respectively. Similar results are observed for the rest of the households (see Figures S4–S6 in the SI). Note that the less frequent emptying and filling cycles in the FF algorithm do not imply a lower rate of water replenishment in the tank, that is, keeping the volume constant does not necessarily imply that the same water remains in the tank during the entire time period. Indeed,



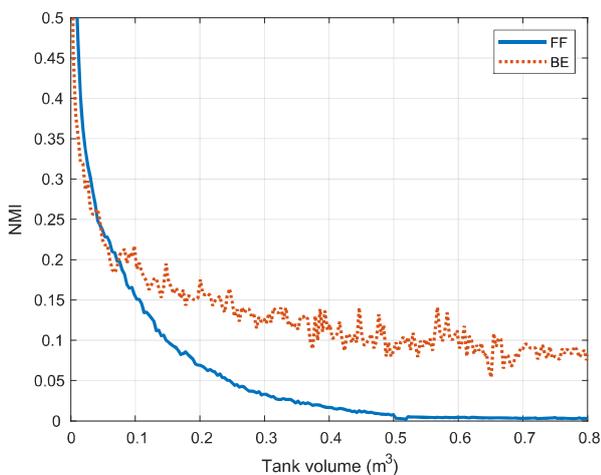
**Figure 7.** Changes in water volume of a  $0.5 \text{ m}^3$  tank resulting from applying the FF (top) and BE (bottom) algorithms to Data Set 3.



**Figure 8.** Demand and changes in smart meter readings for Data Set 3 resulting from applying the FF and BE algorithms with  $0.5 \text{ m}^3$  tank.

the high-frequency fluctuations observed in the FF water volume are the true indication of water replenishment in the tank (see Figure 7).

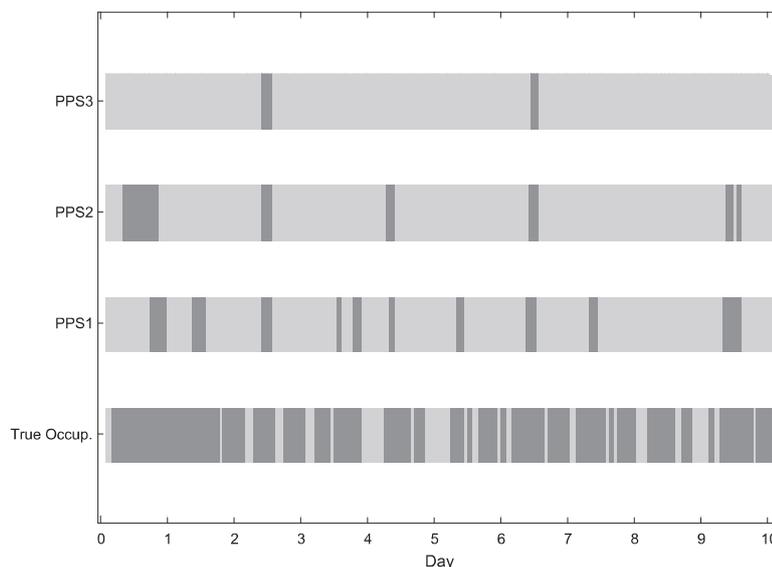
For the same tank size of  $0.5 \text{ m}^3$ , Figure 8 shows the user's demand and the metered inflow,  $Q_{FCV}$ , for the FF (top) and BE (bottom) controls. While the FF flow maintains the flat value of  $Q_{REF}$  the majority of the time, the BE flow fluctuates in a wider range, which, in turn, increases the NMI between the user's demand and the water meter's readings. Similar results are observed for the rest of the households (see Figures S7–S9 in the SI).



**Figure 9.** NMI for the FF and BE algorithms for Data Set 3.

To compare the performance of the FF algorithm with the BE algorithm, we have incrementally increased the size of the water tank between 0 and  $0.8 \text{ m}^3$ , while applying the FF and BE algorithms. Figure 9 presents the resulting NMI for Data Set 3 (and Figures S10–S12 in the SI illustrate similar results for the rest of the data sets). The results show that for very small tank sizes, the BE algorithm provides lower NMI, and thus better privacy hedging, for tanks larger than  $0.08 \text{ m}^3$  (approximately 10% of average daily demand), the FF algorithm significantly outperforms the BE algorithm.

Figure 10 shows the occupancy detection for the three tank sizes, as previously discussed. Comparing the performance of the two privacy moderation algorithms (FF in Figure 6 and BE in Figure 10), we observe that the two algorithms mask occupancy in different ways. For example, examining the PPS3 case, the BE algorithm will show that the household is not occupied most of the time, while the FF algorithm will show occupancy at all the times. On one hand, the two algorithms are comparable in the sense that both mask occupancy. On the other hand, we assert that if our goal is to prevent theft when the household is not occupied then the FF algorithm outperforms the BE algorithm by showing occupancy most of the time.



**Figure 10.** Occupancy detection in Data Set 1 for different tank sizes: (PPS1) 0.075, (PPS2) 0.2, and (PPS3) 0.61 m<sup>3</sup> with the BE moderation algorithm.

#### 4. Discussion

Smart water meters can facilitate the efficient management of vital water resources in the urban environment. Advances in sensing technologies, machine learning techniques, and surging data availability are enabling new ways for realizing smart water metering infrastructures, which can provide improved understanding of consumer behavior and more efficient process management (e.g., leakage detection and improved water demand forecasting for operational purposes). Due to the consensus around the benefits smart water meters provide, many studies (e.g., Clifford et al., 2018; Cominola et al., 2019; Davies et al., 2014) have mainly focused on data analysis and modeling of smart meters from the water utility perspective. Nonetheless, we assert that the social aspects, and more specifically the water user perspective, should be taken into account before wide adoption of smart water systems can be materialized. Introducing user privacy protection technology is vital to relieve public resistance to the adoption of new technology and to promote widespread implementation of smart metering infrastructure, which is needed to enhance the management of urban water systems.

Shifting the paradigm from strictly economic and environmental objectives, social awareness is an increasingly applied concept in planning, management, and operations of water supply systems (Packett et al., 2020; Schaider et al., 2019; Vanderslice, 2011). However, studies including social awareness of assessing privacy concerns as related to advanced water metering infrastructure are scarce. Furthermore, lack of specific information and difficulties in linking the performance of smart metering technology to privacy breach implications on end users hamper the inclusion of privacy awareness in planning and management of urban water supply systems. Water utilities can use the privacy measure proposed in this work in order to integrate the social considerations, which may be otherwise overlooked, into planning and management of smart metering infrastructure. This information can additionally be used by the water utilities to inform consumers about the hidden implications of advanced metering systems. On the consumer side, this information can be used to evaluate their privacy concerns. Furthermore, some users might not oppose to smart meters installation at their household (depending on their subjective valuation of their privacy), while others might oppose to the smart meters realizing the hidden privacy implications. Different states in the United States, as well as countries around the world, employ different opt-out policies for smart meter installation. These include no opt-out policies in place, case-by-case opt-out programs, requiring end user consent, and prohibiting opt-out (Citizens Advice, 2020; King & Jessen, 2014; Shea & Bell, 2019). In the event that the consumer cannot prevent the installation of a smart meter, the PPS proposed in this work can be a potential solution

for the consumer. The current study presents an initial analytical investigation of a privacy protection system; thus, it provides an entry point toward privacy-sensitive design and implementation of smart metering systems.

Our results indicate that the proposed PPS can yield significant privacy hedging even when a small-size tank is used. A significant reduction in NMI (from 1 to 0.1) is achieved with a small tank volume of  $0.2 \text{ m}^3$  (Figure 4) as well as increased OD protection (Figure 6). Since the tank size is directly correlated with the cost of the PPS, there is an advantage for using smaller tanks in the system. The results show that the PPS obtained better privacy hedging when managed by the proposed FF privacy moderation algorithm when compared with the BE algorithm (NMI of 0.07 compared to 0.15 for a tank volume of  $0.2 \text{ m}^3$  as seen in Figure 9). Another key advantage of the FF algorithm is its ease of implementation. Note that the FCV in the FF control rule should take only one of three values  $\{0, Q_{REF}, Q_{MAX}\}$ , while in the BE algorithm it can take any continuous value in the range  $[0, Q_{MAX}]$ . This property impacts the type of the hardware required in the system. That is, the FF control rule requires an actuator that can work with three discrete control inputs, whereas the BE control rule requires an actuator with continuous control inputs. Aside from the higher complexity in the continuous setup, it is also expected to be more prone to failures. Thus, the PPS with a continuous control logic is expected to impose higher capital and maintenance cost.

Noteworthy is that implementing the proposed PPS does not necessarily impair all the benefits of smart meters. Smart meters are expected to improve demand management by utilizing customer activity to alert of abnormal water use such as bursts and leaks in the premises. Readily available smart meters management systems use simple alerts for water overuse (Lloyd Owen, 2018), such benefits of the smart meters are not necessarily lost if the proposed PPS is installed. For example, even in the presence of the PPS, anomaly detection algorithms can still be applied using aggregated water consumption data (e.g., several hours in the day) to infer abnormal water use (it should be noted that small leaks, in the magnitude of the demand hourly variation, may be difficult to detect). On the other hand, the PPS installation could impair the effectiveness of other demand management strategies that require disaggregating of high-resolution water demand data into specific water use events. Several studies (e.g., Jessoe & Rapson, 2014; Zhuang & Sela, 2020) argued that water use events at the end user level could help in understanding the influence of technology-, policy-, and price-based strategies on consumers' behavior and assessing their effectiveness.

Wide adaption of the PPS may have positive externalities for the utilities as well as the end users. From the water utility side, the tanks in the PPS serve as surplus storage in the system. This, microstorage, distributed storage can add a significant storage capacity to the system when widely adapted by end users. Adding storage capacity to the system can advantageously flatten the demand curve, which contributes to a better water-energy nexus management in the context of optimized tariff-based pumping. Moreover, flattened demand patterns dampen peak flows, shift loads to off-peak hours, and, thus, can lead to reduced head losses and pressures, which in turn affect water losses, energy demand for pumping, treatment, and distribution, and associated costs (Plappally & Lienhard V, 2012). In the long run, reducing the peak flow can postpone system expansion and rehabilitation (Beal & Stewart, 2014; Beal et al., 2016; Gurung et al., 2015). On the end user side, this microstorage can contribute to increased reliability in case of water outages and act as a protective physical barrier between the water supply network and the premise plumbing (e.g., during pressure transient events).

Our results build on the NMI as an information-theoretic metric to measure the amount of inherent information available for exploitation by a privacy intruder. Although NMI (and its variant MI) is widely adopted as a privacy measure, due to its distinct feature of measuring privacy regardless of the operational details of the intruder, it does not account for the temporal correlation between the data readings (Ma & Yau, 2015; Yang et al., 2012). Temporal correlation is expected in the end user water demand, as the consumption at time  $t$  is likely to be correlated with consumption at time  $t - 1$ . In general, an attacker might be able to exploit such correlation. To address this issue, Yang et al. (2012) considered the MI between distributions of consecutive pairs of readings instead of individual elements. Nevertheless, noting that our FF algorithm produces flat flow for the majority of the time steps (see Figure 5), the temporal correlation in the metered data is very small, indicating that a low NMI will be achieved in the system even when calculated based on consecutive pairs of readings. Future research directions should address temporal correlations as well as additional privacy measures for evaluating the performance of privacy hedging controls.

## 5. Conclusions

As smart meters installations gain popularity with utilities providers across different sectors, consumers' concerns of possible privacy infringement are growing as well. While the energy sector has been primarily leading the research for smart electricity meters, the water sector is lagging behind. This research aims to promote the discussion and research on smart meters in the water sector. To this end, we propose a new and practical technology coupling a hardware apparatus with a software control model for hedging personal privacy in the presence of a smart water meter. The proposed physical PPS can be also used to investigate different control algorithms in the context of multiobjective analysis that trade-off privacy hedging with other performance metrics (e.g., reliability).

The new approach is demonstrated using real water demand data, and the results show a significant increase in privacy protection even with a small-size water tank. We quantified the trade-off of the level of privacy protection and the size of the system (as surrogate for system's cost) and compare the system performance under different privacy moderation algorithms using information-theoretic (NMI) and empirically based (OD) criteria. Nonetheless, our gross estimate for midsized system is below U.S. \$1,000 (a basic cost estimate of the main components is outlined in Text S1 in the SI). Further research is needed to address the discussed limitations including in-depth economic analysis and the feasibility of the proposed PPS, implementing more interpretable privacy measures, and consumers' willingness to pay for added privacy.

