Article

Generating scenarios of cross-correlated demands for modelling water distribution networks

Roberto Magini [1], Maria Antonietta Boniforti[2], Roberto Guercio [3]

[1] Dept. of Civil, Building, and Environmental Engineering, Sapienza University of Rome, Via Eudossiana 18, 00184 Rome, Italy; roberto.magini@uniroma1.it
[2] Dept. of Civil, Building, and Environmental Engineering, Sapienza University of Rome, Via Eudossiana 18, 00184 Rome, Italy; antonietta.boniforti@uniroma1.it
[3] Dept. of Civil, Building, and Environmental Engineering, Sapienza University of Rome, Via Eudossiana 18, 00184 Rome, Italy; roberto.guercio@uniroma1.it
* Correspondence: roberto.magini@uniroma1.it ; Tel.: +39-06-44585-057

Abstract: This paper presents a methodology for the generation of a limited and representative number of water demand scenarios, taking into account the natural variability and spatial correlation of nodal consumption in a Water Distribution Network (WDN), and estimates their corresponding occurrence probabilities. Scaling laws are used to evaluate the statistics of water consumption at each node as a function of the number of users, considering the main statistical features of the unitary user's demand. Besides, consumption at each node is considered to follow a Gamma probability distribution. A high number of groups of cross-correlated demands, i.e. scenarios, for the entire network were generated using Latin Hypercube Sampling (LHS) and the numerical procedure proposed by Iman and Conover. The Kantorovich distance is used to reduce the number of scenarios and estimate their corresponding probabilities, while keeping the statistical information on nodal consumptions. By hydraulic simulation, the whole number of generated demand scenarios was used to obtain a corresponding number of pressure scenarios on which the same reduction procedure was applied. The probabilities of the reduced scenarios of pressure were compared with the corresponding probabilities of demand.

Keywords: uncertain water demand; scaling laws; scenario generation; scenario reduction; water distribution networks; hydraulic simulation

---

1. Introduction

The conventional modelling of WDNs normally is based on a deterministic approach, not merely with regard to the geometrical and hydraulic features, but also with respect to the demand loadings [1]. However, water demand, being influenced by many factors, i.e. type of users, socio-economic conditions, geographic location with its climate, seasonal fluctuation of weather, water fixtures technology, policies in water management, tariff, is subject to a natural variability. Surely, the variability of the demand represents the major source of uncertainty, which affects the overall reliability of the model for the assessment of the spatial and temporal distribution of pressure heads as well as for the evaluation of the water quality in the different pipes. Uncertainty produced by randomicity of demand assumes a different importance in relation to the spatial and temporal scales that are considered in modelling the network. Obviously, they become more and more relevant as the finer scales are reached, that is small groups of users and instantaneous demands are considered. As stated by Bargiela & Sterling [2], it is possible to obtain accurate predictions for the network as a whole, but estimating nodal consumptions for nodes where the population is low is more difficult.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 January 2019 doi:10.20944/preprints201901.0007.v2

Thus, considering and quantifying uncertainty of water demand could allow to associate an acceptable probability/level of risk to the results from hydraulic and optimization models of WDNs at different temporal and spatial scales: a more robust design and control of these systems can be realized, and obvious opportunities can arise in dimensioning pipes, formulating water balances, controlling system's components, identifying and quantifying leakages.

An approach in dealing with uncertainty of demand consists in explicitly considering different possible realizations of its value at the nodes of a WDN, i.e. different loading scenarios and associating to each of them a measure of their probability. In this way, for example, it is possible to look for a solution which be feasible and as close as possible to the optimum for all the scenarios: this scenario-based approach is known as robust optimization [3]. In the same way it is possible also to derive WDN reliability [4] or to localize leakages [5] or to map pressure-heads for a real-time control [6], all under uncertain demand conditions. Different possible scenarios, which include various aspects, such as peak flows, fire conditions at certain nodes, or pipe breakage, and the corresponding probabilities of occurrence could be obtained by consulting a panel of experts. However, this solution can have strong limitations in such a mathematically sensitive problem and can lead to arbitrary solutions.

