Article

Robust Optimal Operation of Water Distribution Systems

Gal Perelman * , **Avi Ostfeld** and **Barak Fishbain**

Faculty of Civil and Environmental Engineering, Technion—Israel Institute of Technology, Haifa 3200003, Israel
* Correspondence: gal-p@campus.technion.ac.il

Abstract: The operation of water distribution systems (WDS) is an energy-intensive process, which is subject to constraints such as consumer demands, water quality, and pressure domains. As such, tracing an operation policy in which constraints are met while energy costs are minimized, is a foremost objective for water utilities. Given the inherent uncertainties in WDS operation and the importance of supply continuity, it is essential to find an operational strategy that is robust against a wide range of circumstances. One promising approach for optimization under uncertainty is robust optimization (RO), which assures a robust (feasible) solution to realizations of the uncertain parameters, within predefined bounds. This study presents an RO-based method for optimizing pump scheduling under uncertainties of consumer demands and pumping costs. The method can capture various types of correlations between the uncertain parameters, thus better reflecting the uncertain nature of WDS operation. The developed methodology is demonstrated in two case studies with different levels of complexity. The impacts of uncertainty levels and correlation coefficients are analyzed to demonstrate their implications on operation policy. The results show the advantages of using RO with tradeoffs between costs and constraints satisfaction.

Keywords: water distribution systems; pumps scheduling; demand uncertainty; cost uncertainty; robust optimization; temporal–spatial correlation

1. Introduction

The optimal operation of water distribution systems is one of the most researched problems in WDS analysis. The mathematical complexity of the problem and the importance of optimal operation to WDS management result in a large body of work, published in recent decades [1]. While the efforts around this challenge have yielded a variety of optimization models and techniques, most of them account for deterministic approaches where all the problems' parameters are assumed to be known. However, WDS operation consists of several inherent uncertainties. Some of these uncertainties are the result of physical events in the system such as consumers' demands [2], transient hydraulics [3,4], hydrology-related events that affect water availability [5], pump failures or degradation [6], and electricity prices [7]. Other uncertainties are the outcome of missing information causing oversight results of optimization models such as pipes roughness coefficients, and pumps and valves curves. In many cases, the uncertain elements in the system are correlated between them, making the estimation of the joint probabilities highly complex and, therefore, difficult to solve. Thus, there is a great need for finding a methodology for making design and operational decisions under uncertainty [8]. Traditionally, uncertainty is modeled based on probability theory with an assumption that the probability density functions (PDFs) are known. In that case, the performance of systems can be estimated by using probability terms. More recent approaches no longer assume that PDFs exist or are known and treat uncertainty as a set of plausible scenarios [9]. This novel approach has been introduced by Walker, et al. [10] as deep uncertainty and it is commonly used in environmental systems and water resources analysis. In the case of deep uncertainty, robustness is replacing probability to describe the systems' ability to maintain its functionality over different scenarios. Accordingly, robustness is a desirable objective when optimizing WDS under uncertainty.

1.1. WDS Operation under Uncertainty

Few approaches were suggested to tackle the challenge of WDS operation under uncertainty. For example, a scenarios tree stochastic programming was suggested by [11]. This approach was based on generating a tree of discrete scenarios with known probabilities and then optimizing the objective function for the weighted scenarios. Constraints were added to maintain the mass balance in reservoirs between consecutive time steps and scenarios. Thus, the problem dimensions increased exponentially. Napolitano et al. [12] presented a similar scenario-based approach that included a risk term in the objective function to minimize the probability of violating the constraints while minimizing the pumping costs. To overcome the dimensionality challenge in scenarios-based approaches, Schwartz et al. [13] presented a method that limited the scenario's tree size. In their method, each scenario was first solved individually, then the decision variables in each time step were clustered to represent similar decisions. This allowed the pruning of similar branches in the tree that eventually lead to similar solutions. Finally, the reduced tree was solved with stochastic programming. Another method that was used to tackle WDS operation under uncertainty is chance-constrained optimization (CCO). In CCO, uncertain constraints are formulated as soft constraints that are satisfied for a predefined portion of the total scenarios [14]. CCO problems are still an unsolved challenge, for two main reasons: (1) difficulty to estimate probabilities of a point (solution) in the search space, and (2) non-convexity of the search space [15]. Two common methods to tackle CCO are analytic approximations to deterministic, and sampling methods. In the context of WDS optimal operation, CCO was suggested by [16]. In their work, they optimized the operation of a pumping station against a closed-end pipe (with no storage) under uncertain demands. While no storage is available, the pump station flow had to be larger or equal to the uncertain demands. These constraints were relaxed by using a normal distribution approximation and then the problem was solved analytically. Grosso et al. [17] suggested an analytical approximation for the joint probabilities of multiple constraints based on Boole's inequality. Sampling methods were also in use for CCO, for example, Kim et al. [18] used sampling stochastic dynamic programming to optimize the operation of a multi-reservoir system.

1.2. Robust Optimization

Robust optimization (RO), which solves optimization under uncertainties efficiently [19], is well suited for WDS optimization problems. RO is different from the above methods in the sense that it does not return the optimal solution over a probabilistic space (e.g., expected optimal solution) but rather a feasible solution for any realization of the uncertainty within specific bounds. In other words, RO aims at the worst-case scenario and optimizes the decision variables accordingly; hence, the solution is robust (feasible) for any plausible scenario. Recently, RO was used for several water resources problems and proved to have the potential to solve highly complex problems, such as optimal system design [20–22], optimal multi-year water allocation [23], reservoir operation [24], and flood control [25]. A study presented by [26] already pointed to the potential of RO for WDS operation by solving a small case study (Anytown) with demand uncertainty. RO is based on a continuous linear formulation of the WDS operation problem, it is worth mentioning that such approximation is not valid for some conditions as detailed in [27]. To demonstrate the formulation of the RO problem, consider general linear programming with uncertainties:

$$Z = \min_{x} C(\xi)^T x \tag{1}$$

$$Subject\ to:\ A(\xi)^T x - b(\xi) \leq 0 \tag{2}$$

where x is a vector of decision variables, and ξ is a vector of random variables representing perturbations from nominal values. For example, if $a_{i,j}$ is an uncertain parameter, representing the j coefficient in constraint i (entry i,j in the matrix A), then $a_{i,j}(\xi) = a_{i,j}^0 + \Delta_{i,j}\xi_{i,j}$.