## Data Availability Statement

Data are available through DeOreo et al. (2016).

## Acknowledgments

This study was supported by the Israeli Water Authority and by the Center for Cyber Law & Policy at the University of Haifa in conjunction with the Israel National Cyber Directorate in the Prime Minister's Office. This work was also supported by the University of Texas at Austin Startup Grant and by Cooperative Agreement 83595001 awarded by the U.S. Environmental Protection Agency (EPA) to The University of Texas at Austin. This work has not been formally reviewed by EPA. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the Agency. EPA does not endorse any products or commercial services mentioned in this publication.

## References

- Allwood, P. B., Jenkins, T., Paulus, C., Johnson, L., & Hedberg, C. W. (2004). Hand washing compliance among retail food establishment workers in Minnesota. *Journal of Food Protection*, 67(12), 2825–2828. <https://doi.org/10.4315/0362-028X-67.12.2825>
- Anderson, J. L., Warren, C. A., Perez, E., Louis, R. I., Phillips, S., Wheeler, J., et al. (2008). Gender and ethnic differences in hand hygiene practices among college students. *American Journal of Infection Control*, 36(5), 361–368. <https://doi.org/10.1016/j.ajic.2007.09.007>
- Asghar, M. R., Dán, G., Miorandi, D., & Chlamtac, I. (2017). Smart meter data privacy: A survey. *IEEE Communication Surveys and Tutorials*, 19(4), 2820–2835. <https://doi.org/10.1109/COMST.2017.2720195>
- Barth, S., & de Jong, M. D. T. (2017). The privacy paradox—Investigating discrepancies between expressed privacy concerns and actual online behavior—A systematic literature review. *Telematics and Informatics*, 34(7), 1038–1058. <https://doi.org/10.1016/j.tele.2017.04.013>
- Beal, C. D., Gurung, T. R., & Stewart, R. A. (2016). Demand-side management for supply-side efficiency: Modeling tailored strategies for reducing peak residential water demand. *Sustainable Production and Consumption*, 6(January), 1–11. <https://doi.org/10.1016/j.spc.2015.11.005>
- Beal, C. D., & Stewart, R. A. (2014). Identifying residential water end uses underpinning peak day and peak hour demand. *Journal of Water Resources Planning and Management*, 140(7), 1–10. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000357](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000357)
- Chen, D., Kalra, S., Irwin, D., Shenoy, P., & Albrecht, J. (2015). Preventing occupancy detection from smart meters. *IEEE Transactions on Smart Grid*, 6(5), 2426–2434. <https://doi.org/10.1109/TSG.2015.2402224>
- Chen, F., Dai, J., Wang, B., Sahu, S., Naphade, M., & Lu, C.-T. (2011). *Activity analysis based on low sample rate smart meters* (pp. 240–248). Paper presented at Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD'11, Association for Computing Machinery, San Diego, CA. <https://doi.org/10.1145/2020408.2020450>
- Citizens Advice. (2020). *Getting a smart meter installed*. UK: Citizens Advice. Retrieved from <https://www.citizensadvice.org.uk/consumer/energy/energy-supply/your-energy-meter/getting-a-smart-meter-installed/>
- Clifford, E., Mulligan, S., Comer, J., & Hannon, L. (2018). Flow-signature analysis of water consumption in nonresidential building water networks using high-resolution and medium-resolution smart meter data: Two case studies. *Water Resources Research*, 54, 88–106. <https://doi.org/10.1002/2017WR020639>
- 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. *Environmental Modelling and Software*, 102, 199–212. <https://doi.org/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. *Environmental Modelling and Software*, 72, 198–214. <https://doi.org/10.1016/j.envsoft.2015.07.012>
- Cominola, A., Nguyen, K., Giuliani, M., Stewart, R. A., Maier, H. R., & Castelletti, A. (2019). Data mining to uncover heterogeneous water use behaviors from smart meter data. *Water Resources Research*, 55, 9315–9333. <https://doi.org/10.1029/2019WR024897>
- Davies, K., Doolan, C., Van Den Honert, R., & Shi, R. (2014). Water-saving impacts of smart meter technology: An empirical 5 year, whole-of-community study in Sydney, Australia. *Water Resources Research*, 50, 7348–7358. <https://doi.org/10.1002/2014WR015812>
- DeOreo, W. B., Mayer, P., Dziegielewski, B., & Kiefer, J. (2016). *Residential end uses of water, Version 2*. Denver, CO: Water Research Foundation.
- Depuru, S. S. R., Wang, L., & Devabhaktuni, V. (2011). Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft. *Energy Policy*, 39(2), 1007–1015. <https://doi.org/10.1016/j.enpol.2010.11.037>
- European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. *Official Journal of the European Union*. Retrieved from <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

- Giaconi, G., Gunduz, D., & Poor, H. V. (2015). *Smart meter privacy with an energy harvesting device and instantaneous power constraints* (pp. 7216–7221). Paper presented at IEEE International Conference on Communications, IEEE, London, UK. <https://doi.org/10.1109/ICC.2015.7249478>
- Gurung, T. R., Stewart, R. A., Beal, C. D., & Sharma, A. K. (2015). Smart meter enabled water end-use demand data: Platform for the enhanced infrastructure planning of contemporary urban water supply networks. *Journal of Cleaner Production*, *87*(1), 642–654. <https://doi.org/10.1016/j.jclepro.2014.09.054>
- Gurung, T. R., Stewart, R. A., Beal, C. D., & Sharma, A. K. (2017). Smart meter enabled informatics for economically efficient diversified water supply infrastructure planning. *Journal of Cleaner Production*, *163*, S138–S147. <https://doi.org/10.1016/j.jclepro.2017.05.140>
- Gurung, T. R., Stewart, R. A., Sharma, A. K., & Beal, C. D. (2014). Smart meters for enhanced water supply network modelling and infrastructure planning. *Resources, Conservation and Recycling*, *90*, 34–50. <https://doi.org/10.1016/j.resconrec.2014.06.005>
- Horsburgh, J. S., Leonardo, M. E., Abdallah, A. M., & Rosenberg, D. E. (2017). Measuring water use, conservation, and differences by gender using an inexpensive, high frequency metering system. *Environmental Modelling and Software*, *96*, 83–94. <https://doi.org/10.1016/j.envsoft.2017.06.035>
- Jamieson, A. (2009). Smart meters could be “spy in the home”. In *The telegraph, energy*. UK: The Telegraph. Retrieved from <https://www.telegraph.co.uk/finance/newsbysector/energy/6292809/Smart-meters-could-be-spy-in-the-home.html>
- Jawurek, M., Kerschbaum, F., & Danezis, G. (2012). *Privacy technologies for smart grids—A survey of options*. USA: Jerome, S. (2017). *Smart water meter considered evidence in murder case*. USA: Water Online.
- Jessoe, K., & Rapson, D. (2014). Knowledge is (less) power: Experimental evidence from residential energy use. *American Economic Review*, *104*(4), 1417–1438. <https://doi.org/10.1257/aer.104.4.1417>
- Kalogridis, G., Efthymiou, C., Denic, S. Z., Lewis, T. A., & Cepeda, R. (2010). *Privacy for smart meters: Towards undetectable appliance load signatures* (pp. 232–237). Paper presented at 2010 First IEEE International Conference on Smart Grid Communications, IEEE, Gaithersburg, MD. <https://doi.org/10.1109/SMARTGRID.2010.5622047>
- King, N. J., & Jessen, P. W. (2014). For privacy’s sake: Consumer “opt outs” for smart meters. *Computer Law and Security Review*, *30*(5), 530–539. <https://doi.org/10.1016/j.clsr.2014.07.001>
- Koo, J., Lin, X., & Bagchi, S. (2012). PRIVATUS: Wallet-friendly privacy protection for smart meters. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 343–360). Berlin: Springer. [https://doi.org/10.1007/978-3-642-33167-1\\_20](https://doi.org/10.1007/978-3-642-33167-1_20)
- Lloyd Owen, D. A. (2018). *Smart water: Data capture and analysis for sustainable water management* (Ch. 4). USA: Wiley Blackwell. <https://doi.org/10.1002/9781119078678>
- Ma, C. Y. T., & Yau, D. K. Y. (2015). *On information-theoretic measures for quantifying privacy protection of time-series data* (pp. 427–438). Paper presented at ASIACCS 2015—Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security, Association for Computing Machinery, New York, NY. <https://doi.org/10.1145/2714576.2714577>
- McKenna, E., Richardson, I., & Thomson, M. (2012). Smart meter data: Balancing consumer privacy concerns with legitimate applications. *Energy Policy*, *41*, 807–814. <https://doi.org/10.1016/j.enpol.2011.11.049>
- McLaughlin, S., McDaniel, P., & Aiello, W. (2011). *Protecting consumer privacy from electric load monitoring* (pp. 87–98). Paper presented at Proceedings of the ACM Conference on Computer and Communications Security, Association for Computing Machinery, New York, NY. <https://doi.org/10.1145/2046707.2046720>
- Nguyen, K. A., Stewart, R. A., & Zhang, H. (2013). An intelligent pattern recognition model to automate the categorisation of residential water end-use events. *Environmental Modelling and Software*, *47*, 108–127. <https://doi.org/10.1016/j.envsoft.2013.05.002>
- Nguyen, K. A., Stewart, R. A., Zhang, H., & Sahin, O. (2018). Environmental Modelling & Software re-engineering traditional urban water management practices with smart metering and informatics. *Environmental Modelling and Software*, *101*, 256–267. <https://doi.org/10.1016/j.envsoft.2017.12.015>
- Nguyen, K. A., Stewart, R. A., Zhang, H., & Jones, C. (2015). Intelligent autonomous system for residential water end use classification: Autoflow. *Applied Soft Computing Journal*, *31*, 118–131. <https://doi.org/10.1016/j.asoc.2015.03.007>
- Packett, E., Grigg, N. J., Wu, J., Cuddy, S. M., Wallbrink, P. J., & Jakeman, A. J. (2020). Mainstreaming gender into water management modelling processes. *Environmental Modelling and Software*, *127*, 104683. <https://doi.org/10.1016/j.envsoft.2020.104683>
- Pentair. (2020). *Pentair Flotec pressure tanks*. London, UK: Pentair.
- Pittet, D. (2001). Improving adherence to hand hygiene practice: A multidisciplinary approach. *Emerging Infectious Diseases*, *7*(2), 234–240. <https://doi.org/10.3201/eid0702.010217>
- Plappally, A. K., & Lienhard V, J. H. (2012). Energy requirements for water production, treatment, end use, reclamation, and disposal. *Renewable and Sustainable Energy Reviews*, *16*(7), 4818–4848. <https://doi.org/10.1016/j.rser.2012.05.022>
- Quinn, E. L. (2008). *Privacy and the new energy infrastructure*. Rochester, NY: Center for Energy and Environmental Security (CEES). <https://doi.org/10.2139/ssrn.1370731>
- Rajagopalan, S. R., Sankar, L., Mohajer, S., & Poor, H. V. (2011). *Smart meter privacy: A utility-privacy framework* (pp. 190–195). Paper presented at 2011 IEEE International Conference on Smart Grid Communications, SmartGridComm 2011, IEEE, Brussels, Belgium. <https://doi.org/10.1109/SmartGridComm.2011.6102315>
- Rial, A., & Danezis, G. (2011). Privacy-preserving smart metering. In *Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society* (pp. 49–60). New York, NY, USA: ACM. <https://doi.org/10.1145/2046556.2046564>
- Schaidler, L. A., Swetschinski, L., Campbell, C., & Rudel, R. A. (2019). Environmental justice and drinking water quality: Are there socioeconomic disparities in nitrate levels in U.S. drinking water? *Environmental Health: A Global Access Science Source*, *18*(1). <https://doi.org/10.1186/s12940-018-0442-6>
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Shea, D., & Bell, K. (2019). *Smart meter opt-out policies*. Washington, DC: National Conference of State Legislatures (NCSL). Retrieved from <https://www.ncsl.org/research/energy/smart-meter-opt-out-policies.aspx>
- Smart Energy International. (2018). *Global smart water meters market trends*. South Africa: Smart Energy International.
- Solove, D. J. (2010). *Understanding privacy*. Cambridge, MA: Harvard University Press. Retrieved from <http://www.hup.harvard.edu/catalog.php?isbn=9780674035072>
- Souri, H., Dhraief, A., Tlili, S., Drira, K., & Belghith, A. (2014). Smart metering privacy-preserving techniques in a nutshell. *Procedia Computer Science*, *32*(Ramcom), 1087–1094. <https://doi.org/10.1016/j.procs.2014.05.537>