At this aim this paper focuses on defining an objective methodology for the generation of demand scenarios. In the literature the issue of generating demand scenarios has been faced in [7] where uncertain future water consumptions are modeled using probability density functions (PDFs) assigned in the problem formulation phase. Scenarios with correlated nodal demands are generated using LHS and the procedure suggested by Iman and Conover [8]. All nodal demands follow a Gaussian PDF with coefficient of variation Cv = 0.10. The correlation coefficient between any two nodal demands is assumed equal to 0.50, as done by Tolson et al. [9]. In [10] different demand scenarios for WDN are obtained by making different combinations of demand values with specific return periods at each node of the network. The overall probability of each scenario is obtained through a Multivariate Normal Distribution (MVN). This implies that nodal peak demands are assumed to be jointly normal. The correlation between demand was found to significantly affect the occurrence probability of the demand scenarios. Also, Eck et Al. [11] generated demand scenarios considering a MVN distribution. They assumed a prior estimate of the mean values and covariance matrix of water demand from a preliminary analysis. An alternative approach for the estimation of scenario probabilities consists on the use of contingency tables [12]. The contingency tables approach is a non-parametric test in which the different random variables are divided in classes. The marginal probabilities are computed by estimating the frequency of occurrence of each class. The joint probabilities are computed by counting the occurrences of simultaneous classes. However, the estimation of the joint probabilities is a counting process, which is computationally demanding, especially for WDN with a high number of nodes.

Here an approach for the generation of a limited and representative number of demand scenarios is presented, which considers the natural variability and spatial correlation of residential consumptions, also estimating their corresponding occurrence probabilities. Scaling laws [1,13,14] are used to evaluate the statistics of aggregated demand at each node of the network, starting from the statistics of demand of unitary users. The Apulian WDN [15] is considered as a case-study and two different hypotheses were made for the number of users at its nodes. For each of these hypotheses a large number of groups of cross-correlated demands, i.e. scenarios, was generated using LHS from Gamma PDFs and a procedure based on the approach proposed by Iman and Conover [8]. The Kantorovich distance [16] is used to reduce the number of scenarios and estimate their corresponding probabilities. In this way, the limited number of network demand scenarios maintain the statistical information on the demand and the effects on the performance of WDN. In the design and control of

WDNs, demand scenarios are mostly functional to the evaluation of service conditions and specifically pressure-heads in the nodes. In order to evaluate how uncertainty of demand is conveyed to pressure-head field a demand-driven hydraulic model was also applied to solve Apulian network. Many pressure scenarios were obtained, and the same reduction procedure was employed. The probabilities of the reduced scenarios of pressure were compared with the corresponding probabilities of the demand.

2. Description of methodology

2.1. Scaling laws

The first step in the developed approach for the generation of demand scenarios is to characterize total demand at each node of a WDN, i.e., to determine its statistical parameters and probability distributions. Then, it is assumed that the number of households at each node is known beforehand, as well as the demand characteristics of the typical unitary household. Here it can be briefly referred that the unitary user' demand can be obtained either by monitoring water consumption of different types of users, or numerically by descriptive models of water demand such as, for example, End-Uses models [17] or PRP models preserving correlation [18].

The statistics of demand at each node can be obtained using the scaling laws developed in [1,13], which are briefly explained hereafter. Consider a network with $i = 1,2,3,\dots,N$ nodes, with $n_i = n_1, n_2, n_3, \dots, n_N$ households at each node. Assuming that in a time interval T the demand of the typical unitary household of the network is described by an ergodic stationary stochastic process, with mean μ_1, variance σ_1^2 and cross-correlation coefficient between each couple of single-user demands ρ_1, then the nodal demands, which are the aggregated demands of all the users at each node, are finite realizations of a pooled process having the corresponding statistics dependent on the number of households in the node. The expected value for the mean of the pooled process at the *ith* node is given by:

$$E[\mu_{n_i}] = n_i \cdot E[\mu_i], \tag{1}$$

and the expected value for the variance at the same node, neglecting the bias that can be caused when using small the demand series (short observation periods) [13], is given by:

$$E[\sigma_{n_i}^2] = n_i{}^2 \cdot \rho_1 \cdot \sigma_1^2 + n_i \cdot [1 - \rho_1] \cdot \sigma_1^2 \tag{2}$$