Where $a_{i,j}^0$ is the nominal value (mean), $\Delta_{i,j}$ is the uncertainty level, and $\xi_{i,j}$ is the random perturbations. To achieve robust feasibility, which means a feasible solution for any realization of the random variables, the left-hand-side (LHS) of the inequality constraint (2) is maximized over ξ. The rationale behind this is that if the constraint holds for the maximized LHS, it will hold for any value of ξ. Or in other words, if the solution is feasible even against the worst-case scenario, it will meet the constraints in any other (better) scenario. The result is a min-max formulation which is completely deterministic after solving the inner optimization problems over ξ. The equivalent deterministic is obtained by solving the max problems in Equations (3) and (4), and it is referred to as robust counterpart (RC) [19].

$$Z = \min_{x} \left[\max_{\xi \in U} C(\xi)^T x \right] \tag{3}$$

$$\text{Subject to}: \max_{\xi \in U} \left[A(\xi)^T x - b(\xi) \right] \leq 0 \tag{4}$$

where U is a set that defines the feasible region of the random variables ξ. For example, in a box uncertainty such that $\xi_{i,j} \in [-1, 1]$, an uncertain parameter is defined as: $a_{i,j}(\xi) \in \left[a_{i,j}^0 - \Delta_{i,j}\xi_{i,j}, \ a_{i,j}^0 + \Delta_{i,j}\xi_{i,j} \right]$. In this case, the random perturbations of each constraint are independent of other constraints, where the meaning is that no correlation between the random variables is assumed. Although such a modeling strategy has the advantages of simplicity and maintaining the problem linearity, the assumption of no correlation between random variables does not hold in real-life problems and results in over-conservative solutions. To overcome the conservatism in box uncertainty sets, Ben-Tal and Nemirovski [28] suggested an Ellipsoid uncertainty set. The ellipsoid set can be seen as a subset of a box set that does not include all scenarios that may happen. In other words, most extreme scenarios where all the random variables obtain their worst values simultaneously are left out. This approach better reflects the real behavior of uncertainty, where the uncertain parameters are correlated; hence, very extreme scenarios are very rare. A general mathematical description of an Ellipsoid set is described as a stretched ball with a radius of Ω. For example, uncertainty set U for multiple correlated uncertain parameters will be:

$$a(\xi) \in \left\{ a^0 - \Delta\xi, \ a^0 + \Delta\xi, \ \|\xi\| \leq \Omega \right\} \tag{5}$$

where Ω is the Ellipsoid radius, a^0 is a vector of the nominal values, and $\|\xi\|$ is the L-2 norm of the random variables vector ξ. To capture the correlations between random variables, here the uncertainty level Δ is not a scalar but an affine mapping matrix that defines the correlation links between the different terms of ξ. The construction of the mapping matrix Δ is based on the covariance matrix of the different uncertain parameters [29]. More explanations of the technique used for the mapping matrix construction are detailed later. Previous RO studies in water resources research considered simple 1D correlations. For example, demand correlation between different consumers at the same time [21] or extreme rainfall events between two basins [25].

The current research aims to extend the concept of correlation modeling to include cases where different elements of the system are correlated in more than one dimension. This includes temporal correlations between different time steps and spatial correlations of system elements. One example of such correlated uncertainty is consumer demands, which are correlated in time and often to each other [2]. For example, given hot weather, domestic, agriculture, and industrial demands are expected to be higher than the average across different users and across different times of the day.

Another correlated uncertainty that is addressed here is the uncertainty of pumping costs. The cost of pumping water using a given pump is affected by many factors such as pump curve, efficiency curve, suction tank level, discharge tank level, hydraulic conditions, electricity prices, which may be dynamic, technical conditions of the pump, and more. With so many factors, and where some of them are not possible to estimate in real time, the actual

operational costs of pumps are uncertain. Since the hydraulic conditions in the network are temporal–spatial correlated [30], and electricity prices are also temporal correlated [31], we can conclude that the operational costs that depend on these factors are also temporal–spatial correlated. In this work, RO theory is used to develop a model for optimizing the operation of WDS under such temporal–spatial correlated uncertainties. The rest of the paper is organized as follows: below are the main contributions of this paper. Later the methodology is described from a deterministic formulation of WDS operation through uncertain WDS operation and then by the corresponding RC problem. The methodology is illustrated in two case studies. The case study results are analyzed to present the tradeoffs of WDS optimal operation under uncertainty and the impact of parameters such as level of uncertainty and correlation coefficients. Finally, the method's advantages are concluded.

The paper's main contributions are:

- Implementing the RO theory on optimal operation problems with a better approximation of the system's hydraulics.
- Illustrate the robust optimal operation methodology on a large real-life network.
- Address multiple uncertainties in the same problem including both objective and right-hand side (RHS) uncertainties simultaneously.
- Model multi-dimensional correlated uncertainties to capture the temporal–spatial correlations between the elements of the system.

2. Methodology

For formulating the problem, let $x_{t,p,c}$ be the portion of operation duration of pump station (p) at time step (t), when operated with the unit combination (c); let $v_{t,s}$ be state variables representing the volume in tank (s) at time step (t); and $Q_{p,c}$, $P_{p,c}$ are the corresponding flow and power consumption of pump station (p) when operated with the unit combination (c). s_i and s_o are the sets of pumps that, respectively, pump water in and out of tank s. Elec_t is the electricity tariff at time step (t). δt is the duration of time step t (1 h), P is the set of all pumping stations, S is the set of all storage tanks, and T is the number of time steps. Using these notations, the deterministic minimal energy costs operation problem can be described by the following linear programming (LP) presented in Equations (6)–(10). A complete description of the formulation can be found in [32].

$$Z = \min_{x} \sum_{t=1}^{T} \sum_{p=1}^{P} \sum_{c=1}^{C} x_{t,p,c} \cdot P_{p,c} \cdot \text{Elec}_t \tag{6}$$

Subject to:

$$\sum_{p \in s_i}^{p} \sum_{c=1}^{C} x_{t,p,c} Q_{p,c} - \sum_{p \in s_o}^{p} \sum_{c=1}^{C} x_{t,p,c} Q_{p,c} + (v_{t-1,s} - v_{t,s}) = d_{t,s} \delta t \quad \forall t \in T, \ \forall s \in S \tag{7}$$

$$v_{s,\,min} \leq v_{t,s} \leq v_{s,\,max} \quad \forall t \in T, \ \forall s \in S \tag{8}$$

$$v_{s,final} \leq v_{T,s} \quad \forall s \in S \tag{9}$$

$$0 \leq \sum_{c=1}^{C} x_{t,p,c} \leq 1 \quad \forall t \in T, \ \forall p \in P \tag{10}$$

It is possible to eliminate the tanks' volume state variables in Equation (7) and reformulate the problem such that the volume in tank (s) at time (t) is represented with the cumulative sum of inflows and outflows until time (t). The sum of flows until time t is represented by an auxiliary index r such that $\sum_{r=1}^{t} d_{r,s} \delta t$ is the cumulative volume of water consumed from tank (s) from the simulation start until time step (t). Similarly, $\sum_{r=1}^{t} \sum_{c=1}^{C} x_{r,p,c} Q_{p,c}$ is the cumulative volume pumped by pump station p until time t. This

reformulation of constraint (7) also allows the decomposition from equality constraints into two inequality constraints.