- Stewart, R. A., Nguyen, K., Beal, C., Zhang, H., Sahin, O., Bertone, E., et al. (2018). Integrated intelligent water-energy metering systems and informatics: Visioning a digital multi-utility service provider. *Environmental Modelling and Software*, *105*, 94–117. <https://doi.org/10.1016/j.envsoft.2018.03.006>
- Tan, O., Gunduz, D., & Poor, H. V. (2013). Increasing smart meter privacy through energy harvesting and storage devices. *IEEE Journal on Selected Areas in Communications*, *31*(7), 1331–1341. <https://doi.org/10.1109/JSAC.2013.130715>
- United States Court of Appeals. (2018). *Naperville Smart Meter Awareness vs. City of Naperville*. Illinois, USA: United States Court of Appeals.
- Vanderslice, J. (2011). Drinking water infrastructure and environmental disparities: Evidence and methodological considerations. *American Journal of Public Health*, *101*(SUPPL. 1), S109–S114. <https://doi.org/10.2105/AJPH.2011.300189>
- Varodayan, D., & Khisti, A. (2011). *Smart meter privacy using a rechargeable battery: Minimizing the rate of information leakage* (pp. 1932–1935). Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, Prague, Czech Republic. <https://doi.org/10.1109/ICASSP.2011.5946886>
- Vasak, M., Banjac, G., & Novak, H. (2015). *Water use disaggregation based on classification of feature vectors extracted from smart meter data* (Vol. 119, pp. 1381–1390). Paper presented at 13th Computer Control for Water Industry Conference, CCWI 2015, ScienceDirect, Amsterdam, The Netherlands. <https://doi.org/10.1016/j.proeng.2015.08.992>
- Véliz, C., & Grunewald, P. (2018). Protecting data privacy is key to a smart energy future. *Nature Energy*, *3*(9), 702–704. <https://doi.org/10.1038/s41560-018-0203-3>
- Wagner, I., & Eckhoff, D. (2018). Technical privacy metrics: A systematic survey. *ACM Computing Surveys*, *51*(3), 1–38. <https://doi.org/10.1145/3168389>
- Weaver, K. T. (2014). *A perspective on how smart meters invade individual privacy*. Naperville, IL: SkyVision Solutions. Retrieved from <http://skyvisionsolutions.org>
- Wu, T., Xue, Y., & Cui, Y. (2006). *Preserving traffic privacy in wireless mesh networks* (Vol. 2006, pp. 459–461). Paper presented at Proceedings of the 2006 International Symposium on a World of Wireless, Mobile and Multimedia (WoWMoM) Networks, IEEE, Buffalo-Niagara Falls, NY. <https://doi.org/10.1109/WOWMOM.2006.91>
- Yang, W., Li, N., Qi, Y., Qardaji, W., McLaughlin, S., & McDaniel, P. (2012). *Minimizing private data disclosures in the smart grid* (pp. 415–427). Paper presented at Proceedings of the ACM Conference on Computer and Communications Security, Association for Computing Machinery, Raleigh, NC. <https://doi.org/10.1145/2382196.2382242>
- Ząbkowski, T., & Gajowniczek, K. (2013). Smart metering and data privacy issues. *Information Systems Management*, *2*(3), 239–249. <https://www.infona.pl/resource/bwmeta1.element.baztech-6bec03b7-b7eb-4635-9163-aa395fbf253d>
- Zhao, J., Jung, T., Wang, Y., & Li, X. (2014). *Achieving differential privacy of data disclosure in the smart grid* (pp. 504–512). Paper presented at Proceedings of the IEEE INFOCOM, IEEE, Toronto, ON. <https://doi.org/10.1109/INFOCOM.2014.6847974>
- Zhuang, J., & Sela, L. (2020). Impact of emerging water savings scenarios on performance of urban water networks. *Journal of Water Resources Planning and Management*, *146*(1), 04019063. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001139](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001139)
- Zipper, S. C., Stack Whitney, K., Deines, J. M., Befus, K. M., Bhatia, U., Albers, S. J., et al. (2019). Balancing open science and data privacy in the water sciences. *Water Resources Research*, *55*, 5202–5211. <https://doi.org/10.1029/2019WR025080>

## **General Discussion**

### **Research goals**

This research aimed to explore and create new methodologies for optimal and practical real-time operation of WDS. Two control schemes are considered: local and central. The local control scheme will be applied to a small operating zone with a small number of pumps and tanks, thus providing a practical solution for small-scale WDS. It utilizes simple mathematical algorithms which can run on a local PLC thus eliminating the need for high-performance computational hardware and software. The central control scheme is aimed to run at a control center and optimally operate larger WDS. Two main strategies were adapted to seek practical tools. First, considering real-world WDSs, it was shown that the explicit hydraulic behavior can be excluded from the optimization formulation thus eliminating the non-linearity of the problem. Second, the size of the optimization problem was reduced by using wise binary coding of the discrete decision variables in the optimization problem. This approach results in a MILP formulation with a relatively small number of Integer variables. These tools can provide a practical solution that can help large- as well as small-scale water utilities in utilizing advanced optimization methods to gain energy savings and more environmentally friendly operation strategies.

### **Research contribution**

To date, only a limited number of water utilities use a closed-loop optimal pump operations control scheme. Many small water utilities do not have the know-how and the technical personnel to operate sophisticated optimization models. This research contributes to the water sector by developing practical methods for optimal operation of WDSs, the developed tools cover both small-scale and large-scale systems. From an academic perspective, previous work focused mainly on the open-loop operation while this research concentrates on a closed-loop control scheme with online feedback from the system in a rolling horizon mode. The tradeoff between the operation practicality and optimality has not been explored before in the literature since most of the previous work concentrated on the optimal solution for a given operation horizon without simulating the real-time behavior of the system in a close control loop.

## **Limitation and future work**

This research work is inspired by the overgrowing need for water utilities for practical operating tools. To that end, the methods, algorithms, and tools developed herein, aim to satisfy this need. However, this still is academic research, and the developed tools were not implemented, or tested, in a real-world operating environment which is a crucial requirement step before being used by water utilities. The main foreseen limitations of the methodologies developed are associated with the availability and quality of the required operating data in real-time. Data availability and its quality start with the single sensor (e.g., water level, pressure, pumps status, and flows). These digital and analog signals are then transferred to a PLC, or to a remote terminal unit (RTU), which can do some data processing (e.g., transfer the sensors' 4-20 mA output into the correct physical values). Then, the data is transmitted to other sites (e.g., central control center or neighboring facilities) via wired or wireless communication channels. Finally, the data arrives at the remote site or to the SCADA system in which further data processing and archiving are done. Only at this stage, the optimization program at the central control center can access the data. This long chain of data measurement, transfer, processing, and archiving is prone to data loss and distortion which may affect the performance of the optimization and its outcome. Thus, data validation and estimation are required to assure its availability and quality.

In the control loop, after each optimization run, the decision for the first time step (i.e., one hour) is implemented and the system waits for the time step to pass before initiating the next optimization process. However, during this time step, some events may occur in the system that requires intervention. For example, tanks water levels can violate predefined constraints due to unforeseen low or high demand, pump failures, etc. To handle such cases, a system monitor component should be developed.

In recent years, the energy sector has changed in two major aspects. The first is attributed to the growing share of diverse energy sources, mainly renewable sources (e.g., wind, and solar), and the second is attributed to the interconnectivity of large power grids across countries and continents. As a result, an advanced energy market has emerged. Nowadays, energy became a tradable commodity with diverse Spot prices and future contracts. Spot markets are public financial markets where commodities are traded for immediate delivery, while future contracts are agreements to trade commodities at a predetermined price and time in the future. WDSs are identified as one of the largest energy consumers. WDSs are also characterized by their ability to use water storage facilities for shifting energy use patterns through strategic pumping. Thus, water utilities can take advantage of the competitive energy market to reduce their operational

cost and increase profitability. In doing so, water utilities are looking for a delicate balance between the strategic purchase of long-term (e.g., monthly, yearly) energy contracts and real-time scheduling of Spot market energy purchases (e.g., the inter-day and day-ahead). Yet, decision-makers and operators in water utilities lack practical tools to support their actions in face of the new energy pricing structures. As such, the developed methods presented in this research should be extended to facilitate the inclusion of these new energy markets.

## Appendix I

*Sela, L., Salomons, E., and Housh, M. (2019). "Plugin prototyping for the EPANET software." Environmental Modelling & Software, Elsevier Ltd.*

<https://doi.org/10.1016/j.envsoft.2019.05.010>

Journal Impact Factor: 5.288

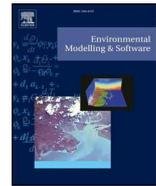
5-Year Impact Factor: 6.036

Publication Status: Published (Accepted 13 May 2019)



Contents lists available at ScienceDirect

## Environmental Modelling &amp; Software

journal homepage: [www.elsevier.com/locate/envsoft](http://www.elsevier.com/locate/envsoft)

## Plugin prototyping for the EPANET software

Lina Sela<sup>a,\*</sup>, Elad Salomons<sup>b</sup>, Mashor Housh<sup>b</sup><sup>a</sup> Department of Civil, Architectural and Environmental Engineering, University of Texas at Austin, United States<sup>b</sup> Department of Natural Resources and Environmental Management, University of Haifa, Israel

## ARTICLE INFO

## Keywords:

EPANET  
Plugins  
Graphical user interface

## ABSTRACT

Hydraulic simulation tools, such as EPANET, are the primary tools for evaluating water distribution systems performance. This work presents a first step towards demonstrating a simple and straightforward implementation of plugins in the new EPANET GUI to facilitate plugins development by the water systems modeling community. The paper shows the code structure and the basic functionalities of a custom plugin demonstrated using three specific examples: *Count*, *FireFlow*, and *Elevations* plugins. A prototype repository, where developers and users can share and download EPANET plugins is presented and discussed. EPANET plugins framework can support knowledge transfer by increasing the visibility and usability of developed analytical tools and software, thus providing benefits for researchers and practitioners. The proposed plugins are freely available through GitHub.

## Software availability

EPANET plugins are available from GitHub repository <https://github.com/eladsal/EPANET-Plugins/>

EPANET plugins prototype repository is available at [plugins.epanet.net](http://plugins.epanet.net)

System requirements EPANET-UI-MTP4r2.exe available at [https://github.com/USEPA/SWMM-EPANET\\_User\\_Interface/releases/tag/MTP4r2](https://github.com/USEPA/SWMM-EPANET_User_Interface/releases/tag/MTP4r2)

## 1. Introduction

Hydraulic simulation tools, such as EPANET (Rossman, 1994), are the primary tools for evaluating Water Distribution Systems (WDS) performance. The open-source EPANET is the most widely used software for WDS hydraulic and water quality analysis for commercial and research purposes. EPANET Graphical User Interface (GUI) enables the end-user to create and edit a water network model, run steady state and extended period simulations of the hydraulic and water quality dynamics in WDS. The first official release of EPANET by the U.S. Environmental Protection Agency (USEPA) was in 1993 and the last official release of EPANET was in 2008 (version 2.00.12). This version allows the user to interact with EPANET and perform hydraulic simulations through the GUI as well as through a Dynamic Link Library (DLL) of functions that enables the developers programmatically link EPANET engine with external software. In 2016, the Open Water

Analytics (OWA) community released a new EPANET version 2.1 under the Open Source Project, which provides some performance improvements, bug fixes, and usage features for the computational engine (Water Analytics, 2018).