Equation (2) shows that the expected value of the variance of the aggregated process depends on ρ_1 : if demands are perfectly correlated in space, i.e. ρ_1 is equal to one, equation (2) is simplified into:

$$E[\sigma_{n_i}^2] = n_i{}^2 \cdot \sigma_1^2 \tag{2a}$$

if demands are uncorrelated in space, i.e. ρ_1 is equal to zero, equation [2] is simplified into:

$$E[\sigma_{n_i}^2] = n_i \cdot \sigma_1^2 \tag{2b}$$

For partially correlated demands a power law $E[\sigma_{n_i}^2] = n_i^{\alpha}\sigma_1^2$, with α scaling variance exponent, has been also derived [1,14]. Therefore, a complete statistical characterization of demand requires not only the definition of its mean and variance, but also the definition of the correlation between demands of each couple of users and groups of users. The cross-correlation refers to the similarity between demand patterns from different consumers or from different nodes. This parameter was

 doi:10.20944/preprints201901.0007.v2

proved to be not negligible [19] and to affect the hydraulic performance of a WDN as well as its cost to achieve a desired level of reliability: it was verified that higher cross correlations lead to higher pressure fluctuations, which have negative impacts on the reliability of the WDN [20]. Following the same assumption and notation as for the mean and variance, the expected value for the cross-correlation between all nodes of the network is represented by a N-by-N square matrix, whose elements are given by:

$$E\left[\rho_{n_i n_j}\right] = \frac{E\left[cov_{n_i n_j}\right]}{E\left[\sigma_{n_i}\right]\cdot E\left[\sigma_{n_j}\right]} = \frac{n_i n_j \rho_1}{\left[n_i(1+\rho_1(n_i-1))\right]^{1/2}\cdot\left[n_j\left(1+\rho_1(n_j-1)\right)\right]^{1/2}} \quad (3)$$

with $i,j = 1,2,3,\ldots,N$, and where, for example, $\rho_{n_1 n_2}$ is the cross-correlation coefficient between the demand of n_1 aggregated households at node 1, and the demand of n_2 aggregated households at node 2.

Through equations (1,2,3) the nodal demand statistics of the network and the correlation structure between them are entirely defined.

2.2. Generation of scenarios

In simulation and optimization problems several methods exist to cope with uncertainty. If we do not know exactly input data, in our case water demand at the nodes of a WDN, because they can assume different values and then many combinations of them are possible, we are dealing with scenarios. But, if the statistical features of uncertain data are known, numerical solutions can be obtained by approximating the probability distribution function with discrete distributions having a finite number of outputs, again referred to as scenarios.

Then, the second phase of the present approach consists in the generation of a large number of water demand scenarios for a WDN, based on the statistics estimated by the scaling laws and making the hypothesis of Gamma-distributed water demand at each node. The knowledge of the scaling laws of the statistical moments and the type of the probability distributions of water demand in relation to the number of users, prove to be a useful tool to face the inherent uncertainty of demand and in particular to address the optimization problems. Using Gamma distribution for demand is supported, when the number of aggregated users is high enough or time resolution is greater than five minutes, by measurements in Latina case-study [1]. Furthermore, in a recent work Kossieris and Makropoulos [21] investigated the performance of ten probabilistic models showing that both Gamma and Weibull distributions can be used to adequately describe the nonzero water demand recorded at different time scales.

Scenarios can be generated following different methods: by matching a small set of statistical properties, e.g. moments [22,23] or simulating some defined mathematical process (e.g. Brownian motion) or sampling from known distributions [24]. For scenarios with a large number of variables correlated in between, sampling from the joint distribution is not usual for the difficulty in defining the distribution itself. An alternative to specifying the joint distribution is to make use just of the marginal distributions and the linear or rank correlation matrix. If nothing is known about the form of the joint distribution a coupling procedure can be used: in this case the generated scenarios will respect an arbitrary dependency structure based on the procedure followed.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 January 2019 doi:10.20944/preprints201901.0007.v2

Each demand scenario D_u is defined here as a set or combination of nodal demand values occurring simultaneously in the WDS. It can be represented by the N dimensional vector:

$$D_u = [d_{1,u}, d_{2,u}, \dots d_{N,u}], \qquad \forall\, i = 1,2,\dots,N \;\; and \;\; u = 1,2,\dots,S \tag{4}$$

where $u = 1,2,\dots,S$ is the index identifying the different scenarios, and $d_{i,u}$ is the demand at node i for the uth scenario and D_u depicts a one-dimensional stochastic data process.