$$Z = \min_x \sum_{t=1}^{T} \sum_{p=1}^{P} \sum_{c=1}^{C} x_{t,p,c} \cdot P_{p,c} \cdot Elec_t \tag{11}$$

Subject to:

$$-\sum_{r=1}^{t} \sum_{p \in s_i}^{p} \sum_{c=1}^{C} x_{r,p,c} Q_{p,c} + \sum_{p \in s_o}^{p} \sum_{c=1}^{C} x_{t,p,c} Q_{p,c} \leq v_{0,s} - v_{s,\,min} - \sum_{r=1}^{t} d_{r,s} \delta t \quad \forall t \in T, \quad \forall s \in S \tag{12}$$

$$\sum_{r=1}^{t} \sum_{p \in s_i}^{p} \sum_{c=1}^{C} x_{r,p,c} Q_{p,c} - \sum_{p \in s_o}^{p} \sum_{c=1}^{C} x_{t,p,c} Q_{p,c} \leq v_{s,\,max} - v_{0,s} + \sum_{r=1}^{t} d_{r,s} \delta t \quad \forall t \in T, \quad \forall s \in S \tag{13}$$

$$v_{s,final} \leq v_{0,s} + \sum_{t=1}^{T} \sum_{p \in s_i}^{p} \sum_{c=1}^{C} x_{t,p,c} Q_{p,c} - \sum_{p \in s_o}^{p} \sum_{c=1}^{C} x_{t,p,c} Q_{p,c} - \sum_{t=1}^{T} d_{t,s} \delta t \quad \forall s \in S \tag{14}$$

$$0 \leq \sum_{c=1}^{C} x_{t,p,c} \leq 1 \quad \forall t \in T, \forall p \in P \tag{15}$$

The current study addresses multiple uncertainties in the problem properties. The first uncertainty is in consumer demand and the second is the uncertainty in pumping costs. These two types of uncertainties are found in two parts of the problem. Demand uncertainty is in the term d which is in the constraints RHS. The pumping cost uncertainty is in the objective function coefficients. Accordingly, the constants representing the demands and costs in problems (11)–(15) are replaced with affine functions as follows:

$$d_{t,s}(\xi) = d_{t,s}^0 + \Delta_{dem} \xi \tag{16}$$

$$C_{t,p,c}(\xi) = C_{t,p,c}^0 + \Delta_{cost} \xi , \quad where \quad C_{t,p,c}^0 = P_{p,c} \cdot Elec_t \tag{17}$$

where d_t^0 and $C_{t,p,c}^0$ are the nominal coefficients, Δ is the affine mapping matrix and ξ is the random variables representing the perturbation from the nominal values. Δ_{dem} is a mapping matrix for uncertain demands. It captures the correlations between time steps (temporal) and between consumers (spatial). Similarly, Δ_{cost} is a mapping matrix for the uncertain pumping costs. We note that each of the uncertain parameters can be drawn from a different probability distribution with different properties (mean and standard deviation). The uncertainty properties of each element in the system together with the temporal–spatial correlations are all captured within the mapping matrices to describe the influence of random perturbations on the actual values of demands and costs. The next section describes the construction of uncertainty sets based on these properties to accommodate all the correlations and properties of the uncertain elements.

2.1. Uncertainty Sets

As mentioned above, the RO method does not require any knowledge of the PDFs of the random variables ξ. Instead, the random variables are drawn from uncertainty sets U which can be described by minimal stochastic information as the mean and covariance of the nominal values. RO requires defining only the ranges (feasible area) for each random variable. Ben-Tal and Nemirovski [28] presented the tractability advantages of using Ellipsoid uncertainty sets and proved probabilistic bounds to meet the constraints for a family of probabilistic distributions. Another advantage of ellipsoidal uncertainty sets is that it is appropriate to model the correlation between different random variables. Therefore, ellipsoidal sets were chosen to model uncertainty in this paper. It was shown above, in Equations (16) and (17), that an uncertain parameter is described by a linear combination of nominal values, and a mapping matrix multiplied by the random variable.

Moreover, it is noted that the nominal values and mapping matrix can be predefined with the first two statistical moments that usually can be concluded from historical records even while the PDF itself is unknown.

The nominal value is often defined as the mean of the uncertain parameter, while the mapping matrix is constructed according to the parameter covariance. Lastly, the Ellipsoid radius Ω is a parameter that controls the level of risk, such that any value of ζ that fulfills Equation (5) inequality is inside the Ellipsoid and hence satisfies the constraint. Accordingly, large values of Ω will result in more robust solutions.

For example, constructing the demand uncertainty set of a single consumer, the covariance matrix is constructed according to the demands standard deviation (STD) vector σ. Where every entry in σ is the STD of demands at a single time step t, across different days.

$$\sigma = [\sigma_1, \sigma_2, \ldots, \sigma_{24}] \tag{18}$$

To account for temporal correlation, a matrix ρ is introduced such that $\rho_{i,j}$ is the correlation coefficient between consumption at time step i and time step j. Accordingly, the covariance matrix can be calculated by ρ and σ as follows:

$$\Sigma = \sigma\rho\sigma^T = \begin{bmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \cdots & \rho_{1,2}\sigma_1\sigma_T \\ \rho_{2,1}\sigma_1\sigma_2 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \rho_{n,1}\sigma_1\sigma_T & \cdots & \cdots & \sigma_T^2 \end{bmatrix} \tag{19}$$

The mapping matrix, noted Δ, is derived from the covariance matrix by using the Cholesky decomposition:

$$\Sigma = \Delta\Delta^T \tag{20}$$

If the spatial correlation is to be added, the above covariance matrix needs to be extended to include the STD of the other consumers and correlation coefficients between every pair of consumers and every pair of time steps. For that purpose, a matrix of size $(nT \times nT)$ is initiated, where n is the number of correlated elements (consumers), and T is the number of the simulation's time steps. The matrix is organized such that the first T rows and columns represent the first consumer, the next T rows and columns represent the second consumer, and so forth. The diagonal block's matrices are composed of the covariance matrices of each consumer as described in (19), each with the shape of $T \times T$. The rest of the entries are the correlation coefficients between consumers and between time steps. For example, in the following matrix (21), $r_{12,11}$ is the correlation coefficient between consumer 1 in time step 1 to consumer 2 in time step 1. Such a matrix (21) allows the representation of the full relationships between consumers and time steps.