Since its original release, EPANET has been used across a multitude of applications in WDS analysis, such as optimal design and operation (Kapelan et al., 2005; Murphy et al., 1994; Ormsbee and Lansey, 2006; Perelman et al., 2013; Savic and Walters, 2002; Xie et al., 2014; Zierolf et al., 2002), leak detection and localization (Boulos and Aboujaoude, 2011; Martínez-Solano et al., 2017; Whitman et al., 2018), sensor placement (Eliades et al., 2014; Phillips et al., 2008) and system security (Housh and Ohar, 2017, 2018). Over the years, several independent efforts have been administrated to improve and extend the modeling and simulation capabilities of EPANET. These efforts can be classified as enhancing the: (1) computational engine and (2) GUI of EPANET. This paper focuses on *enhancing EPANET capabilities through the GUI*. The majority of research efforts have been predominantly focused on enhancing the modeling and computational engine. Some examples include the EPANET-MSX (Shang et al., 2008) for modeling multi-species dynamics, EPANET-RTX (Hatchett et al., 2011) for real-time modeling extension and EPANET-PDX (Siew and Tanyimboh, 2012) for pressure-driven analysis in WDS. However, these and many other extensions were not incorporated into the core EPANET libraries nor into the GUI, and as a result, these projects are not widely adopted. In addition to enhancing the computational engine, several prior works created augmented versions of EPANET by creating new GUIs that resemble the

\* Corresponding author.

E-mail addresses: [linasela@utexas.edu](mailto:linasela@utexas.edu) (L. Sela), [selad@optiwater.com](mailto:selad@optiwater.com) (E. Salomons), [mshous@univ.haifa.ac.il](mailto:mshous@univ.haifa.ac.il) (M. Housh).<https://doi.org/10.1016/j.envsoft.2019.05.010>

Received 22 October 2018; Received in revised form 22 March 2019; Accepted 13 May 2019

Available online 16 May 2019

1364-8152/ © 2019 Elsevier Ltd. All rights reserved.

original EPANET GUI and include the new features, e.g. EPANET-BAM (Ho and Khalsa, 2009) to model incomplete mixing at cross-junctions, EPANET-Z (Zonum Solutions, 2009) to include online maps display, and IrrigatePlus (2017) to optimize irrigation system management. However, the usability of these tools is limited since some tools have not been released as open source and each tool requires a dedicated GUI, hence requiring many different versions of EPANET for each specialized functionality (Iglesias-Rey et al., 2017).

Noteworthy exception is the work by Iglesias-Rey et al. (2017), which presented an architecture to exchange information between EPANET and third-party applications, following the prototype of SWMM software architecture (Rossman, 2015). Another noteworthy effort is the new re-engineered EPANET GUI first released by the USEPA in March 2016 (SWMM-EPANET UI, 2018a; b). The new EPANET GUI, developed using Python scripting language, is intended to maintain the original functionalities of the EPANET software (e.g. run hydraulic and water quality simulations) as well as provide additional functionalities such as *scripting* and *plugins* management. This paper explores the *new plugins framework of EPANET GUI*.

Plugins are pieces of software, which extend the capabilities and features of an already existing software. Predominantly, the existing software is already compiled, and its source code is not affected by the added plugins. Plugins are available for many common desktop and web software. Web browsers (Google Chrome, 2018; Microsoft Edge, 2018) are a prominent example of software applications that allow users to extend its basic functionalities with additional features such as block ads, better protect privacy, or just personalize the look of the browser. Many free and commercial engineering software utilize a plugin platform, including ArcGIS, QGIS, and AutoCAD (AutoDesk.Inc 2015; Environmental Systems Research Institute (ESRI) 2011). Each software hosts its plugins in a different way but largely the plugins are centralized in one location. For example, the WordPress plugins are hosted in a plugins repository with over 56,000 items (WordPress, 2018), similarly to Mozilla Firefox (Mozilla, 2018) and the QGIS plugins (QGIS, 2016) that are hosted in a centralized repository. By having plugins in a centralized location, it makes it easier, on one hand, for the users to find, review and download plugins, and on the other hand, for the developers to maintain and support their plugins. Plugins framework has become a standard approach in many applications to add new features and enhance the basic capabilities.

The main two reasons for an application to support a plugin framework is to keep its size and complexity to a minimum and at the same time, to allow the addition of new features and capabilities. Furthermore, not all users need all the capabilities of all the plugins, as such users can select plugins relevant to their specific application or task. That is, plugins are *optional* and are not required to run the original application. Plugins framework enables third-party developers to develop new features thus reducing the burden of the main application developers. Plugins may even have a different software license than the main application. While the main application may be an open source software with a permissive software license (for example the MIT license (Open Source Initiative, 2018)), a plugin may have a more restrictive one, such as the GPL license (Negus, 2015). Furthermore, plugins for commercial software can be either commercial such as the Analytic Solver Optimization (Frontline Solvers, 2015) which extends the basic features of the Excel Solver Add-in, or free such as the YALMIP toolbox (Lofberg, 2004), which adds mathematical programming capabilities for Matlab. In general, plugins framework offers a systematic way to separate the software license and the business model of the main application from these of the added plugins.

Advanced hydraulic modeling and analytics tools coupled with a plugin framework within a user-familiar environment of EPANET create an opportunity to transform how new models and techniques are being developed, shared, and used in the water systems modeling community. Plugin framework has already transformed a number of other software applications in engineering, but the water systems modeling

community has been slow to adopt. Although myriad of modeling and analytics tools evolving around EPANET and hydraulic modeling have been developed, these typically remained limited to the domain of their developers and are not being widely used. Furthermore, none of the modeling advances have been incorporated in any official version of EPANET and little efforts have been made to make these advances accessible to the typical end-user that relies on the EPANET GUI. Hence, there is a gap between scientific progress and practical needs (Uber et al. (2018) Sela and Housh, 2019). The changing paradigm of a community-driven software development, as opposed to an individual-driven product development, increases the need and motivates a more collaborative development environment of models and software for water systems modeling. The new EPANET GUI plugins framework can help in bridging this gap and facilitates the integration of different tools developed in the water systems modeling community under the EPANET umbrella. By so doing, EPANET will leverage the different models and tools, which are continuously being developed by the water distribution systems analysis (WDSA) community, to benefit researchers and practitioners and make these tools accessible to a wider community of potential EPANET users.

The abovementioned new EPANET GUI (SWMM-EPANET UI, 2018a; b), developed as an open source project, is the first version of the software to include *plugins support framework*. The goal of this paper is two-fold: to motivate users/developers to use/develop plugins using the new framework and to motivate the further development and improvement of the plugin framework. We achieve our goal by developing three custom plugins that demonstrate the capabilities of the EPANET plugins framework for advanced hydraulic analysis and provide the complete codes for testing, reproduction, and development of new plugins functionalities. Initial applications utilizing the new EPANET GUI plugin environment were recently presented and discussed in the first joint WDSA-CCWI conference, Canada 2018 (Kandjani et al., 2018; Salomons et al., 2018). This paper presents the plugin framework through detailed implementation examples. Then, a prototype plugins repository is outlined, followed by conclusions and further development needs.

## 2. The plugin framework in the EPANET GUI

The new EPANET GUI is being developed in Python, (2018), a high-level, open-source programming language, coupled with QGIS, a free and open-source geographic information system application. The new EPANET GUI download and installation instructions as well as system design overview and source code can be found on the USEPA GitHub repository (SWMM-EPANET UI, 2018a) The new GUI, shown in Fig. 1,

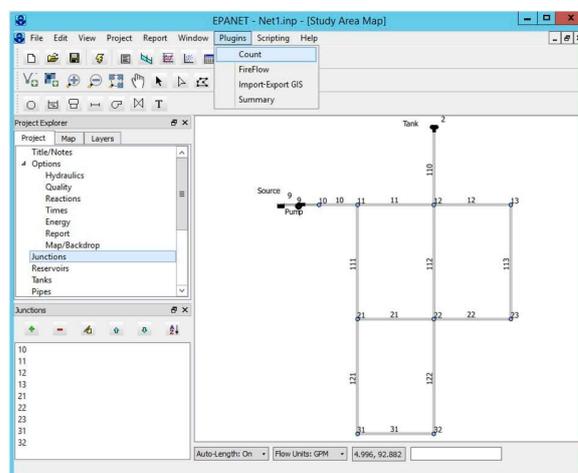


Fig. 1. Screenshot of the new EPANET GUI.

resembles that of EPANET 2.00.12 with only few minor exceptions, such as the navigation menu on the left-hand side of the screen and two new *Plugins* and *Scripting* options in the main menu toolbar. The downloadable files include several components related to the software, which are stored in relevant sub-directories under the top-level source directory, *EPANET-UI*, including *examples*, *scripts* and *plugins*.

The plugins are placed in the *plugins* sub-directory and each plugin is placed in a dedicated folder. Each plugin contains a set of mandatory settings and files and all plugins share a common set of management options that prescribe the core program how to communicate and control the plugin functionalities upon execution. Upon the initialization of the EPANET application, a search for available plugins is made in the *plugins* directory under the application path. Each found plugin is automatically added to the application main toolbar, *Plugins*, as shown in Fig. 1. For illustration purposes, four available plugins are shown (Fig. 3): *Count*, *FireFlow*, *Import-Export GIS*, and *Summary*. A plugin may be activated by clicking its name on the *Plugins* dropdown menu. Each plugin sub-folder must have a Python file named *\_init\_.py* (note the double underscores before and after the *init* string). This *init* file should include at least two variables: *plugin\_name* (String) and *plugin\_create\_menu* (Boolean). The first is the plugin's name, also the same as the name of its sub-directory, while the second variable indicates whether the plugin will have its own menu once activated. In such case, a dictionary for the menu items must be defined using a variable named *\_all\_* (note the double underscores before and after the *all* string).

Next, we briefly describe the data and model objects that are available for the plugin. In the following sections, we provide snippets of plugin codes demonstrating the available functionalities and provide the complete codes in GitHub repository (<https://github.com/eladsal/EPANET-Plugins/>) to inform similar efforts. The plugin framework supported by EPANET is shown in Fig. 2. When the EPANET GUI program is started, the main map form is initialized and loaded. The main map object holds the entire data model of the EPANET project and, after the hydraulic and water quality simulations are executed, the output results. A reference to the main map object is passed to all the plugins by the *session* object. The *session* object contains the *project* and the *output* objects, which hold the network input and output data, respectively, as shown in Fig. 2. The *project* object is conveniently structured by sections according to the traditional EPANET *.inp* input file format. For example, the project object includes the *pipes* sub-class, which corresponds to the [PIPES] section in the EPANET input file, and defines pipes' properties such as name, inlet\_node, outlet\_node, diameter, length, roughness, loss\_coefficient, initial\_status, and description. A full list of the EPANET input file sections format can be found in the EPANET user manual (Rossman, 2000). The *output* object interacts with the simulation engine and contains the simulation results for each time step of the simulation period for all network elements organized by the *nodes* and *links* sub-classes. The time series results can be extracted with the *nodes.get\_series()* and the *links.get\_series()* methods.

These results include demand, head, pressure, and quality values for nodes, and flow, headloss, quality, status, and velocity for links. A detailed description of available results from EPANET hydraulic and water quality simulations is included in the program user manual (Rossman, 2000). In addition to the *project* and *output* objects, the *session* object includes a set of basic methods for opening and saving network models and running hydraulic and water quality simulations.

Before demonstrating the plugins, we show some of the basic functionalities available to the developer to perform hydraulic simulations as well as setting and extracting model parameters using a prototype code listed in Table 1. In lines 1 and 2, the main session and the project objects are defined. In line 3, *net1.inp* file is read into the current session. Line 4 retrieves the base demand for node 22 using the *session.project.junctions.value* method and line 5 assigns a new base demand for node 22. The hydraulic simulation is performed in line 6 and results are stored in the *session.output* object in line 7. Finally, in line 8, the simulated pressures at node 22 are retrieved using the *session.output.get\_series()* method.

### 3. Examples for developing custom plugins

#### 3.1. Plugin 1: Count

For illustrative purpose, a snippet of Python code for a simple plugin named *Count* is given in Table 2. When activated, this plugin adds a new menu to the main toolbar in the EPANET GUI with two sub-menus that report the number of junctions and pipes in the water network that is currently loaded in the GUI. Lines 1 and 2 in Table 2 are the name of the plugin and menu creation, respectively, as mentioned above. Line 3 includes a dictionary of the plugin's sub-menus, as shown in Fig. 3. Line 4 is a standard Python declaration to import a "Message Box" object for reporting. The main function of the plugin, *run()*, in line 5, is called when the user clicks on one of the sub-menus. According to the user's selection, one of the choices in lines 6 or 9 is executed, the number of junctions or pipes are retrieved in lines 7 and 10, respectively, and reported in lines 8 and 11. This code calls the *session.project.junctions* object that retrieves the list of junctions in the open project of the current session. To integrate the *Count* plugin with the main EPANET application, a sub-directory *Count* needs to be created with the code listed in Table 2 saved as *\_init\_.py* and placed under the *Plugins* sub-directory under the main EPANET-UI directory (Fig. 4).

