Real-time optimal operation of water systems under demand uncertainty and maximum water age constraints

Elad Salomons[1,*], *Hao Cao*[2], *Kristina Korder*[2], *Avi Ostfeld*[1], *Pu Li*[2]

[1] Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel

[2] Process Optimization Group, Technische Universität Ilmenau, Ilmenau 98693, Germany

*Corresponding author: eladsa@technion.ac.il

Abstract

The inherent uncertainty in water demand poses significant challenges to water distribution systems' (WDSs) efficiency and quality. This study introduces a model predictive control framework tailored for real-time optimal operation of WDSs under uncertain demand and maximum water age constraints to ensure water quality requirements. The methodology presented utilizes a scenario-based energy-cost optimization approach to account for demand uncertainties. As water age is unmeasurable, a model linearly related to some measured/observed network variables (e.g., flows, water levels, etc.) is proposed to infer water age values. Then, a scenario-based mixed-integer linear programming (SB-MILP) problem is formulated and solved repeatedly online to adjust operational strategies for minimizing energy operation costs while satisfying water age constraints. The outcome of this model is a feasible operation scheme for all demand scenarios, providing a cost-effective decision that meets water age limits. The proposed model is tested on a real-world-based test case and validated through a series of sensitivity analyses.

Plain Language Summary

The uncertainty in water demand makes it hard to maintain the efficiency and quality of water distribution systems. This study presents a control framework designed for the real-time optimal operation of these systems, taking into account unpredictable demand and water age requirements to ensure water quality. The method uses scenario-based optimization to handle demand uncertainties. Since water age can't be directly measured, the study proposes a model that estimates it based on

certain observable variables in the network. A scenario-based mixed-integer linear programming (SB-MILP) problem is then repeatedly solved online to adjust operational strategies, aiming to minimize costs while meeting water age requirements. This model provides a practical operation plan for all demand scenarios, offering a cost-effective solution that adheres to water age limits. The proposed model is tested on a real-world scenario and validated through various sensitivity analyses.

Keywords

Water demand, water distribution, real-time operation, demand uncertainty, water age, model predictive control, scenario-based optimization

Highlights

- The paper presents a real-time control framework for water distribution systems under demand uncertainty.
- It uses scenario-based optimization to minimize costs and maintain water quality.
- Validation on the C-Town network model effectively performed under varying demand conditions.

Introduction

Water distribution systems (WDSs) are critical infrastructures that deliver potable water to residential, commercial, and industrial consumers. Ensuring a continuous supply of high-quality water requires operators to confront multiple challenges, among the most pressing of which are uncertainties in water demand, adherence to regulatory requirements, and cost-effective resource allocation. Indeed, water demand is inherently stochastic, influenced by factors such as diurnal consumption patterns, weather conditions, and consumer behavior. These fluctuations, coupled with stringent water quality regulations, complicate the task of maintaining system efficiency and reliability. A primary water quality concern relates to the control of water age, which is defined as the time water spends within the distribution network before reaching end-users (Dandy et al., 2013). Water age is a critical indicator of overall water quality because it is directly tied to the decay of disinfectant residuals and

the formation of disinfection by-products in the network. Consequently, prolonged water age can result in water that does not meet stipulated safety standards, making effective control strategies for water age an essential component of modern WDS management (Mala-Jetmarova et al., 2017).

Traditional approaches to WDS operation and management often rely on rule-based methods or deterministic optimization models that do not explicitly consider the stochastic nature of water demand. While these methods can be straightforward to implement, they may yield suboptimal solutions when demand deviates from expected values, thereby undermining both efficiency and water quality goals. Model-based predictive methods have emerged as a promising alternative for handling the complexities inherent in large-scale water networks. Among such methods, Model Predictive Control (MPC) has attracted significant attention due to its capacity to address multivariable control problems and manage constraints in a systematic manner (Camacho & Bordons, 2007). By predicting the future states of the system, such as tank levels, flow rates, and pressures, MPC can proactively determine control actions (for pumps, valves, and other actuators) that balance cost, system reliability, and quality considerations (Salomons & Housh, 2020b; Tang et al., 2014). Since a future time horizon is considered for dynamic optimization, it benefits both constraint satisfaction and cost minimization in a dynamic way. For instance, when it is known (from the statistical point of view) that the demand will considerably decrease in the future, the pump will decrease the water supply in advance. The amount of the pumping reduction is optimized based on the model-based prediction, such that the water age will not be higher than its upper bond. To minimize the operation cost, we would like to pump more water in the off-peak tariff period. For this purpose, the dynamic optimization will empty the tanks in advance. Otherwise, the water level will be over its upper bound.

An especially powerful feature of MPC is its ability to incorporate water quality constraints directly into the decision-making process. For instance, constraints on water age or disinfectant residual can be formulated as part of the MPC optimization, ensuring that operational decisions meet both quantity and quality criteria (Cherchi et al., 2015). This approach contrasts with conventional strategies in which water quality is treated separately or secondarily, often through post-processing checks rather

than real-time control adjustments. Furthermore, many regulatory standards for WDSs emphasize not just meeting average quality measures but also ensuring that water quality remains within acceptable limits even under adverse or fluctuating conditions. Consequently, the flexibility nature of MPC make it an attractive tool for system operators seeking to fulfill regulatory obligations without incurring excessive operational costs.

One challenge, however, is that MPC, in its basic (deterministic) form, may not be fully equipped to handle high degrees of uncertainty in demand and water quality parameters. Demand uncertainty is particularly critical, as deviations in forecasted consumption can lead to suboptimal operation, such as overfilling or underfilling storage tanks, increased pumping costs, or degraded water quality due to longer residence times. To address these issues, researchers have proposed scenario-based MPC frameworks (Bernardini & Bemporad, 2009; Lucia et al., 2013; Maiworm et al., 2015; Xu et al., 2022). Under this paradigm, demand uncertainty is modeled by generating a finite set of possible demand profiles (or scenarios) over the prediction horizon. These scenarios capture a range of possible future demand patterns, reflecting the uncertainty inherent in water consumption. While these scenarios include normal demand variations, it should be noted that extreme conditions, such as partial system malfunction, fires, etc., are not part of this framework. In such cases, the objective of system operators focuses on supplying as much of the demand and prioritizing specific zones in the system rather than energy cost savings. Our approach evaluates the performance of the system by considering multiple demand scenarios simultaneously, leading to robust operations.

Within each scenario, the model tracks both hydraulic and water quality variables, including water age. Through a mixed-integer linear programming (MILP) or similar optimization formulation, scenario-based MPC ensures that the selected control actions are robust, that is, they maintain acceptable performance across the range of demand scenarios considered (Creaco et al., 2019). Wytock et al. (2017) demonstrated the effectiveness of scenario-based robust MPC in dynamic energy management applications, providing a foundation for extending similar methodologies to the water sector. By addressing uncertainties proactively, scenario-based MPC can balance multiple objectives,

such as minimizing operational costs, maintaining adequate pressure, and controlling water age below regulatory thresholds.

The concept of water age itself has long been recognized as an essential surrogate for water quality in distribution systems. Rossman & Boulos (1996) suggested detailing numerical methods for modeling water quality in complex pipe networks, focusing on the transport and decay of contaminants as well as the importance of water age in capturing disinfectant residual behavior. In their approach and in widely used hydraulic solvers such as EPANET (Rossman et al., 2020), water age is typically treated as a conservative tracer starting at zero at each source, then increasing over time as it travels through pipes and mixes at junctions and tanks. This simulation framework has made it possible for subsequent studies to incorporate water age into optimization and control algorithms.

Indeed, recent research has explored the simultaneous minimization of water age and other system performance metrics. Korder et al. (2024) proposed an optimization approach aimed at reducing both water age and pressure through strategic placement and operation of pressure reducing valves. Such integrative approaches highlight the value of considering multiple objectives (e.g., pressure, energy cost, and water age) in a single optimization framework, especially given that reducing excessive pressure may also mitigate leakage but could inadvertently increase residence times in certain parts of the network. Similar trade-offs arise in many real-world networks, necessitating sophisticated modeling and control techniques.

Beyond demand uncertainties, variability in water quality parameters, including initial disinfectant concentrations, decay kinetics, and by-product formation rates, can further complicate operational decisions. Basupi & Kapelan (2015) highlighted the importance of robust system design for future demand scenarios by stressing resilience and flexibility. Such principles can be extended to operational control, where real-time adjustments must consider not only shifting demands but also the evolving chemical and biological processes in the water. Roach et al. (2016) compared robust optimization with info-gap methods in water resource management, offering insights into how different uncertainties can be addressed depending on the complexity of the system and the tolerance

for risk. Ultimately, the goal is to ensure reliability and service quality even under the broadest plausible range of operating conditions.

EPANET (Rossman et al., 2020) and other hydraulic solvers have proven invaluable in capturing water age dynamics. EPANET's conservative tracer model assumes complete mixing in tanks, which can sometimes underestimate local water age gradients but remains a practical compromise for real-time or near-real-time computations. As water moves from sources to consumers, the age accumulates, and any deviation from the assumed demand schedule affects the predicted water age distribution (Braun et al., 2020). When demand is higher than expected, turnover in the network occurs more rapidly, potentially lowering water age, whereas lower-than-expected demand can lead to extended residence times. Hence, the uncertainty in demand manifests directly as uncertainty in the water age. By employing a scenario-based approach, operators can envision both "high-demand" and "low-demand" situations as well as intermediate conditions, creating an envelope within which water age values fluctuate. The control algorithm then identifies a robust operational strategy (e.g., pump schedules and valve settings) that maintains water quality within acceptable limits across all scenarios. This robust approach contrasts with deterministic optimization, which might only target an expected demand profile and fail to account for variability. In real-world operations, such flexibility is key for compliance with water quality regulations, especially when consumer demand exhibits sudden changes or seasonal shifts.

Building on these foundational insights, the present study proposes a scenario-based Linear Model Predictive Control (LMPC) framework that explicitly addresses demand uncertainty while controlling water age within acceptable bounds. The framework begins with the development of a linear model that relates observable states (e.g., tank levels, pressures, and flow rates) to water age, serving as an inferential sensor. In real time, this sensor is updated based on actual measurements, allowing water age to be predicted over a future horizon (e.g., 24–48 hours). Scenarios capturing different plausible demand profiles is then constructed, and a scenario-based Mixed-Integer Linear Programming (SB-MILP) problem is formulated. The solution to this problem yields control actions (pump switches, valve settings, etc.) that minimize operational costs subject to water age constraints for each scenario.

The linear structure of the model ensures computational tractability, making it suitable for time-critical applications where decisions must be made with minimal delay.

To validate the proposed scenario-based LMPC, this study applies it to the C-Town benchmark network (Ostfeld et al., 2012). By leveraging EPANET's hydraulic and water quality simulation capabilities, as a surrogate to the real WDS, the study demonstrates how real-time adjustments in pumping schedules and valve manipulations can keep water age within acceptable thresholds across multiple demand scenarios. The results confirm that the methodology not only reduces operational costs but also achieves robust adherence to water quality standards, thus underscoring its potential for deployment in full-scale water utilities.

In summary, the integration of MPC with scenario-based optimization offers a sophisticated and robust framework for real-time WDS operation under uncertain demand and water quality conditions. Water age serves as a critical surrogate for water quality, linking disinfectant residual decay and by-product formation with system hydraulics. By proactively addressing the stochastic nature of demand, scenario-based LMPC ensures that water age remains within acceptable limits, even under shifting consumption patterns. This paper's contribution lies in unifying these concepts into a single, linearized control structure and validating the approach through a comprehensive real-world case study. Such methodology paves the way for broader industry adoption, as water utilities increasingly seek data-driven, cost-effective solutions that guarantee safe and reliable service to consumers.

