ADVANCING EARTH AND SPACE SCIENCES

Water Resources Research

RESEARCH ARTICLE
10.1029/2023WR034867

Multi-Objective Optimization Models for the Design of Water Distribution Networks by Exploring Scenario-Based Approaches

M. C. Cunha[1], R. Magini[2], and J. Marques[1]

[1]University of Coimbra, CEMMPRE, Department of Civil Engineering, Coimbra, Portugal, [2]Sapienza, University of Rome, DICEA, Rome, Italy

Key Points:

- Nodal demands are critical inputs and the main source of uncertainty for modeling and designing water distribution networks (WDNs)
- Currently available water demand data sets offer opportunities to improve deterministic design by using statistical and optimization methods
- A statistical methodology for generating demand scenarios is essential to solve WDNs robust design optimization problems

Correspondence to:
M. C. Cunha,
mccunha@dec.uc.pt

Citation:
Cunha, M. C., Magini, R., & Marques, J. (2023). Multi-objective optimization models for the design of water distribution networks by exploring scenario-based approaches. *Water Resources Research*, *59*, e2023WR034867. https://doi.org/10.1029/2023WR034867

Received 31 MAR 2023
Accepted 2 JUL 2023

Author Contributions:
Conceptualization: M. C. Cunha, R. Magini
Data curation: M. C. Cunha, R. Magini, J. Marques
Formal analysis: M. C. Cunha, R. Magini, J. Marques
Funding acquisition: M. C. Cunha
Investigation: M. C. Cunha, R. Magini
Methodology: M. C. Cunha, R. Magini
Writing – original draft: M. C. Cunha, R. Magini, J. Marques
Writing – review & editing: M. C. Cunha, R. Magini

Abstract Demand in a water distribution network (WDN) is an aleatory variable, owing to the unpredictable behaviors of water users. Therefore, it is one of the main reasons for uncertainty in the design of this infrastructure. The increasing number of water demand data sets offers opportunities to improve the traditional deterministic design approaches of WDNs by combining statistical and optimization methods. Robust optimization (RO) takes demand uncertainty into account by studying solutions that perform well under any possible demand scenario, that is, any possible realizations of this variable in the lifetime of a WDN. The right choice of scenario is therefore essential to ensure the reliability of the designed network. This paper presents a statistical methodology for generating scenarios to be used to solve a robust design optimization problem. It involves three steps: (a) descriptive analytics of historical data to derive the marginal distributions of peak hour demand in each node of the WDN, (b) generation of a very large number of snapshots by stratified sampling from the correlated marginal distributions of nodal peak demand, (c) generation of the peak demand scenarios by reducing the number of snapshots. Two heuristic techniques are proposed to reduce the number of snapshots, and for each of them, two different numbers of scenarios are derived. Two multi-objective RO models are solved: the first model includes cost minimization and a mean-variance Generalized Resilience and Failure index maximization objectives, and the second one additionally considers the minimization of the maximum undelivered demand, formalized using a regret function.

1. Introduction

Water distribution networks (WDNs) are essential infrastructure that should be planned to take into account the uncertain future in which they will operate. According to Creaco et al. (2021), nodal demands are critical inputs for modeling WDNs and their variability is one of the main sources of uncertainty that affects network sizing (Magini et al., 2019). For this reason, several researchers propose that WDN design should be developed taking several possible demand conditions into consideration. But when many demand scenarios are used, design definition becomes more complex. Powerful tools are needed to find optimal designs, particularly when more than one objective is pursued. Taking multiple demand scenarios into account can avoid the outcome being an under- or over-sized network, which can happen when only a single under- or over-estimate of demand is made.

The expansion of smart water meter technology at service connections in WDNs in recent years has made many high-frequency measurements of water consumption available. The literature is rich in heterogeneous consumption data sets at various spatial scales, ranging from city district to a single household or individual water fixture, and several temporal sampling frequencies, from monthly up to sub-daily: hour, minute, or second (Di Mauro et al., 2021). The availability of so much data supports planners and water utility managers in choosing the best design solution and operational strategy but, also shows up the high degree of variability that marks water demand, mainly due to the unpredictable behavior of human beings. Statistical and data mining tools support descriptive and predictive analytics, which can capture demand variability at different scales in space and time, derive statistical moments, define appropriate probability density functions, or model suitable stochastic processes.

There is a large body of literature dealing with the "statistical uncertainty" (Walker et al., 2003) of water demand. In some cases, the main purpose of the studies is to analyse demand, for example, for time pattern recognition, end-use disaggregation, simulation, and forecasting. In other cases, they consider water demand uncertainty when addressing WDN design and management issues, such as for optimal network sizing (Salcedo-Díaz et al., 2020), WDN management (Creaco et al., 2017), leak detection (Jahanpour & Tolson, 2018), district metering areas' boundaries (Di Nardo et al., 2018), and water quality management (Giudicianni et al., 2022).

Regarding water demand as an uncertain quantity, a particular point concerns peak demand. Peak demand is the forcing parameter in the sizing of WDNs, where it is traditionally estimated by defining the peaking factor, that is, the ratio of the maximum flow during some specified time interval to the average annual flow. It is usually assessed by deterministic expressions based on hourly datasets and engineering judgment that has no theoretical basis. However, considering the hourly time frame could lead to an underestimation of demand and compromise the reliability of the network. The availability of data sets measured with small sampling intervals, from 1 min to 1 hr, and over long periods of 1–2 years or more, make it possible to statistically describe the uncertainty of peak demand at the different time scales, thus developing estimates of this factor based on the extreme value theory. Furthermore, the availability of peak demand probability distributions, with parameters depending on the number of users and on the sampling interval, enables the generation of instantaneous states of the demand in the different nodes, hereinafter denoted as snapshots, that is, simultaneous values of the demand for water in all nodes. For this purpose, the cross-correlation between pairs of nodal demands is crucial.

The increasing number of descriptive and predictive analytics on demand datasets that we can find in the literature offers opportunities to improve the traditional deterministic design approaches of WDNs by combining statistical and optimization methods.

Including uncertainty demand issues in decision-making models gave rise to various applications. Classically, uncertainty was embraced through surrogate approaches (such as resilience or robustness measures), or stochastic approaches using chance-constrained formulations, worst-case analysis, or an aggregated objective for optimizing expected values, as in stochastic linear programming that builds on Dantzig's recourse concept (Dantzig, 1955).

Greenberg and Morrison (2007) mentioned the limitations of some classical models, commending the robust optimization (RO) approach to provide designs that are fundamentally robust against errors that may be the cause of harmful failures in important infrastructure systems. In fact, there is a clear need for methods that can find solutions for optimizing an objective function while robustly taking into account risks related to modelling uncertainties (Hart et al., 2007).

The important contribution of this paper in this field is the application of multi-objective robust optimization models for handling the design of WDNs taking statistical uncertainty in nodal demands into account. Compared to previous work available in the literature, the proposed approach requires an additional step after the sampling procedure to develop water demand snapshots for creating scenarios and assigning their probabilities. However, this allows to provide clear and organized insights that facilitate an open discussion with decision makers. This paper proposes a novel conceptual framework to embrace these issues. Scenarios, that is, demand snapshots having a fixed weight/probability, are defined using statistical analysis of historical data scaled according to the number of users. The methodology described by Magini et al. (2019) is the basis for the generation of a great number of demand snapshots. Furthermore, in this paper two heuristic techniques are employed to reduce the snapshots and assign them a weight/probability, retaining only a small number which are still statistically representative of the entire set. The reduced snapshots constitute the scenarios to be employed in solving the RO design problem. They embody the demand spectrum of variation while still maintaining a level of tractability to prevent computational burden. As far as the authors are aware, this is the first study about WDN optimization that directly incorporates statistical uncertainty in RO models based on the type of scenario generation developed. The framework is very versatile to being explored by the users and to being adapted to the real-wold problem at stake, to the availability of data and to the stakeholders involved. Deterministic solutions, solutions obtained through two-objective and three-objective robust models are analysed and compared to give insights to decision makers on the best way of dealing with statistical uncertainty on demand for the design of WDNs.

The remainder of the paper, after this introduction, is organized as follows. The next section consists of a literature review. In Section 3 the problem is stated and corresponding multi-objective optimization models are set out, the case study is presented, and the scenario generation is developed. Section 4 includes the results and the main discussions. The paper closes with the conclusions.

2. Literature Review

2.1. Peak Demand and Scenarios Generation

With regard to peak demand uncertainty, Zhang and Buchberger (2005), considering residential water use and fitting results from a Poisson rectangular pulse stochastic model with the Gumbel distribution, developed a theoretical reliability-based estimate of the peaking factor. More recently, Balacco et al. (2017) applied local parameters to the equation of Zhang and Buchberger (2005) and suggested a relationship between the peak factor and the number of users. Probabilistic estimation of the peak coefficients for residential users was also proposed by Pallavicini and Magini (2007), using statistical inference and relating the main statistics of consumption data to the number of users. These authors showed that the Gumbel and log-normal were the most suitable distributions to fit the available measures. Practical expressions of the peak coefficients were derived as a function of the return time, the temporal resolution of consumption data, and the number of aggregated users. Tricarico et al. (2007), Gato-Trinidad and Gan (2012), and Gargano et al. (2017), analyzed peaking factors in real networks. They pointed out that the traditional time interval of 1 hr could result in an underestimation of peak demand, conversely, a time scale that is too fine, less than 5 min, could be excessive. Moreover, Gargano et al. (2017) used large data sets of water demand relative to different numbers of residential users and obtained with various sampling intervals, and they confirmed the effectiveness of the log-normal, Gumbel, and log-logistic distributions in describing peak water demand.

As regards the generation of water demand snapshots, the cross-correlation between the nodal demands plays a fundamental role. Cross-correlation is usually expressed by the Pearson coefficient and depends on the number of users and the sampling interval (Vertommen et al., 2015). The role of nodal demand correlation in WDN was first highlighted by Filion et al. (2007), who studied its influence on the hydraulic performance of these systems. He used the multivariate flow model proposed by Fiering (1964) to generate snapshots of stationary demand with nodal values differently correlated to each other. Considering two WDNs from the literature, he found that the standard deviation of pressure head and capital costs were sensitive to the level of cross-correlation between nodal demands.

More recently, Magini et al. (2019) presented a bottom-up approach in which snapshots of demand were generated using Latin hypercube sampling (LHS) and considering gamma marginal probability distributions. All statistics, including the cross-correlation coefficient, were treated as depending on the number of users, in agreement with the scaling laws approach (Magini et al., 2008; Vertommen et al., 2015). To respect cross-correlation, a procedure that combines the NORTA (Cario & Nelson, 1997) and the Iman-Conover method (Iman & Conover, 1982) was performed. A heuristic technique was also proposed to reduce the number of snapshots (further details are presented in Section 3).

Creaco et al. (2021) faced a similar issue for reconstructing snapshots of peak demand with a two-step methodology which imposes the respect of the rank cross-correlations. A beta probability distribution model with tuneable bounds was used for the demand generation at different scales of aggregation to reproduce the statistics not only in terms of mean and standard deviation but also skewness, and to provide a better representation of extreme values. However, the beta distribution with tuneable limits, besides two shape parameters, contains two more, which respectively define the upper and lower limit of the distribution. This increases the model parameterization burden and introduces the need for a subjective choice.

2.2. Optimization Models

The literature for dealing with demand uncertainty management contains many different attempts at optimizing WDN design.

A chance constrained formulation can be found in Lansey et al. (1989) for a least-cost design problem, converting the constraints into deterministic equivalents based on the parameters of the demand and roughness coefficients distribution.

A non-probabilistic robust least-cost design of WDNs was studied by Perelman et al. (2013a) in which no assumption was made about the probability distribution of water demand, but demand uncertainty was quantified by a deterministic user-defined ellipsoidal uncertainty set; Perelman et al. (2013b) explicitly take into account different correlations between water demand at nodes added to the optimization model of the previous paper. In

10.1029/2023WR034867

both cases the problems were reformulated as deterministic equivalents of the stochastic problem using the robust counterpart concept (Bertsimas et al., 2019).

Eck et al. (2015), facing the WDN design as a robust optimization problem, modeled the uncertain nodal demands as a multivariate Gaussian distribution with mean vector and covariance matrix defined from measured data sets. Demand snapshots corresponding to a chosen cumulative probability were generated using the Cholesky decomposition of the covariance matrix. Because at a chosen cumulative probability there is an infinite number of snapshots, cluster analysis methods were applied, or extreme total demands picked up for selection. The selected snapshots, relating to a cumulative probability of 90%, were then used in an optimization model.

The previous optimization models are directly based on values picked from samples obtained through Monte Carlo or Latin hypercube methods or reformulated as deterministic equivalents of the stochastic problem.

The seminal paper by Mulvey et al. (1995) was a historical turning point in the optimization of complex infrastructure systems in that it introduced some new concepts for developing RO models. Robust optimization is very helpful when modeling uncertainty in optimization problems, exploring solutions that perform well under any possible realization of uncertain input variables: the scenarios. Scenarios are particular representations of how reality might behave, and how "well" robust solutions perform is a challenge when it comes to the formulation of robust optimization models. Stochastic linear programming can be considered as a special case of RO. Specifically, stochastic programming requires the replacement of the multivariate probability distribution of the uncertain variables with a discrete probability measure just through scenarios, that is, a finite number of realizations to be used for optimizing expected value, not considering the decision maker's preferences regarding risk (Mulvey et al., 1995).