$$\Sigma = \begin{pmatrix} \begin{bmatrix} & \Sigma_1 & \end{bmatrix} & \begin{bmatrix} r_{12,11} & \cdots & r_{12,1T} \\ \vdots & \ddots & \vdots \\ r_{12,T1} & & r_{12,TT} \end{bmatrix} & \cdots & \begin{bmatrix} r_{1n,11} & \cdots & r_{1n,1T} \\ \vdots & \ddots & \vdots \\ r_{1n,T1} & & r_{1n,TT} \end{bmatrix} \\ \begin{bmatrix} r_{21,11} & \cdots & r_{21,1T} \\ \vdots & \ddots & \vdots \\ r_{21,T1} & & r_{21,TT} \end{bmatrix} & \begin{bmatrix} & \Sigma_2 & \end{bmatrix} & \cdots & \begin{bmatrix} & & \end{bmatrix} \\ \begin{bmatrix} & \vdots & \end{bmatrix} & \begin{bmatrix} & \vdots & \end{bmatrix} & \ddots & \begin{bmatrix} & \vdots & \end{bmatrix} \\ \begin{bmatrix} r_{n1,11} & \cdots & r_{n1,1T} \\ \vdots & \ddots & \vdots \\ r_{n1,T1} & & r_{n1,TT} \end{bmatrix} & \begin{bmatrix} & & \end{bmatrix} & \cdots & \begin{bmatrix} & \Sigma_n & \end{bmatrix} \end{pmatrix} \tag{21}$$

The next step is implementing the constructed uncertainty sets in the uncertain problem (11)–(17) and converting it to an equivalent deterministic problem (RC formulation).

2.2. Robust Counterpart

Robust optimization is based on converting an uncertain problem into a deterministic equivalent problem at the worst-case point of the unknown parameters, this problem is named RC. To achieve the RC formulation, the problem is reformulated with "hard constraints", that cannot be violated, at their worst point of the unknown parameters. Then, the original problem is solved over the decision variables space. The meaning of such formulation is that our solution will satisfy the constraints for any realization of the uncertainty (any scenario). In this formulation, a solution that satisfies the worst scenario would satisfy any other (better) scenario; hence, it would be in the feasible space. To convey the problem into the worst case, inner optimization problems are introduced. For example, given a general constraint, with uncertainty in constraint coefficients, a new maximization problem is formulated:

$$\left(a^0 + \Delta\xi\right)^T x \le b \quad \text{Original constraint} \tag{22}$$

$$\max_{\|\xi\| \le \Omega} \left(a^0 + \Delta\xi\right)^T x \le b \quad \text{RC constraint} \tag{23}$$

If the LHS of (23) is maximized, the inequality will hold for any realization of ξ. The solution for problem (23) has been formulated by [19]:

$$\left(a^0\right)^T x + \Omega\|\Delta x\| \le b \quad \text{RC constraint at the worst realization of } \xi \tag{24}$$

An equivalent deterministic constraint is received, it is a second-order conic constraint; thus, the problem is no longer linear. However, the problem remains convex and therefore tractable [19]. To find the RC of the WDS operation under uncertainty, the uncertainty sets described in (21) are integrated into problems (11)–(17), and the random variables are maximized as shown in (22)–(24). The result is a deterministic optimization problem with linear constraints and a second-order objective function. It is noted that the uncertainty in the constraints is only in the RHS vector, which means only in parameters that are not multiplied by the decision variables. Hence, the constraints system remains linear. On the other hand, the uncertainty in the objective coefficients does multiply the decision variables and results in a deterministic objective function that contains the L-2 norm of our decision variables.

$$Z = \min_x \sum_{t=1}^{T} \sum_{p=1}^{P} \sum_{c=1}^{C} C_{t,p,c}^0 x_{t,p,c} + \Omega\|\Delta_{cost} x_{t,p,c}\| \tag{25}$$

Subject to:

$$-\sum_{r=1}^{t} \sum_{p \in s_i} \sum_{c=1}^{C} x_{r,p,c} Q_{p,c} + \sum_{p \in s_o} \sum_{c=1}^{C} x_{t,p,c} Q_{p,c} + \sum_{r=1}^{t} (d_{r,s}^0 + \Omega\|\Delta_{dem}\|)\delta t \le v_{0,s} - v_{s,\,min}$$
$$\forall t \in T, \ \forall s \in S \tag{26}$$

$$\sum_{r=1}^{t} \sum_{p \in s_i} \sum_{c=1}^{C} x_{r,p,c} Q_{p,c} - \sum_{p \in s_o} \sum_{c=1}^{C} x_{t,p,c} Q_{p,c} + \sum_{r=1}^{t} (-d_{r,s}^0 - \Omega\|\Delta_{dem}\|)\delta t \le v_{s,\,max} - v_{0,s}$$
$$\forall t \in T, \ \forall s \in S \tag{27}$$

$$v_{s,\,min} \le v_{t,s} \le v_{s,\,max} \quad \forall t \in T, \quad \forall s \in S \tag{28}$$

$$-\sum_{t=1}^{T} \sum_{p \in s_i} \sum_{c=1}^{C} x_{t,p,c} Q_{p,c} + \sum_{t=1}^{T} (d_{r,s}^0 + \Omega\|\Delta_{dem}\|)\delta t \le v_{0,s} - v_{final,s} \quad \forall s \in S \tag{29}$$

$$0 \leq \sum_{c=1}^{C} x_{t,p,c} \leq 1 \quad \forall t \in T, \ \forall p \in P \tag{30}$$

2.3. Case Studies

The method was tested against two case studies. The first case study is a small illustrative network, and the second case study is a real-life network.

2.3.1. Case Study 1: Illustrative Network

The first case study is a small illustrative network as presented in Figure 1. The network has two water sources, a pump station with two parallel pumps, and a well. The system has a single tank of 3000 m^3 volume, where min and max allowed volumes are 500 and 2800 m^3, respectively. A single aggregative consumer represents the demand in the network.
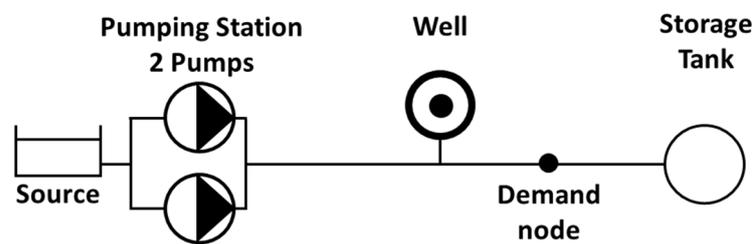
Figure 1. Illustrative network layout.

The network is optimized for a 24 h period to find the minimum operational electricity costs. Electricity tariffs contain two rates: on-peak and off-peak with respective costs of 1.25 and 1 € per kWatt-h. The on-peak tariff occurs during working hours (08:00–17:00) and the off-peak tariff during the rest of the day. The pump station and well data are detailed in Tables 1 and 2. By analyzing the specific energy of the different hydraulic states, we can conclude that operating the pump station (units 1 and 2) is preferred due to slightly lower energy consumption per pumped volume. However, when considering the uncertainty of the costs, the variance of the station's costs is higher than the well's costs variance, thus making the pump station more sensitive to dynamic conditions in the network (uncertainty).

Table 1. Pump station hydraulic operational states.