The remainder of this paper is organized as follows. The methodology section details the water age estimation modeling and the proposed MPC framework. The case study and results section presents the application of the framework to the C-Town network and discusses its outcomes. Finally, the conclusion section summarizes our findings and suggests directions for future research.

Methodology

Water age estimation

Water quality models, including water age, are based on the principles of conservation of mass coupled with reaction kinetics (Rossman et al., 1993; Rossman & Boulos, 1996). Both principles are governed by nonlinear equations and require the hydraulics to be solved first. With the aim of preserving a linear model for the purpose of a real-time application, we propose a multivariate regression model for water age estimation, Eq. (1).

$$wa_j = B^T X \tag{1}$$

where, wa_j is the water age at junction j for a given time, X is a vector of observations of state variables (e.g., water levels at tanks, flows, etc.) associated with junction j, and B is a vector of parameters containing the regression coefficients. In this paper, a junction can also be a tank. A least-squares method is employed to determine the optimal coefficient values by minimizing the sum of squared residuals. That is, for each junction of interest, we formulate the quadratic optimization problem depicted in Eqs. (2)-(3).

$$\min \sum \left(wa_t^{sim} - wa_t^{est} \right)^2 \quad \forall t \in T \tag{2}$$

s.t.,

$$wa_t^{est} = \sum b_p x_{f,t} \quad \forall t \in T, p \in P \tag{3}$$

where, wa_t^{sim} is the water age which is provided by simulation using EPANET at time t, wa_t^{est} is the estimated water age at time t (for the given junction), T is the time span of the optimization period ($t \in T$), b_p is the p-th regression coefficient ($b \in B$), $x_{p,t}$ is the p-th observation variable at time t, and P is the set of regression coefficients ($p \in P$). It should be noted that this formulation does not restrict the estimated water age to be positive to avoid over-constraining the problem. This is because we are interested in the higher values of the water age, not around zero. At some junctions, as will be shown for the test case, the water age time series has sharp peaks and, solving problem Eq. (3), large residuals will remain at the peak points. Therefore, for these junctions we add a second stage of

a weighted least-squares optimization. The time based weights, k_t, are calculated using the first stage results, Eq. (4).

$$k_t = \left| wa_t^{sim} - wa_t^{est} \right| \tag{4}$$

Then, the first stage objective function with Eq. (2) is replaced as by Eq. (5).

$$\min \sum k_t \left(wa_t^{sim} - wa_t^{est} \right)^2 \quad \forall t \in T \tag{5}$$

Noteworthy is that k_t is calculated with the results of the first stage while the error in the water age in Eq. (5), $wa_t^{sim} - wa_t^{est}$, is calculated during the second stage. Although the two-stage optimization problem is nonlinear, the number of regression parameters is relatively small. As will be seen in our case study, less than ten regression terms are usually enough for water age estimation. As a result, we obtain Eq. (3) as a model to estimate water age at junctions in the network, based on the online observation of the variables X. This makes it possible to design a model-based optimal control of WDSs in which a direct measurement of water age is not available.

Formulation of the scenario-based linear MPC

In this study, the aim of optimal operation is defined, under demand uncertainty, as cost minimization of WDS, i.e. to minimize the energy costs of the pump stations and meanwhile restrict the water age in a user-defined range by controlling pumps and valves. We generate several demand scenarios to describe the demand uncertainty. Therefore, the objective function, Eq. (6), comprises of two terms: the first is the energy operating cost of the system and the second is a penalty term to minimize the water age deviation in the system from the required values.

$$\min \sum_{n \in SCN} \sum_{t \in T} \sum_{s \in S} \sum_{c \in C_s} u_{s,c,t}^n \cdot q_{s,c} \cdot e_{s,c} \cdot EC_t \cdot \Delta t + \sum_{j \in J} \sum_{t \in T} M \cdot \Delta wa_{j,t}^n \tag{6}$$

The first term loops over the set of demand scenarios ($n \in SCN$), time ($t \in T$), pump stations ($s \in S$), and pump combinations within each pump station ($c \in C_s$). T is the user-defined prediction

horizon. $u^n_{s,c,t}$ is a decision variable determining whether a specific pump combination is operating or not. $q_{s,c}$ is the flow of pumping combination c of station s, $e_{s,c}$ is its specific energy, EC_t is the electricity cost (known tariff from the energy market) at time t, and Δt is the time step duration. Thus, the operation cost is a linear function of the decision variable u which can be integer or continuous, expressed as follows:

$$u \in \{0,1\} \quad \forall t \in T^{\text{int}} \tag{7}$$

$$0 \le u \le 1 \quad t \in T - T^{\text{int}} \tag{8}$$

$$\sum_{t \in T} u^n_{s,c,t} \le 1 \qquad \forall s \in S, c \in C_s, n \in SCN \tag{9}$$

This decision variable u, in its integer form, Eq. (7), represents whether a specific pumping combination is operated ($u = 1$) or not ($u = 0$) during time step t. While in its continuous form, Eq. (8), it represents the fraction of time in which the pumping combination is operating during time step t. The set of time steps, in which u as an integer variable is T^{int}. The size of T^{int} is referred to as the *binarization level* (Salomons & Housh, 2020a). A binarization level of one means that only in the first-time step the decision variables are integer while for the rest of the time steps they are continuous. The aim of introducing the binarization level is to relax the MILP problem to reduce the computational burden. The regionalization of the binarization level approach is that, in the MPC loop, only the first-time step's decisions are implemented to the system, as described above. Thus, during a single time step, $u^n_{s,c,t}$ is limited by one (see Eq. (9)), since only one pumping combination can be operated in a station at any given time.

The second term of the objective function Eq. (6) aims to minimize the water age deviation, Δwa in the system from the upper bound values $wa^{\max}$, by imposing a penalty parameter M on the deviation. This is over the set of all junctions, j being evaluated ($j \in J$). Here, Δwa is a positive decision variable which will be zero if the water age at the junction is within the valid range, i.e.:

$$wa_{j,t}^{n} \le wa_{j}^{\max} + \Delta wa_{j,t}^{n} \tag{10}$$

where $wa_{j,t}^{n}$ is the water age in junction j at time t for scenario n and $wa_{j}^{\max}$ is the upper bound of the water age at junction j. The formulation of Eq. (6) and Eq. (11) leads to a *soft-constraint* for the water age to avoid infeasibility.

As described above in Eq. (3), the water age is a linear function of the observation variables, briefly expressed as:

$$wa_{j,t}^{n} = f\left(wa_{j,t-1}^{n}, q(u_t^n), v_t^n\right) \qquad \forall j \in J, n \in SCN \tag{11}$$

where $wa_{j,t-1}^{n}$ is the water age at the junction in the previous time step to describe the water age dynamics. $q(u_t^n)$ denotes the flows from pumps corresponding to the pumping combinations selected by the optimization, and v_t^n are the volumes of the water tanks. The mass balance of the tanks is given in Eq. (12).

$$v_{r,t}^{n} = v_{r,t-1}^{n} + \sum_{s \in S_{in,r}} \sum_{c \in C_s} u_{s,c,t}^{n} \cdot q_{s,c} \cdot \Delta t - \sum_{s \in S_{out,r}} \sum_{c \in C_s} u_{s,c,t}^{n} \cdot q_{s,c} \cdot \Delta t - d_{r,t}^{n} \quad \forall r \in R, t \in T, n \in SCN \tag{12}$$

where $v_{r,t}^{n}$ and $v_{r,t-1}^{n}$ are the water volumes of tank r in the demand scenario n during the current and previous time steps, respectively. The second and third terms in Eq. (12) are the volumes of water pumped into and out of the tank, where $s_{in,r}$ and $s_{out,r}$ are the set of stations pumping water in and out of the tank, respectively. The last term in Eq. (12) is the demand associated with the pressure zone of the tank r. The water volume in the tanks is constrained as follows

$$V_{r,t}^{\min} \le v_{r,t}^{n} \le V_{r,t}^{\max} \qquad \forall r \in R, t \in T, n \in SCN \tag{13}$$

$$v_{r,0}^{n} = V_{r}^{init} \quad \forall r \in R, n \in SCN \tag{14}$$

$$v^{n}_{r,t_{\max}} \geq V_{r,t_{\max}} \quad \forall r \in R, n \in SCN \tag{15}$$

where $V^{\min}_{r,t}$ and $V^{\max}_{r,t}$ are the volume upper and lower bound of tank r in time t, respectively. Finally, to describe the demand uncertainty, we define several demand scenarios by demand multipliers. For the first time step ($t = 0$), we assume for all scenarios the same value of the decision variables, Eq. (16), since this is a viable decision to be implemented for the pump stations.

$$u^{n}_{s,c,0} = u^{1}_{s,c,0} \quad \forall s \in S, c \in C_s, n \in SCN \tag{16}$$

As a result, our scenario-based mixed-integer linear MPC is formulated by Eqs. (6)-(16). The structure of its realization is shown in Figure 1. The WDS to be controlled consists of pump stations and a water distribution network (WDS). In our case study, the WDS is simulated by EPANET. In the MPC, at each time point, the observation variables X will be measured from the WDS and supplied to the water age estimator. The estimated water age WA is then used in the solution of the scenario-based MILP problem. Its result U will be supplied for the optimal operation of the pump stations.

Figure 1: Structure of scenario-based MPC

Case study and results

The developed approach is applied on a portion of the real-world based C-Town network (Salomons, 2025), as shown in Figure 2a, (Ostfeld et al., 2012). The network consists of 186 junctions, 210 pipes (of about 30 Km in length), two tanks (T1 and T5, with volumes of 5000 and 500 cubic meters respectively), and two pumping stations (S1 and S4 with a maximum capacity of about 685 and 215 m^3/hr, respectively) in which there are five pumps in total.

Figure 2: C-Town (a) and the test part (b) networks

As a first step, the water age, according to Eqs. (2) and (3), is required to be estimated. As representative nodes of the network, Tanks T1, T5, and junction J1056 are selected for the water age estimation. The logic behind this selection is that most of the water flows through the two tanks, and J1056 is a junction in the middle of the main pressure zone supplied by S1 and T1. An EPANET simulation with a duration of 1,000 hours is performed for the water age to be stabilized and less affected by the initial condition of the network. We consider this state as the initial state for testing our approach. The observation variables ($x \in X$) used for the estimation of the water age for Tank 1, Tank 5 and J1056 are given in Table 1, along with the resulted regression coefficients ($b \in B$), respectively.

Table 1: Water age model estimation results

The quality of the estimation of the water age for T1 and T5 is shown in Figure 3. It can be seen that the water age prediction is quite good with some deviations of 5 to 10 hours. The RMSEs for T1 and T5 are 4.7 and 3.5 hours, respectively. However, it should be noted that the accuracy of the prediction is relevant only for the “peaks”, or high values, of the water age as this estimation is used for the maximum water age constraint. As will be seen in the MPC optimization results, the water prediction accuracy is well fitted for this purpose. However, applying the procedure for the estimation of the water age at junction J1056 is not satisfactory, as shown in Figure 4a. It can be seen that the procedure fails to estimate the high peaks of the water age. As described in the water age estimation section, for such cases we add a second stage for the estimation algorithm (see Eqs. (4) and (5)). The results of the second stage are shown in Figure 4b, indicating a significant improvement at the peak points. It is noteworthy that, although the peaks in the water age values are now estimated accurately, the accuracy of the low values of the water age is limited. As mentioned above, we are concerned about the range or the limit of the water age estimation and the error at low levels of water age has no impact on our closed-loop control. For example, if the water age maximum levels allowed are in the range of 30 hours, then inaccuracy in estimating low water age levels are not meaningful.