Additional features of RO formulation allow the introduction of higher moments of the statistical distributions (Mulvey et al., 1995). Robust optimization is a flexible framework capable of managing noisy and/or uncertain data (Mulvey et al., 1995), using formulations particularly appropriate to handling asymmetric distributions and more prone to risk averse decision maker's needs. This is an attractive approach, even for the design of large complex infrastructure systems such as water networks that include discrete decisions and uncertainty issues in water demand. In fact, the corresponding optimization RO scenario-based models can be solved effectively and efficiently if a tractable uncertainty set is selected (Ben-Tal et al., 2000). Two main concepts were introduced by Mulvey et al. (1995) for the RO formulations. The formulation contains a term expressing the idea of solution robustness, to model the uncertainty of the objective function, driving the solution to be "close" to the optimum for each scenario. Another term represents model robustness used to penalize violation of constraints, to model the uncertainty of the feasibility of the solution, driving to solutions "almost" feasible for each of the scenarios.

Therefore, to develop a robust optimization model, a critical step is to derive a suitable "set of uncertainty," that is, a set of scenarios and their probabilities.

However, in robust optimization, statistical methods based on historical data analytics are not the only way to generate scenarios (an advantage over stochastic optimization where probability distributions are required (Gabrel et al., 2013)). Other approaches are possible such as consulting a panel of experts or forecasting methods or exploring future alternatives more generally. But not all methods make it possible to assign each scenario its weight/probability of occurrence objectively. We have to note that there are examples of such strategies in the literature on robust optimization problems for WDN sizing. For single objective models, Cunha and Sousa (2010) faced the problem of a gravity flow WDN design considering seven different abnormal working conditions related to firefighting flows and pipe breaks but without uncertainty in nodal demand. The weights/probabilities of each scenario can be obtained from a "panel of experts" using methods like the so-called Delphi methods (Green et al., 2007). The same case study and scenarios of the previous work were employed by Marques et al. (2012), who extended the RO model with the insertion of a pumping station to cope with the abnormal conditions and avoid oversizing the pipes.

There are also some examples in the literature where multi-objective approaches to the design of WDNs are proposed to embrace the idea of robustness. The advantages of using multi-objective formulations date back to the paper by Todini (2000) where a surrogate index approach can be found: an optimization model embracing two objectives, one for cost minimization and another based on the concept of resilience. In fact, resilience was then proposed to overcome reliability issues. Resilience would represent the capacity of the WDN to overcome failures without the need for uncertainty description. As such, a designer could understand the trade-off between the costs and preserving a certain level of resilience to cope with possible failures.

A new attempt to develop a robust multi-objective design of WDNs was proposed later by Kapelan et al. (2005). It contemplated uncertainties related to future water demand and pipe roughness coefficients. The two objectives were to minimize the total WDNs design cost and maximize WDNs robustness, with robustness defined as the probability of satisfying minimum pressure head constraints at all nodes. A sampling-based technique was used to propagate the WDNs model input uncertainties to the outputs. Uncertain variables were modeled with a Gaussian distribution whose first and second-order moments were predefined. The procedure suggested by Iman and Conover (1982) enabled correlated random samples to be generated. Furthermore, different cases were tested: one with all nodal demands independent of each other and another in which the correlation coefficient between any two nodal demands was assumed to be equal to 0.50, as described by Tolson et al. (2004). The different correlation was considered to depend on weather conditions which can affect extra consumption for garden irrigation.

Jung et al. (2012) also dealt with demand and roughness uncertainty for the multi-objective optimization of WDNs considering cost minimization and maximization of robustness as the two objectives. The robustness is defined by a disturbance index that measures the pressure variation at the critical network node or the pressure variation at all network nodes.

In these two earlier examples, again, the inclusion of uncertainty into the optimization model is dealt with by taking values directly from the samples already built. To keep the representativeness, a large number of sampled values have to be used and the problem can very easily become intractable. None of the examples mentions probabilities assigned to the scenarios, as defined for the implementation of robust optimization models (Mulvey et al., 1995).

The design of WDNs was also studied by Creaco et al. (2015) with a two-step methodology. In the first step, the WDN is designed with the multi-objective of maximizing the resilience index and minimizing costs. Then in a second phase, the solutions found for a benchmark demand scenario are loaded with more demand scenarios to find out if all the required demand is met for these new conditions. The results show that by increasing the index of resilience the delivered demand also increases. This can be seen as a sensitivity analysis used as a post-optimization tool.

Taking advantage of the features of robust optimization formulations, this paper will contribute with a framework to improve solutions when compared to the classical optimization formulations. In fact, the whole spectrum of the distributions can be embraced in an organized way, while still keeping the problem computationally tractable. Scenarios and their probabilities are used instead of point estimates (Mulvey et al., 1995). Magini's work is fundamental for this and represents a novelty for the integration of uncertainty in optimization models. Optimization robust models can be shaped to respond to the needs of decision makers. In fact, the properties of the solution can be characterized and how well it performs in different scenarios can be managed through "solution robust" and "model robust" concepts. This will create the flexibility to manage the level of infeasibility acceptance and their trade-off with the costs involved.

3. Materials and Methods for WDNs

3.1. Problem Statement

This work aims to analyze how different demand scenarios affect the design of WDN. The pipe sizes obtained are analyzed and compared to understand how they work in a set of possible operational situations. Scenarios are generated with a statistical procedure that involves three stages, the first of which is determination of the peak demand marginal distributions in each node of the network using scaling laws derived from historical data sets that allow the estimation of first and second order statistical moments, including nodal cross-correlation as a function of the type and number of users. Also the marginal probability law is derived from the historical data, so the second stage is the generation of a very large number of snapshots by stratified random sampling from the correlated marginals of nodal peak demand; the third stage is the generation of the peak demand scenarios by reducing the snapshots to a number still able to describe the statistical uncertainty of peak demand.

The influence of the number of reduced scenarios on the outcome of the RO design problem is analyzed. Furthermore, a new snapshot reduction technique is presented which is compared with that described in Magini et al. (2019).

3.2. Multi-Objective Models for the Design of Water Distribution Networks

Two optimization models are solved. In the first, a model for the optimization of two objectives is proposed: cost minimization (Equation 1) and Mean-Variance Generalized Resilience and Failure (Mean-Var GRF) index maximization (Equation 2).

$$\text{Min CT} = \sum_{i=1}^{\text{NPI}} (C\text{pipe}_i(\text{Dc}_i) \times L_i) \tag{1}$$

$$\text{Max Mean} - \text{Var GRF} = \text{MeanGRF} - \lambda \sum_{s=1}^{\text{NS}} (\text{GRF}_s - \text{MeanGRF})^2 \times \text{prob}_s \tag{2}$$

where

$$\text{MeanGRF} = \sum_{s=1}^{\text{NS}} (\text{GRF}_s) \times \text{prob}_s \tag{3}$$

$$\text{GRF}_s = \text{Ir}_s + \text{If}_s \tag{4}$$

$$\text{Ir}_s = \frac{\sum_{i=1}^{\text{NN}} \max(\text{Cn}_{i,s} H_{i,s} - \text{Dd}_{i,s} H\text{des}_i, 0)}{\sum_{r=1}^{\text{NR}} Q_{r,s} H_{r,s} + \sum_{p=1}^{\text{NP}} Q_{p,s} H_{p,s} - \sum_{i=1}^{\text{NN}} \text{Dd}_{i,s} H\text{des}_i} \quad s \in \text{NS} \tag{5}$$

$$\text{If}_s = \frac{\sum_{i=1}^{\text{NN}} \min(\text{Cn}_{i,s} H_{i,s} - \text{Dd}_{i,s} H\text{des}_i, 0)}{\sum_{i=1}^{\text{NN}} \text{Dd}_{i,s} H\text{des}_i} \quad s \in \text{NS} \tag{6}$$

where CT—total cost; NPI—number of pipes; $C\text{pipe}_i$ (Dc_i)—unit cost of pipe i as a function of the commercial diameter Dc_i; Dc_i—commercial diameter of pipe i chosen in the set of commercial sizes; L_i—length of pipe i; GRF—generalized resilience/failure index; λ—weight assigned to the variability of GRF; NS—number of scenarios; Ir_s—resilience index for scenario s; If_s—failure index for scenario s; prob_s—probability of scenario s; NN—number of nodes; $\text{Cn}_{i,s}$—outflow delivered to the users of node i in scenario s; $H_{i,s}$—nodal heads for node i in scenario s; $\text{Dd}_{i,s}$—nodal demands for node i in scenario s; $H\text{des}_i$—desired heads for node i; NR—number of source nodes; $Q_{r,s}$—water discharges leaving the source node r in scenario s; $H_{r,s}$—heads of source node i in scenario s; NP—number of pumps; $Q_{p,s}$—water discharge of pump p in scenario s, and $H_{p,s}$—head of pump p in scenario s.

In this model, the first objective (Equation 1) is to minimize the network construction costs given by the sum for all pipes of the unit costs of the commercial diameters to be used multiplied by the length of the pipe. The second objective is to maximize the mean-variance of GRF index (Equation 2). This is an index proposed by Creaco et al. (2016) which is suitable for pressure-driven modeling and is based on the original resilience index proposed by Todini (2000). The GRF is the sum of the resilience index (Equation 5) with the failure index (Equation 6). The GRF value is a measure of the power surplus/deficit of WDNs and is equal to the resilience index when it is greater than 0 or is equal to failure index when this is less than 0. Details about this GRF index can be found in Creaco et al. (2016).

The Mean-Var GRF objective is obtained for all demand scenarios under analysis and their probability, and tends to address risk averse behavior by reducing the chance of solutions that are particularly weak in some scenarios to be selected. Indeed, using expected outcome performance as an objective is not suitable for "moderate and high risk decisions under uncertainty" (Mulvey et al., 1995), because decision makers are very often risk averse. For handling this type of situation a surrogate of risk can be given by the variance of the performance. Therefore, the formulation used in Equation 2 represents the maximization of the objective for a given level of risk. This means that this value is highly dependent on the problem in question and the decision-making context. These mean-variance type models (Markowitz, 1991) started to be applied, some years ago, in several areas:

Table 1
Number of Users and Average Peak Demand at the Nodes of the Network

Node ID	Users	Peak demand (L/s)	Node ID	Users	Peak demand (L/s)
1	126	0.49	19	483	1.89
2	267	1.04	20	240	0.94
3	261	1.02	21	246	0.96
4	210	0.82	22	249	0.97
5	162	0.63	23	222	0.87
6	204	0.80	24	174	0.68
7	69	0.27	25	198	0.77
8	150	0.59	26	435	1.70
9	141	0.55	27	366	1.43
10	285	1.11	28	78	0.31
11	450	1.76	29	159	0.62
12	234	0.92	30	141	0.55
13	297	1.16	31	231	0.90
14	141	0.55	32	264	1.03
15	282	1.10	33	198	0.77
16	312	1.22	34	192	0.75
17	327	1.28	35	297	1.16
18	519	2.03	36	123	0.48

Hodder and Dincer (1986) for location of capacitated facilities, Watkins and McKinney (1997) for finding robust solution to water resources problems, Zeferino et al. (2012) for wastewater treatment management at regional level, and Vieira and Cunha (2016) for capacity expansion of multisource water-supply sysntems. This type of models is well documented in the literature (Snyder, 2006). Including λ, serves to define the importance given to the value of the variance term across scenarios. Defining the λ value should take these concerns into account and provide an adequate balance (which may be a decision maker's choice) between the importance of the mean and the variance terms of the objective function. Very low λ values are not suitable, as solutions become very similar compared to those obtained by only including the MeanGRF (first term of Equation 2). The effect of the variability in the objective function can be practically non-existent if the magnitude of the variance term is very low compared to the MeanGRF term. High values for λ can also be unsuitable as the results lead to solutions mainly controlled by the second term of Equation 2.

The second model includes three objectives, the previous cost minimization (Equation 1) and Mean-Var GRF maximization (Equation 2), and an additional objective: the minimization of UD_{max}, that is, the maximum undelivered demand across all the scenarios (Equation 7):

$$\text{Min UD max} = \sum_{s=1}^{\text{NS}} \max_{i}^{\text{NN}} (\text{Dd}_{i,s} - \text{Cn}_{i,s}) \times \text{prob}_s \quad \forall s \in \text{NS}, \quad \forall \in \text{NN} \quad (7)$$

This is used to minimize the maximum undelivered demand while taking into account the maximum value for all nodes and all demand scenarios. This is a regret-based decision made under uncertainty, measuring the impact of not fulfilling the ideal decision without uncertainty.

The pressure driven simulator is the EPANETpdd (Morley & Tricarico, 2008), because the robust formulation used leads to pressure requirements not being totally satisfied in some circumstances. This is an upgrade of the original EPANET (Rossman, 2002) simulator that uses a head-flow relationship suggested by Wagner et al. (1988). It is applied to simulate the hydraulic behavior of the network and validate the usual hydraulic constraints of WDN design optimization models of nodal mass balance and the head loss in pipes.

3.3. Case Study

Both optimization models are applied to the Fossolo network (FOS) studied by Bragalli et al. (2008, 2012). Twenty-two commercial diameters are available to size the 58 pipes. The minimum pressure at the 36 demand nodes is set to 40 m and water demand is not fully satisfied below this value. For pressure values below 10 m, water request is not delivered in the node. The network is supplied by a single reservoir whose head is 121 m. Network data can be consulted at Bragalli et al. (2008).

Statistical uncertainty is modeled by making use of the residential water consumption data set of the Latina case study set out by Magini et al. (2008). It is therefore assumed that the types of users are comparable. Demand data are aggregated at the 5 min sampling interval. The number of users and the mean demand at peak hour in each node, Table 1, are estimated using demand data by Bragalli et al. (2008) and the consumption statistics for individual users derived from the historical data. In detail, since d_i is the demand at peak hour in the ith node of the network, as described by Bragalli et al. (2008) and $\mu_{1,\text{ peak}}$ the expected value of single-user consumption at peak hour, assuming a linear dependence between consumption and number of users, the number of users was calculated as $d_i/\mu_{1,\text{peak}}$.