A sampling method from the marginal distribution based on the LHS and the Iman Conover approach [8] is followed. The restricted pairing technique by Iman and Conover induces rank correlation between the given marginals by shuffling finite-size samples from each of them. The appropriate shuffle is determined by ranking the input samples the same as in a reference sample with the desired rank correlation. The complete procedure, that will be described in the following, is quite straightforward because it requires only the Cholesky decomposition, some matrix algebra, and the final rearrangement of the original uncorrelated sample

2.2.1. Procedure for generating scenarios

Step 1. Create a random (*S, N)* dimensional matrix **Z*,** containing S Latin Hypercube Samples of size *N* from a standardized normal distribution, where *S* is the number of scenarios and *N* the number of the demand nodes in WDN. For this purpose the Matlab function *lhsnorm* was used. Owing to the finite size of the samples their correlation matrix **R*** does not coincide with the identity matrix **I**, that is they are not independent. Then, the lower triangular Cholesky decomposition is applied to induce the desired correlation. Specifically:

$$\mathbf{I} = \mathbf{C} \cdot \mathbf{C}^{\mathbf{T}}$$
$$\mathbf{E} = \mathbf{R}^{*} \cdot \mathbf{R}^{*\mathbf{T}}$$
$$\mathbf{Z} = \mathbf{Z}^{*} \cdot \mathbf{C} \cdot \mathbf{E}^{-1}$$

and the (*S, N*) dimensional matrix **Z** of perfectly independent *S* samples of size *N* from a standardized normal distribution is obtained. In order to obtain the Cholesky root the Matlab function *chol* was used.

Step 2. Create a random (*S, N*) dimensional matrix containing *S* standardized normal samples with the correlation matrix **Corr** from the scaling laws for nodal demand. To this aim the desired correlation is induced in **Z** also applying the lower triangular Cholesky decomposition, that is:

$$\mathbf{Corr} = \mathbf{P} \cdot \mathbf{P}^{\mathbf{T}}$$
$$\mathbf{E} = \mathbf{R}^{*} \cdot \mathbf{R}^{*\mathbf{T}}$$
$$\mathbf{G} = \mathbf{Z} \cdot \mathbf{P} \cdot \mathbf{C}^{-1}$$

Step 3. Transform matrix **G** in the (*S, N*) dimensional matrix **D*** complying with the desired marginal distributions at each demand node. Transformation is based on the inverse Cumulative Distribution Function, CDF, of the desired marginals, F_i. Specifically, for N non-normal random samples **D*** = [$\mathbf{D}^*_1$, $\mathbf{D}^*_2$,…,$\mathbf{D}^*_i$] , 1 ≤ i ≤ N, with desired CDF, the following equation holds:

$$D_i^* = F_i^{-1}\left(\Phi(G_i)\right) \tag{5}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 January 2019 doi:10.20944/preprints201901.0007.v2

where $\Phi(G_i)$ is the CDF of the i*th* samples of G and is uniformly distributed. $\Phi(G_i)$ can also be interpreted as a realization from the Gaussian copula. Applying the inverse CDF F_i^{-1} to the uniform random variable $\Phi(G_i)$ ensures that D_i^* is distributed according to Φ_i. Unfortunately, transformation in Equation 1 is non-linear, therefore the correlation matrix **Corr*** of **D*** is no equal to the desired correlation matrix **Corr**.