319

320 Figure 3: water age estimation for (a) Tank1, and (b) Tank 5

321

322 Figure 4: Water age estimation for junction J1056, (a) first stage, and (b) second stage

323 Based on the water age model, the scenario-based linear MPC is now used for real-time optimization.
324 The flow and specific energy of the combinations of the pump stations, $q_{S,C}$ and $e_{S,C}$ respectively,
325 are given in Table 2. These input values were derived from the operation history of the pumping
326 stations. For the energy cost, we assume that the electricity tariff, EC, has a three-tier structure: on-
327 peak, mid-peak, and off-peak with values as 1.0676, 0.6182, and 0.3407 NIS/kWh (NIS – New Israeli
328 Shekel), respectively. In addition, the duration of the peaks needs to be defined. We assume that
329 during weekdays, Sunday through Thursday, the mid-peak hours are 6-8, on-peak 14-21, and the rest
330 is off-peak. On Friday, the mid-peak hours are 16-20, and the rest is off-peak (no on-peak on Friday).
331 On Saturday, the on-peak hours are 17-19, mid-peak 19-21, and the rest is off-peak, respectively.

332 Table 2: Flow and specific energy of the combinations of the pump stations

333

334 The base optimization run is performed over one week. That is, the run consists of 168 optimization
335 steps, each with a prediction horizon $T = 48h$ with a time step $\Delta t = 1h$. The binarization level, T^{int},
336 is set to five hours. The constraint (upper bound) on the water age is set as 50 hours for the two tanks
337 (T1 and T5), and 40 hours at junction J1056. It should be noted that, since the estimation of the water
338 age is not fully accurate, we use a safety level (i.e., a setback) to ensure the real water age stays inside
339 the specified range. Thus, we define a safety level of at least 10 hours.

340 The penalty parameter in the objective function (see Eq.(6)) for deviations from the water age
341 constraints, M, is set to 100 NIS per hour of deviation. The lower bound of the tank volumes is
342 1,000 and 50 m^3 and their upper bound is 4,700 and 450 m^3 for tanks 1 and 5, respectively. To

describe the demand uncertainty, we use two demand scenarios which are set as 5% below and above the expected demand. It should be noted that future expected demands are not known and need to be forecasted. However, in this study, we assume that any forecast error will be included in the scenario-based 5% spread.

The initial water age at the tanks and junction J1056 is set to 20 hours which we achieved from simulation with EPANET. Each run of the optimization provides an optimal operation profile for the pump stations over the prediction horizon. The decision for the first hour will be realized in the real WDS. For the case study, the EPANET model is used as the real WDS with the actual demands, to simulate the state of the system. At the end of the first time step, the values of the state variables will be used as the initial condition for the next optimization run.

All optimization runs were performed with the Gurobi solver (Gurobi Optimization, 2014) and the OpenSolver framework (Mason, 2012). The problem is a mixed-integer linear programming (MILP) and the default solver settings were used, with a stop criteria gap of 0.5%. OpenSolver is an Excel VBA add-in that extends Excel's built-in Solver with more powerful free and commercial solvers. The resulting optimization problem has 1248 decision variables and 874 constraints. The results of the one-week optimization runs are shown in Figure 5 and Figure 6 (the "base-run", see the sensitivity analysis below). Figure 5 shows the resulting water volumes in the two tanks and the water flows at the pumping stations by realizing the optimized operation profile. It can be seen that, as expected, the tank volumes are within the lower and upper bounds. Furthermore, from the economic point of view, they follow a daily pattern which corresponds to the energy tariff change during the day, i.e., the tanks are filled during the off-peak tariff hours and drained during the mid-peak and on-peak periods. Similarly, the pump stations pump water into the tanks, Station 1 to Tank 1 and Station 4 to Tank 5, with higher flows during the off-peak hours and lower flows during the on-peak periods.

Figure 5: Tank volume and station flow trajectories by the scenario-based MPC

The resulting water age trajectories, as calculated by EPANET, for Tank 1, Tank 5, and junction J1056 are given in Figure 6a, Figure 6b, and Figure 6c, respectively. As can be seen, the water age is within the specified range, i.e., there are no violations of the water age upper bound. The total operating cost for the week, for this base run is 7,027 NIS. It should be noted that the current operating cost of this system is unavailable. However, in the next section, we compare different scenarios to the base run, and show the effect of demand variability on the operating costs and the water age.

Figure 6: Water age trajectories by the scenario-based MPC (EPANET calculated)

Figure 7: Run-time cumulative probability

For the real-time operation of a water distribution system, the computation time for solving the optimization problem is essential since the update of the decision variables should be made available within the defined sampling time $\Delta t = 1hr$. Therefore, a run-time experiment with multiple optimization runs is performed by using a personal computer (Lenovo 12th Gen Intel® Core™ i7-1260P 2.1GHz with 32 GB of RAM) and its result is presented in Figure 7. As can be seen, the maximum run-time of the optimization procedure does not exceed 45 seconds, which is sufficient for the real-time control, specifically when using a time-step of one hour.

Sensitivity analysis

To investigate the impact of different demand scenarios, a set of sensitivity analysis (SA) runs were performed, which are summarized in Table 3. First, in SA1, we use the true demands (Ostfeld et al., 2012) instead of the two demand scenarios in the base run (BR). It yields a slightly lower cost of 7,002 NIS, which shows a reduction of 0.4% from the BR. Although the cost reduction is not significant, a lower cost is expected due to the decreased uncertainty in the demand.

Table 3: Sensitivity analysis results

In SA2-SA4, we use different degrees of demand scenarios, i.e., 3%, 8%, and 10% below and above the expected demand (compared to the 5% of the BR). The resulting tank volume and water age trajectories for Tank 1 are given in Figure 8 and Figure 9, respectively. As expected, it can be seen in Table 3 that the total cost of operation slightly increases with the increase of the uncertainty level (with a small exception of the 8% spread). From Figure 8 it can be seen that, as the uncertainty increases, the use of the water volume increases as well so as to be "prepared" for possible increases in the demand. Consequently, as the water volume in the tank increases, the water age increases as well, with small violations of the 50h water age constraint, as shown in Figure 9.

Figure 8: Water volume trajectories in Tank 1 in different demand scenarios (SA1-SA4)

Figure 9: Water age trajectories at Tank 1 in different demand scenarios (SA1-SA4)

Furthermore, the effect of the maximum volume of Tank 1 is examined. Specifically, we define its maximum volume as 2,000, 2,300, and 2,500 m^3, denoted as SA5-SA7, respectively, compared to the maximum volume of 4,700 m^3 in the BR. The optimal tank volume and water age trajectories in Tank 1 are given in Figure 10 and Figure 11, respectively. It can be seen that the effect of a smaller tank on the total operating cost is significant, i.e. as the water volume decreases, the operating costs increase, as shown in Table 3. It should be noted that the minimum volume for all the cases remained at 1,000 m^3. The reduction of the operating costs with the increase of the available storage is expected as more pumping can be shifted for the on-peak to the off-peak hours and for water to be stored in the tank for on-peak use. From the water age perspective, the water turnover volume ratio of the tank increases for the larger volumes (i.e., improving the mixing). For example, in the base run,

417 the water level reaches 3,000 m^3 and drops to the minimum level of 1,000 m^3, which is a turnover of
418 about two-thirds of the total volume. Contarary, when the maximum volume is set at 2,000 m^3, the
419 turnover is only half. Consequently, the water age is lower for the larger volumes, as shown in Figure
420 11. It should be noted that for SA5 (tank maximum volume as 2,000 m^3) the water age slightly
421 exceeds the upper bound. This is because in our problem formulation, a *soft-constraint* for the water
422 age is used to avoid the issue of infeasibility.

423

424 Figure 10: Tank 1 water volume trajectories for SA5-SA7 (BR: Base Run)

425

426 Figure 11: Tank 1 water age trajectories for SA5-SA7

427 In SA8, a case of a smaller tank (the upper bound of Tank 1 as 2,000 m^3) without the water age
428 constraints is studied. The results show that the effects are minor in this test case. The total operating
429 cost is reduced with less than 1% (see Table 3), the water volume is similar (see Figure 12), and
430 Tank1 water age is slightly higher, as expected, since no constraint is posed on water age, as shown in
431 Figure 13.

432

433 Figure 12: Tank water volume trajectories for SA8 (WA: Water Age)

434

435 Figure 13: Tank water age trajectories for SA8 (WA: Water Age)

436 Finally, in SA9, we examine the robustness of the proposed algorithm, with the Tank 1 maximum
437 volume of 2,000 m^3 (i.e. a considerably smaller tank) as in SA5, in the case where the actual demand
438 is increased by 5%. The optimization results show that, as expected, higher demands require
439 additional pumping, thus the operating costs increased by 8.8% (see Table 3 for SA5 and SA9).

Figure 14 shows the water age at the two tanks. It can be seen that the water age with the higher demands are slightly lower, this is due to the increased pumping and rapid water change in the tanks.

Figure 14: Tank water age trajectories for SA9 (BR: Base Run)

Conclusions

This study proposed a multivariable linear regression model to predict water age across nodes within a water distribution network based on measurable variables like tank levels and water flows. We use a first- and second-stage optimization scheme to refine the estimates, particularly for high water age peaks. Second, a scenario-based MPC framework was developed to operate WDSs with demand uncertainty. The formulated MPC solves a mixed-integer programming problem in real time, minimizing the operating costs according to the electricity tariff and mostly satisfying the specified water age constraints. The results of a case study demonstrated that the system will indeed maintain water age within the acceptable limits while minimizing operational costs and meeting various constraints such as tank volumes. Furthermore, a sensitivity analysis demonstrated the robustness of the proposed control strategy for the test case under various demand scenarios and tank volume constraints, indicating only minor deviations from the expected performance even under unsuitable conditions. Using the proposed approach, the overall operational performance was effectively managed, and its flexibility in response to different scenarios was confirmed, illustrating its potential for cost efficiency and system adaptability under demand uncertainties and variable operational conditions.

Since the test case used in this study was based on a simulated, real-world-inspired water distribution system, future work should include validating the proposed methodology on actual operational networks. Testing the approach on a real-world network would provide deeper insights into the practical applicability of the approach. Another direction would be the extension of the framework to

larger and more complex networks that include multiple critical junctions with water age constraints. This may require enhancements in both the water age estimation model and the MPC formulation to ensure scalability and computational efficiency. Developing a systematic methodology for identifying water-age-critical nodes based on topological, hydraulic, or demand-based criteria would also strengthen the framework's applicability across diverse networks. In addition, an automated procedure for selecting an appropriate model complexity of water age (e.g., the number and type of variables in the regression model) could be explored using machine learning or adaptive modeling techniques. This would allow the estimation model to balance accuracy and computational load depending on a given system's specific needs. Lastly, the economic impacts of different tariff structures and the inclusion of renewable energy sources (e.g., solar-powered pumps) could be evaluated to further improve cost efficiency and sustainability.

Acknowledgment

This study was funded by the German Research Foundation (DFG) under project number 327870500 (LI 806/20-2), and by the Israeli Water Authority (budget number 2033800).

Data Availability Statement

The data of the test case network used in this research have been uploaded to the FAIR-compliant Zenodo online repository (Salomons, 2025).