3.4. Demand Scenarios

The three phases for the determination of the peak demand scenarios relating to FOS case study are developed below.

Table 2
Statistics and Parameters Derived From the Available Data Set

Statistical parameter	Value
$\mu_{1,\,\text{mean}}$ (L/min)	0.440
$\mu_{1,\,\text{peak}}$ (L/min)	0.704
$\sigma_{1,\,\text{peak}}$ (L/min)	1.01
scaling law exponent α	1.23
Pearson coefficient $\rho_{1\text{-}1}$	0.05

3.4.1. Statistical Characterization of Peak Demand Using the Scaling Laws

The stochastic modeling of water demand requires knowledge of the statistical features of the demand data at different spatial and temporal aggregations. Analysis of real data has revealed the presence of a non-trivial scaling of the second order moments with the number of customers. In this context Magini et al. (2008) and Vertommen et al. (2015) estimated the scaling laws on two residential indoor water demand data sets, from measurements in the case-studies of Latina, Italy and Cincinnati (Magini et al., 2008), Ohio (Buchberger & Wells, 1996). The main findings of these studies were the linearity of the mean with the number of users, the non–linearity of second order statistics depending on the cross-correlation between the demand of the different couples of users. Two limit cases were detected: linearity in the case of no correlation, and a quadratic relation in the case of total correlation between users. Partial correlations give intermediate exponents between 1 and 2. Therefore the expected value μ_{n_i} for the mean of the aggregated process at the ith node is given by:

$$\mu_{n_i} = n_i \cdot \mu_1 \tag{8}$$

and the expected value for the variance $\sigma^2_{n_i}$ at the same node, neglecting the bias due to short observation periods (Vertommen et al., 2015):

$$\sigma^2_{n_i} = n_i^{\alpha}\sigma_1^2 \tag{9}$$

with α scaling exponent for variance, μ_1 and σ_1^2, respectively, mean and variance of the single user. The cross-correlation between couples of nodal demands was also found to depend on the number of users and, following the same assumption, according to Vertommen et al. (2012) its expected value is:

$$\rho_{n_i n_j} = \frac{cov_{n_i n_j}}{\sigma_{n_i}\cdot\sigma_{n_j}} = \frac{n_i n_j \rho_{1-1}}{\left[n_i(1+\rho_{1-1}(n_i-1))\right]^{1/2}\cdot\left[n_j(1+\rho_{1-1}(n_j-1))\right]^{1/2}} \tag{10}$$

with $i,j = 1,2,3,\ldots,$NN and where, for example, $\rho_{n_i n_j}$ is the cross-correlation coefficient between the demand of n_i aggregated users at node i, and the demand of n_j aggregated users at node j, and $\rho_{1-1}$is the Pearson's coefficient between couples of single users, which is assumed to be equal on all nodes.

The nodal demand statistics of the network and the correlation structure are entirely defined through Equations 8–10.

Table 2 shows the mean water demand $\mu_{1,\text{mean}}$, the mean peak demand $\mu_{1,\text{peak}}$ and the standard deviation $\sigma_{1,\text{peak}}$ of the individual user's consumption obtained from historical data. In the same table the exponent α of the scaling law for variance and the Pearson's coefficient ρ_{1-1} are also reported.

3.4.2. Water Demand Snapshots Generation

The scaling laws make it possible to determine the statistics of the water demand in each node of the network as a function of the number and type of users. A further hypothesis concerns the probability distribution that best represents peak demand data. In this regard, the gamma distribution is chosen on the basis of the experimental data of the Latina case study and with the support of the literature (Kossieris et al., 2019). Based on this, a probabilistic sampling technique that respects the cross-correlation of water demands can be developed for generating statistically representative snapshots.

Each demand snapshot D_u, defined as a set or combination of nodal demand values occurring simultaneously in the WDN, is represented by the *NN* dimensional vector:

$$D_u = \left[d_{1,u}, d_{2,u}, \ldots d_{\text{NN},u}\right] \tag{11}$$

where NN is the number of nodes, S the total number of snapshots, $u = 1,2,\ldots,S$ the index identifying the different snapshot and, $d_{i,u}$ is the demand at node i for the u_{th} snapshot.

For this purpose, an algorithm that combines the LHS (McKay et al., 1979) from the nodal marginal distributions and the Iman-Conover method (Iman & Conover, 1982) to induce correlation between samples was employed.

10.1029/2023WR034867

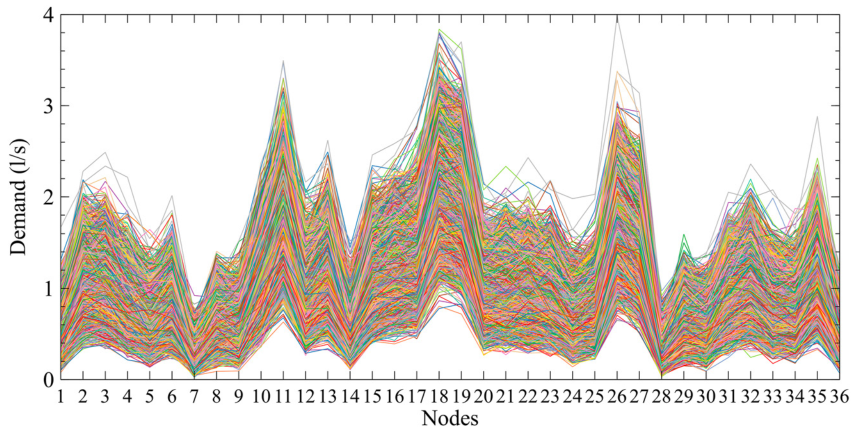

Figure 1. 10,000 snapshots of water demand.

LHS is a "stratified sampling" technique that produces a good description of the input data probability distribution with fewer iterations than simple random sampling. The full procedure is quite simple to implement as it requires only the Cholesky decomposition, some matrix algebra, and the final rearrangement of the uncorrelated original sample. It is described in detail in Magini et al. (2019).

Figure 1 shows the 10,000 snapshots of nodal peak demand generated in the FOS network.

3.4.3. Peak Demand Scenarios Generation by Snapshots Reduction

Each of the generated snapshots defines a picture of the simultaneous nodal water demand, therefore the greater the number of snapshots, the better the description of the variability that characterizes statistical uncertainty. However, too many scenarios cannot be handled when addressing robust optimization problems. Furthermore, the generated snapshots are not associated with a probability/weight value; it can only be assumed that each of them is equiprobable with probability/weight equal to 1/*S*, where *S* is the total number of snapshots. Instead, the objective to be pursued is to consider a limited number of scenarios, each with a probability/weight that represents its possibility of fulfillment. Then, the probability mass function of the generated snapshots must be described by the probability mass function of a small number of suitable snapshots.

For instance, if $\mathbb{P}$ is the probability mass of the *NN*-dimensional stochastic data process described by the S snapshots (Equation 11) each of which with probability p_u, with $\sum_{u=1}^{S} p_u = 1$, it must be approximated by the probability mass $\mathbb{Q}$, as close as possible to $\mathbb{P}$, having a smaller number of elements each of which with probability p'_u, with $\sum_{u=1}^{S} p'_u = 1$.

To approximate $\mathbb{P}$ with $\mathbb{Q}$ in probabilistic terms, the Wasserstein-Kantorovich distance between these distributions was used. For discrete probability distributions with a finite number of elements this is just the optimal value of a linear transportation problem in which a cost function, c_u, defined in a given metric space, is introduced as a measure of the distance between pairs of elements (Dupacová et al., 2002). More simply, this is a measure of the minimal effort required to reconfigure the probability mass of one distribution in order to obtain the other distribution. The first question that arises is: what probability distribution $\mathbb{P}$ do we want to approximate with the reduced number of snapshots? In this work we consider the probability distribution of the total flow entering the network, that is, the sum of the simultaneous demands delivered in the nodes in each snapshot (Figure 2). The problem is not easy, therefore in the literature heuristic reduction algorithms have been proposed that may employ fast-back or fast-forward strategies (Heitsch & Romisch, 2003). Here an iterative algorithm based on the fast-forward selection is developed in which the cost function c_u is defined by the NN-dimensional Euclidean norm (Magini et al., 2019). Each snapshot is assumed to be equiprobable with probability/weight equal to 1/*S*. At each iteration, from the set of unselected snapshots, initially composed of all generated snapshots, the one that has the minimum Wasserstein-Kantorovich distance is taken and inserted into the set of reduced snapshots which is initially empty. The optimal selection of individual snapshots can be repeated recursively until the prescribed number, S, of elements is reached. At each step, the cost matrix composed of the cost functions of the unselected snapshot pairs is updated. Once the desired number of snapshots is reached, the algorithm finishes by transferring

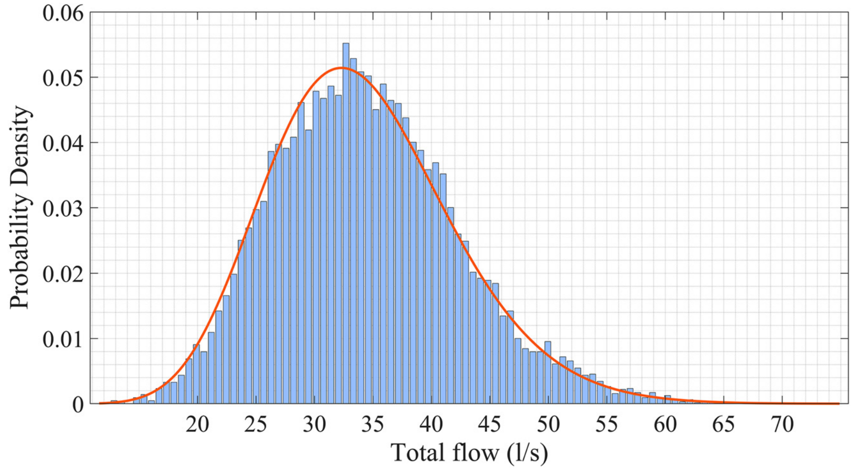

Figure 2. Probability mass function of the total flow entering the network. The red line shows type gamma pdf, which is the best distribution fitting empirical data.

the probability from the unselected snapshots to the selected ones that are closest to them according to the initial cost function, c_u. The probability of each reduced snapshot is set equal to the sum of its initial probability and all the probabilities of the unselected snapshots that are closest to it. The heuristic algorithms do not guarantee the optimal solution to the original transportation problem, but they work well in practice (Heitsch & Romisch, 2003).

Alternative distributions could be considered, for example, the water requirement in a specific node to respect more strict service conditions.

Two different approaches for reducing snapshots are proposed below.

The first applies the reduction technique to the overall set of snapshots. In this case, most of the reduced snapshots naturally pile up away from the tails, in a position closer to the central values of P (Figures 3a and 4a).

In the second approach, the probability mass P is divided into M sub-intervals, covering the entire field of the flow entering the WDN and containing a different number of snapshots. In this case, the reduction technique is applied separately to each sub-interval until a single snapshot is obtained for each of them, having probability $\sum_{u=1}^{M} p_u$, where M is the number of snapshots in the sub-interval and $p_u = 1/S$ is the probability/weight initially assumed for each snapshot generated (Figures 3 and 4b). The fast-forward reduction algorithm is applied separately for each sub-interval until a single snapshot is obtained to which the probability of the corresponding sub-interval can then be attributed. In the robust design of the WDN, the subdivision into sub-intervals makes it possible to consider even the extreme inlet flow rates, in particular those belonging to the right tail. The effectiveness of the partitioned reduction technique is evaluated in the next section, by comparison the results of robust optimization. Specifically, a robust design procedure is followed by first considering the reduced scenarios from the entire P distribution and then the partitioned distribution.

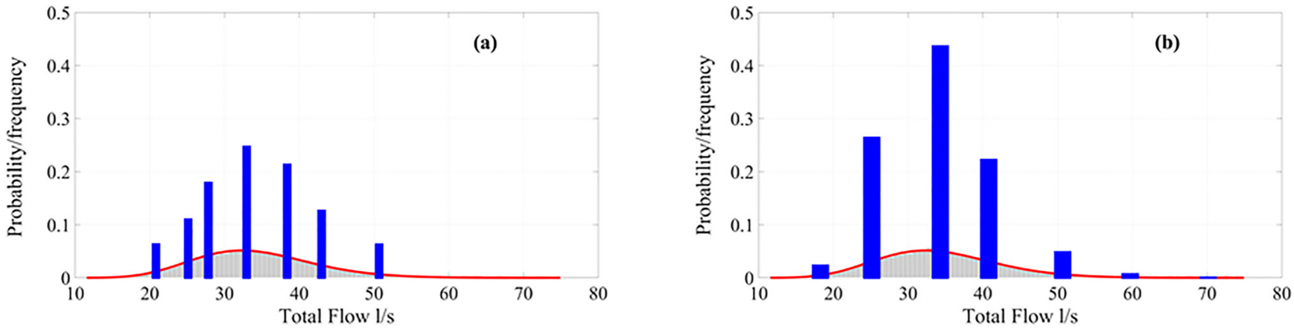

Figure 3. Probability mass of the seven reduced scenarios (blue), probability mass of the 10,000 generated scenarios (light gray), type gamma pdf (red) (a) reduction without sub-intervals, (b) reduction with sub-intervals.