Flow (m^3/h)	Mean Power (kWatt)	STD Power (kWatt)	S. Energy (kWatt-h/m^3)	Unit 1	Unit 2
250	100	10	0.4	1	0
250	95	10	0.38	0	1
400	172	10	0.43	1	1

Table 2. Well hydraulic operational states.

Flow (m^3/h)	Mean Power (kWatt)	STD Power (kWatt)	S. Energy (kWatt-h/m^3)	Well
300	126	5	0.42	1

Two uncertainty sets are constructed according to the types of uncertainty in the problem. The first is the consumer demand uncertainty. The simulation is for 24 h, for each hour the mean and standard deviation are calculated according to the demand at the same hour over different days. The nominal demand values are the mean demands and the demands covariance is a 24 × 24 matrix according to Equations (18) and (19). The mean and

STD demands are detailed in data file S5. The temporal correlation coefficients are set such that each time step is correlated with the previous time steps according to Equation (31):

$$\rho_{t_1, \ t_2} = 1 - e^{(t_2 - t_1 - 1)C} \tag{31}$$

where C is a parameter that determines the decline rate. For the demand uncertainty, the rate was set to be 0.4. Thus, $\rho_{t,t+1} = 1$ and the correlation decline at an exponential rate. The result is that demand is correlated to a few preceding time steps. Next, the operational cost uncertainty set is constructed. The costs of the pumping station and well are correlated in time and between them. Here, the decline rate parameter is also assumed to be 0.4. The pumping elements are correlated between them (spatially correlated). Therefore, the operational costs covariance matrix is constructed according to Equation (21). The diagonal block matrices are calculated as the covariance matrix of every single element by its STD values and the declined temporal correlation as shown in Equations (18) and (19). Then, the matrices outside the diagonal represent the correlation between every time step and element to every time step of another element. The matrices for both uncertainty sets are presented in Figures S1 and S2 in the Supplementary Materials of this paper.

2.3.2. Case Study 2: Sopron Network

The second case study is based on a real regional network of the city of Sopron, Hungary [33]. The network consists of eight pressure zones (tanks), where five of which have demands. The network is fed by five wells, each well has a booster pump, where three of the wells' pumps are variable speed pumps (VSP). The network also consists of eight constant-speed pumping stations. Each of the pumping stations' power is supplied from a different power station. The problem includes additional constraints that were not part of the illustrative example. Each of the wells has a min and max required volume to pump within a 24 h operation. The changes in VSP flows are limited such that flows are constant within the electricity tariff period. Lastly, the available power from the power stations is limited. To address the additional constraints the mathematical Formulations (25)–(30) were extended as follows:

$$Q_i(t) = Q_i(t+1) \quad \forall t \in TP, \ \forall i \in VSP \tag{32}$$

$$v_{i, \ min} \leq \sum_{t=1}^{T} Q_i(t)\delta t \leq v_{i, \ max} \quad \forall i \in VSP \tag{33}$$

$$\sum_{k=1}^{K} P_k(t) \leq P_{ps, \ max} \quad \forall k \in ps, \ \forall ps \in PS, \ \forall t \tag{34}$$

where TP is tariff periods, meaning the flow of every VSP is constant during every period of the same tariff. $P_k(t)$ is the power consumed by pump station (k) in time (t), where (k) points to a pump station that is connected to the power station (ps). The full description of the second case study and its data are detailed in [33]. The network topology is presented in Figure 2.

As mentioned above, the Sopron network has five consumers, which in this paper are assumed to have uncertain demands such that all consumers are temporal–spatial correlated. The mean and STD demands are detailed in data file S6. As a first run, the temporal correlation decline is assumed to be C = 0.4 For the spatial correlation, a 0.8 correlation was assumed between every pair of consumers. For operational cost uncertainty, the example considered uncertainty in two pumping stations, PS4 and PS5. From a pre-analysis, it was found that these two stations face the largest variance in terms of suction and discharge heads; hence, their operational points are less expected. The covariance matrix of the operational costs was constructed according to STD of 10 and 5 for PS4 and PS5, respectively, with a temporal correlation of 0.4 (same as the illustrative

network). The matrices for both uncertainty sets are presented in Figures S3 and S4 in the Supplementary Materials of this paper.
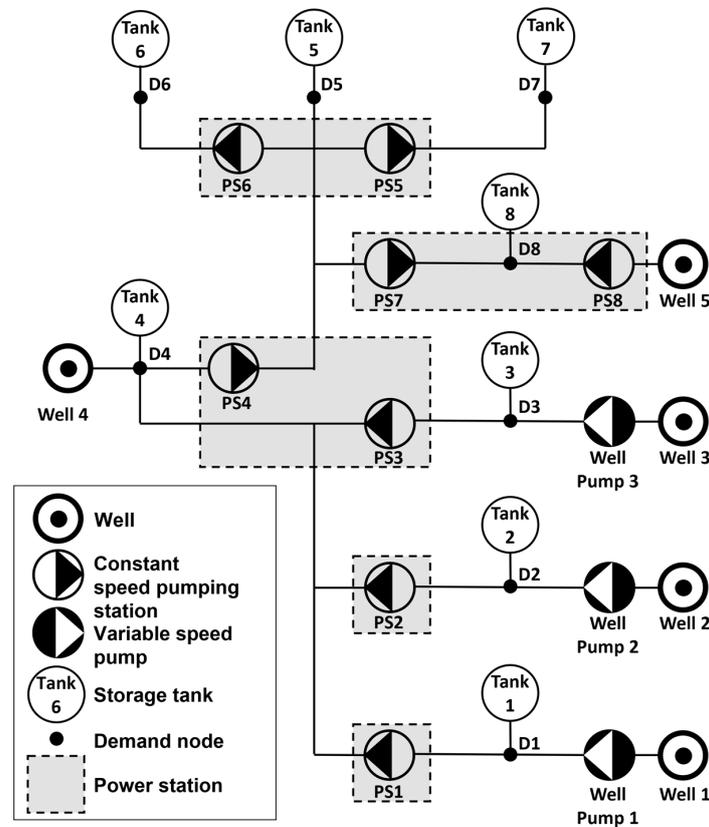

Figure 2. Sopron network layout.

3. Results

3.1. Case Study 1: Illustrative Network Results

The network was analyzed with different values of robustness (Ω). Where Ω represents the ellipsoid size. In other words, Ω is a quantitative measure for the extremeness of the scenarios that the solution is immunized against. A complementary stage was held after solving the optimization problem, which estimates the probability of violating the constraints. A Monte Carlo simulation of 1000 samples drawn from a multivariate normal distribution with the same parameters was used for constructing the uncertainty set. The estimated probability is the joint probability which means that valid samples are only those that did not violate any of the constraints. The results of the first case study are shown in Figure 3. The lower subplot shows the operational costs (objective values) as a function of the robustness, and the upper subplot shows the joint probability to violate one or more of the constraints.