References

Basupi, I., & Kapelan, Z. (2015). Evaluating Flexibility in Water Distribution System Design under Future Demand Uncertainty. *Journal of Infrastructure Systems*, *21*(2), 04014034. https://doi.org/10.1061/(asce)is.1943-555x.0000199

Bernardini, D., & Bemporad, A. (2009). Scenario-based model predictive control of stochastic constrained linear systems. *Proceedings of the IEEE Conference on Decision and Control*, 6333–6338. https://doi.org/10.1109/CDC.2009.5399917

Braun, M., Piller, O., Deuerlein, J., Mortazavi, I., & Iollo, A. (2020). Uncertainty quantification of water age in water supply systems by use of spectral propagation. *Journal of Hydroinformatics*, *22*(1), 111–120. https://doi.org/10.2166/hydro.2019.017

Camacho, E. F., & Bordons, C. (2007). *Model predictive control (book)*. *Studies in Systems, Decision and Control* (Vol. 197). Retrieved from http://link.springer.com/10.1007/978-3-030-11869-3_4%0Ahttp://link.springer.com/10.1007/978-3-030-11869-3

Cherchi, C., Badruzzaman, M., Oppenheimer, J., Bros, C. M., & Jacangelo, J. G. (2015, April 15). Energy and water quality management systems for water utility's operations: A review. *Journal of Environmental Management*. Academic Press. https://doi.org/10.1016/j.jenvman.2015.01.051

Creaco, E., Campisano, A., Fontana, N., Marini, G., Page, P. R., & Walski, T. (2019, September 15). Real time control of water distribution networks: A state-of-the-art review. *Water Research*. Pergamon. https://doi.org/10.1016/j.watres.2019.06.025

Dandy, G., A, M., Qiu, M., G., S., & Z., Y. (2013). Advantages and Limitations of the Use of Water Age as a Surrogate for Water Quality in Water Distribution Systems.

Gurobi Optimization, I. (2014). Gurobi Optimizer Reference Manual.

Korder, K., Hao Cao, ·, Salomons, E., Ostfeld, · Avi, Li, · Pu, Li, P., et al. (2024). Simultaneous Minimization of Water Age and Pressure in Water Distribution Systems by Pressure Reducing Valves. *Water Resources Management 2024*, 1–19. https://doi.org/10.1007/S11269-024-03828-6

Lucia, S., Finkler, T., & Engell, S. (2013). Multi-stage nonlinear model predictive control applied to a semi-batch polymerization reactor under uncertainty. *Journal of Process Control*, *23*(9), 1306–1319. https://doi.org/10.1016/j.jprocont.2013.08.008

Maiworm, M., Bäthge, T., & Findeisen, R. (2015). Scenario-based Model Predictive Control: Recursive feasibility and stability. In *IFAC-PapersOnLine* (Vol. 28, pp. 50–56). Elsevier. https://doi.org/10.1016/j.ifacol.2015.08.156

Mala-Jetmarova, H., Sultanova, N., & Savic, D. (2017). Lost in optimisation of water distribution systems? A literature review of system operation. *Environmental Modelling & Software*, *93*, 209–254. https://doi.org/10.1016/j.envsoft.2017.02.009

Mason, A. J. (2012). OpenSolver - An Open Source Add-in to Solve Linear and Integer Progammes in Excel (pp. 401–406). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-29210-1_64

Ostfeld, A., Salomons, E., Ormsbee, L., Uber, J. G., Bros, C. M., Kalungi, P., et al. (2012). Battle of the Water Calibration Networks. *Journal of Water Resources Planning and Management*, *138*(5), 523–532. https://doi.org/10.1061/(asce)wr.1943-5452.0000191

Roach, T., Kapelan, Z., Ledbetter, R., & Ledbetter, M. (2016). Comparison of Robust Optimization and Info-Gap Methods for Water Resource Management under Deep Uncertainty. *Journal of Water Resources Planning and Management*, *142*(9), 04016028. https://doi.org/10.1061/(asce)wr.1943-5452.0000660

Rossman, L. A., & Boulos, P. F. (1996). Numerical Methods for Modeling Water Quality in Distribution Systems: A Comparison. *Journal of Water Resources Planning and Management*, *122*(2), 137–146. https://doi.org/10.1061/(asce)0733-9496(1996)122:2(137)

Rossman, L. A., Boulos, P. F., & Altman, T. (1993). Discrete Volume-Element Method for Network Water-Quality Models. *Journal of Water Resources Planning and Management*, *119*(5), 505–517. https://doi.org/10.1061/(asce)0733-9496(1993)119:5(505)

Rossman, L. A., Woo, H., Tryby, M., Shang, F., Janke, R., & Haxton, T. (2020). *EPANET 2.2 User Manual*. Washington, DC.

Salomons, E. (2025). mini C-Town water network model [Model]. *Zenodo*. https://doi.org/10.5281/ZENODO.14759988

Salomons, E., & Housh, M. (2020a). A Practical Optimization Scheme for Real-Time Operation of

Water Distribution Systems. *Journal of Water Resources Planning and Management*, *146*(4), 04020016. https://doi.org/10.1061/(asce)wr.1943-5452.0001188

Salomons, E., & Housh, M. (2020b). Practical real-time optimization for energy efficient water distribution systems operation. *Journal of Cleaner Production*, *275*, 124148. https://doi.org/10.1016/j.jclepro.2020.124148

Tang, Y., Zheng, G., & Zhang, S. (2014). Optimal control approaches of pumping stations to achieve energy efficiency and load shifting. *International Journal of Electrical Power and Energy Systems*, *55*, 572–580. https://doi.org/10.1016/j.ijepes.2013.10.023

Wytock, M., Moehle, N., & Boyd, S. (2017). Dynamic energy management with scenario-based robust MPC. In *Proceedings of the American Control Conference* (pp. 2042–2047). https://doi.org/10.23919/ACC.2017.7963253

Xu, B., Sun, Y., Huang, X., Zhong, P. an, Zhu, F., Zhang, J., et al. (2022). Scenario-Based Multiobjective Robust Optimization and Decision-Making Framework for Optimal Operation of a Cascade Hydropower System Under Multiple Uncertainties. *Water Resources Research*, *58*(4), e2021WR030965. https://doi.org/10.1029/2021WR030965

Figure 1.

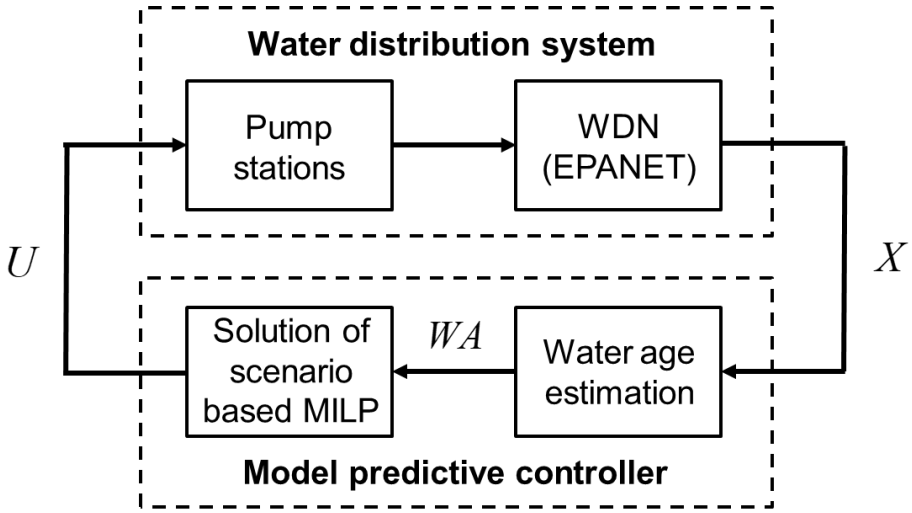

Water distribution system
Pump stations
WDN (EPANET)
U
X
Solution of scenario based MILP
WA
Water age estimation
Model predictive controller

Figure 2.

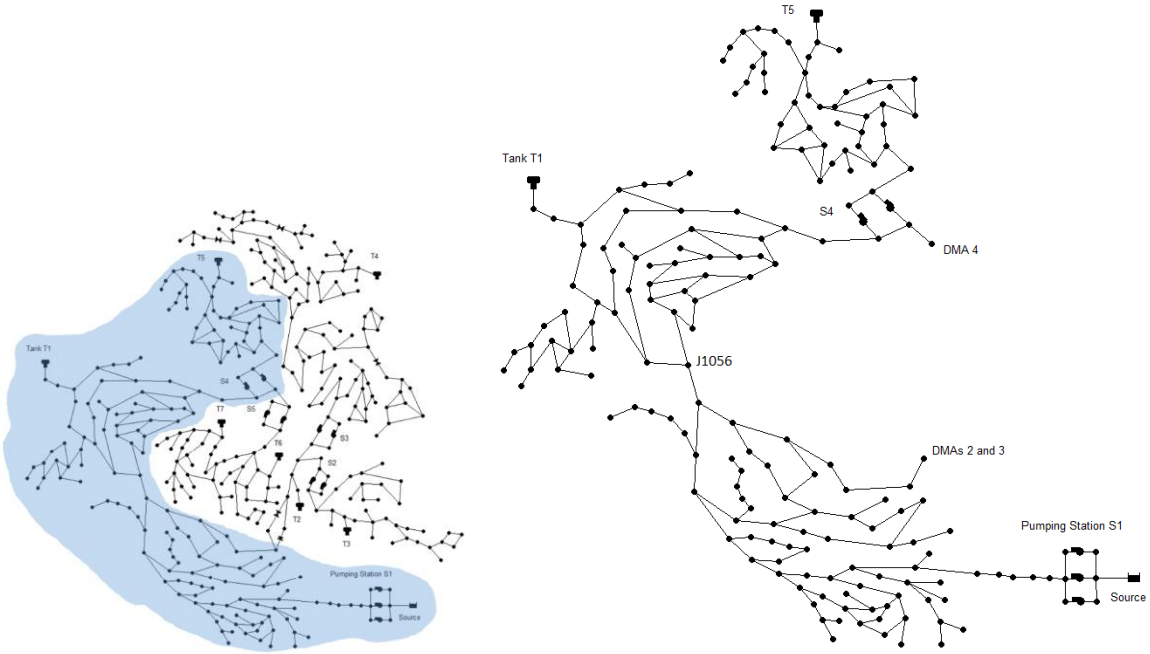

T5
Tank T1
S4
DMA 4
J1056
DMAs 2 and 3
Pumping Station S1
Source
T4
S5
T7
T6
S3
S2
T2
T3

Figure 3.