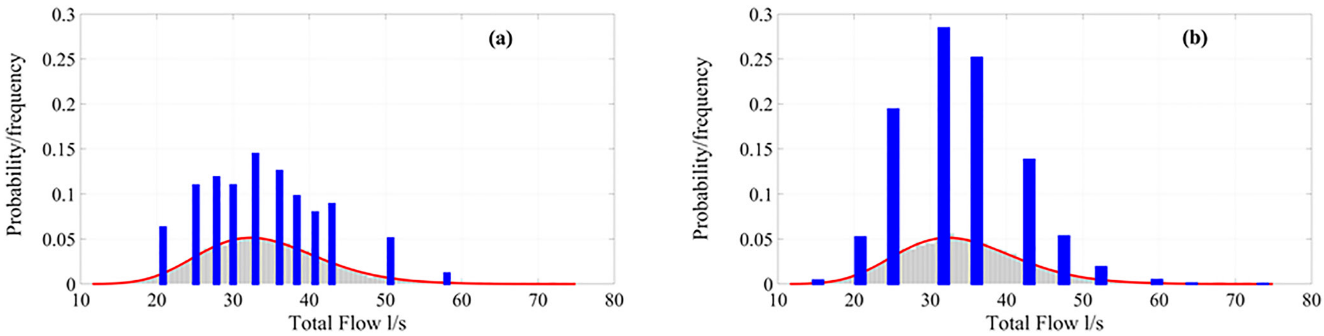

Figure 4. Probability mass of the 11 reduced scenarios (blue), probability mass of the 10,000 generated scenarios (light gray), type gamma pdf (red) (a) reduction without sub-intervals, (b) reduction with sub-intervals.

The number of sub-intervals and consequently of scenarios is a critical factor. In both cases, the choice of the reduced number of snapshots should derive from a trade-off between greater detail in the representation of the probability distribution and lower computational burden in robust optimization. The comparison is made by reckoning a different number of scenarios, 7 and 11 (Figure 5), to also check how much this number influences the robust design solution in each of the two reduction approaches.

3.5. Optimization Algorithm

Both robust multi-objective optimization models, including the demand uncertainty described in Section 3.2, are solved by the algorithm MOSA-GR (MultiObjective Simulated Annealing with new Generation and

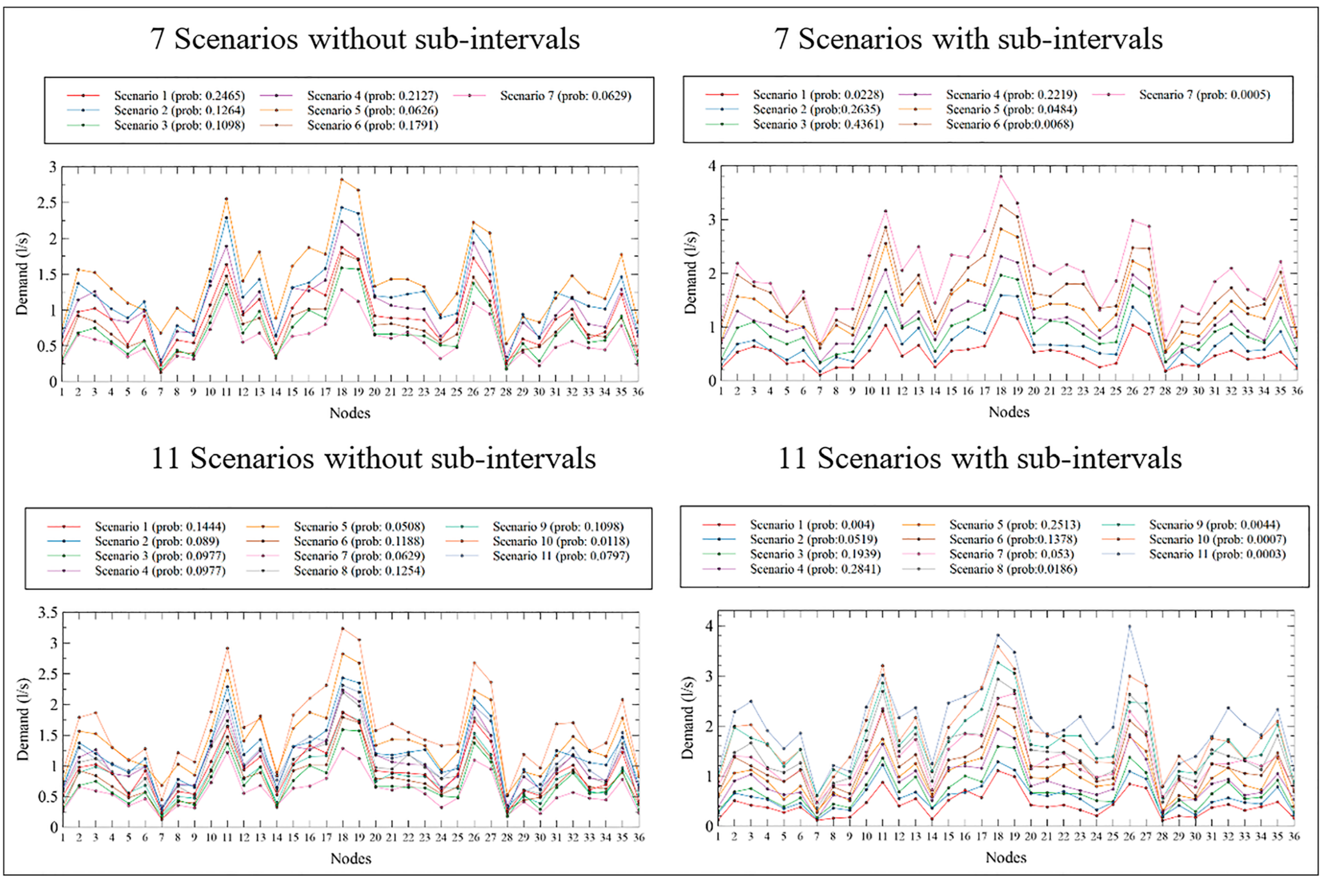

Figure 5. Nodal demand for the case of 7 scenarios with and without sub-intervals and 11 scenarios with and without sub-intervals.

10.1029/2023WR034867

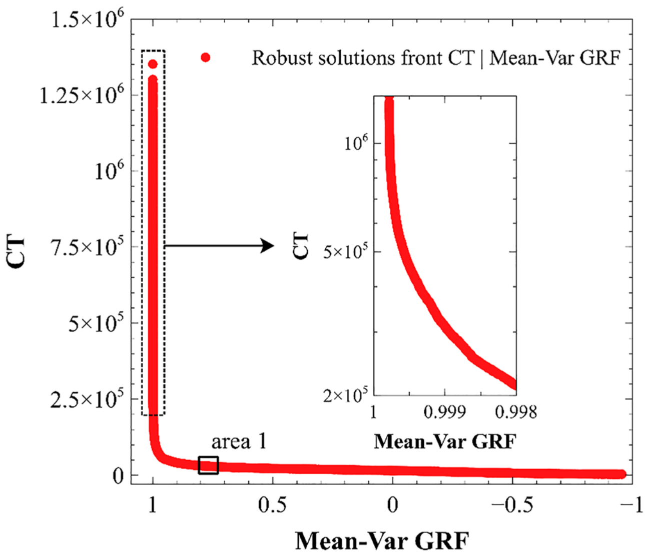

Figure 6. Robust design solutions for the two-objective model.

Reannealing procedures) developed by Cunha and Marques (2020). This is a trajectory-based heuristic that builds on the simulated annealing concept (Kirkpatrick et al., 1983). It is an upgrade of the original simulated annealing algorithm that includes features to promote diversity and uniformity during the convergence process to obtain the best possible set of non-dominated solutions. This was done by implementing different generation strategies to build new candidate solutions and using a reannealing process to intensify the search in the last stage of the algorithm. This means that dense Pareto fronts covering the whole spectrum of trade-offs between solution are obtained. In Cunha and Marques (2020) 12 benchmark problems (including the FOS case) are used to show the large number of additional solutions for reshaping and augmenting the Pareto front obtained compared to the existing solutions from several evolutionary algorithms.

4. Results and Discussion

4.1. Comparison of the Deterministic and Robust Solutions of the Two-Objective Model

The FOS network is sized using the two-objective model proposed in Section 3.2. for the set of seven demand scenarios (this is the robust design) without considering sub-intervals (defined in Section 3.4). The results are all obtained for the same value assigned to the parameter λ in Equation 2. This value is used to assign a level of importance to the variance of GRF and contributes to represent how risk-averse decision-makers are (Section 3.2). It is case-dependent (it may be a decision maker's choice) and defines how much the variance term will control the GRF objective. Of the tested values, $\lambda = 0.1$ was a good compromise between the term MeanGRF and the importance attributed to the variance of GRF between scenarios. It was used in all the comparisons that are drawn in this results section. Figure 6 includes the solutions obtained. These robust solutions will be compared with a network design sized for a single demand condition according to the original case study (results presented in Cunha & Marques, 2020), in which scenarios are not taken into account (this is the deterministic design). Furthermore, in this case, no pressure deficits were allowed and therefore minimum pressures must be met at all nodes in the network for the deterministic design.

The graph in Figure 6 shows the shape of the front with robust design solutions and how increasing the value of the Mean-Var GRF implies increasing costs. The zoom highlighting a part of the front shows a slight variation of Mean-Var GRF in solutions with higher costs. To compare the differences between solutions obtained for the robust design with solutions of the deterministic front, a particular robust solution is selected in the front (part highlighted in area 1 of Figure 6). A deterministic solution with CT value very similar to the robust one is selected from the set of non-dominated solutions found in Cunha and Marques (2020) for the deterministic approach. This pair of specific solutions is analyzed in detail next.

The robust solution selected from the front part highlighted in area 1 comes at a cost of €28,698. The deterministic solution with a cost of €28,679 comes closest. The deterministic solution includes 25 equal, 24 smaller and 9 larger pipe diameters compared to the robust case (diameters in the boxes for both cases in Figure 7). The network nodes are numbered in the deterministic solution with a prefix (N) and pipes are in the robust solution with prefix (P).

In area 1, the selected solution is in a part of the front characterized by low cost and Mean-Var GRF values, therefore some lack of fulfillment of nodal pressures could be expected. In this area, there is no guarantee that the desirable pressures will be satisfied at all nodes for the set of demand scenarios. The comparison of the designs between the deterministic and robust solution shown in Figure 7 is aided by colored rectangles to locate the diameters of the robust solutions that are larger (blue) or smaller (yellow) than the deterministic ones. The robust design reinforces 24 pipes that supply the most distant nodes. This strengthens the network by improving the pressure levels in these nodes. The deterministic design, however, tends to reinforce the pipe near the reservoir, but this is not enough to deliver pressures at the same levels as the robust ones.

For comparison purposes, the two solutions are loaded with the seven demand scenarios so that nodal pressures can be compared in comparable circumstances. The nodal pressures are shown in Figure 8 for four (out of seven)

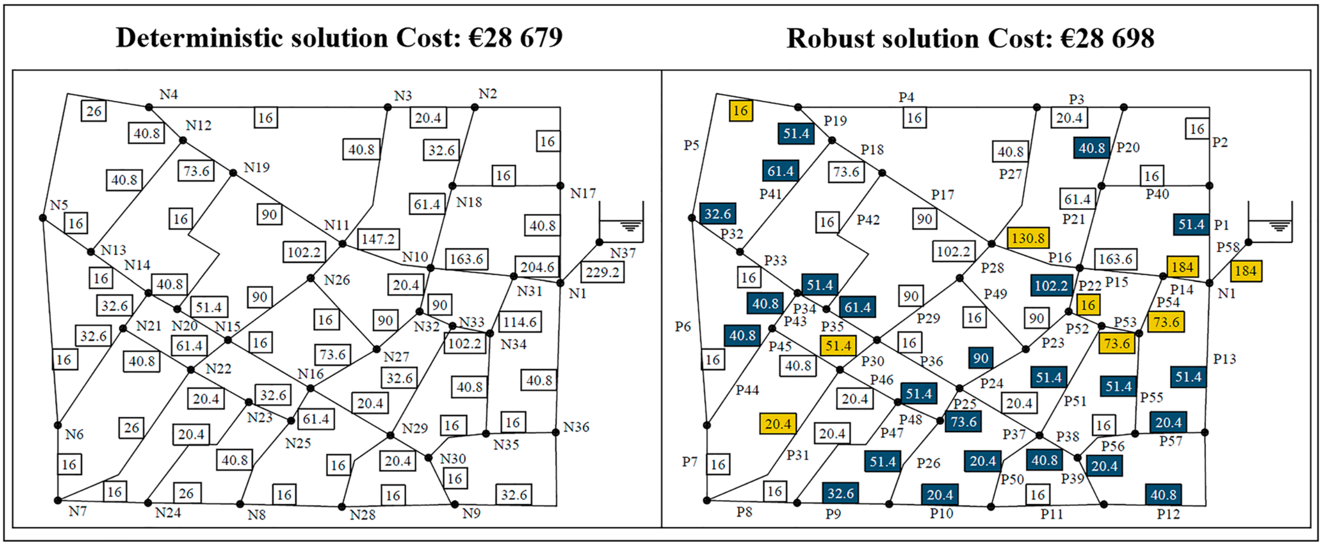

Figure 7. Comparisons of designs for the deterministic and robust solution. Pipe diameters (D) in (mm).

different demand scenarios for the deterministic and robust solution (the most probable scenarios in the right side of the distribution corresponding to the higher demands, see Figure 5). In brief, across all scenarios there are 20 critical situations for the deterministic solution where minimum pressures of 40 m are not confirmed and in the robust case there are just four. According to Figure 5, scenario 1 is the most probable (0.25), and in this case all required pressures are confirmed in both designs. Scenario 4 is the second most probable (0.21), and in this situation five nodes in the deterministic solution include pressures below the required level. In the robust case,