#### 3.2. Plugin 2: Fire Flow

In this section, we demonstrate the *FireFlow* plugin, which conducts a fire flow analysis in a WDS. Fire flow analysis is a common practice used by water engineers to ensure protection is provided during fire emergencies (Boulos et al., 2006; Xiao et al., 2014). The aim of a fire flow analysis is to determine whether the required flow is available at

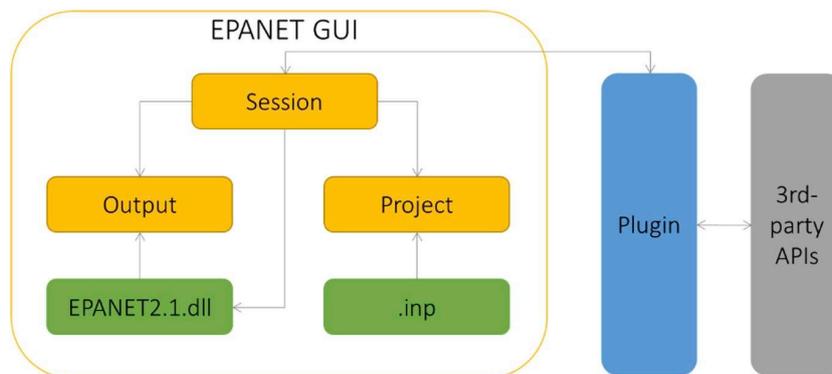


Fig. 2. EPANET plugins framework.

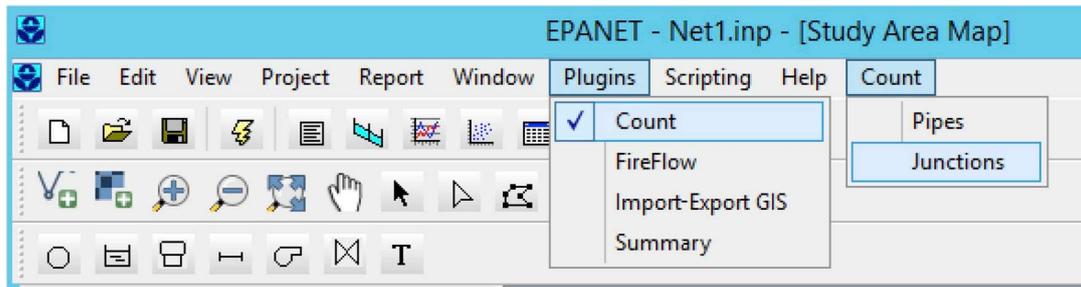


Fig. 3. EPANET plugins menus and the Count plugin sub-menu.

Table 1  
Hydraulic simulation prototype code.

Line	Code
1	s = session
2	p = s.project
3	s.open_project_quiet("D:\net1.inp")
4	demand = p.junctions.value["22"].base_demand_flow
5	p.junctions.value["22"].base_demand_flow = float(demand) * 1.1
6	i = s.run_simulation()
7	o = s.output
8	pressures = o.nodes["22"].get_series(o,AttributePressure)

fire hydrants while adequate pressures are maintained in the WDS during the stress conditions. Most of the commercial hydraulic simulation software include a fire flow analysis tool (Bentley, 2018; HCP, 2018; InfoWater, 2018). A basic feature of any fire flow analysis is to determine the relationship between the available fire flow discharge at a specific node in the network and the network-wide minimum pressure. The pressure-demand relationship can be evaluated using a rating curve. The rating curve is achieved by performing a series of hydraulic simulations each time increasing the fire flow discharge at a given network location and recording the minimum pressure in the WDS. This process is repeated for each fire flow node in the network. Given a rating curve, the network engineer or operator can evaluate the performance of the WDS under different conditions. The current version of EPANET GUI is designed to perform only a single hydraulic simulation and does not offer a way to automate or run multiple hydraulic simulations. Hence, for an EPANET GUI user, fire flow analysis is a tedious process, which requires the engineer to repeatedly change the boundary conditions, perform hydraulic simulations, and manually record and process the pressures in response to changes in the fire flow conditions. The fire flow analysis can be easily automated to perform multiple simulations to calculate the fire flow rating curve through a plugin extension.

An open source *FireFlow* EPANET plugin was developed herein and can be freely downloaded from a GitHub EPANET-Plugins repository (FireFlow, 2018). After the *FireFlow* plugin folder is downloaded and

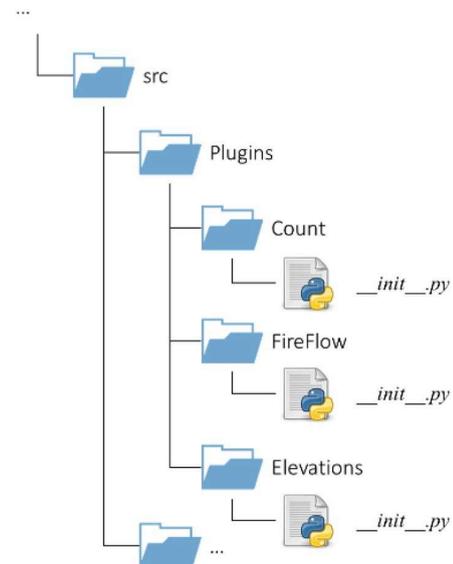


Fig. 4. Schematic architecture of the Plugins sub-directory and the relevant custom plugins: Count, FireFlow, and Elevations.

Table 2  
Count plugin code snippet.

Line	Code
1	plugin_name = "Count"
2	plugin_create_menu = True
3	__all__ = {"Junctions":1, "Pipes":2}
4	from PyQt4.QtGui import QMessageBox as Qmb
5	def run(session=None, choice=None):
6	if choice == 1:
7	n = format(len(session.project.junctions.value))
8	Qmb.information(None, '', "Number of Junctions: " + n, Qmb.Ok)
9	elif choice == 2:
10	n = format(len(session.project.pipes.value))
11	Qmb.information(None, '', "Number of Pipes: " + n, Qmb.Ok)

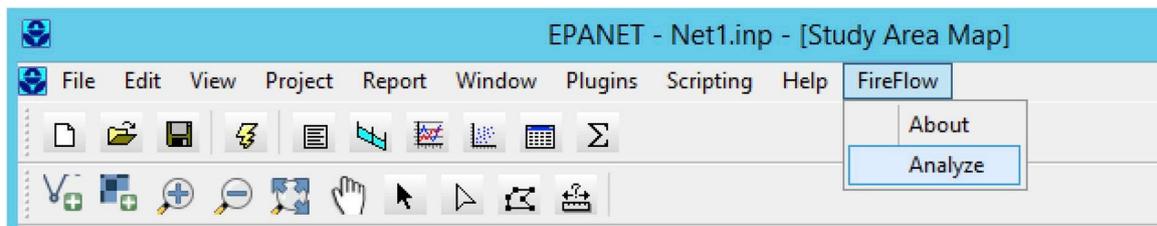


Fig. 5. New *FireFlow* drop-down menu which is created for performing automated fire flow analysis in EPANET GUI using the open source *FireFlow* plugin.

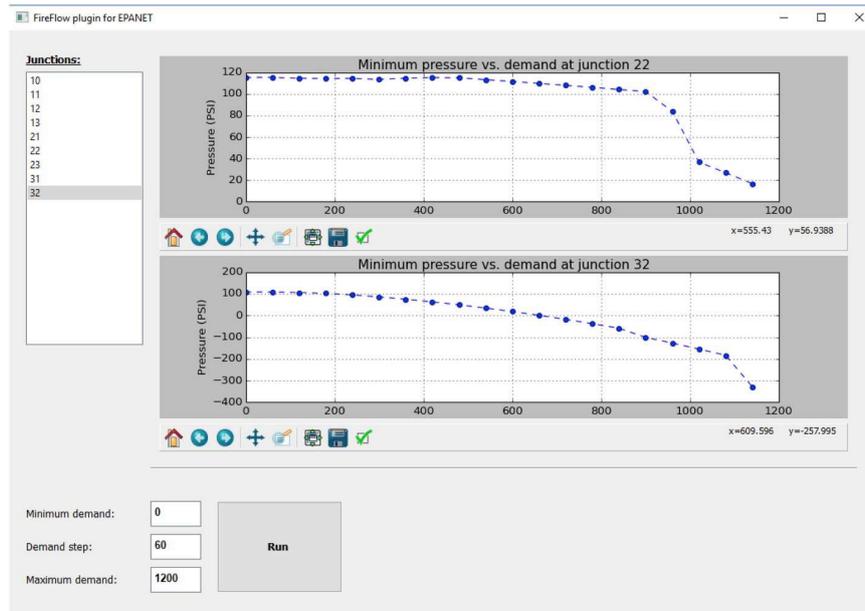


Fig. 6. GUI of the *FireFlow* plugin analysis.

placed in the EPANET *Plugins* directory, a new *FireFlow* menu will appear in the main EPANET GUI, as shown in Fig. 5.

After the user selects the “Analyze” option, a new window will appear (Fig. 6) with the list of the junctions included in the network model (in this example *Net1.inp* was used, see Fig. 1). To run the fire flow analysis, the user selects the fire flow node to conduct the analysis, along with three additional parameters: the minimum and maximum fire flow discharge to analyze and the incremental step between the minimum and maximum flows. After selecting the node and setting the parameters, the user can click the “Run” button to perform the simulations. Once the simulation runs are completed, the demand-pressure rating curve is presented, the user may select additional nodes for the fire flow analysis. Fig. 6 shows the *FireFlow* analysis window. Two analyses were performed for nodes 22 and 32, respectively, ranging the fire flow discharge from 0 to 1200 gpm. The top plot demonstrates the change in the minimum pressures in the network as a response to increasing the discharge in node 22. The minimum pressures gradually decrease from 120 psi to 100 psi as the discharge increases to 900 gpm, then the minimum pressure rapidly decrease to 10 psi, as fire flow discharge increases to 1200 gpm. Similar analysis is performed for node 32 (bottom plot in Fig. 6), however the results show a rapid decrease in minimum pressures reaching negative values as fire flow discharge increases to 600 gpm. These results indicate that node 22 satisfies the fire flow upper limit, whereas node 32 is sensitive to fire flow conditions and does not satisfy the fire flow upper limit.

Table 3 shows a code snippet from the *FireFlow* plugin main code. Line 1 is the main loop over the number of requested hydraulic simulations based on the range of the fire flow discharge. In lines 2–3, the

current demand is calculated and assigned to the analyzed junction. Then, a full hydraulic simulation is performed by calling the *session.run\_simulation()* method (line 4), which performs both the hydraulic and water quality simulations. The system's pressures are extracted (lines 5–6) using the *get\_series()* method. Finally, the demand and minimum pressure are recorded in lines 7 and 8, respectively, which are then plotted for the user. The *FireFlow* plugin is a simple example of how the basic EPANET software may be extended via the plugins framework in order to add new capabilities to the program with new algorithms and graphical user interface. The *FireFlow* plugin relies on EPANET hydraulic simulator but does not alter the main code of EPANET. Furthermore, the *FireFlow* plugin is *optional* and is not required to run the original application.