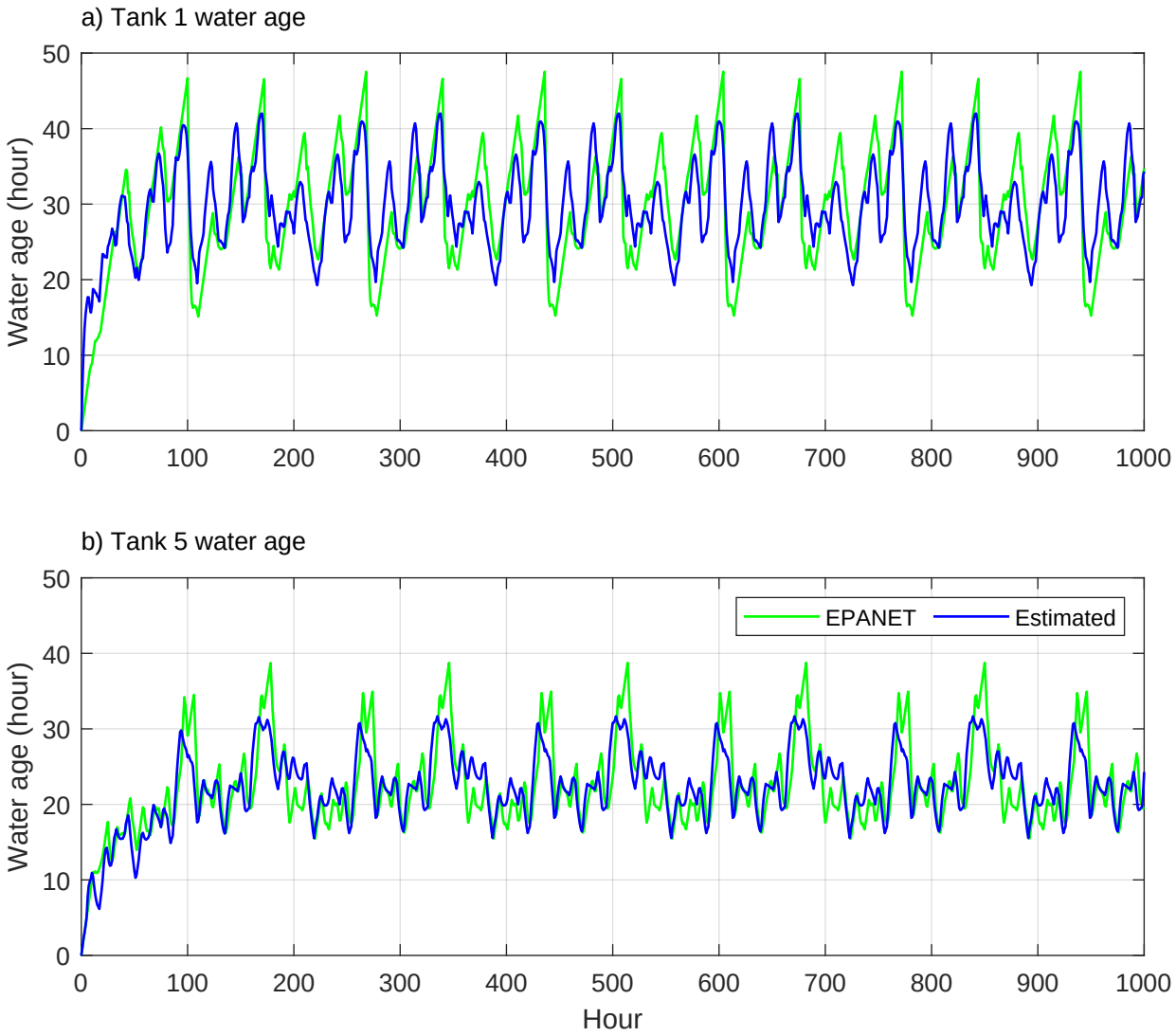

a) Tank 1 water age
Water age (hour)
0
10
20
30
40
50
0
100
200
300
400
500
600
700
800
900
1000
b) Tank 5 water age
Water age (hour)
EPANET
Estimated
Hour

Figure 4.

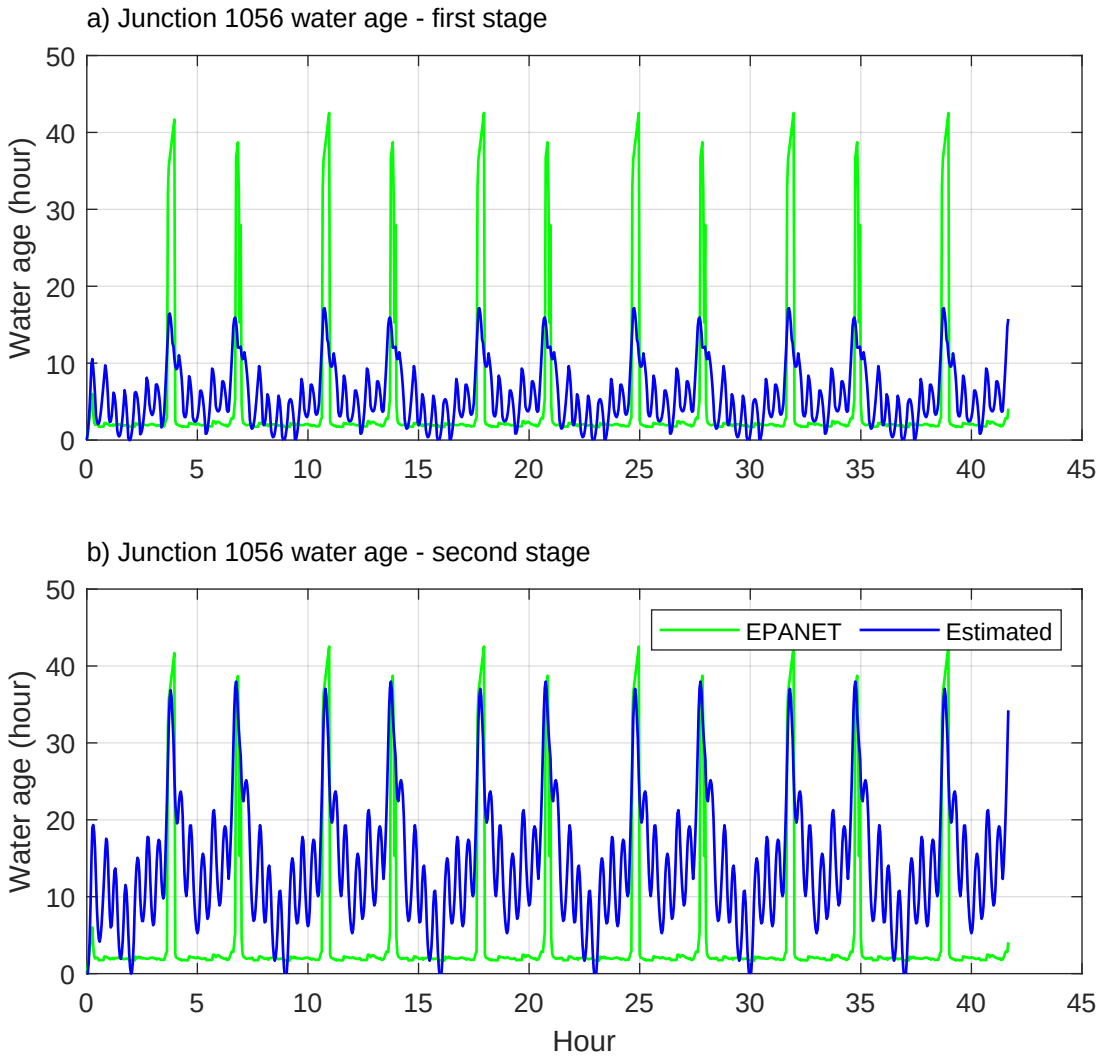

a) Junction 1056 water age - first stage
Water age (hour)
0
10
20
30
40
50
0
5
10
15
20
25
30
35
40
45
b) Junction 1056 water age - second stage
EPANET
Estimated
Water age (hour)
Hour

Figure 5.

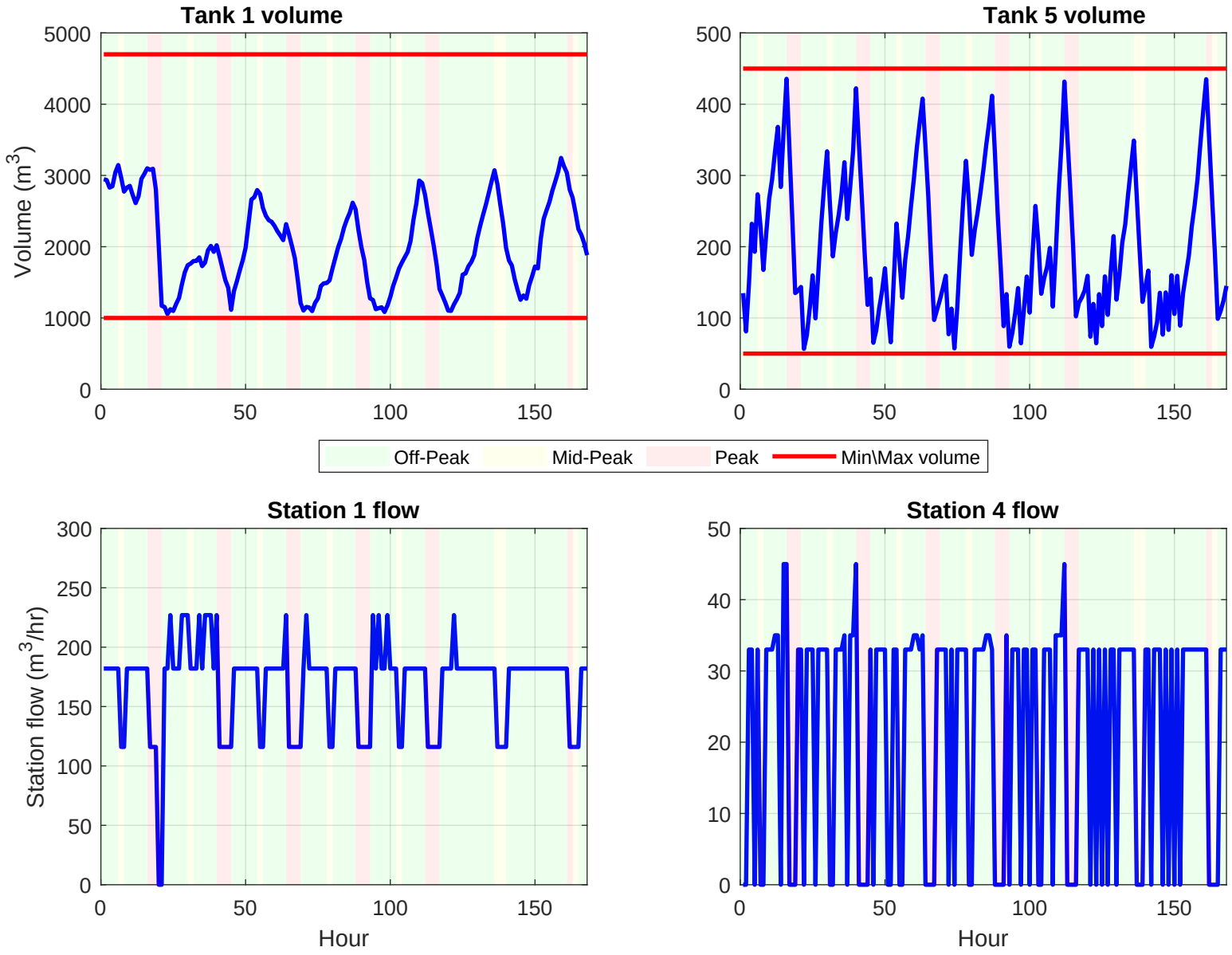

Tank 1 volume
Volume (m3)
5000
4000
3000
2000
1000
0
0
50
100
150
Tank 5 volume
500
400
300
200
100
0
0
50
100
150
Off-Peak
Mid-Peak
Peak
Min\Max volume
Station 1 flow
Station flow (m3/hr)
300
250
200
150
100
50
0
0
50
100
150
Hour
Station 4 flow
50
40
30
20
10
0
0
50
100
150
Hour

Figure 6.