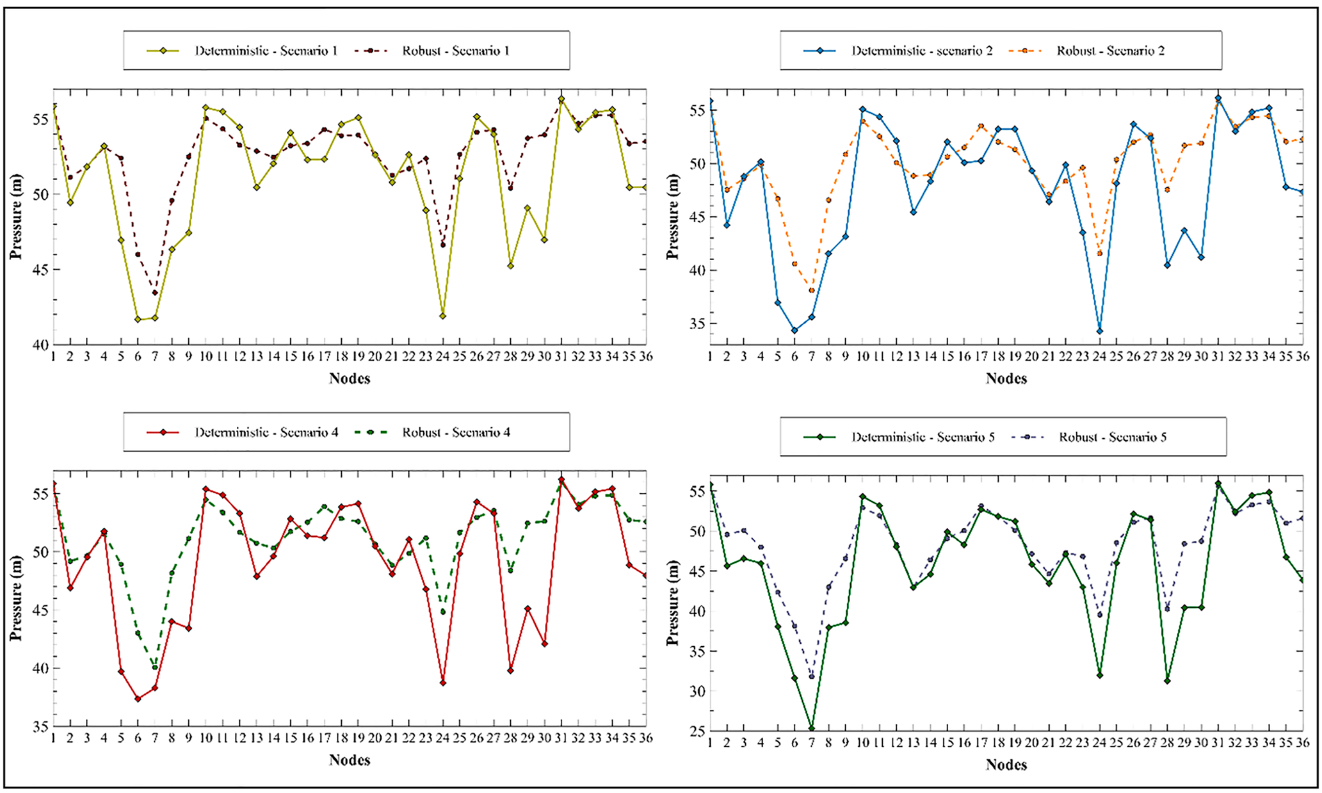

Figure 8. Comparisons of nodal pressures (m) for scenarios 1, 2, 4, and 5 for a pair of deterministic and robust solutions with similar costs.

all pressures are validated. The probability of scenario 2 is 0.13, and in this demand condition the robust solution does not satisfy the pressure in critical node 7 (38.1 m). The deterministic solution performs worse as it does not validate the pressure in four nodes and the lowest pressure of 34.3 m is in node 24. Scenario 5 is the least probable (0.06) and presents the most critical demand condition. The pressure limits fail in seven nodes in the deterministic and in three nodes in the robust designs. The minimum pressure of 25.3 m for the deterministic case is much lower than the minimum of 31.8 m for the robust case. These minimum pressures occur in one of most distant nodes from the reservoir (node 7). As the robust design reinforces pipes that supply the most distant nodes (Figure 7), it is possible to improve the pressure levels in these nodes even under extreme demand situations like scenario 5. This is not the case for the deterministic design.

From these results, the benefits of embracing multiple demand scenarios in the robust optimization model are evident. For the same budget, better solutions can be identified using a robust model that provides superior hydraulic capacity than the deterministic case. This is due to a smarter arrangement of pipe sizes that increase the network capacity where it is effectively needed to simultaneously handle a set of different possible operating conditions.

Most probable scenarios in the left side of the gamma distribution do not show any limitations, because they correspond to lower demands.

4.2. Comparison of Solutions of the Two-Objective Model With the Three-Objective Model

The importance of including an additional objective (minimization UD_{max} as proposed in Equation 7) in the optimization model is evaluated in this section. Despite the relation of the undelivered demand with the pressure deficits, the addition of this new objective will bring extra insights for decision makers. This will increase the complexity of solving the problem, but this objective may lead to positive effects in the performance of the network solutions.

Figure 9 contains four charts. The top graph depicts a view of the Pareto front (with regard to the first (CT) and second (Mean-VAR GRF) objectives) obtained for the two models, indicating zooms 1, 2, and 3 that are shown in the three other graphs of the figure. In these graphs, two pairs of particular solutions are highlighted by black circles.

The two- and three-objective model solutions are obtained for the same set of seven demand scenarios without considering sub-intervals (Section 3.4). The non-dominated solutions for the three-objective problem are above the solutions determined for the two-objective case (Figure 9). A practical conclusion from this situation is that selecting a solution obtained from the three-objective problem with a similar Mean-Var GRF value to a solution from the two-objective model requires additional investment. Or, for a similar cost level, the three-objective model solutions have lower Mean-Var GRF values. But, counting a third objective UD could present advantages that outweigh the disadvantages that these conclusions might suggest.

The particular pairs of solutions in the black circles of Figure 9 have similar costs. These pairs of solutions from the two- and three-objective model are loaded with the same set of seven demand scenarios. As in the previous comparison, nodal pressures are shown for the scenarios in the right side of the distribution corresponding to the higher demands.

For zoom 1, located in an area of high cost and Mean-Var GRF, despite the very similar costs (€37,003 for the three-objective and €37,010 for the two-objective model), the design schemes are very different. In fact, there are only 17 pipes with equal diameters while 41 are different. The three-objective solution tends to reinforce the peripheral network pipes, such as P1, P2, P4, P5, P6, P7, P8, P9, P10, and P11 (P2, e.g., increases from 16 to 51.4 mm, which is five levels up), which connect the nodes furthest from the reservoir to reduce the undelivered demand at these nodes, as is the case of node 7 (the node with the lowest pressure value for scenario 5). The increase can also take place in inner pipes that take the flow to the most critical nodes (nodes 6, 7, and 24). Pipes P7, P8, and P31 (incident pipes of node 7) increased the diameter thus: P7 from 16 to 32.6 mm (three levels of increase), P8 from 20.4 to 40.8 mm (three levels of increase) and P31 from 20.4 to 32.6 mm (two levels of increase). This is important to minimize the maximum undeliverable demand objective, and in fact, in this case, the three-objective solution delivers all demand required (the third object UDmax is equal to zero), which does not happen with the solution for the two-objective model. This is because the nodal pressure at node 7 is

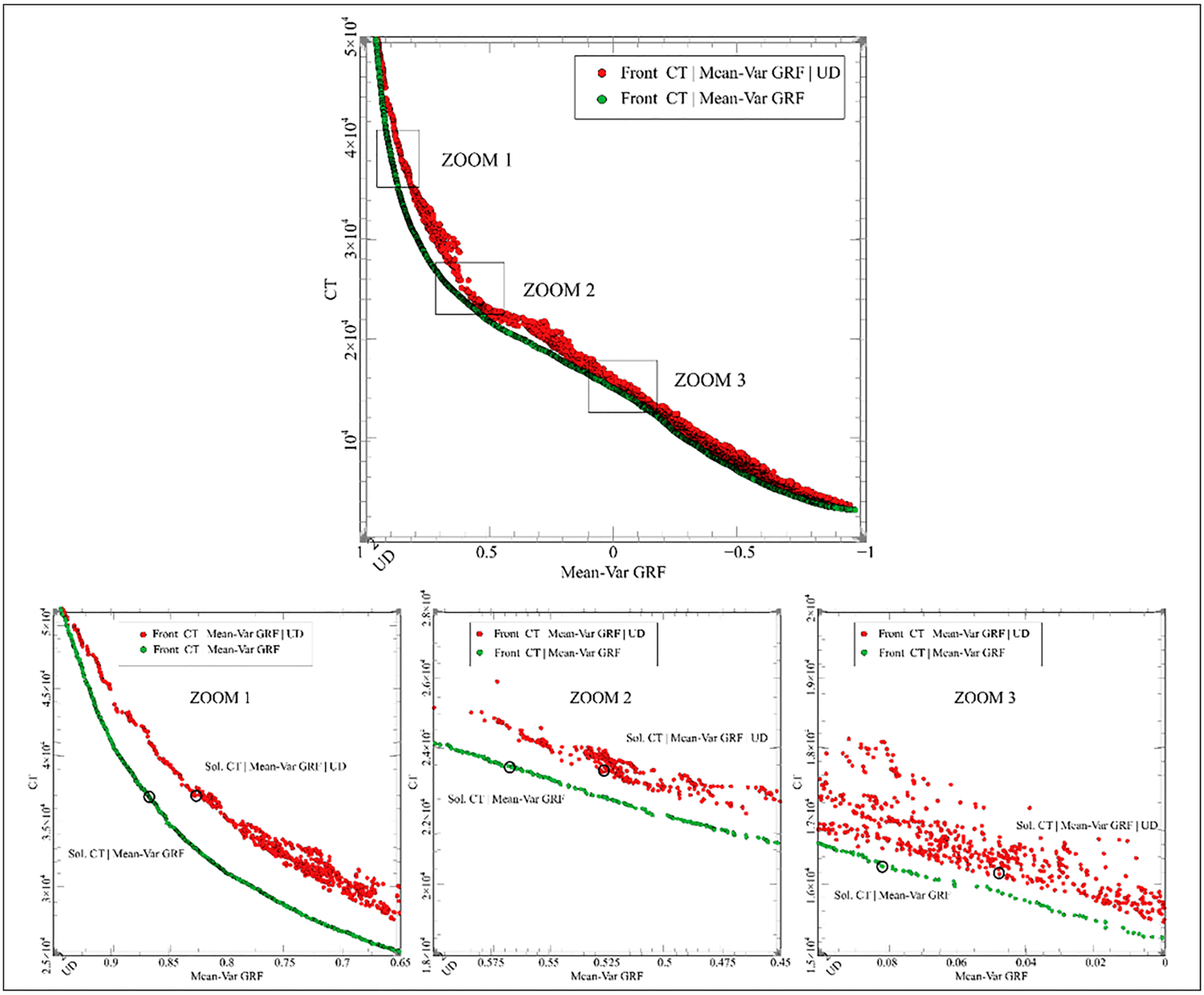

Figure 9. Comparisons of solutions obtained by the two-objective model (green points for Front CT | Mean-Var GRF) with solution obtained by the three-objective model (red points for the Front CT | Mean-Var GRF | UD).

not enough to meet all demand required and thus there is a demand deficit equal to 0.03 L/s (the third objective UDmax = 0.002 L/s, this means the value of the regret function of Equation 7 across all the scenarios taking into account their probabilities). But, by reducing the UDmax objective in the three-objective problem, the Mean-Var GRF of the two-objective solution (0.868) turns out to be better than the value of 0.813 obtained for three objectives. Of the 41 pipe diameters that were changed for the three-objective solution of zoom 1, 24 have higher diameters and 17 are sized with lower diameters to maintain the same cost level. As stated, the peripheral pipes are reinforced in the three-objective solution but the diameter sizes of pipes near the reservoir (such as P14, P15 P22, P23, P24, and P58) are lower than in the two-objective solution. This reduction of pipe capacity near the reservoir means that pressures are lower for three-objective solutions in most of the nodes in the central area of the network (even when they comply with the pressure constraints). This explains why the solutions for the three-objective problem are above the solutions determined for the two-objective case (dominated in terms of Pareto front in Figure 9).

The nodal pressures and the undelivered demand for the pairs of solutions of zoom 2 and zoom 3 are shown in Figures 10 and 11, respectively.