Step 4. Apply the Iman-Conover algorithm proposed by Ekström [25] in order to get a better approximation of the desired correlation matrix **Corr** for the (*S*, *N*) matrix of nodal demand scenarios **D***. The algorithm is described in the following steps:

4.1 Calculate lower triangular Cholesky decomposition **V** of **Corr**, i.e. **Corr** = **V**·**V**^{T} .
4.2 Calculate lower triangular Cholesky decomposition **Q** of **Corr***, i.e. **Corr*** = **Q**·**Q**^{T}.
4.3 Obtain **T** such that **Corr** = **T**·**Corr**·**T**^{T} , can be calculated as **T** = **V**·**Q**$^{-1}$.
4.4 Obtain the matrix **ScoreD*** by rank-transforming **D** and convert to van der Waerden scores defined as $F_i^{-1}\left(\Phi(\mathrm{i}/(\mathrm{N}+1)\right)$where ϕ is the CDF of the standard normal distribution, i is the assigned rank and N is the total number of samples.
4.5 Calculate the target scores matrix **ScoreD** = **ScoreD***·**T**^{T}.
4.6 Match up the rank pairing in **D*** according to **ScoreD,** obtaining the new (*S*,*N*) dimensional matrix **D** containing the *S* scenarios of the *N* nodal demand in the WDS. The N samples are distributed according to the desired marginals and their correlation matrix is close to the correlation matrix derived from the scaling laws.

2.3. Scenario reduction

With the scenario generation procedure, we obtain a great number of pictures each of which represents a single snapshot of the whole water demand in the network. The higher the number of scenarios generated the better is the description of the variability of water demand in the WDS. But it is not possible to manage such a large number of scenarios to deal with stochastic or robust optimization problems. Moreover, the probability associated with each of them is not very significant. We have to reduce the scenarios and at the same time determine, for the reduced scenarios, a significant weight representative of their possibility to be realized. Then, the goal of scenario reduction is to approximate the discrete distribution of the generated scenarios with another discrete distribution having fewer elements. At this point, the choice of the number of scenarios becomes a critical step in obtaining meaningful solutions taking into account the system performance and the robustness of the solution to variations in the uncertain data.

It is assumed that the probability distribution $\mathbb{P}$ of the N-dimensional stochastic data process is approximately given by many scenarios

$$D_u = \left[d_{1,u}, d_{2,u}, \dots d_{N,u}\right],\ \forall\, i = 1, 2, \dots, N\ \ and\ \ u = 1, 2, \dots, S \tag{6}$$

to which the probabilities p_u are associated and $\sum_{u=1}^{S} p_u = 1$.

In order to approximate the probability distribution $\mathbb{P}$ with another $\mathbb{Q}$ distribution, with a smaller number of elements, so that $\mathbb{Q}$ will be as close as possible to $\mathbb{P}$ in terms of probabilistic distance, we use the Kantorovich distance, K, which is the most common probability distance used in stochastic optimization. It is defined between two probability distributions $\mathbb{P}$ and $\mathbb{Q}$ and represent

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 January 2019 doi:10.20944/preprints201901.0007.v2

the optimal value to a linear transportation problem [26]. In the case where $\mathbb{P}$ and $\mathbb{Q}$ are finite distributions, the Kantorovich distance is obtained by solving the following problem (see [24])

$$K(\mathbb{P},\mathbb{Q}) = inf\left\{\sum_{u=1}^{S}\sum_{w=1}^{\tilde{S}}\eta_{uw}c_N(d^u,\tilde{d}^w):\ \eta_{uw}\geq 0, \sum_{u=1}^{S}\eta_{uw}=q_w, \sum_{w=1}^{\tilde{S}}\eta_{uw}=p_u\ \forall u,\forall w\right\} \qquad (7)$$

where $c_N(d^u,\tilde{d}^w) := \sum_{\tau=1}^{i}|d_\tau^u - \tilde{d}_\tau^w|$, $i = 1,\dots,N$ with $|.|$ some norm on $\Re^n$, with w=1,..,$\tilde{S}$, and ,$\tilde{S}$ is the number of reduced scenarios. The cost function c_N measures the distance between scenarios, p_u and q_w are respectively the probabilities of scenarios in $\mathbb{P}$ and in $\mathbb{Q}$.