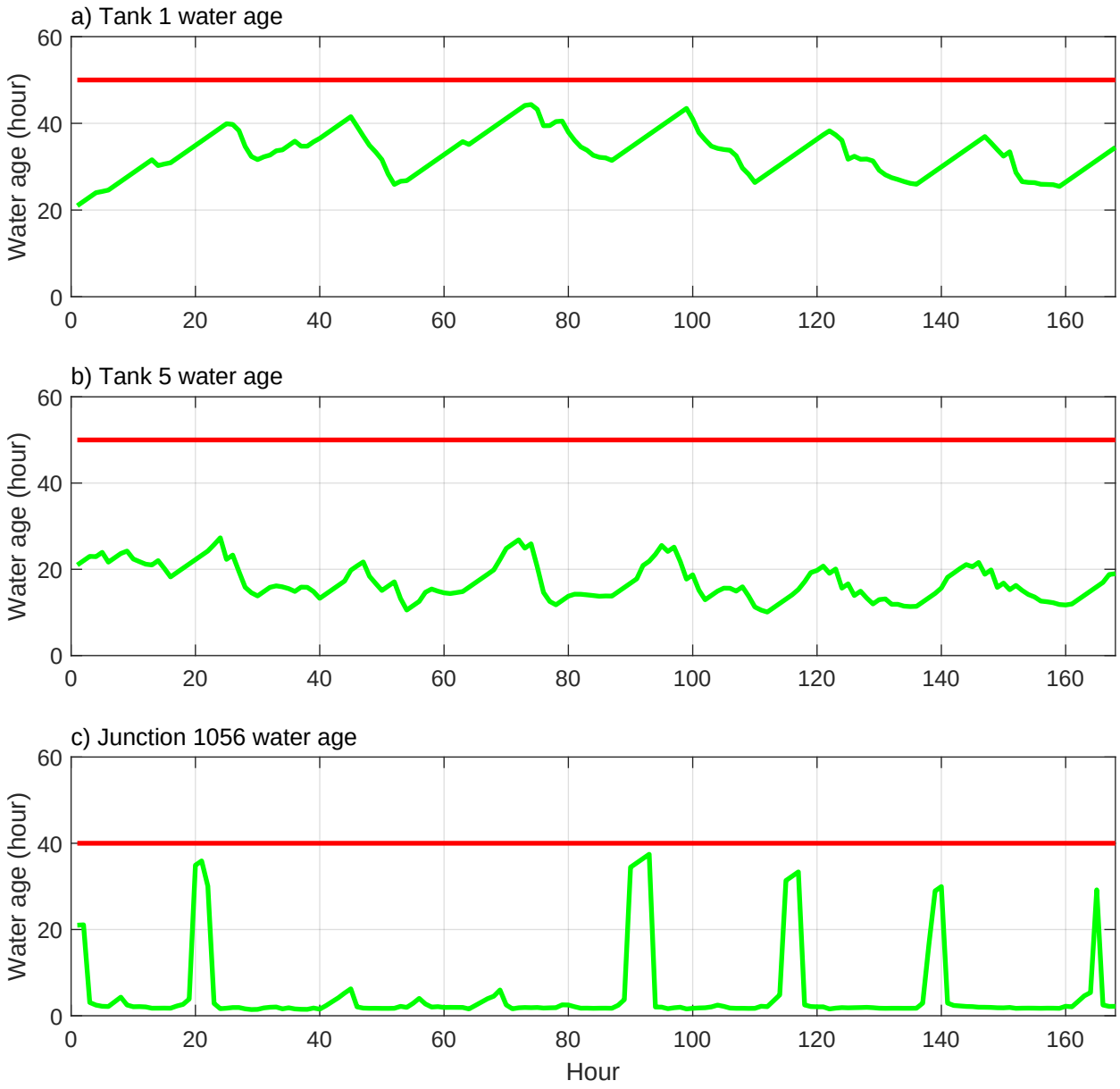

a) Tank 1 water age
Water age (hour)
0
20
40
60
0
20
40
60
80
100
120
140
160
b) Tank 5 water age
Water age (hour)
0
20
40
60
0
20
40
60
80
100
120
140
160
c) Junction 1056 water age
Water age (hour)
0
20
40
60
0
20
40
60
80
100
120
140
160
Hour

Figure 7.

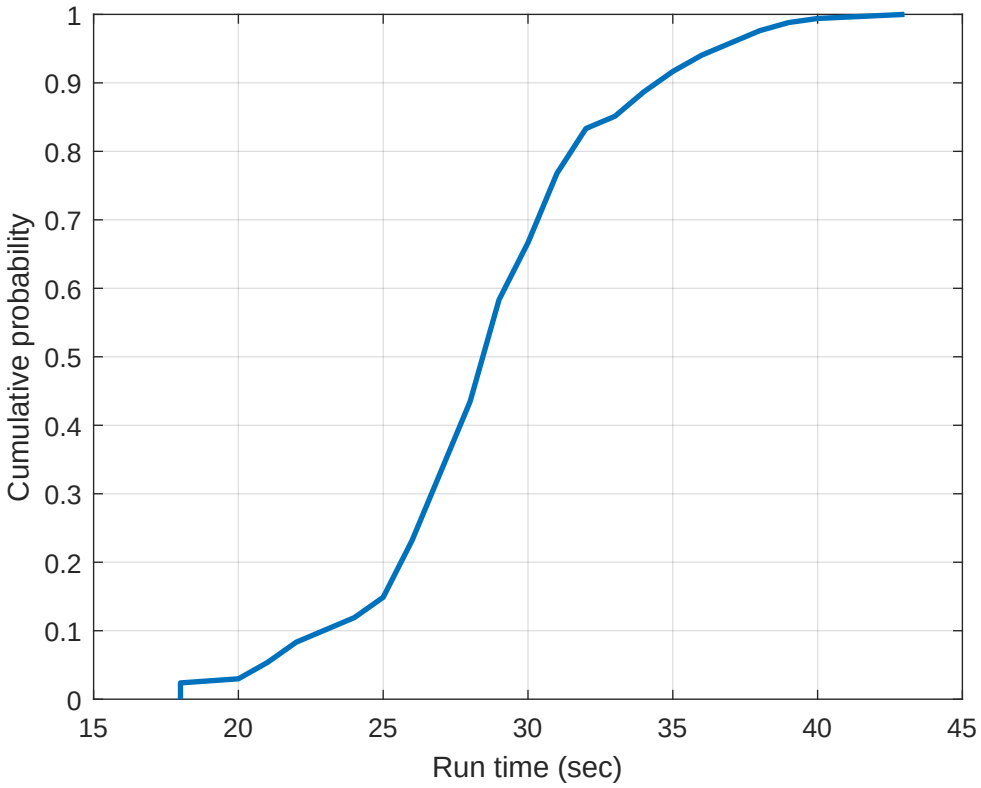

Cumulative probability
Run time (sec)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
15
20
25
30
35
40
45

Figure 8.

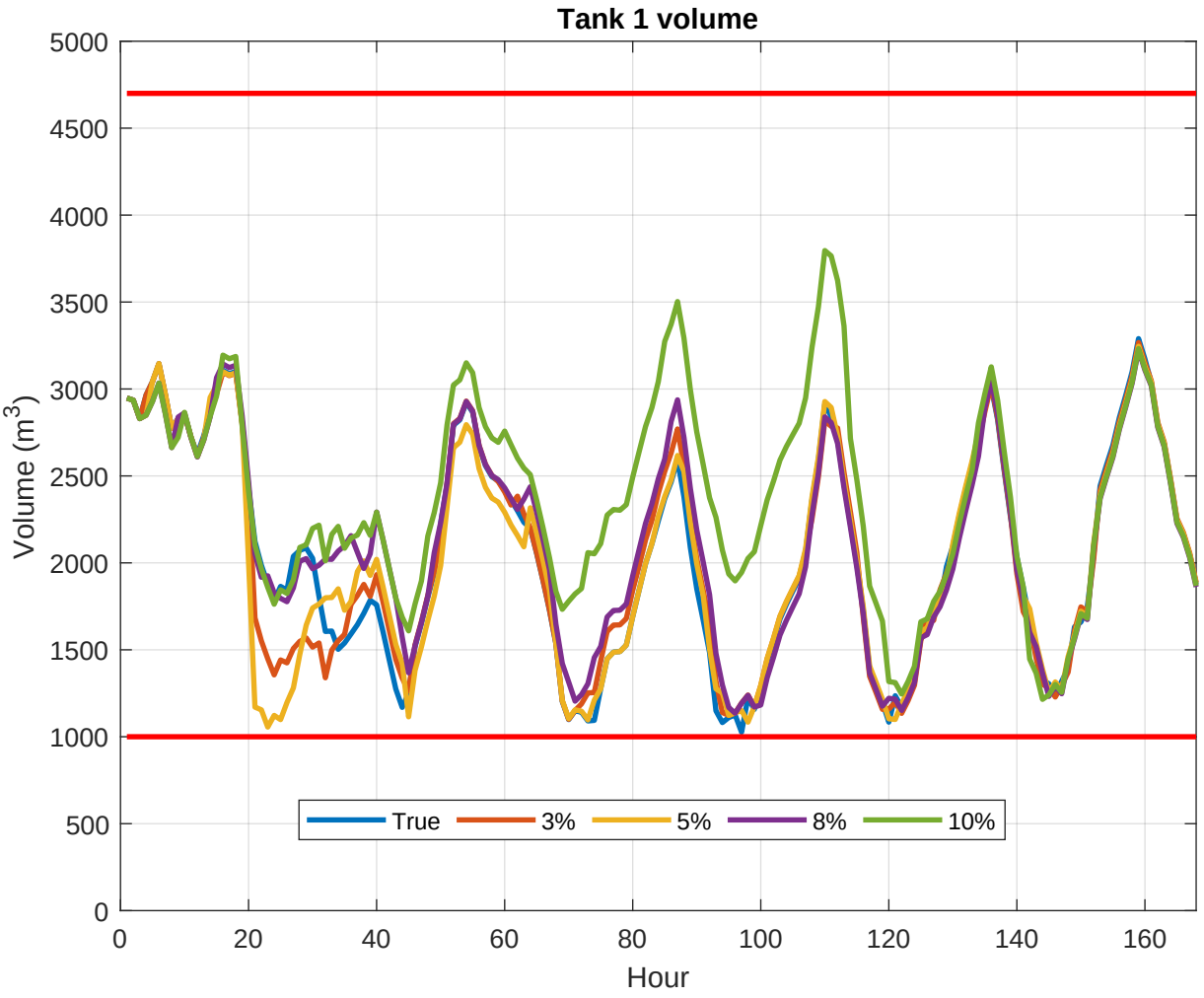

Tank 1 volume
Volume (m3)
Hour
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
0
20
40
60
80
100
120
140
160
True
3%
5%
8%
10%

Figure 9.

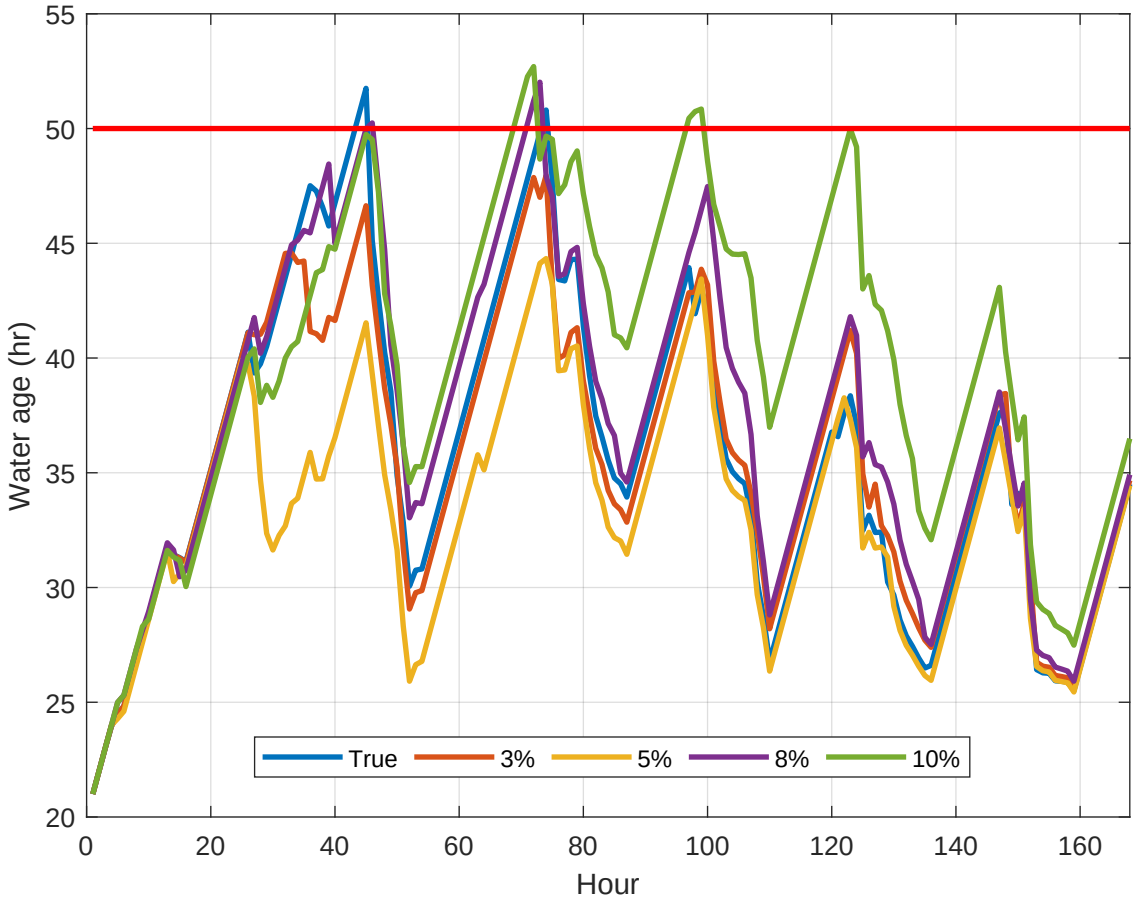

55
50
45
40
35
30
25
20
Water age (hr)
0
20
40
60
80
100
120
140
160
Hour
True
3%
5%
8%
10%

Figure 10.