For zoom 2, the network design of the three-objective model shows a Mean-Var GRF value of 0.515 less than the value of 0.57 obtained for the two-objective model for a similar investment cost (€23,495 for the three-objective

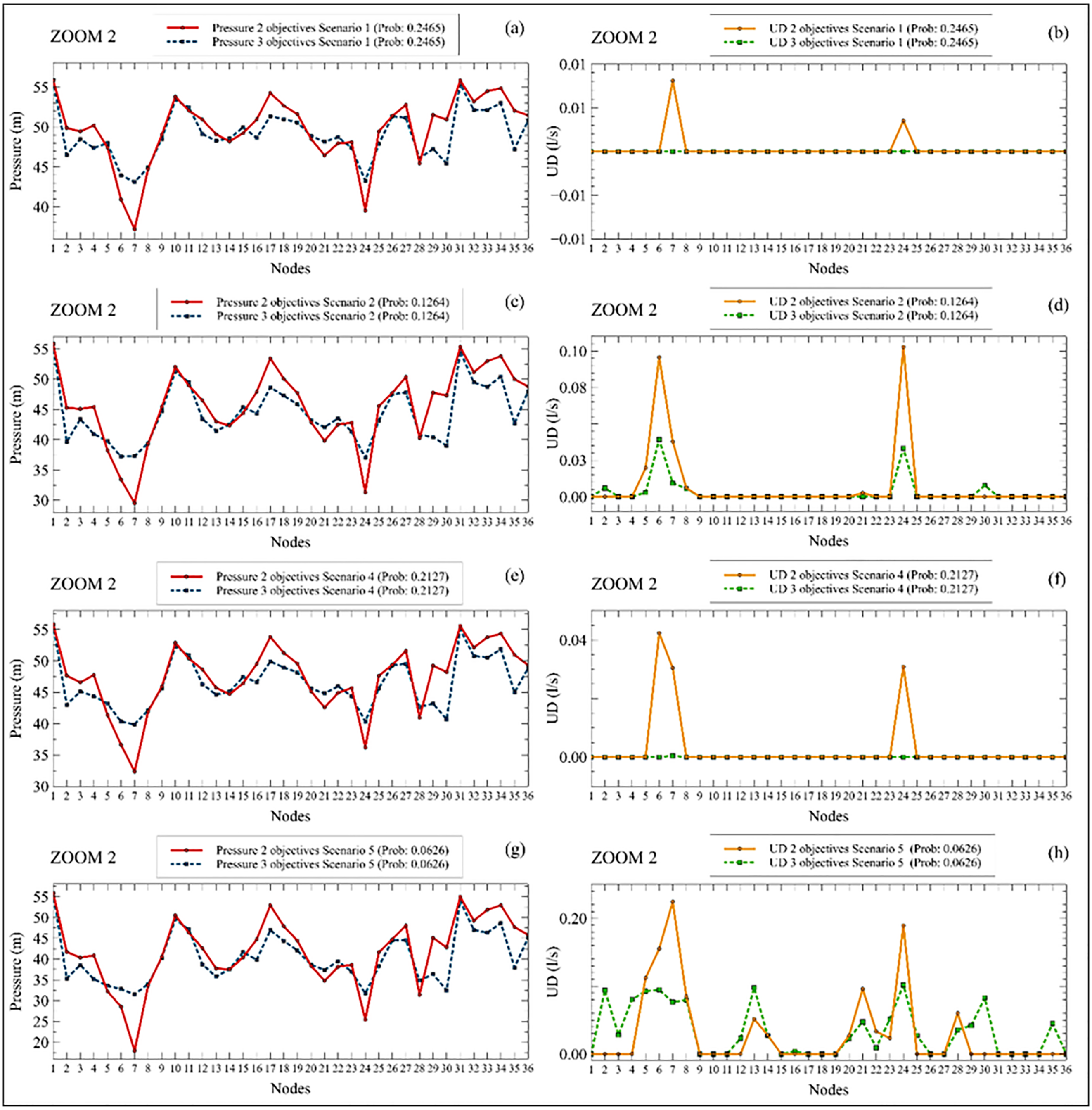

Figure 10. Comparisons of nodal pressures (m) and the nodal undelivered demand (L/s) in the scenarios 1, 2, 4, and 5 for pairs of solutions selected in zoom 2.

model and €23,481 for the two-objective model) and an UDmax for the three-objective of 0.011 L/s lower than the 0.038 L/s of the two-objective model. The three-objective solution reinforce pipes in the bottom part of the network as pipes P7, P9, P10, P11, P12, P13, P31 that increase the hydraulic capacity to supply the nodes 5, 6, 7, 24 and 28. This increases the pressure levels in these nodes and reduce the undelivered demand volumes (Figure 10).

The comparison of nodal pressures of zoom 2 in Figure 10a illustrates that the three-objective solution for the scenario 1 (blue squares) avoids low pressures at critical nodes 7 and 24 (nodes with the lowest pressures and this tends to decrease the UDmax objective). This is not the case of the two-objective solution that includes low pressures in these nodes (red circles). The same kind of conclusions can be taken for the other scenarios 2, 4, and

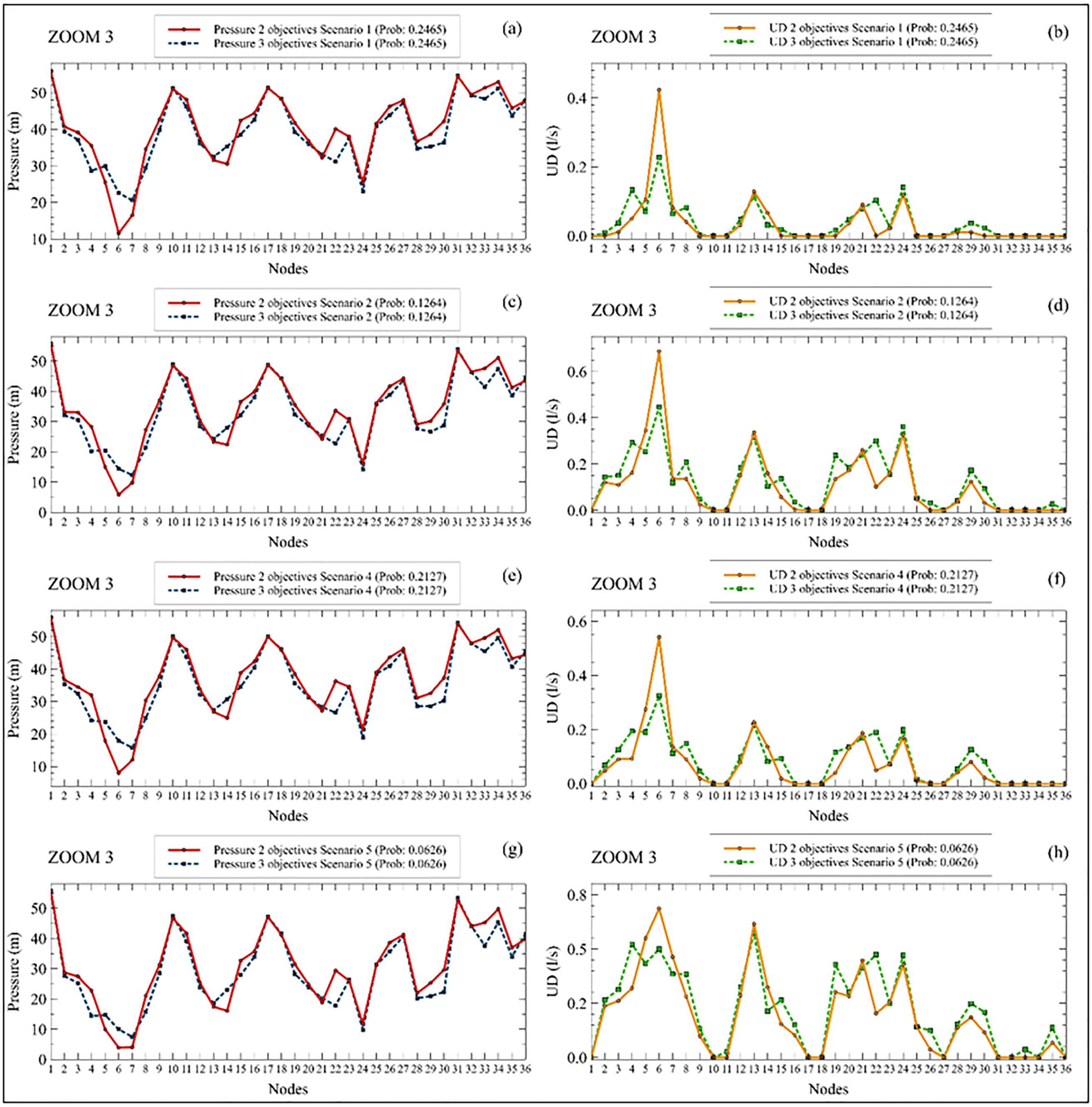

Figure 11. Comparisons of nodal pressures (m) and the nodal undelivered demand (L/s) in scenarios 1, 2, 4, and 5 for pairs of solutions selected in zoom 3.

5 in Figures 10c, 10e, and 10g. The undelivered demand in the three-objective model solution is zero for all nodes in scenario 1 (Figure 10b) and in scenario 4 this is also true for all nodes except node 7 with a residual undelivered demand. In scenarios 2 and 5 (corresponding to the lowest probabilities), there is a number of nodes registering undelivered demand, but mostly inferior in the three-objective model solution compared with the two-objective case (only just slightly higher in some nodes close to the reservoir). For scenario 5 (Figure 10h) the maximum undelivered demand is equal to 0.10 L/s in the three-objective model solution. If the UDmax was not used as objective, as in the two-objective solution, the maximum undelivered demand would be 0.22 L/s (an increment to more than the double in this scenario). It is interesting to observe in Figure 10h that there are several nodes with an undelivered demand value very similar to the maximum (0.1 L/s) for the three-objective solution. This is

achieved by optimizing a network design configuration that tries to avoid large undelivered demands because a regret objective function is pursued (Equation 7). The reduction in the UD is achieved by the deterioration of the Mean-Var GRF index (for pairs of solutions with the same cost level). Figure 10 attest this situation as pressures in critical nodes are higher for the three-objective case, but in some other nodes of the network, the pressures tend to be smaller (easily observable at these figures where the red circles appear above the blue squares in many of these nodes) and this reduce the Mean-Var GRF. The same cost level between this pair of solutions is maintained by reducing diameters of a set of other pipes (P14, P51, P53, P54, P55, P56, P57) including a larger diameter in two-objective solutions than solutions of the three objectives and, as such, the nodal pressures in N29, N30, N33, N34, and N35 become higher. The same occurs with pipes that supply nodes N16, N17, N18, N19, N27, N32. This corresponds to a reorganization of the network, dropping energy excess in some nodes to reduce the undelivered demand in other nodes.

Finally, in zoom 3 (Figure 11), an additional pair of low-cost similar solutions of €16,293 for the two-objective and €16,286 of the three-objective model is selected. In this area of the front, the hydraulic capacity of the network is not enough to supply the demand required even in the most likely operating conditions.

For all the four scenarios represented in Figure 11, a large part of the nodes does not verify the minimum pressures. The three-objective solution includes higher pressures at nodes 6 and 7 than the two-objective solution for all the four scenarios (Figures 11a, 11c and 11e and 11g) that allows lowering the UD at these nodes. From Figure 11b, for scenario 1, the UD in node 6 of the two-objective solution (0.42 L/s) is reduced to almost half the value for the three-objective case (0.23 L/s). Large reductions of UD in this node also occur in scenarios 2, 4, and 5. The three-objective solution uses a larger pipe diameter size than the two-objective one in almost all pipes inside a hypothetical triangle with vertices in nodes N5, N9, and N7. These pipes increase the hydraulic capacity to supply nodes 5, 6, 7, and 14.

The three-objective solution includes 23 pipes of equal size, 22 larger and 13 smaller than the two-objective solution. As in the previous case of zoom 2, the three-objective solution tends to reinforce the peripheral network pipes, which connect the most distant nodes of the reservoir (nodes 6 and 7). These pipe reinforcements are crucial in the three-objective solution so that the undelivered demand at these nodes is less than in the two-objective solution for scenarios 1, 2, 4, and 5 (Figures 11b, 11d, 11f, and 11h).

The Mean-Var GRF index is higher in the two-objective solution (0.082) than for the three-objective solution (0.049), as is also the case in the previous zoom 1 and 2. This decrease in Mean-Var GRF in the three-objective solution is needed to reduce the UDmax from 0.396 L/s in the two-objective to 0.238 L/s in the three-objective solution. The increase in the Mean-Var GRF index of the two-objective solution is achieved by increasing the pressures on a set of network nodes compared to the three-objective solution. In some nodes there is a small increase in pressure as in N20, N25, and N27 (less than 1m for all scenarios in Figure 11) and in some others the growth is greater as in Nodes N4, N22, N30, N33 with a value greater than 7 m for scenario 5 (Figure 11). Even though, these are not critical nodes in terms of the UDmax and therefore they do not negatively influence the value of this objective. Pressure increases at these nodes in the two-objective solution because there is an increase in pipe diameters to form reinforced paths for their supply (such as pipes P54, P55, and P56, as an example of reinforced path for supply node N30).

In the seven scenarios under evaluation (Figure 3a), three (scenarios 3, 6, and 7) are on the left side of the distribution (left of scenario 1) and include low demands; another three are on the right side (scenarios 2, 4 and 5). For zooms 1 and 2, all nodes check the minimum pressures in the left tail (scenarios 3, 6, and 7) but not in the case of zoom 3 solutions. These are lower cost solutions and have insufficient hydraulic capacity to satisfy pressures even for low demand scenarios.

Scenario 5 is the one with highest number of nodes that do not check for nodal pressures. Scenario 7 is the one with fewest number of nodes with problems, three nodes for two- and three-objective solutions. These are nodes 6, 7, and 24 that do not check the minimum pressures in all seven scenarios. Nodes 5, 8, 13, and 21 are also problematic because in five of the seven scenarios they do not check the minimum pressures for the two compared solutions either.

The low demand scenarios contribute 8% to the UDmax objective (Equation 7) compared to the 79% contribution of high demand scenarios (scenarios 2, 4 and 5) and scenario 1 with 13% remaining for the three-objective solution.

For all three pairs of solutions analyzed, it is possible to conclude that the designs are very different even for very similar costs between the pairs. There is only one pipe (P27) for which the same diameter is assigned for the pair of

Table 3
Objective Range Where There Is Trade-Off Between All the Three Objectives

	Range Ct (€)	Ct max(€)	Ct min(€)	Range M-V GRF	M-V GRF max	M-V GRF min	Range UD(L/s)	UD max(L/s)	UD min(L/s)
7_{all}	39,960	43,154	3,194	1.85	0.89	−0.96	1.92	1.92	0
7_{sub}	46,481	49,689	3,208	1.88	0.93	−0.95	1.91	1.91	0
11_{all}	43,258	46,452	3,194	1.87	0.91	−0.96	1.94	1.94	0
11_{sub}	50,216	53,412	3,196	1.90	0.95	−0.95	1.95	1.95	0

solutions of two and three-objective models analyzed, however changing from zoom to zoom (61.4 in zoom 1, 40.8 in, zoom 2 and 32.6 mm in zoom 3). In the other 57 pipes, there is at least one different pipe diameter size in one of the zooms. It is also possible to conclude that pipes P7 and P31 have larger pipe diameters in the three-objective solutions than for the two-objective ones, in all the three zooms. This reduces the maximum undelivered demands of the critical nodes 6 and 7. P55, near the reservoir, is always smaller in the three-objective solutions than in the two-objective ones.