The previous problem is not easy to solve [27] and in order to overcome to these difficulty heuristic algorithms have been developed, in particular fast backward and fast forward strategies have been implemented. In this paper, we make use of the forward selection algorithm. It defines an iterative process which starts with an empty set. At each iteration, from the set of the non-selected scenarios, the scenario minimizing the Kantorovich distance between the reduced and original sets is selected and inserted in a reduced set . The optimal selection of a single scenario may be repeated recursively until the prescribed number S of elements is reached.

Actually, the forward selection algorithm does not guarantee that the reduced set of scenarios is the closest in the Kantorovich distance with respect to the original set and represents the optimal solution of the original problem described by equation (7). However, the empirical results described in the literature [28] indicate that the forward selection algorithm works well in practice.

2.3.1. Procedure for reducing scenarios

In this paper a scenarios reduction algorithm based on Kantorovich distance was used. In particular, the fast forward selection algorithm as described in [26] was applied. Starting from an empty set, an iterative process was followed until the required number of selected scenarios was reached. Below is a brief description of this methodology.

First, the high number of generated demand scenarios at each node have been assumed to be equiprobable. Thus, at each iteration, using the Euclidean norm ℓ^2, the distances between all the possible pairs of scenarios were calculated for each node. Then, by summing the corresponding distances for all the different nodes, a cost function matrix c_N was derived. This function allows to evaluate the Kantorovich distances matrix between pairs of scenarios taking into account their probability of occurrence.

The scenario corresponding to the minimum value of the Kantorovich distance is then selected and the cost matrix is updated by replacing each element with the minimum value between the original element and the one corresponding to the selected scenario. At this point, the procedure is repeated, and a new scenario is added to the reduced scenario set until the number of requested scenarios is reached. In the end, an optimal redistribution of probabilities was carried out by adding the probabilities of non-selected scenarios to the probabilities of those in the reduced set, that is the probability of each non-selected scenario was summed to the probability of the closest selected scenario according to the cost function.

Therefore, according to equation, the new probability of a preserved scenario is equal to the sum of its former probability and all the probabilities of the deleted scenarios that are closest to it according to c_N. Obviously, the deleted scenarios have probability zero.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 January 2019 doi:10.20944/preprints201901.0007.v2

3. Application example

3.1 Theoretical WDN: the Apulian network

For a better understanding of the developed approach an application using the Apulian network layout [15] is elaborated as an example. The Apulian network layout comprises 1 reservoir, 23 demand nodes and 34 pipes. (Figure 1). The network is used only with the objective of defining a scene and providing visual help to the application example.

The geometrical features of the WDN are summarized in Table.1.

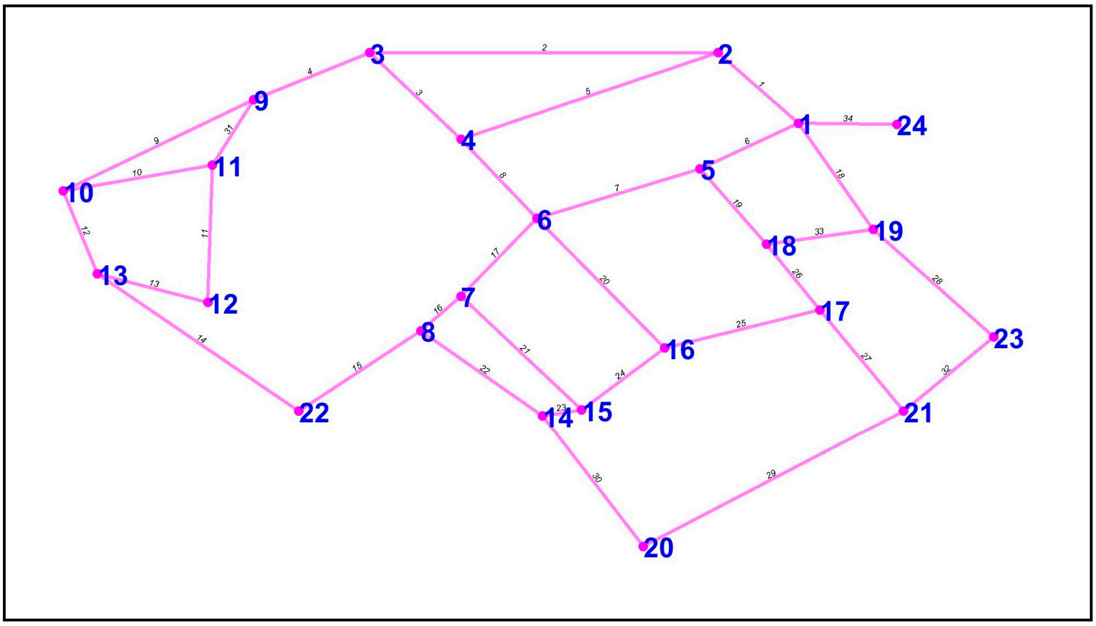


Figure 1. Apulian Network layout.