As expected, the operational costs increase with the increase in robustness, and at the same time, the probability to violate the constraints decreases. The demand uncertainty affects the cost indirectly by increasing the total pumped volume or by shifting pumping to on-peak periods. However, the uncertainty of the operational costs affects the result directly as it is the costs themselves that are uncertain. In this example, a robustness value larger than 15 will guarantee the constraint satisfaction in 95% of the scenarios. Nevertheless, such robustness magnitude might be too high for the cost uncertainty. In such cases, different robustness values can be considered for different parts of the problem. For example, if the demand uncertainty is examined with values of $[0, 2 \ldots, 20]$ the robustness of the costs can be $[0, 1 \ldots, 10]$. The method found a solution ($\Omega = 16$) that satisfies all constraints with

a probability of 96%, and operational costs of 1983.7 €. The deterministic solution value is 1906.7 € which means the robust solution is 4% more expansive.
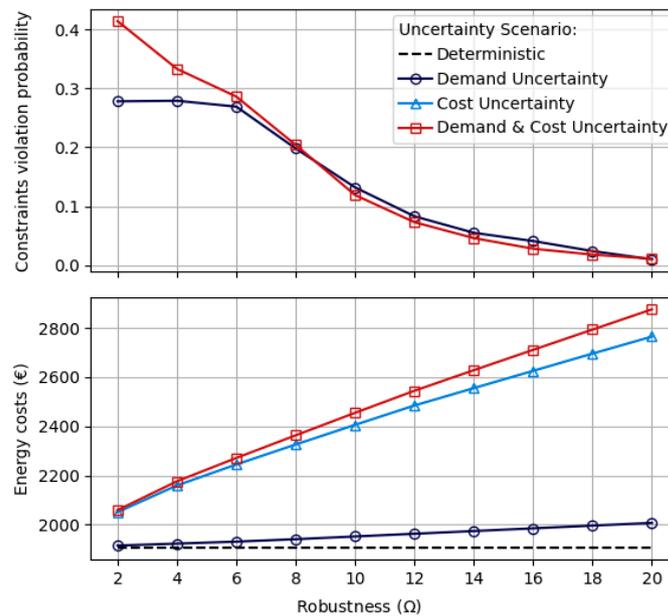

Figure 3. Illustrative network results.

Another analysis of the method examined the influence of the level of cost uncertainty. The level of uncertainty is expressed in the model through the cost standard deviations (STD). The example consists of two water sources where one of them (pump station) is slightly more expansive, yet it is also less certain. In this analysis, the prices STD values of the two sources were changed to examine the effect of the STD on the water sources combination. Regardless of the STD, the correlation between the two sources remains unchanged. The results are shown in Figure 4, it can be seen that when both sources have low STD, the pump station is preferred due to its lower cost. In this case, the station will supply approximately 80% of the total volume. As the station's prices' STD increases and the well's STD decreases, the well becomes the main source of the system with 95% supply. When the station's prices STD is low, and the well's prices STD is high, the station is pumping the entire volume.
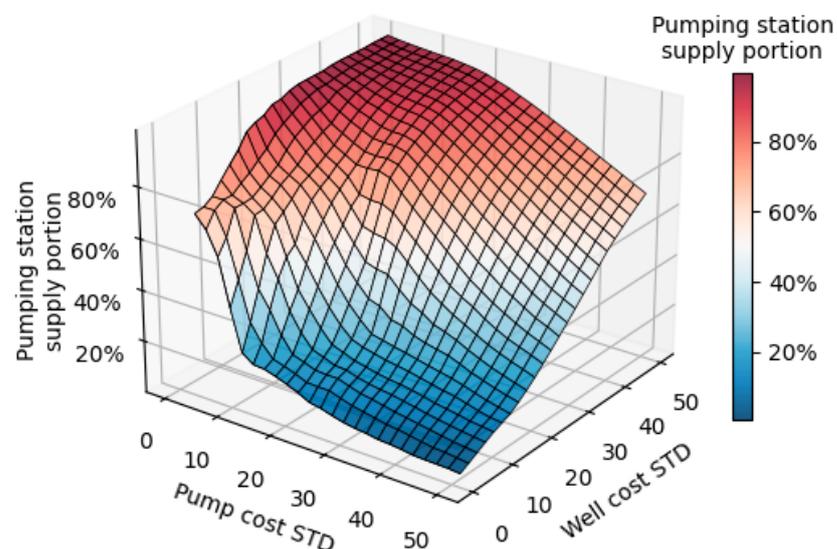
Figure 4. Uncertainty level effect on sources combination.

3.2. Case Study 2: Sopron Network Results

As outlined above, the method was used for a set of different robustness values to analyze the tradeoffs between operational costs and robustness. Figure 5 depicts the results of the Sopron network. Similar to the previous example, the cost uncertainty is more dominant but the difference between the two types of uncertainties is less significant. The method found a robust solution that satisfies the constraints with a probability of 90% and operational costs of 7078 €. The deterministic solution value is 6688 € which indicates that the robust solution is 5.8% more expansive. The constraints satisfaction rate is shown in the upper subplot of Figure 5. The probabilities were calculated based on 1000 Monte Carlo samples drawn from a normal multivariate distribution with nominal demands as presented by [33] as the mean values and the same covariance matrix used for the uncertainty set.
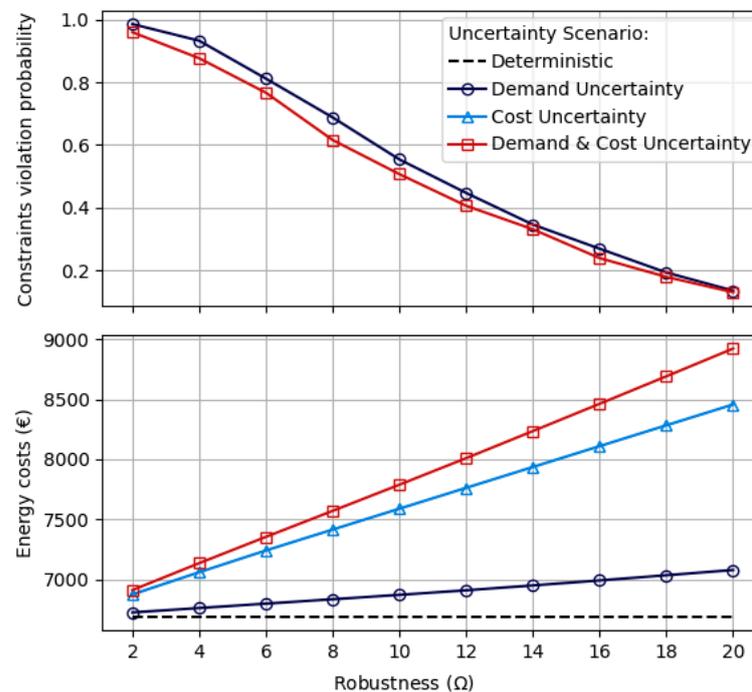
Figure 5. Sopron network results.