The analysis of the solutions from the three zoomed areas shows the rearrangement of the design of the network whenever the UDmax values have to be taken into account. The reinforcements differ for the three cases. The inclusion of the third objective (Equation 7), leads to designs of the network along the Pareto front so that a balance is established between the three objectives (this really corresponds to assigning different sets of weights to the objectives). It is clear that compared to the results obtained for the two-objective model, this reorganization gives rise to solutions where UDmax objective is always lower.

4.3. Comparison of Solutions Determined for 7 Scenarios and 11 Scenarios With and Without Considering Sub-Intervals

So far, only the case of the set of seven demand scenarios without considering sub-intervals has been evaluated. We will now indicate scenarios obtained without a sub-interval adding suffix *"all"*' to their number, that is, 7*all*, 11*all* and scenarios obtained with sub-intervals adding suffix *"sub"* to their number, that is, 7*sub*, 11*sub*. The effect of considering sub-intervals or not in obtaining network demand scenarios is analyzed in this section, as well as the increase in the number of scenarios to represent nodal demand uncertainty. For this, Table 3 includes the objective ranges for the 7 scenarios and for the 11 scenarios, considering and not considering sub-intervals. This objective range is determined for the set of non-dominated solutions where there is a trade-off between all the three objectives.

Using sub-intervals to determine the demand conditions of the network has the impact of increasing the ranges of objectives for both cases of 7 and 11 scenarios. Sub-intervals in determining demand scenarios has a larger impact than using a larger number of scenarios (from 7 to 11).

In fact, the employment of more scenarios (from 7 to 11) tends to use conditions with a slight increase of the nodal demand in the most extreme situations (Figure 5), but this is accompanied by a decrease in the probabilities associated with each scenario, as they must be diluted through a higher number of scenarios.

Table 3 shows that the cost range increases if sub-intervals are applied. If a comparison is made for the range of 7*all* scenarios (€39,960) with the range of 7*sub* scenarios (€46,481) an increase of 16% is reached. Comparison of the range of 11*all* (€43,258) with the range of 11*sub* (€50,216) shows a similar increase of 16%.

Figure 12 presents the set of graphs with the non-dominated solutions obtained for 7*sub* and 11*sub* scenarios, from different perspectives. The extension of the fronts is not only due to the consideration of subintervals to define the demands and their probabilities, but also due to the increase in the number of scenarios.

The top two graphs show, in the same perspective, the determined borders where there is a compromise between the three objectives to be optimized. The front obtained for 11*sub* increases the range of solutions from cost values of €49,689 in the case of 7*sub* to values of €53,216 for 11*sub*. For each of the problems, three graphs are also presented that correspond to the border formats, analyzing the perspective of the CT VS UD, CT VS M-V GRF and M-V GRF VS UD. There is a broader spectrum of solutions, not only at the level of the costs as already mentioned, but also at the level of the M-V GRF and the UD objectives. Despite this change in limits, the two fronts are very similar in shape.

When solutions from the three-objective model, determined for 7*all* scenarios (zooms 1, 2, and 3 analyzed in Section 4.3) are loaded with 7*sub* scenarios demands, it is noticed that solutions presenting similar costs do not

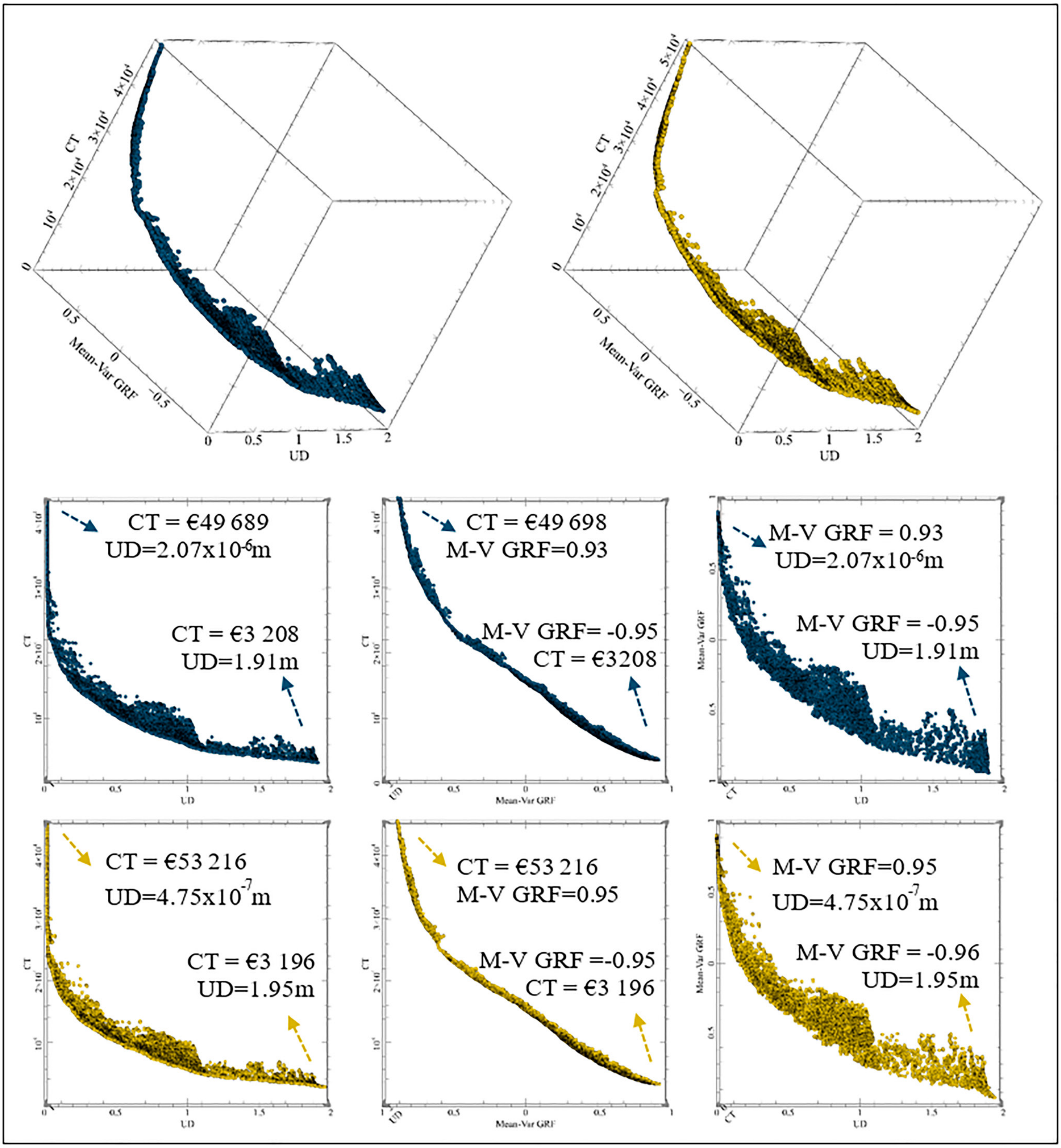

Figure 12. Comparisons of non-dominated solutions from different perspectives for 7*sub* (blue) and 11*sub* (yellow) scenarios.

show significative changes for Mean-VarGRF and UDmax. In spite of the higher demands available for 7*sub* scenarios (Figure 5), the decrease of probabilities for critical scenarios (high demand scenarios) give rise to this conclusion. It should be noted that the solutions analyzed are among the solutions in the range of variation that is common to both fronts, otherwise they could not be compared.

When solutions presenting similar costs obtained for 7*all* and 11*all* scenarios are loaded with the demands of 11*all* and 7*all* scenarios respectively (as before, solutions analyzed are among the solutions in the range of

10.1029/2023WR034867

variation i.e., common to both fronts), there is a slight increase of UDmax for the solution obtained for the 11*all* conditions. This can be explained by the lower probabilities assigned to the critical scenarios (those presenting the highest demands) for the solution designed for 11*all* compared to the solutions from 7*all* scenarios.

5. Conclusions

User demand is the major source of uncertainty in the design of urban WDNs. Uncertain parameters, if not correctly considered, can compromise the robustness of the infrastructure, that is, its ability to guarantee service functionality for different demand scenarios. In this work, therefore, a multi-objective RO model solved by the algorithm MOSA-GR (MultiObjective Simulated Annealing with new Generation and Reannealing procedures), developed by Cunha and Marques (2020), was used to take water demand uncertainty into account, ensuring the best trade-off between the least cost and the highest network robustness. In one case this was measured through a Mean-Var GRF (GRF is the Generalized Resilience and Failure index) and in another by considering both the Mean-Var GRF index and the maximum undelivered nodal demand across all the scenarios, UD_{max}.

Demand uncertainty was modeled using a data-driven method, generating demand scenarios with different probabilities of occurrence. Preliminarily, this method includes the statistical characterization of the available consumption data and the determination of scaling laws linking the consumption statistics to the number and type of the users. In this way, a very large number of snapshots were generated by stratified random sampling from the correlated marginals. Two different heuristic techniques were applied to reduce the snapshots to a number that at the same time decreased the computational burden and secures an appropriate description of the water demand variability. The reduction technique also made it possible to associate a probability/weight to each of the remaining snapshots. The scenarios and the corresponding probabilities are thus defined. The procedure was applied to the Fossolo network studied by Bragalli et al. (2008) and the results were compared with those from the same network, sized for a single demand condition by Cunha and Marques (2020) using the MOSA-GR algorithm according to the original case study. The comparisons of the solutions obtained by the two-objective model with the three-objective one, including the additional objective of reducing UDmax, also show an improvement in the performance of the network.

An additional aim of the paper is to show how the different parts work together and what results can be expected. This is why a real case study is considered to analyze the role of the main ingredients needed to implement the framework and respective results. When we are dealing with uncertainty we cannot establish definitive rules applicable to all types of problems, regardless of the data and the circumstances of the decision-making procedure.

The main takeaways from this work can be summarized as follows.

- The development of a framework that includes uncertainty issues regarding demand to improve the design of WDNs compared with the classical optimization formulations is of a great interest. Indeed, the features of multi-objective RO models allow account to be taken of statistical uncertainty in nodal demands by considering scenarios, that is, demand snapshots having a fixed weight/probability.
- For this, a novel methodology (based on Magini's work) boosts the integration of uncertainty in optimization models. This means the whole spectrum of demand distribution is embraced in an organized way, while still keeping the problem computationally tractable. Scenarios and their probabilities are used instead of point estimates (Mulvey et al., 1995). This means that data collection is a crucial aspect. The expansion of smart water meter technology in recent years will contribute to this endeavor.
- The generation of scenarios depicting uncertainty of water demand definitely represents the most original element of the work. Scenarios are defined using a statistical analysis of historical data scaled according to the number of users. The results highlight the advantage of applying the reduction methodology based on reducing the Wasserstein-Kantorovich distance to sub-intervals of the probability distribution of the flow rates entering the network. This methodology, like stratified sampling techniques, divides the statistical population into sub-populations and broadens the range of all the objectives of the optimization model. Using sub-intervals in determining demand scenarios proves to have a greater impact than using a larger number of scenarios without partitioning.
- The results also highlight the reorganization of the hydraulic design not only when we move from a deterministic approach to a robust approach, but also when we move from a two-objective model to the use of a three-objective model:
  - The comparision of results of the two-objective model using a deterministic approach and the robust model here proposed in which the same scenarios are loaded show a better performance of the RO model. The

robust design strengthens the network pipes that supply the most distant nodes and this improves the pressure levels in these nodes. The number of nodes that do not satisfy minimum pressures of 40 m is much higher in the deterministic solution.

- The use of the three-objective model, including UDmax as a third objective for the optimization procedure, instead of the two-objective model, affects the design configuration by trying to prevent large undelivered demands. This is achieved by reinforcing peripheral network pipes that connect nodes furthest from the reservoir in order to reduce the undelivered demand at these nodes. But this UD reduction is achieved through the decline of the Mean-Var GRF for pairs of solutions with the same cost level as pressures in critical nodes are greater for the three-objective case, but in other nodes of the network the pressures tend to be smaller due to the reorganization of the network design. When UDmax comes into play in guaranteeing robustness for the network, undelivered demand is reduced in all the critical nodes.

- It is clear that with the implementation of the proposed framework, systematized insights can be provided which facilitate an open discussion with decision makers and robust models can be shaped to respond to their needs.
- RO models are established and ready to be explored. Other aspects for their implementation, such as the number of scenarios to be used and the level of risk aversion of decision makers, are problem-dependent. They have to be well thought out by the users, given the specific conditions of the application.

To sum up, we can state that a statistically based scenario generation proves useful in favoring the design of more robust distribution networks and can find increasing applications given today's large availability of consumption data and the recent advances in big data analytics. More, and more complex, networks should analyzed. To the best of our knowledge, this is the first study that explicitly incorporates statistical uncertainty in RO models based on the scenario generation approach proposed.

Furthermore, this paper contributes to a better informed decision-making process. The analysis of trade-off between different objectives, embracing investments needed to gain infrastructure robustness and including social aspects related to the undelivered demands, will allow decision makers to consider people's willingness to support more sustainable infrastructure systems among the choices that could be implemented.

Data Availability Statement

The hydraulic model of the FOS network can be accessed in Bragalli et al. (2012). Relevant information for the analysis of case studies is openly available in Cunha et al. (2023).