### 3.3. Plugin 3: Elevations

The process of building a hydraulic model typically originates from a water utility's records on the location of network elements and characteristics that are maintained in Geographic Information Systems (GIS) (Deuerlein et al., 2015; Roma et al., 2015). However, most utilities do not have detailed enough records of their pipeline infrastructure and, additionally, some loss of information is inevitable when transforming GIS pipeline records into hydraulic models. Hence, when required information is not available, some estimates must be made in order to assess the network pressures and hydraulic grade lines. Traditionally, commercial vendors are able to provide very accurate Digital Elevation Model (DEM) of any given area, however, usually at a high cost. Other companies, such as Google and Microsoft, provide a less

**Table 3**  
FireFlow plugin code snippet.

Line	Code
1	for i in range(0, n):
2	dem = min_demand + i * delta_demand
3	j.base_demand_flow = dem
4	i = session.run_simulation()
5	nod = session.output.nodes[junc]
6	pres = nod.get_series(session.output, AttributePressure)
7	d.append(dem)
8	p.append(min(pres))

**Table 4**  
EPANET Elevations plugin code snippet.

Line	Code
1	junctions_list = session.project.junctions.value[:]
2	for j in junctions_list:
3	lat = str(j.y); lon=str(j.x)
4	txt = lat + ', ' + lon
5	request =
	Request('https://maps.googleapis.com/maps/api/elevation/xml?locations='
	+ txt + '&key=' + API_KEY)
6	response = urlopen(request).read()
7	#custom function to retrieve XML elevation data to elev
8	j.elevation = elev

accurate DEM for a small fee, or in some cases, for free via their mapping products, such as Google Maps and Bing Maps, respectively. The third EPANET plugin developed in this work, the *Elevations* plugin, uses the Google Maps elevations API to retrieve the elevation data for the network nodes (Google Maps Platform, 2018). Table 4 shows a code snippet from the main code of the *Elevations* plugin (the full plugin code is available on GitHub). First, the list of junctions is extracted from the *session.project* object (line 1). Then, the X and Y (longitude and latitude) coordinates are extracted for each junction (line 3). In line 5 the URL request with the Google Maps API is constructed and opened (line 6). It should be noted that a private API Key is used for authentication. Line 7 executes a custom function, which extracts the elevation data from the XML data structure returned by the Google Maps API to the *elev* variable. Finally, the elevation data is assigned to the junction's elevation property (line 8). This plugin demonstrates the possibility for the EPANET software to interact with web-services to get and set data.

#### 4. A prototype repository for EPANET plugins

Three prototype plugins were presented in the previous sections. However, for a successful and sustainable transition to the plugins framework there is a need for a *centralized place where developers and users can upload and download plugins and share information*. A centralized plugins repository is common across many well-known applications such ArcGIS (2018), QGIS (2016), and Autodesk (2016). Fig. 7 shows a prototype repository for EPANET plugins, which can be used to easily share plugin files between developers and users. The user could find a list of available plugins, description, installation instructions, test cases, and reviews by other users. The developer can find documentation, code snippets, and submission instructions. The users can evaluate the published plugins, by providing feedback to the developers and the WDSA community, and help determine how robust and reliable these plugins are. Such a repository can provide a viable pathway towards research transfer and dissemination of new scientific tools developed by the WDSA community and help distributing these tools to the multitude of EPANET users to enhance their hydraulic modeling and analysis capabilities. This is especially true for users that rely on the user interface who often cannot take advantage of advanced modeling tools because these typically require advanced programming skills.

Moreover, such a repository can create new opportunities for exchanging information, ideas, and benchmarking related to computing and analysis of WDSs as well as simplifying plugins implementation by providing documentation, tutorials, and templates.

As the expectations for more transparent and reproducible research are rising, there is an increasing pressure on the WDSA community to adopt a more open and collaborative research and software development. See for example the publication data policy in the leading journals of Environmental Modeling and Software (Elsevier, 2018), Water Resources Research (American Geophysical Union, 2013), and Journal of Water Resources Planning and Management (Rosenberg et al., 2018). The EPANET-plugins framework can support achieving this goal, thus providing benefits for the researchers and the practitioners. Certainly, sharing research tools and making them accessible to the wider water systems modeling community through plugins will not resolve all modeling limitations of EPANET. Nevertheless, if successful, it could result in hydraulic simulation improvements, active community engagements and interest in EPANET software development (Uber et al., 2018). The *Count*, *FireFlow*, and *Elevations* plugins presented here are three simple examples of enhancement to current capabilities of EPANET. Other pressing examples include automatic demand assignment from raw billing data and smart meters into demand pattern format for hydraulic modeling, robust GIS tools to transform GIS data sets into hydraulic models, and pressure-driven modeling.

#### 5. Conclusions

This work presents a first step towards demonstrating a straightforward implementation of plugins in the new EPANET GUI. The plugin framework supported by EPANET through Python scripting program is responsible for setting up the plugin environment. The plugins are independent components from the main EPANET software that can be developed and distributed separately, and are not required to execute the original application. Plugins integrated with the main EPANET application can provide additional modeling and analysis functionalities that are not available in the current EPANET software, e.g. fire flow analysis demonstrated in this work. Advanced hydraulic modeling and analytics tools coupled with a plugin framework within a user-familiar environment of EPANET create an opportunity to transform how

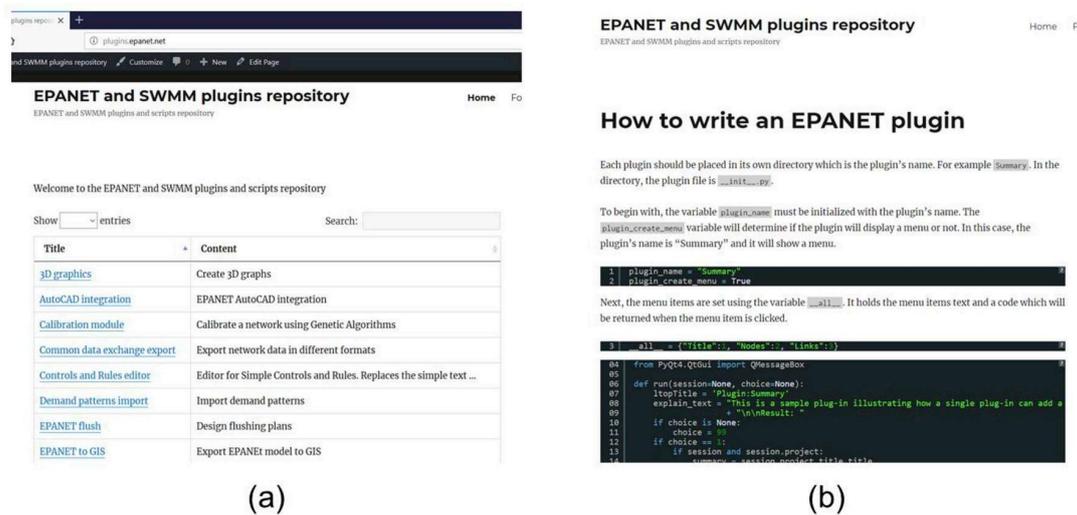


Fig. 7. Prototype repository for EPANET plugins: (a) list of available plugins, (b) plugin documentation for the developer.

new models and techniques are being developed, shared, and used in the water systems modeling community. As the expectations for more transparent and reproducible research are rising, EPANET-plugins framework can support knowledge transfer thus providing benefits for the researchers and the practitioners.

### Acknowledgments

This work was supported by the University of Texas at Austin New Faculty Startup Grant and by the Israeli Water Authority Grant 4501284516.

### References

- American Geophysical Union, 2013. AGU publications data policy. <http://publications.agu.org/author-resource-center/publication-policies/data-policy/> Oct. 20, 2018.
- ArcGIS, 2018. ArcGIS code sharing. <http://codesharing.arcgis.com> Oct. 20, 2018.
- Autodesk, 2015. AutoCAD - overview. <http://www.autodesk.com/products/autocad/overview#> Oct. 20, 2018.
- Autodesk, 2016. Autodesk App Store - FAQ. Autodesk. <https://apps.autodesk.com/en/Public/FAQ> Oct. 20, 2018.
- Bentley, 2018. Water distribution analysis and design software. <https://www.bentley.com/en/products/product-line/hydraulics-and-hydrology-software/watergems> Oct. 20, 2018.
- Boulos, P.F., Aboujaoude, A.S., 2011. Managing leaks using flow step-testing, network modeling, and field measurement. *J. Am. Water Work. Assoc.* 103 (2), 90–97.
- Boulos, P.F., Lansley, K.E., Karney, B.W., 2006. Comprehensive Water Distribution Systems Analysis Handbook for Engineers and Planners. Comprehensive Water Distribution Systems Analysis Handbook for Engineers and Planners.
- Deuerlein, J., Piller, O., Arango, I.M., Braum, M., 2015. Parameterization of Offline and Online Hydraulic Simulation Models. *Procedia Engineering*, Elsevier B.V., pp. 545–553.
- Eliades, D.G., Kyriakou, M., Polycarpou, M.M., 2014. Sensor Placement in Water Distribution Systems Using the S-PLACE Toolkit. *Elsevier B.V. Procedia Eng.*, vol. 70, 602–611 2010.
- Elsevier, 2018. Data Statement. <https://www.elsevier.com/authors/author-services/research-data/data-statement> Oct. 20, 2018.
- Environmental Systems Research Institute (ESRI), 2011. ArcGIS Desktop: Release 10. (Redlands).
- FireFlow, 2018. EPANET plugins-FireFlow. <https://github.com/eladsal/EPANET-Plugins/tree/master/FireFlow>.
- Frontline Solvers, 2015. Analytic solver platform. <https://www.solver.com/premium-solver-platform> Oct. 20, 2018.
- Google Chrome, 2018. Get more done with the new Chrome. <https://www.google.com/chrome/> Oct. 20, 2018.
- Google Maps Platform, 2018. Elevation API. <https://developers.google.com/maps/documentation/elevation/start> Oct. 20, 2018.
- Hatchett, S., Uber, J., Boccellini, D., Haxton, T., Janke, R., Kramer, A., Matracia, A., Panguluri, S., 2011. Real-Time distribution system modeling: development, application and insights. In: International Conference on Computing and Control for Water Industry (CCWI), pp. 1–6.
- HCP, 2018. Fire flow capacity simulation. <http://www.hcp.net.au/realtime/>

- [simulateFireFlows.html](#) Oct. 20, 2018.
- Ho, C.K., Khalsa, S.S., 2009. EPANET-BAM: water quality modeling with incomplete mixing in pipe junctions. *Water Distribution Systems Analysis* 1–11.
- Housh, M., Ohar, Z., 2017. Integrating Physically Based Simulators with Event Detection Systems: Multi-Site Detection Approach, vol. 110. *Water Research*, Elsevier Ltd, pp. 180–191.
- Housh, M., Ohar, Z., 2018. Model-based approach for cyber-physical attack detection in water distribution systems. *Water Res.* 139, 132–143.
- Iglesias-Rey, P.L., Martínez-Solano, F.J., Ribelles-Aguilar, J.V., 2017. Extending EPANET Capabilities with Add-In Tools. *Procedia Engineering*, pp. 626–634 The Author(s).
- InfoWater, 2018. Fireflow Part 1: basics. <http://blog.innovyze.com/2017/03/31/fireflow-part-1-basics/> Oct. 20, 2018.
- Irrigateplus, 2017. Design irrigation systems in a quick and efficient way. <http://www.irrigateplus.com/?c=home> Oct. 20, 2018.
- Kandjani, E.M., Sela, L., Faust, K., Leite, F., Hodges, B.R., 2018. Extending the capabilities of EPANET user interface with the optimal design plugin tool. In: WDSA/CCWI Joint Conference Proceedings.
- Kapelan, Z.S., Savic, D.A., Walters, G.A., 2005. Multiobjective design of water distribution systems under uncertainty. *Water Resour. Res.* 41 (11), 1–15.
- Lofberg, J., 2004. YALMIP: a toolbox for modeling and optimization in MATLAB. In: 2004 IEEE International Conference on Robotics and Automation (IEEE Cat. No.04CH37508), pp. 284–289.
- Martínez-Solano, F.J., Iglesias-Rey, P.L., Arce, S.X.M., 2017. Simultaneous calibration of leakages, demands and losses from measurements. Application to the Guayaquil network (Ecuador). *Procedia Engineering* 397–404.
- Microsoft Edge, 2018. "Microsoft Edge- the faster way to get things done on the web." <https://www.microsoft.com/en-us/windows/microsoft-edge> Oct. 20, 2018.
- Mozilla, 2018. Extensions. <https://addons.mozilla.org/en-US/firefox/> Oct. 20, 2018.
- Murphy, L.J., Dandy, G.C., Simpson, A.R., 1994. Optimum Design and Operation of Pumped Water Distribution Systems Reuse of Effluent View Project Better Data-Driven Decision-Making under Future Climate Uncertainty View Project. *Hydraulics in Civil Engineering*.
- Negus, C., 2015. GNU general public license. *Linux® Bible*, pp. 860–872.
- Open Source Initiative, 2018. The MIT license | open source initiative. <https://opensource.org/licenses/MIT> Oct. 20, 2018.
- Ormsbee, L.E., Lansley, K.E., 2006. Optimal control of water supply pumping systems. *J. Water Resour. Plan. Manag.* 120 (2), 237–252.
- Perelman, L., Housh, M., Ostfeld, A., 2013. Robust optimization for water distribution systems least cost design. *Water Resour. Res.* 49 (10), 6795–6809.
- Phillips, C.A., Kapelan, Z., Salmons, E., VanBriesen, J., Wu, Z.Y., Trachtman, G.B., Eliades, D., Piller, O., Gueli, R., Hart, W.E., Walski, T., Preis, A., Propato, M., Isovitich, S., Guestrin, C., James, W., Ostfeld, A., Xu, J., Leskovec, J., Berry, J.W., Fischbeck, P., Uber, J.G., Savic, D., Huang, J.J., Small, M., Dorini, G., Ghimire, S.R., McBean, E.A., di Pierro, F., Polycarpou, M., Barkdoll, B.D., Watson, J.-P., Jonkergouw, P., Krause, A., Khu, S.-T., 2008. The battle of the water sensor networks (BWSN): a design challenge for engineers and algorithms. *J. Water Resour. Plan. Manag.* 134 (6), 556–568.
- Python, 2018. Python. <https://www.python.org/> Oct. 20, 2018.
- QGIS, 2016. QGIS Python Plugins Repository Plugins Tagged with: Openlayers. *QGIS*. <https://plugins.qgis.org/plugins/tags/openlayers/> Oct. 20, 2018.
- Roma, J., Pérez, R., Sanz, G., Grau, S., 2015. Model calibration and leakage assessment applied to a real Water Distribution Network. *Procedia Engineering* 603–612.
- Rosenberg, D.E., Ph, D., Asce, A.M., 2018. New Policy to Specify Availability of Data , Models , and Code, vol. 144. pp. 1618001 9.
- Rossman, L., 2015. Storm Water Management Model User's Manual Version 5.1. United States Environment Protection Agency.