One contribution of the current paper is the modeling of spatial–temporal correlation as part of the robust optimization formulation. Here, the properties of the correlations are analyzed to examine the solution's consequence from various types of correlations. For the temporal correlation, the analysis considered three scenarios of no temporal correlation, moderate decline correlation with $C = 0.25$, and fast decline correlation with $C = 0.5$. These scenarios were integrated with four spatial correlation scenarios of zero correlation (ZC) which means all coefficients are zero, negative correlation (NC) where all coefficients are -0.8, positive correlation (PC) where all coefficients are 0.8, and finally, zones correlation (ZONES) where arbitrarily the consumers were divided into two zones by left and right sides of the network as it presented in Figure 2. On the left side of the network are consumers D4, D5, and D6. These consumers were assumed to be correlated between them with a coefficient of 0.9. On the right side of the network, consumers D7, and D8 were assumed to be correlated between them with coefficients of 0.8. All the other coefficients between the two zones are zero. The zone correlation illustrates a situation where the network supplies two different zones which behave differently from each other. The results obtained from the correlation analysis are shown in Figure 6, where it can be seen that the operational costs are not dramatically affected by either temporal or spatial correlations. The scenario of no temporal correlation is the least expensive, but the differences are quite

small. In the case of negative correlation, there is a gap of 1.6% between no temporal correlation (7757 €) and fast decline temporal correlation (7887 €). On the other hand, the temporal correlation impacts the constraints violations rate more dramatically. It can be seen that with no temporal correlation, the probability to violate the constraints is much higher than when a temporal correlation exists. The most prominent case is the combination of no correlation at all (no temporal correlation and zero spatial correlation) presented in the most left bar in Figure 6. In this case, the solution will violate at least one constraint with a probability of 99%.
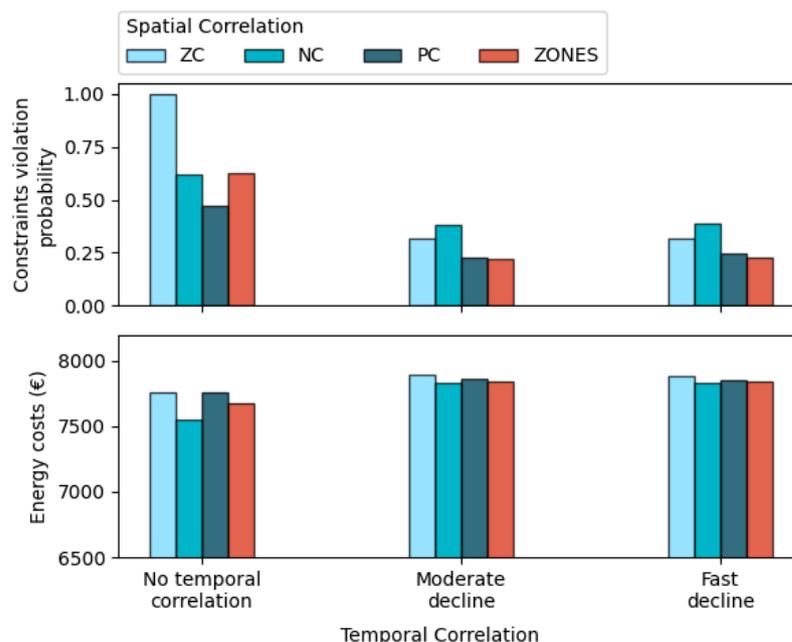
Figure 6. Sopron network correlation analysis.

4. Conclusions

The optimal operation of WDS is a mature research area, yet the subject of operation under uncertainty is still unevolved. Robust optimization (RO) is one approach for optimization under uncertainty, and it holds several advantages such as the guaranteed feasibility for any scenario within a predefined uncertainty set. It does not require probability density functions of the unknown parameters, and it results in tractable formulations that can be solved in short run times even for large problems. Based on RO theory, an optimization model was developed to solve WDS optimal operation problems with uncertainty in both consumer demands (RHS) and operational costs (objective coefficients). The uncertainty sets of the model were constructed to include plausible scenarios while extreme scenarios were left out. The justification for such an approach is the reduction in the solution conservatism. On the other hand, in very extreme scenarios, the system will not be able to satisfy the water demands. The contraction of the uncertainty sets was performed by addressing the correlations between different uncertain parameters. The model mathematical formulation implemented Ellipsoid uncertainty sets, which captured temporal and spatial correlations of the uncertain parameters.

Two case studies were examined in this paper and illustrated how the different types of uncertainties impact the operation policy, thus emphasizing the importance of considering uncertainties in WDS operation. The method supports systems operators by presenting the tradeoff between cost and reliability. For example, the suggested approach found solutions that increase the probability to satisfy all constraints (above 90%) with a penalty in energy costs of 4–6%. It was shown that the uncertainty level of pumping costs can reshape the combination of water sources. While in deterministic approaches the less expansive sources are exhausted first; here, importance is also given to the level of certainty of the source. This

insight might have possible implications for water availability, water quality, etc. Finally, the importance of correlation analysis was presented by examination of different temporal and spatial correlation scenarios. A deterministic approach cannot handle such correlations as it considers only known values. In theory, a classic stochastic approach can model uncertainty, yet in real problems, the dimensionality and non-convexity of the stochastic problems are intractable; here, the RO theory suggests a valuable strategy that can find a feasible solution, while considering correlated uncertainties such that the solutions are not too conservative. Future research opportunities are extending this work to dynamic folding horizon optimization with adjustable robust optimization and comparing the RO performance to other optimization under uncertainty techniques.

References

1. Mala-Jetmarova, H.; Sultanova, N.; Savic, D. Lost in Optimisation of Water Distribution Systems? A Literature Review of System Operation. *Environ. Model. Softw.* **2017**, *93*, 209–254. [CrossRef]
2. Avni, N.; Fishbain, B.; Shamir, U. Water Consumption Patterns as a Basis for Water Demand Modeling. *Water Resour. Res.* **2015**, *51*, 8165–8181. [CrossRef]
3. Meniconi, S.; Maietta, F.; Alvisi, S.; Capponi, C.; Marsili, V.; Franchini, M.; Brunone, B. A Quick Survey of the Most Vulnerable Areas of a Water Distribution Network Due to Transients Generated in a Service Line: A Lagrangian Model Based on Laboratory Tests. *Water* **2022**, *14*, 2741. [CrossRef]
4. Marsili, V.; Meniconi, S.; Alvisi, S.; Brunone, B.; Franchini, M. Stochastic Approach for the Analysis of Demand Induced Transients in Real Water Distribution Systems. *J. Water Resour. Plan. Manag.* **2022**, *148*, 04021093. [CrossRef]
5. Maiolo, M.; Mendicino, G.; Pantusa, D.; Senatore, A. Optimization of Drinking Water Distribution Systems in Relation to the Effects of Climate Change. *Water* **2017**, *9*, 803. [CrossRef]
6. Aghapoor Khameneh, P.; Miri Lavasani, S.M.; Nabizadeh Nodehi, R.; Arjmandi, R. Water Distribution Network Failure Analysis under Uncertainty. *Int. J. Environ. Sci. Technol.* **2020**, *17*, 421–432. [CrossRef]
7. Housh, M.; Salomons, E.; Sela, L.; Simpson, A.R. Water Distribution Systems on the Spot: Energy Market Opportunities for Water Utilities. *J. Water Resour. Plan. Manag.* **2022**, *148*, 02522002. [CrossRef]
8. McPhail, C.; Maier, H.R.; Westra, S.; van der Linden, L.; Kwakkel, J.H. Guidance Framework and Software for Understanding and Achieving System Robustness. *Environ. Model. Softw.* **2021**, *142*, 105059. [CrossRef]
9. Maier, H.R.; Guillaume, J.H.A.; van Delden, H.; Riddell, G.A.; Haasnoot, M.; Kwakkel, J.H. An Uncertain Future, Deep Uncertainty, Scenarios, Robustness and Adaptation: How Do They Fit Together? *Environ. Model. Softw.* **2016**, *81*, 154–164. [CrossRef]