The data from two-years long water demand measurements of 82 single household users in the town of Latina, Italy [1] were considered for the definition of the statistical parameters of a typical residential consumption and the calibration of the scaling laws. Table 3 summarizes the average water demand of each household unit and its relevant statistics at peak hour (7-8 am), considering a five-minute time step in data monitoring. The users were considered all the same type and for this reason the same statistical parameters were employed. The number of users at each node of the WDN is listed in Table 1.

Regarding the correlation between pairs of single households a very low value of the Pearson coefficient, i.e. $\varrho = 0.0043$, was considered, in agreement with most of the experimental data from the case study of Latina. Table 3 also shows the number of household units for each water demand node. In order to highlight how the number of users per node and their relationships affect the generated demand scenarios two different frameworks were examined to which correspond respectively *DemandA* and *DemandB* column. The *DemandA* values are the same used by Giustolisi et Al. [29] for Apulian network and the users' number is consequent. Differently, *DemandB* values were defined assuming a smaller total number of users and a greater variability of the number of users per node.

 doi:10.20944/preprints201901.0007.v2

330 **Table 1.** WDN geometrical features, number of users and water demand at nodes

PIPES											
Pipe number	**Start Node**	**End Node**	**Length** *(m)*	**C Hazen Williams**	**D** *(m)*	**Pipe number**	**Start Node**	**End Node**	**Length** *(m)*	**C Hazen Williams**	**D** *(m)*
1	1	2	348.5	100	0.327	**18**	1	19	583.9	100	0.164
2	2	3	955.7	100	0.29	**19**	5	18	452	100	0.229
3	3	4	483	100	0.1	**20**	6	16	794.7	100	0.1
4	3	9	400.7	100	0.29	**21**	7	15	717.7	100	0.1
5	2	4	791.9	100	0.1	**22**	8	14	655.6	100	0.258
6	1	5	404.4	100	0.368	**23**	15	14	165.5	100	0.1
7	5	6	390.6	100	0.327	**24**	16	15	252.1	100	0.1
8	6	4	482.3	100	0.1	**25**	17	16	331.5	100	0.1
9	9	10	934.4	100	0.1	**26**	18	17	500	100	0.204
10	11	10	431.3	100	0.184	**27**	17	21	579.9	100	0.164
11	11	12	513.1	100	0.1	**28**	19	23	842.8	100	0.1
12	10	13	428.4	100	0.184	**29**	21	20	792.6	100	0.1
13	12	13	419	100	0.1	**30**	20	14	846.3	100	0.184
14	22	13	1023.1	100	0.1	**31**	9	11	164	100	0.258
15	8	22	455.1	100	0.164	**32**	23	21	427.9	100	0.1
16	7	8	182.6	100	0.29	**33**	19	18	379.2	100	0.1
17	6	7	221.3	100	0.29	**34**	24	1	158.2	100	0.368

331

NODES					
node ID	**elevation** *(m)*	**users A**	***DemandA*** *(l/s)*	**users B**	***DemandB*** *(l/s)*
1	6.4	932	10.86	155	1.8
2	7	1461	17.03	427	4.98
3	6	1282	14.95	48	0.56
4	8.4	1224	14.28	129	1.5
5	7.4	869	10.13	1270	14.81
6	9	1316	15.35	486	5.67
7	9.1	782	9.11	63	0.73
8	9.5	901	10.51	766	8.93
9	8.4	1045	12.18	63	0.73
10	10.5	1249	14.57	45	0.52
11	9.6	848	9.88	964	11.24
12	11.7	650	7.58	354	4.12
13	12.3	1303	15.2	122	1.42
14	10.6	1162	13.55	185	2.15
15	10.1	791	9.23	81	0.94
16	9.5	960	11.2	55	0.64
17	10.2	984	11.47	968	11.29
18	9.6	928	10.82	55	0.64
19	9.1	1258	14.68	45	0.52
20	13.9	1142	13.32	416	4.85
21	11.1	1255	14.63	567	6.61
22	11.4	1030	12.01	137	1.59
23	10	886	10.33	920	10.73
24(Reservoir)	36.4	24258	282.86	8321	96.97