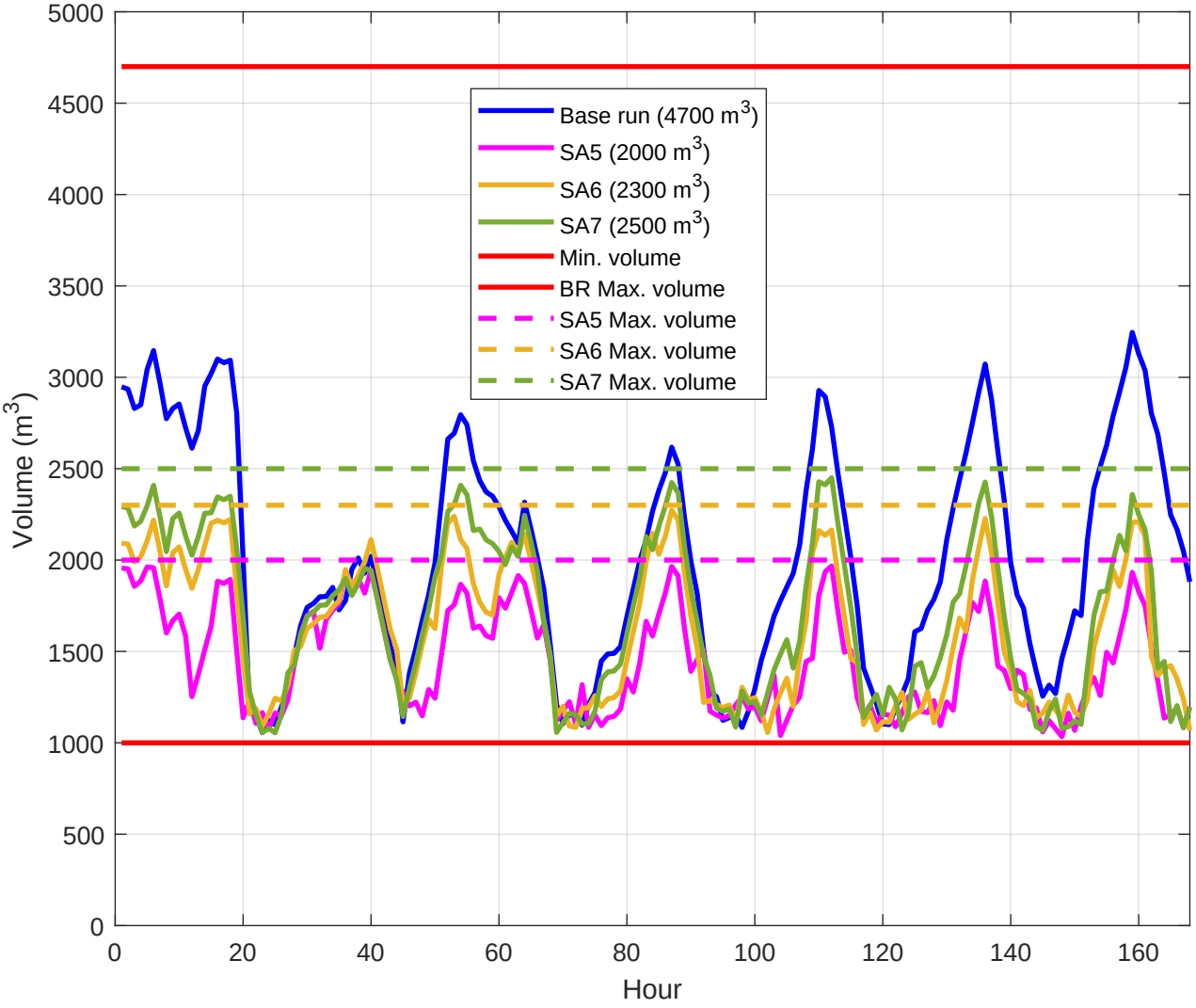

Base run (4700 m3)
SA5 (2000 m3)
SA6 (2300 m3)
SA7 (2500 m3)
Min. volume
BR Max. volume
SA5 Max. volume
SA6 Max. volume
SA7 Max. volume
Volume (m3)
Hour
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
0
20
40
60
80
100
120
140
160

Figure 11.

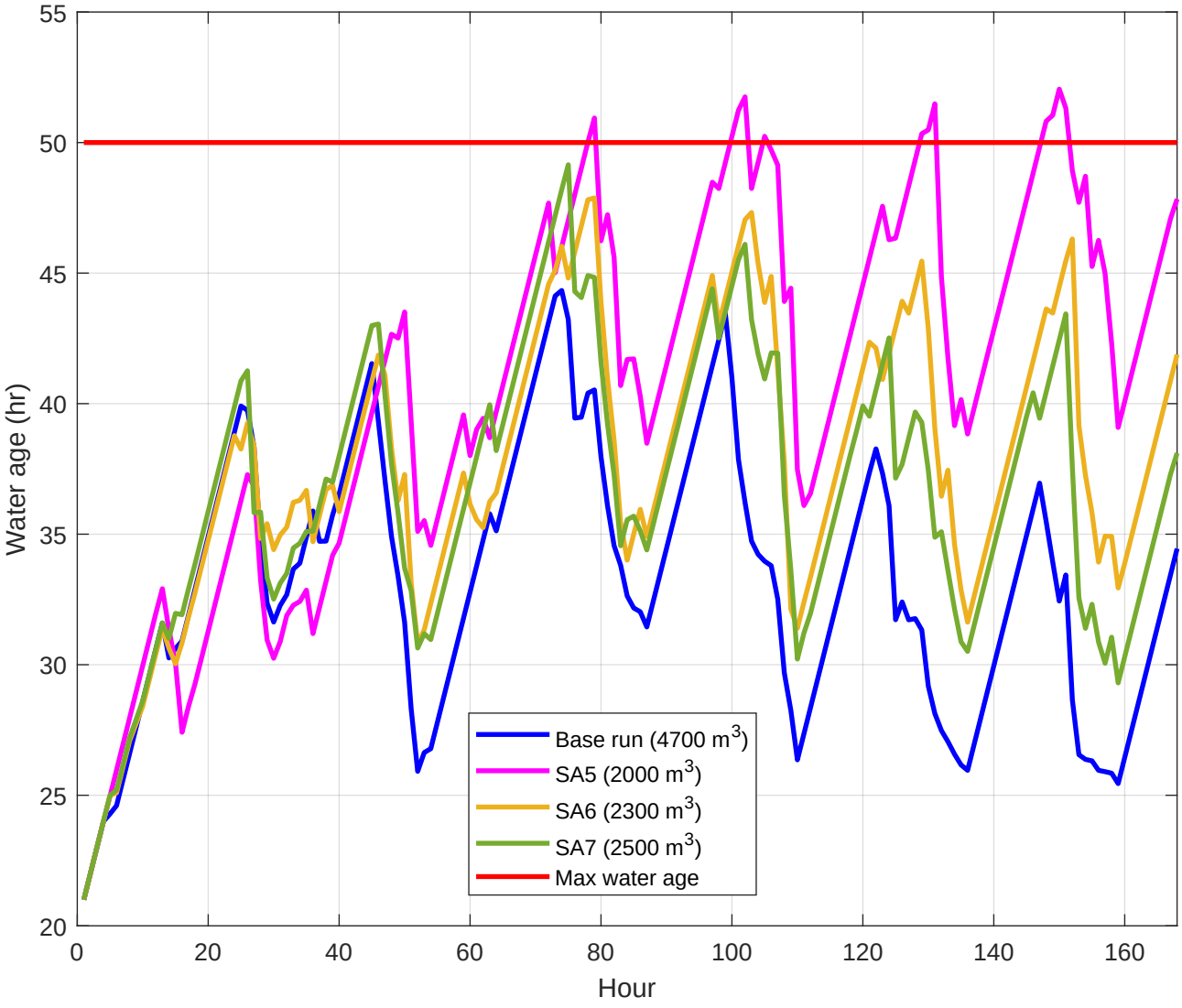

Water age (hr)
Hour
55
50
45
40
35
30
25
20
0
20
40
60
80
100
120
140
160
Base run (4700 m3)
SA5 (2000 m3)
SA6 (2300 m3)
SA7 (2500 m3)
Max water age

Figure 12.