Acknowledgments
The authors MCC and JM would like to acknowledge the support of national funds through FCT, under the project UIDB/ 00285/2020. JM was acknowledged a research grant awarded by the University of Coimbra.

References

Balacco, G., Carbonara, A., Gioia, A., Iacobellis, V., & Piccinni, A. F. (2017). Evaluation of peak water demand factors in Puglia (Southern Italy). *Water*, *9*(96), 96. https://doi.org/10.3390/w9020096

Ben-Tal, A., Margalit, T., & Nemirovski, A. (2000). Robust modeling of multistage portfolio problems. *High Performance Optimization*, 303e328. Kluwer Academic Publishers.

Bertsimas, D., Dunn, J., Pawlowski, C., & Zhuo, Y. W. (2019). Robust classification. *INFORMS J. Optim*, *1*(1), 2–34. https://doi.org/10.1287/ijoo.2018.0001

Bragalli, C., D'Ambrosio, C., Lee, J., Lodi, A., & Toth, P. (2008). Water network design by MINLP.

Bragalli, C., D'Ambrosio, C., Lee, J., Lodi, A., & Toth, P. (2012). On the optimal design of water distribution networks: A practical MINLP approach [Dataset]. Optimization and Engineering, *13*, 219–246. https://doi.org/10.1007/s11081-011-9141-7

Buchberger, S. G., & Wells, G. J. (1996). Intensity, duration, and frequency of residential water demands. *Journal of Water Resources Planning and Management*, *122*(1), 11–19. https://doi.org/10.1061/(asce)0733-9496(1996)122:1(11)

Cario, M. C., & Nelson, B. L. (1997). *Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix*. Tech. Rep. Department of Industrial Engineering and Management Sciences, Northwestern University.

Creaco, E., Fortunato, A., Franchini, M., & Mazzola, M. R. (2015). Water distribution network robust design based on energy surplus index maximization. *Water Science & Technology: Water Supply*, *15*(6), 1253–1258. https://doi.org/10.2166/ws.2015.091

Creaco, E., Franchini, M., & Todini, E. (2016). Generalized resilience and failure indices for use with pressure-driven modeling and leakage. *Journal of Water Resources Planning and Management*, *142*(8), 04016019. https://doi.org/10.1061/(asce)wr.1943-5452.0000656

Creaco, E., Galuppini, G., Campisano, A., & Franchini, M. (2021). Bottom-up generation of peak demand scenarios in water distribution networks. *Sustainability*, *13*(1), 1–18. https://doi.org/10.3390/su13010031

Creaco, E., Pezzinga, G., & Savic, D. (2017). On the choice of the demand and hydraulic modeling approach to WDN real-time simulation. *Water Resources Research*, *53*(7), 6159–6177. https://doi.org/10.1002/2016wr020104

Cunha, M., Magini, R., & Marques, J. (2023). Data_Multi-objective optimization models for the design of water distribution networks by exploring scenario-based approaches [Dataset]. Mendeley Data, V2. https://doi.org/10.17632/hc8cstzcxx.2

Cunha, M., & Marques, J. (2020). A new multiobjective simulated annealing algorithm—MOSA-GR: Application to the optimal design of water distribution networks. *Water Resources Research*, *56*(3). https://doi.org/10.1029/2019wr025852

Cunha, M., & Sousa, J. (2010). Robust design of water distribution networks for a Proactive risk management. *Journal of Water Resources Planning and Management*, *136*(2), 227–236. https://doi.org/10.1061/(asce)wr.1943-5452.0000029

Dantzig, G. (1955). Linear programming under uncertainty. *Management Science*, *1*(3–4), 197–206. https://doi.org/10.1287/mnsc.1.3-4.197

Di Mauro, A., Cominola, A., Di Nardo, A., & Castelletti, A. (2021). Urban water consumption at multiple spatial and temporal scales. A review of existing datasets. *Water*, *13*(36). https://doi.org/10.3390/w13010036

Di Nardo, A., Di Natale, M., Gargano, R., Giudicianni, C., Greco, R., & Santonastaso, G. F. (2018). Performance of partitioned water distribution networks under spatial-temporal variability of water demand. *Environmental Modelling & Software*, *101*, 128–136. https://doi.org/10.1016/j.envsoft.2017.12.020

Dupacová, J., Gröwe-kuska, N., & Römisch, W. (2002). Scenario reduction in stochastic programming, an approach using probability metrics. *Mathematical Programming*, *95*, 493–511.

Eck, B., Fusco, F., & Taheri, N. (2015). Scenario generation for network optimization with uncertain demands. In *Proeedings of World Environmental and Water Resources Congress, 17th Water Distribution Systems Analysis Symposium* (pp. 844–852).

Fiering, M. B. (1964). Multivariate technique for synthetic hydrology. *Journal of the Hydraulics Division - Civil Engineering Database*, *90*(5), 43–59. https://doi.org/10.1061/jyceaj.0001123

Filion, Y., Adams, B., & Karney, B. (2007). Cross correlation of demands in water distribution network design. *Journal of Water Resources Planning and Management*, *133*(2), 137–144. https://doi.org/10.1061/(asce)0733-9496(2007)133:2(137)

Gabrel, V., Murat, C., & Thiele, A. (2013). Recent advances in robust optimization: An overview, 2012. Retrieved from http://www.optimization-online.org/DB_FILE/2012/07/3537.%20pdf

Gargano, R., Tricarico, C., Granata, F., Santopietro, S., & De Marinis, G. (2017). Probabilistic models for the peak residential water demand. *Water*, *9*(417), 417. https://doi.org/10.3390/w9060417

Gato-Trinidad, S., & Gan, K. (2012). Characterizing maximum residential water demand. *Urban Water Journal*, *122*, 15–24.

Giudicianni, C., Campisano, A., Di Nardo, A., & Creaco, E. (2022). Pulsed demand modelling for the optimal placement of water quality sensors in water distribution networks. *Water Resources Research*, *58*, e2022WR033368. https://doi.org/10.1029/2022WR033368

Green, K. C., Armstrong, J. S., & Graefe, A. (2007). Methods to elicit forecasts from groups: Delphi and prediction markets compared. *International Journal of Forecasting*, *8*, 17–20.

Greenberg, H. J., & Morrison, T. (2007). Chapter 1 robust optimization. *Biology*, 51.

Hart, W. E., Carr, R. D., Phillips, C. A., & Watson, J. P. (2007). *LDRD final report: Robust analysis of large-scale combinatorial applications, SAND2007-5983*. Sandia National Laboratories.

Heitsch, H., & Romisch, W. M. (2003). Scenario reduction in stochastic programming. *Computational Optimization and Applications*, *24*(2), 187–206.

Hodder, J. E., & Dincer, M. C. (1986). A multifactor model for international plant location and financing under uncertainty. *Computers & Operations Research*, *13*(5), 601–609. https://doi.org/10.1016/0305-0548(86)90053-5

Iman, R. L., & Conover, W. J. (1982). A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics*, *11*(3), 311–334. https://doi.org/10.1080/03610918208812265

Jahanpour, M., & Tolson, B. A. (2018). Pressure-driven leak and burst diagnosis under demand uncertainty in water distribution networks. *1st International WDSA / CCWI 2018 Joint Conference, Kingston, Ontario, Canada*, (Vol. 9).

Jung, D., Chung, G., & Kim, J. H. (2012). Optimal design of water distribution systems considering uncertainties in demands and roughness coefficients. In *Water Distribution Systems Analysis 2010 - Proceedings of the 12th International Conference on Water Distribution Systems Analysis (WDSA)* (pp. 1390–1399).

Kapelan, Z. S., Savic, D. A., & Walters, G. A. (2005). Multiobjective design of water distribution systems under uncertainty. *Water Resources Research*, *41*(11), W11407. https://doi.org/10.1029/2004wr003787

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*(4598), 671–680. https://doi.org/10.1126/science.220.4598.671

Kossieris, P., Tsoukalas, I., Makropoulos, C., & Savic, D. (2019). Simulating marginal and dependence behaviour of water demand processes at any fine time scale. *Water*, *11*(5), 885. https://doi.org/10.3390/w11050885

Lansey, K. E., Duan, N., Mays, L. W., & Tung, Y. (1989). Water distribution system design under demand uncertainty. *Journal of Water Resources Planning and Management*, *115*(5), 630–645. https://doi.org/10.1061/(asce)0733-9496(1989)115:5(630)

Magini, R., Boniforti, M. A., & Guercio, R. (2019). Generating scenarios of cross-correlated demands for modelling water distribution networks. *Water*, *11*(3), 493. https://doi.org/10.3390/w11030493

Magini, R., Pallavicini, I., & Guercio, R. (2008). Spatial and temporal scaling properties of water demand. *Journal of Water Resources Planning and Management*, *134*(5), 276–284. https://doi.org/10.1061/(asce)0733-9496(2008)134:3(276)

Markowitz, H. M. (1991). Portfolio selection, efficiency diversification of investments. *Cowles Foundation Monograph*, *16*. Yale University Press: Basil Blackwell.

Marques, J., Cunha, M. C., Sousa, J., & Savić, D. (2012). Robust optimization methodologies for water supply systems design. *Drinking Water Engineering and Science*, *5*(1), 173–192. https://doi.org/10.5194/dwes-5-31-2012

McKay, M. D., Conover, W. J., & Beckman, R. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *21*(2), 239–245. https://doi.org/10.1080/00401706.1979.10489755

Morley, M. S., & Tricarico, C. (2008). Pressure driven demand extension for EPANET (EPANETpdd).

Mulvey, J. M., Vanderbei, R. J., & Zenios, S. (1995). Robust optimization of large-scale systems. *Operations Research*, *43*(2), 264–281. https://doi.org/10.1287/opre.43.2.264

Pallavicini, I., & Magini, R. (2007). Experimental analysis of residential water demand data: Probabilistic estimation of peak coefficients at small time scales. In B. Ulanicki, K. Vairavamoorthy, D. Butler, P. L. M. Bounds, & F. A. Memon, (Eds.), *Water management challenges in global change*.

Perelman, L., Housh, M., & Ostfeld, A. (2013a). Least cost design of water distribution systems under demand uncertainty: The robust counterpart approach. *Journal of Hydroinformatics*, *15*(3), 737–750. https://doi.org/10.2166/hydro.2013.138

Perelman, L., Housh, M., & Ostfeld, A. (2013b). Robust optimization for water distribution systems least cost design. *Water Resources Research*, *49*(10), 6795–6809. https://doi.org/10.1002/wrcr.20539

Rossman, L. (2002). In B. Coulbeck & O. Chun-Hou (Eds.), *EPANET2 user manual*. *System analysis and simulation*. US Environmental Protection Agency, National Risk Management Research Laboratory. Retrieved from **Error! Hyperlink reference not valid.**

Salcedo-Díaz, R., Ruiz-Femenia, R., Caballero, J. A., & Ravagnani, M. A. S. S. (2020). Water distribution network optimization considering uncertainties in the nodes demands. *Computer Aided Chemical Engineering*, *48*, 1183–1188.

Snyder, L. V. (2006). Facility location under uncertainty: A review. *IIE Transactions*, *38*(7), 537–554. https://doi.org/10.1080/07408170500216480

Todini, E. (2000). Looped water distribution networks design using a resilience index based heuristic approach. *Urban Water*, *2*(2), 115–122. https://doi.org/10.1016/s1462-0758(00)00049-2

Tolson, B. A., Maier, H. R., Simpson, A. R., & Lence, B. J. (2004). Genetic algorithms for reliability-based optimisation of water distribution systems. *Journal of Water Resources Planning and Management*, *130*(1), 63–72. https://doi.org/10.1061/(asce)0733-9496(2004)130:1(63)

Tricarico, C., de Marinis, G., Gargano, R., & Leopardi, A. (2007). A. Peak residential water demand. *Proceedings of the ICE - Water Management J.*, *160*(2), 115–121. https://doi.org/10.1680/wama.2007.160.2.115

Vertommen, I., Magini, R., & Cunha, M. (2015). Scaling water consumption statistics. *Journal of Water Resources Planning and Management*, *141*(5), 04014072.

Vertommen, I., Magini, R., Cunha, M. C., & Guercio, R. (2012). Water demand uncertainty: The scaling law approach. In D. A. Ostfeld (Ed.), *Water supply systems analysis: Selected topics* (pp. 1–25). InTech.

Vieira, J., & Cunha, M. C. (2016). Systemic approach for the capacity expansion of multisource water-supply systems under uncertainty. *Journal of Water Resources Planning and Management*, *142*(10), 04016034. https://doi.org/10.1061/(asce)wr.1943-5452.0000668

Wagner, J. M., Shamir, U., & Marks, D. H. (1988). Water distribution reliability: Simulation methods. *Journal of Water Resources Planning and Management*, *114*(3), 276–294. https://doi.org/10.1061/(asce)0733-9496(1988)114:3(276)

Walker, W. E., Harremoes, P., Rotmans, J., van der Sluijs, J. P., van Asselt, M. B. A., Janssen, P., & Krayer von Krauss, M. P. (2003). Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, *4*(1), 5–17. https://doi.org/10.1076/iaij.4.1.5.16466

Watkins, D. W., & McKinney, D. C. (1997). Finding robust solutions to water resources problems. *Journal of Water Resources Planning and Management*, *123*(1), 49–58. https://doi.org/10.1061/(asce)0733-9496(1997)123:1(49)

Zeferino, J., Cunha, M. C., & Antunes, A. P. (2012). Robust optimization approach to regional wastewater system planning. *Journal of Environmental Management*, *109*, 113–122. https://doi.org/10.1016/j.jenvman.2012.05.008

Zhang, X., & Buchberger, S. G. (2005). A new look at peaking factors. In *Proceedings of CCWI 2005 "Water Management for the 21st Century" Exeter, UK 5-7 Sept 2005*.