- Rossman, L.A., 1994. EPANET User's Manual. Version 1.1. US Environmental Protection Agency, Cincinnati, OH.
- Rossman, L.A., 2000. Epanet 2 Users Manual.
- Salomons, E., Hatchett, S., Eliades, D.G., 2018. The Epanet Open Source Initiative. pp. 0–7 M(June 2000).
- Savic, D.A., Walters, G.A., 2002. Genetic algorithms for least-cost design of water distribution networks. *J. Water Resour. Plan. Manag.* 123 (2), 67–77.
- Sela, L., Housh, M., 2019. Increasing usability of water distribution analysis tools through plug-in development in EPANET. *J. Hydraul. Eng.* 145 (5), 02519001.
- Shang, F., Uber, J.G., Rossman, L. a., 2008. Modeling reaction and transport of multiple species in water distribution systems. *Environ. Sci. Technol.* 42 (3), 808–814.
- Siew, C., Tanyimboh, T.T., 2012. Pressure-dependent EPANET extension. *Water Resour. Manag.* 26 (6), 1477–1498.
- SWMM-EPANET UI, 2018a. SWMM-EPANET\_User\_Interface. [https://github.com/USEPA/SWMM-EPANET\\_User\\_Interface](https://github.com/USEPA/SWMM-EPANET_User_Interface) Oct. 20, 2018.
- SWMM-EPANET UI, 2018b. Application features requirements document: Re-engineering SWMM/EPANET user interface application software. [https://github.com/USEPA/SWMM-EPANET\\_User\\_Interface/tree/master/doc](https://github.com/USEPA/SWMM-EPANET_User_Interface/tree/master/doc) Oct. 20, 2018.
- Uber, J.G., Boccelli, D.L., Hatchett, S., Kapelan, Z., Saldarriaga, J., Simpson, A.R., ... van Zyl, J.E., 2018. Let's get moving and write software: An open source project for EPANET. *J. Water Resour. Plan. Manag.* 144 (4).
- Water Analytics, 2018. Release Notes for EPANET 2.1. [http://wateranalytics.org/EPANET/release\\_2.1.html](http://wateranalytics.org/EPANET/release_2.1.html) Oct. 20, 2018.
- Whitman, B.E., Weir, M., Posluszny, E.T., Walski, T., Bezts, W., 2018. Modeling leakage reduction through pressure control. *J. Am. Water Work. Assoc.* 98 (4), 147–155.
- Wordpress, 2018. Extend your WordPress experience with 54,753 plugins. <https://wordpress.org/plugins> Oct. 20, 2018.
- Xiao, C., Li, B., He, G., Sun, J., Ping, J., Wang, R., 2014. Fire flow capacity analysis based on hydraulic network model. *Procedia Engineering* 386–394.
- Xie, X., Nachabe, M., Zeng, B., 2014. Optimal scheduling of automatic flushing devices in water distribution system. *J. Water Resour. Plan. Manag.* 141 (6), 04014081.
- Zierolf, M.L., Tryby, M.E., Polycarpou, M.M., Rossman, L.A., Boccelli, D.L., Uber, J.G., 2002. Optimal scheduling of booster disinfection in water distribution systems. *J. Water Resour. Plan. Manag.* 124 (2), 99–111.
- Zonum Solutions, 2009. EPANET Z v0.5. <http://www.zonums.com/epanetz.html> Oct. 20, 2018.

## References

- Abdelmeguid, H., and Ulanicki, B. (2012). “Feedback rules for operation of pumps in a water supply system considering electricity tariffs.” *Water Distribution Systems Analysis 2010 - Proceedings of the 12th International Conference, WDSA 2010*, 1188–1205.
- Alvisi, S., and Franchini, M. (2017). “A robust approach based on time variable trigger levels for pump control.” *Journal of Hydroinformatics*, 1–12.
- Alvisi, S., Franchini, M., and Marinelli, A. (2007). “A short-term, pattern-based model for water-demand forecasting.” *Journal of Hydroinformatics*, 9(1), 39.
- Blinco, L. J., Simpson, A. R., Lambert, M. F., and Marchi, A. (2016). “Comparison of Pumping Regimes for Water Distribution Systems to Minimize Cost and Greenhouse Gases.” *Journal of Water Resources Planning and Management*.
- Copeland, C., and Carter, N. T. (2017). *Energy - Water Nexus: The Water Sector's Energy Use R43200*. Washington, D.C.
- Coulbeck, B., and Orr, C. H. (1989). “Real-Time Optimized Control of a Water Distribution System.” *IFAC Proceedings Volumes*, Elsevier, 22(6), 441–446.
- Creaco, E., Lanfranchi, E., Chiesa, C., Fantozzi, M., Carrettini, C. A., and Franchini, M. (2016). “Optimisation of leakage and energy in the Abbiategrasso district.” *Civil Engineering and Environmental Systems*, Taylor & Francis, 33(1), 22–34.
- Donkor, E. A., Mazzuchi, T. A., Soyer, R., and Alan Roberson, J. (2014). “Urban Water Demand Forecasting: Review of Methods and Models.” *Journal of Water Resources Planning and Management*, 140(2), 146–159.
- Goldstein, R., and Smith, W. (2002). “U.S. Electricity Consumption for Water Supply & Treatment - The Next Half Century.” *Technical Report*, 4(Water and Sustainability).
- Herrera, M., Torgo, L., Izquierdo, J., and Pérez-García, R. (2010). “Predictive models for forecasting hourly urban water demand.” *Journal of Hydrology*.
- IBM Corp. (2009). “V12. 1: User's Manual for CPLEX.” *International Business Machines Corporation*, 12(1), 481.
- Jamieson, D. G., Shamir, U., Martinez, F., and Franchini, M. (2007). “Conceptual design of a generic, real-time, near-optimal control system for water-distribution networks.” *Journal of Hydroinformatics*, 9(1), 3.

- Lougee-Heimer, R. (2003). “The Common Optimization INterface for Operations Research: Promoting open-source software in the operations research community.” *IBM Journal of Research and Development*, 47(1), 57–66.
- Mala-Jetmarova, H., Sultanova, N., and Savic, D. (2017). “Lost in optimisation of water distribution systems? A literature review of system operation.” *Environmental Modelling & Software*, Elsevier Ltd, 93, 209–254.
- Marchi, A., Simpson, A. R., and Lambert, M. F. (2016). “Optimization of Pump Operation Using Rule-Based Controls in EPANET2: New ETTAR Toolkit and Correction of Energy Computation.” *Journal of Water Resources Planning and Management*, 142(7), 04016012.
- Ormsbee, L. E., and Lansey, K. E. (1994). “Optimal Control of Water Supply Pumping Systems.” *Journal of Water Resources Planning and Management*, 120(2), 237–252.
- Pacchin, E., Alvisi, S., and Franchini, M. (2017). “A Short-Term Water Demand Forecasting Model Using a Moving Window on Previously Observed Data.” *Water*, 9(3), 172.
- Paschke, M., Spencer, K., Waniarcha, N., Simpson, A. R., and Widdop, T. (2001). “Genetic algorithms for optimising pumping operations.” *19th Federal Convention, Australian Water Association, Canberra, Australia*.
- Rao, Z., and Salomons, E. (2007). “Development of a real-time, near-optimal control process for water-distribution networks.” *Journal of Hydroinformatics*, 9(1), 25.
- Rossman, L. A., Woo, H., Tryby, M., Shang, F., Janke, R., and Haxton, T. (2020). *EPANET 2.2 User Manual*. Washington, DC.
- Sanders, K. T., and Webber, M. E. (2012). “Evaluating the energy consumed for water use in the United States.” *Environmental Research Letters*, 7(3).
- Sanks, R. L., and Tchobanoglous, G. (1998). *Pumping Station Design Second Edition Editor-in-Chief Co-Editors Boston Oxford Johannesburg Melbourne New Delhi Singapore*.
- Shamir, U., and Salomons, E. (2008). “Optimal Real-Time Operation of Urban Water Distribution Systems Using Reduced Models.” *Journal of Water Resources Planning and Management*, 134(2), 181–185.
- Zhou, S. L., McMahon, T. A., Walton, A., and Lewis, J. (2002). “Forecasting operational demand for an urban water supply zone.” *Journal of Hydrology*, 259(1–4), 189–202.

# תפעול אופטימלי בזמן אמת של מערכות אספקת מים באופן מעשי

מאת : אלעד זלומנס

## תקציר

בארצות הברית קיימות כ- 52,000 מערכות מים. כ- 5% מהמערכות האלו מספקות מים ל- 85% מהתושבים בארה"ב בעוד 95% מהמערכות מספקות מים לפחות מ- 3,300 תושבים. יחסים דומים קיימים גם באירופה ובמזרח הרחוק. בישראל ישנם כ- 55 תאגידי מים בגדלים שונים ובנוסף יש כ- 1,000 ספקים מים נוספים. חלק לא מבוטל מהתאגידים וספקי המים הם בעלי מערכות אספקת מים קטנות יחסית עם מעט מתקני שאיבה ואגירה פשוטים. תפעול מערכות אספקת מים באופן אופטימלי ואוטומטי, אפילו עבור מערכות פשוטות, עשוי להיות תהליך מורכב שכן הוא מצריך הערכת צריכת המים, ניסוח ופתרון בעיה מתמטית מורכבת בתדירות גבוהה יחסית. בנושא זה, האתגרים העומדים בפני אנשי התאגידים הינם:

- זמינות מידע – רוב השיטות לחיזוי הצריכה דורשות סדרות נתונים היסטוריים ארוכות. אפילו אם הנתונים זמינים קיימת בעייתיות בזמן שינויי צריכה גדולים.
- קושי חישובי – הבעיות המתמטיות הנדרשות לפתרון הן לא לינאריות ולא קמורות הכוללות אופטימיזציה בשלמים לייצוג משתני פעולת המשאבות (פועל/דומם).
- אי קיום חדר בקרה מרכזי ומחסור בכוח-אדם מקצועי.