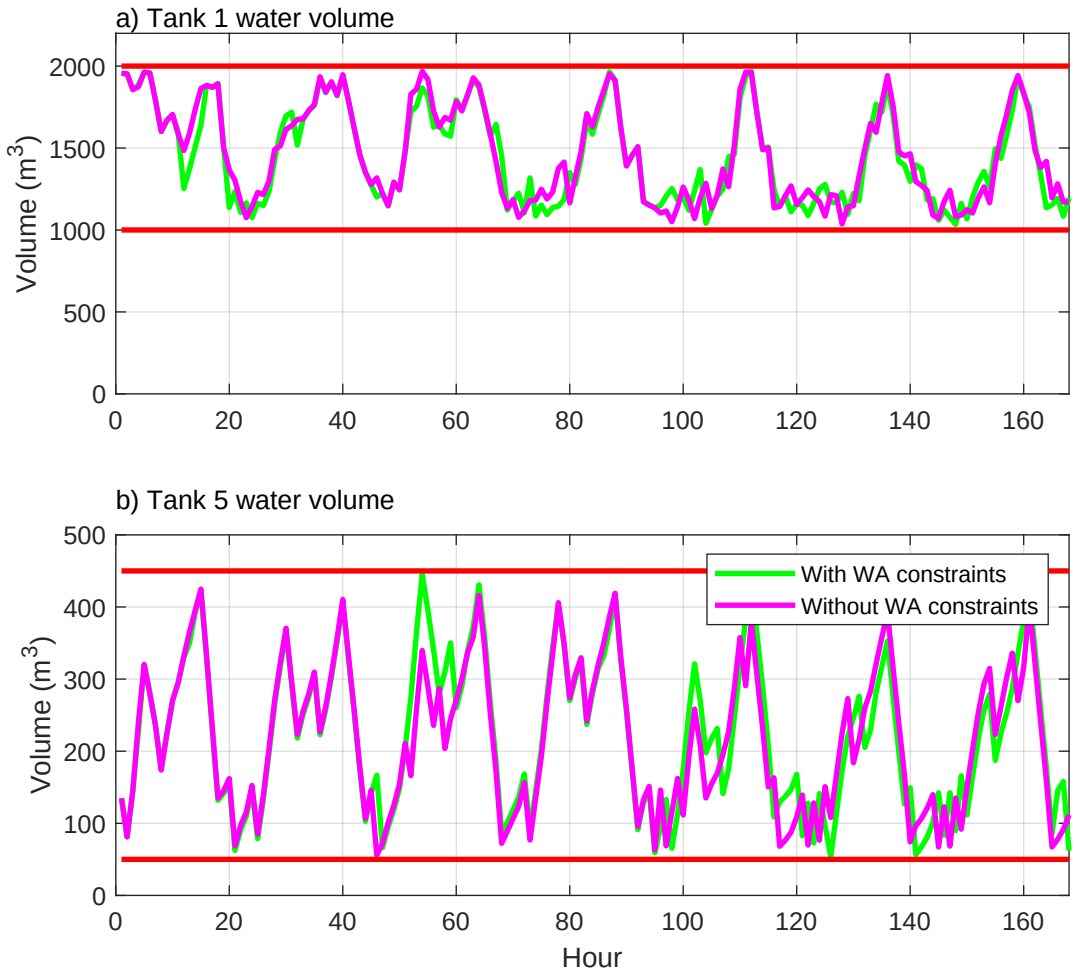

a) Tank 1 water volume
Volume (m3)
2000
1500
1000
500
0
0
20
40
60
80
100
120
140
160
b) Tank 5 water volume
Volume (m3)
500
400
300
200
100
0
With WA constraints
Without WA constraints
Hour

Figure 13.

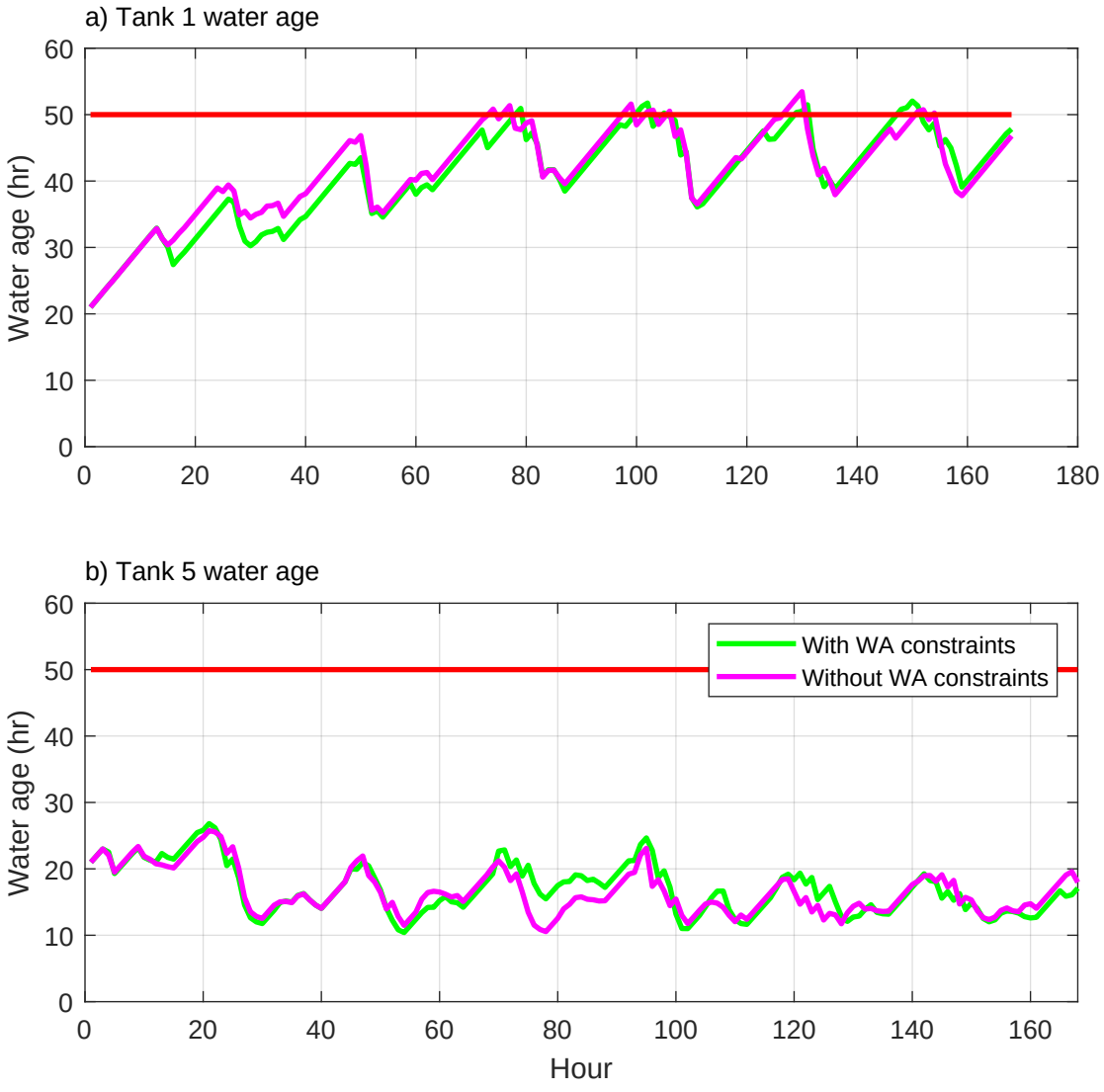
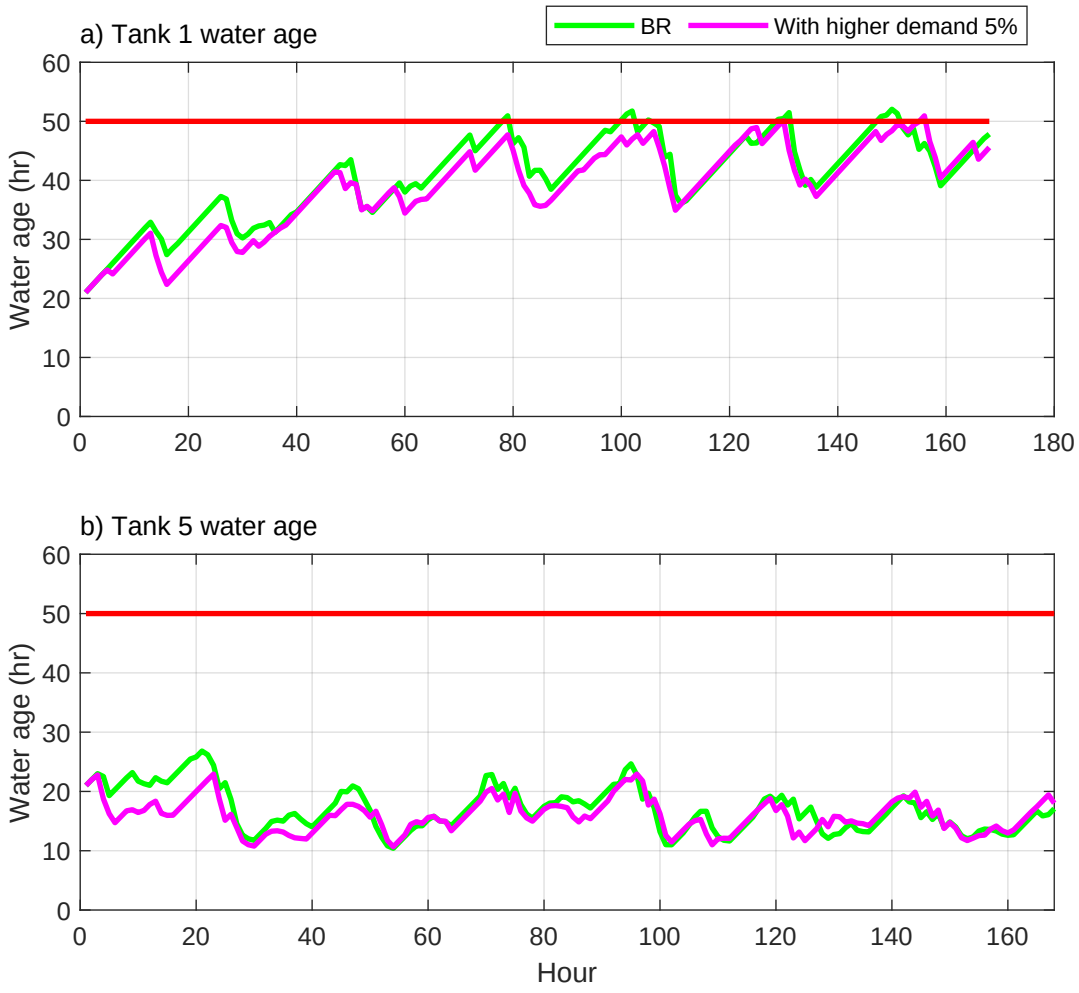

a) Tank 1 water age
Water age (hr)
0
10
20
30
40
50
60
0
20
40
60
80
100
120
140
160
180
b) Tank 5 water age
Water age (hr)
With WA constraints
Without WA constraints
0
10
20
30
40
50
60
0
20
40
60
80
100
120
140
160
Hour

Figure 14.


a) Tank 1 water age
BR
With higher demand 5%
Water age (hr)
0
10
20
30
40
50
60
0
20
40
60
80
100
120
140
160
180
b) Tank 5 water age
Water age (hr)
0
10
20
30
40
50
60
0
20
40
60
80
100
120
140
160
Hour

Observation variable	Tank 1	Tank 5	Junction 1056
S1 flow (LPS)	-0.01609	-0.0168	-0.0456
Tank 1 level (m)	0.34517	0.0703	0.7907
S4 flow (LPS)	0.04721	-0.0273	0.0762
Tank 1 previous water age (hr)	0.92365	-	-
DMA 1 demand (LPS)	0.03362	0.0160	0.0380
Tank 5 previous water age (hr)	-	0.9701	-
Tank 5 level (m)	-0.23300	0.4952	-
DMA 5 demand (LPS)	-0.05407	0.0052	-
Junction 1056 previous water age (hr)	-	-	0.8936

Pump station	Combination (#)	Flow (lps)	Specific energy (kW / m^3)
S1	1	116	0.1
	2	100	0.14
	3	94	0.12
	4	180	0.15
	5	182	0.13
	6	170	0.16
	7	227	0.17
S4	1	35	0.32
	2	33	0.3
	3	45	0.4

Run	Change from base run	Total cost (NIS)	Change from BR (%)
BR	None – base run (demand scenarios spread of 5%)	7,027	-
SA1	Using true demands	7,002	-0.36
SA2	Demand scenarios spread of 3%	6,996	-0.44
SA3	Demand scenarios spread of 8%	6,986	-0.58
SA4	Demand scenarios spread of 10%	7,097	0.99
SA5	Maximum volume for Tank 1 of 2,000 m^3	7,402	5.33
SA6	Maximum volume for Tank 1 of 2,300 m^3	7,145	1.67
SA7	Maximum volume for Tank 1 of 2,500 m^3	7,059	0.46
SA8	Demand scenarios spread of 5%, Tank 1 maximum volume of 2,000 m^3 and no WA constraint	7,387	5.12
SA9	Demand scenarios spread of 5%, Tank 1 maximum volume of 2,000 m^3 and increase "real" demand 5%	8,056	14.64