10. Walker, W.E.; Lempert, R.J.; Kwakkel, J.H. Deep uncertainty. In *Encyclopedia of Operations Research and Management Science*, 3rd ed.; Grass, S.I., FU, M.C., Eds.; Springer: New York, NY, USA, 2013; Volume 1, pp. 395–402.

11. Pallottino, S.; Sechi, G.M.; Zuddas, P. A DSS for Water Resources Management under Uncertainty by Scenario Analysis. *Environ. Model. Softw.* **2005**, *20*, 1031–1042. [CrossRef]

12. Napolitano, J.; Sechi, G.M.; Zuddas, P. Scenario Optimisation of Pumping Schedules in a Complex Water Supply System Considering a Cost–Risk Balancing Approach. *Water Resour. Manag.* **2016**, *30*, 5231–5246. [CrossRef]

13. Schwartz, R.; Housh, M.; Ostfeld, A.; Asce, F. Limited Multistage Stochastic Programming for Water Distribution Systems Optimal Operation. *J. Water Resour. Plan. Manag.* **2016**, *142*, 06016003. [CrossRef]

14. Charnes, A.; Cooper, W.W. Chance-Constrained Programming. *Manag. Sci.* **1959**, *6*, 73–79. [CrossRef]

15. Ahmed, S.; Shapiro, A. Solving Chance-Constrained Stochastic Programs via Sampling and Integer Programming. *INFORMS Tutor. Oper. Res.* **2008**, 261–269. [CrossRef]

16. Khatavkar, P.; Mays, L.W. Model for Optimal Operation of Water Distribution Pumps with Uncertain Demand Patterns. *Water Resour. Manag.* **2017**, *31*, 3867–3880. [CrossRef]

17. Grosso, J.M.; Ocampo-Martínez, C.; Puig, V.; Joseph, B. Chance-Constrained Model Predictive Control for Drinking Water Networks. *J. Process Control* **2014**, *24*, 504–516. [CrossRef]

18. Kim, Y.-O.; Eum, H.-I.; Lee, E.-G.; Ko, I.H. Optimizing Operational Policies of a Korean Multireservoir System Using Sampling Stochastic Dynamic Programming with Ensemble Streamflow Prediction. *J. Water Resour. Plan. Manag.* **2007**, *133*, 4–14. [CrossRef]

19. Ben-Tal, A.; el Ghaoui, L.; Nemirovski, A. *Robust Optimization*; Princeton University Press: Princeton, NJ, USA, 2009; ISBN 1400831059.

20. Chung, G.; Lansey, K.; Bayraksan, G. Reliable Water Supply System Design under Uncertainty. *Environ. Model. Softw.* **2009**, *24*, 449–462. [CrossRef]

21. Perelman, L.; Housh, M.; Ostfeld, A. Robust Optimization for Water Distribution Systems Least Cost Design. *Water Resour. Res.* **2013**, *49*, 6795–6809. [CrossRef]

22. Boindala, S.P.; Ostfeld, A. Robust Multi-Objective Design Optimization of Water Distribution System under Uncertainty. *Water* **2022**, *14*, 2199. [CrossRef]

23. Housh, M.; Ostfeld, A.; Shamir, U. Optimal Multiyear Management of a Water Supply System under Uncertainty: Robust Counterpart Approach. *Water Resour. Res.* **2011**, *47*, 1–15. [CrossRef]

24. Pan, L.; Housh, M.; Liu, P.; Cai, X.; Chen, X. Robust Stochastic Optimization for Reservoir Operation. *Water Resour. Res.* **2015**, *51*, 409–429. [CrossRef]

25. Housh, M. Non-Probabilistic Robust Optimization Approach for Flood Control System Design. *Environ. Model. Softw.* **2017**, *95*, 48–60. [CrossRef]

26. Goryashko, A.P.; Nemirovski, A.S. Robust Energy Cost Optimization of Water Distribution System with Uncertain Demand. *Autom. Remote Control* **2014**, *75*, 1754–1769. [CrossRef]

27. Jowitt, P.W.; Germanopoulos, G. Optimal Pump Scheduling in Water-Supply Networks. *J. Water Resour. Plan. Manag.* **1992**, *118*, 406–422. [CrossRef]

28. Ben-Tal, A.; Nemirovski, A. Robust Solutions of Uncertain Linear Programs. *Oper. Res. Lett.* **1999**, *25*, 1–13. [CrossRef]

29. Yuan, Y.; Li, Z.; Huang, B. Robust Optimization under Correlated Uncertainty: Formulations and Computational Study. *Comput. Chem. Eng.* **2016**, *85*, 58–71. [CrossRef]

30. Gomes, S.C.; Vinga, S.; Henriques, R. Spatiotemporal Correlation Feature Spaces to Support Anomaly Detection in Water Distribution Networks. *Water* **2021**, *13*, 2551. [CrossRef]

31. Eck, B.; McKenna, S.; Akrhiev, A.; Kishimoto, A.; Palmes, P.; Taheri, N.; van den Heever, S. Pump Scheduling for Uncertain Electricity Prices. In Proceedings of the World Environmental and Water Resources Congress 2014: Water Without Borders, Portland, OR, USA, 1–5 June 2014; pp. 426–434. [CrossRef]

32. Perelman, G.; Fishbain, B. Critical Elements Analysis of Water Supply Systems to Improve Energy Efficiency in Failure Scenarios. *Water Resour. Manag.* **2022**, *36*, 3797–3811. [CrossRef]

33. Selek, I.; Bene, J.G.; Hs, C. Optimal (Short-Term) Pump Schedule Detection for Water Distribution Systems by Neutral Evolutionary Search. *Appl. Soft. Comput.* **2012**, *12*, 2336–2351. [CrossRef]