בהינתן האתגרים לעיל, מטרת מחקר זה הינה לפתח מתודולוגיות מעשיות וחדשות לתפעול אופטימלי ויעיל של מערכות מים תוך עמידה באילוצי הצריכה ואילוצים הידראוליים במערכת הפיזית. באופן כללי, ניתנה עדיפות למתודולוגיות פשוטות, שאינן דורשות נתונים רבים והגדרות פרמטרים רבים. כמו-כן, השיטות שפותחו הינן בעלות "חתימה מחשובית" נמוכה אשר לא יחייבו מערכות מחשוב ותקשורת מורכבות. למשל, במקום להסתמך על סדרות נתונים היסטוריים ארוכות לביצוע חיזוי צריכה, ניתן להסתמך על שיטה פשוטה שמבוססת על נתוני צריכת המים בשבועות הקודמים. שימוש בשיטת חיזוי פשוטה כזאת בשילוב עם טכניקות אופטימיזציה לטיוב כללי התפעול מהווה פתרון פרקטי לבעיית התפעול המורכבת, פתרון זה לא מחייב מערכות מחשוב ותקשורת מורכבות.

תפעול מערכת אספקת מים היא דוגמה ברורה לתחלופה בין עלות תפעול לבין אמינות המערכת. מצד אחד, בהיבט של אמינות, שואפים מפעילי המערכת לשמור את מפלסי המים בבריכות גבוהים ככל שאפשר על מנת להתמודד היטב עם מצבי חירום ומצבי קיצון כגון תקלות, שריפות, הפסקות חשמל וכו'. מהצד השני, בהיבט של חיסכון בהוצאות האנרגיה, חברות המים מעוניינות לנצל את נפחי האיגום הקיימים על מנת להסיט את הפעלת המשאבות לשעות בהן תעריפי האנרגיה נמוכים יותר. בניסיון לאזן את שתי המטרות הסותרות הללו, מפלסי המים המינימליים והמקסימליים מוגדרים בדרך כלל כדי להבטיח את האמינות הנדרשת תוך מתן אפשרות למפעיל לשנות את מפלסי המים של הבריכות בין המינימום והמקסימום המוגדרים כדי למזער את עלויות התפעול. הפעלה בתוך גבולות מוגדרים מראש יכולה להתבצע במערכי תפעול ובקרה שונים. אחד

מהמאפיינים של מערכים אלה הוא המיקום הפיזי של הבקר. במובן זה, ניתן לשקול את שתי התצורות הקיצוניות (הבאות: 1) מערך בקרה מקומי בו בקר הפיקוד מותקן ונמצא באתר כגון תחנת השאיבה או הבריכה, או 2) מערך בקרה מרכזי בו בקר הפיקוד מותקן במקום מרכזי כגון חדר בקרה ומפקד על מערכת גדולה יותר.

עבור מערך של תחנת שאיבה ובריכת מים, לולאת הבקרה הפשוטה ביותר מבוססת על הגדרות זמנים לצורך הפעלת המשאבה וערכי לחץ לצורך הדממתה. בצורת הפעלה לא שכיחה זו, מוגדר מראש זמן הפעלת המשאבה (כגון תחילתו של פרק הזמן בו תעריפי החשמל נמוכים). בעת השאיבה עולה מפלס המים בבריכה עד למפלס עליון עבורו נסגר מגוף הכניסה לבריכה. סגירת המגוף גורמת לעליית הלחץ בסניקת המשאבה עד לערך בו מופסקת פעולתה. היתרון המרכזי של אופן הפעלה זה הוא בכך שאין צורך במערך תקשורת כלל בין תחנת השאיבה והבריכה וכל הפיקוד נעשה בתחנה עצמה. מצד שני, בקרה מבוססת זמן לא מאפשרת תגובה למקרים בהם ישנם שינויים בצריכה ולכן ייתכן ואופן הפעלה זה ידרוש עידכונים ידניים תכופים יחסית. בימינו, מערכות פיקוד ובקרה (SCADA) נפוצות מאד וכן עלות התקנתן ותחזוקתן נמוכה יחסית. גורם זה הקטין את השימוש במערך הבקרה הקודם ואפשר שימוש במערכי בקרה מתקדמים יותר המבוססים על תקשורת בין מתקני רשת אספקת המים. במערך זה, מפלס המים בבריכה משודר באופן קבוע לבקר הפיקוד הנמצא בתחנת השאיבה אשר מתרגם את מצב הבריכה לפקודות הפעלה והדממה לתחנה על-פי אלגוריתם הבקרה שנקבע בו מראש. מערך הבקרה הפשוט ביותר במצב זה עושה שימוש במפלסי הפעלה והדממה קבועים. יחד עם זאת, מפלסי הפעלה והדממה קבועים אלו אינם מאפשרים חיסכון בהוצאות האנרגיה כאשר קיים תעריף מבוסס זמן (כגון תעו"ז, תעריף עומס זמן). על מנת להתמודד עם חוסר היעילות של מפלסי הפעלה והדממה הוצע בעבר השימוש במפלסי הפעלה והדממה אשר משתנים בהתאם לזמני תעריפי החשמל השונים. מערך בקרה זה מאפשר הורדת מפלסי הבריכות בזמנים בהם התעריפים גבוהים ומילוי הבריכות בזמנים בהם הם נמוכים. יחד עם זאת, שיפור ביעילות האנרגטית בשיטה זו עלולה לגרום למקרים של הפעלות והדממות תכופות אשר אינם רצויים במערכת אספקת המים. בנוסף, הוצעו מפלסי הפעלה והדממה משתנים בצורת פונקציות לינאריות או אחרות בתקופות התעריף השונות. מערך הבקרה המוצג לעיל הוא למקרה בו בקר הפיקוד מותקן מקומית (כגון בתחנת השאיבה). למרות שלמערך זה יתרונות מסויימים, החיסרון העיקרי שלו הוא אי היכולת לראות את המערכת בכללותה דבר היכול להתאפשר במקרה של מערך בקרה מרכזי. כאשר בקר הפיקוד נמצא במקום מרכזי הוא יכול לעשות שימוש באלגוריתמים מתוחכמים הדורשים יכולת מחשוב גבוהה יותר.

מטרת מחקר זה הינה לפתח מתודולוגיות מעשיות וחדשות לתפעול אופטימלי ויעיל של מערכות מים תוך עמידה באילוצי הצריכה ואילוצים הידראוליים במערכת הפיזית.

במסגרת מחקר זה הושם דגש על ביצועי מערך הבקרה בכללותו במקום על כל אחד ממרכיביו בנפרד כפי שנעשה ברבים ממחקרים אחרים. במילים אחרות, ניתנה עדיפות למתודולוגיה פשוטה ומעשית, אשר נותנת ביצועים טובים אך לא מושלמים, על פני שיטות פתרון מורכבות בעלות חתימה מחשובית גבוהה הנותנות תועלת שולית קטנה. דרכי ההתמודדות עם האתגרים לעיל הינם:

- יישום של אלגוריתמים פשוטים לחיזוי צריכה – ניתן דגש על שיטות לחיזוי הצריכה אשר מצריכות סדרות נתונים היסטוריות קצרות, פשוטות לחישוב ואשר מאפשרות תגובה לשינויים בצריכה בזמנים קצרים.
- פיתוח מערכי בקרה מקומיים ומרכזיים – בהשוואה למערכי בקרה מקומיים אשר הינם בעלי מידע מועט לגבי המערכת בכללותה וכן בעלי חתימה מחשובית נמוכה, מערכי הבקרה המרכזיים עשויים

להיות בעלי יתרון בכך שהם "רואים" את כלל המערכת ויכולים לקחת בחשבון את מצב המערכת על כלל האלמנטים בה. יחד עם זאת, בדרך כלל מערכי בקרה מרכזיים נסמכים על שיטות אופטימיזציה בכדי לפתור את בעיית התפעול אשר דורשים חתימה מחשובית גבוהה על מנת לרוץ בזמן אמת. למרות שעלויות האנרגיה יכולות להיות ממוזערות באמצעות שיטות מתוחכמות אלו, רמת הסיבוכיות שלהן, בהשוואה למערך בקרה מקומי, הופך אותן לפחות נפוצות. מערכי הבקרה המקומיים נפוצים יותר בזכות פשטותם ועמידותם מכיוון שאינם נסמכים על פרוטוקולי תקשורת מורכבים שכן רוב המידע הנדרש עבורם נמצא במקום זמין עבורם באתר. יתרה מכך, מערכי הבקרה המקומיים מפותחים מראש ונשלחים אל בקר הפיקוד ליישום בתחנות השאיבה. במקרה זה, תכניות התפעול עשויות לפעול זמן רב ללא עדכון לפחות עד שיש שינוי מהותי במערכת אספקת המים (כגון שינויים במשאבות, שינוי מהותי בצריכת המים וכו'). לפיכך, פותחו שיטות עבור שני מערכי הבקרה הללו:

- בקרה מקומית – ראשית פותח מערך בקרה מקומי אשר יכול להיות מיושם על בקר פיקוד מקומי ללא פיקוד מאתר מרכזי בזמן אמת. מערך זה הנו עבור אזור הידראולי קטן, כגון עבור תחנה אחת או שתיים ובריכה. מקרה זה הנו נפוץ מאד בתאגידים בעולם בכלל ובישראל בפרט. המאפיינים המרכזיים של מערך בקרה זה הם דרישות נמוכות של המידע הדרוש וחתימה מחשובית נמוכה אשר אינה דורשת תוכנות אופטימיזציה מורכבות.
- בקרה מרכזית – במיקום מרכזי, כגון חדר הבקרה, ישנה אפשרות להשתמש במשאבים מחשוביים חזקים וקיים מידע רב יותר על מערכת אספקת המים. דבר זה מאפשר שימוש בשיטות חישוב ואלגוריתמים מתוחכמים יותר כולל בעזרת מערכות אופטימיזציה ייעודיות, מסחריות או כאלו בקוד פתוח עבור רשתות אספקת מים גדולות. יחד עם זאת, למערכי בקרה המנוסחים כבעיות NLP, MILP ו-LP ישנם חסרונות מסוימים. במחקר נבחן השימוש בניסוחי MILP ברמות אומדן שונות (עם תוצאות מיטביות שונות) והוראה כי עבור תפעול בזמן אמת, התחלופה בין תוצאות מיטביות לבין יעילות הפתרון, מוטה. כלומר, שיטה מקורבת עשויה לתת פתרון מעשי אך יחד עם זאת, לא לאבד רבות מטיב הפתרון.

- רוב המחקרים בנושא תפעול מערכות אספקת מים התרכזו בתכנון אופטימלי עבור תקופה נתונה כגון 24 שעות ולא בלולאת הבקרה הסגורה עם משוב מהמערכת בתצורה של חלון זמן רציף. כמתואר לעיל, רק תוצאות צעד הזמן הראשון בתוכנית התפעול מבוצעות בפועל שכן לאחר מכן תהליך התכנון חוזר על עצמו עבור צעד הזמן הבא. ניתן לומר שהמשאבים הרבים המושקעים במציאת פתרון לתקופה ארוכה למעשה "מתבזבזים" שכן הם לא מגיעים לעולם לידי יישום. לפיכך, ובניגוד למחקרים קודמים, ניתן דגש על תהליך הבקרה הכולל בהיבט מעשי.

תרומת מחקר זה היא בפיתוח שיטות מעשיות לתפעול מערכות אספקת מים תוך התמקדות במאפיינים הייחודיים של תאגידי וספקי מים, וכן פיתוח כלים לתפעול יעיל של מערכות מים תוך התאמה למערכות המחשוב והתקשורת אשר בשימוש תאגידי וספקי המים. כלים אלו, בעלי "חתימה מחשובית", נמוכה צפויים לאפשר תפעול יעיל גם ללא ידע נרחב במערכות מחשוב ואוטומציה וללא השקעה רבה בצידוד מחשוב נוסף. ובכך, כלים אלה מאפשרים לתאגידיים ליהנות מתפעול מיטבי ללא השקעה גדולה במערכת הפיזית.

# תפעול אופטימלי בזמן אמת של מערכות אספקת מים באופן מעשי

מאת : אלעד זלמונס

בהנחיית ד"ר משהור חוש

חיבור לשם קבלת התואר "דוקטור לפילוסופיה"  
דוקטורט פרסומים

אוניברסיטת חיפה  
הפקולטה למדעי החברה  
החוג לניהול משאבי טבע וסביבה

אוקטובר, 2021

# תפעול אופטימלי בזמן אמת של מערכות אספקת מים באופן מעשי

מאת : אלעד זלמונס

חיבור לשם קבלת התואר "דוקטור לפילוסופיה"  
דוקטורט פרסומים

אוניברסיטת חיפה  
הפקולטה למדעי החברה  
החוג לניהול משאבי טבע וסביבה

אוקטובר, 2021