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 January 2019 doi:10.20944/preprints201901.0007.v2

Table 2. Relevant statistics and parameters derived from the consumption data of Latina.

Statistical parameter	value
$\mu_{average}$ (L/min)	0.365
$\mu_{peak\ hour}$ (L/min)	0.700
$\sigma_{peak\ hour}$ (L/min)	0.870
scaling law exponent α	1.230

3.2 Generation of demand scenarios

The methodology presented in this work was applied to generate scenarios of contemporary water demands in the supply nodes of the Apulian distribution network. Two different hypotheses have been made on the number of users at the demand nodes. In the first one, *DemandA*, the number of users was determined in order to obtain the demand values assumed by Giustolisi et al. [29], which are appropriate for the subsequent hydraulic simulation of the network. Instead, the second hypothesis, *DemandB*, considers a lower number of total users, but, above all, considerably differentiates the number of users at each node. This was done with the aim of highlighting how the correlation matrix obtained from the scaling laws is influenced by the number of users in the nodes and by their mutual relations, and the proposed methodology can manage complex scenarios. The statistical parameters describing the unitary user' demand are obtained from the experimental data of a group of users in the case study of Latina [1]. Data are referred to peak hour and their sampling interval is equal to five minutes.

3.2.1 DemandA

Twelve-thousand demand scenarios were generated using the statistics estimated by the scaling laws and making the hypothesis of Gamma-distribution at each node, Figure 2. The large number of scenarios makes it difficult to distinguish them. But above all, nothing can be said about their probability.

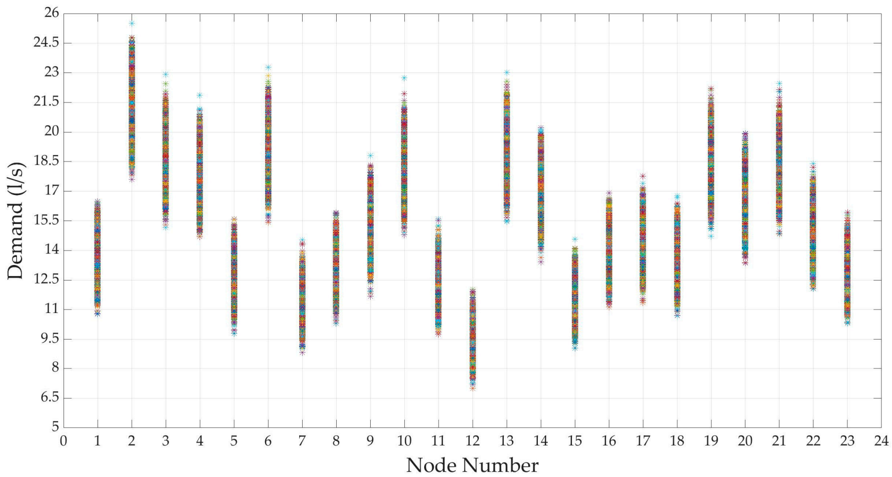


Figure 2. Generated demand scenarios, *DemandA*.

Gamma distribution proves to be the best in fitting the generated data in all nodes of the WDN, Figure 3. The scale and shape parameters of the input distributions $\Gamma(a,b)$, estimated by the scaling laws, well agree with the corresponding parameters of the output ones, Table 3.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 January 2019 doi:10.20944/preprints201901.0007.v2

Regarding the correlation matrix of the generated scenarios, it almost perfectly matches with the input correlation matrix obtained from the scaling laws. Table 4 compares the minimum, average and maximum value of the input and output correlation matrices.

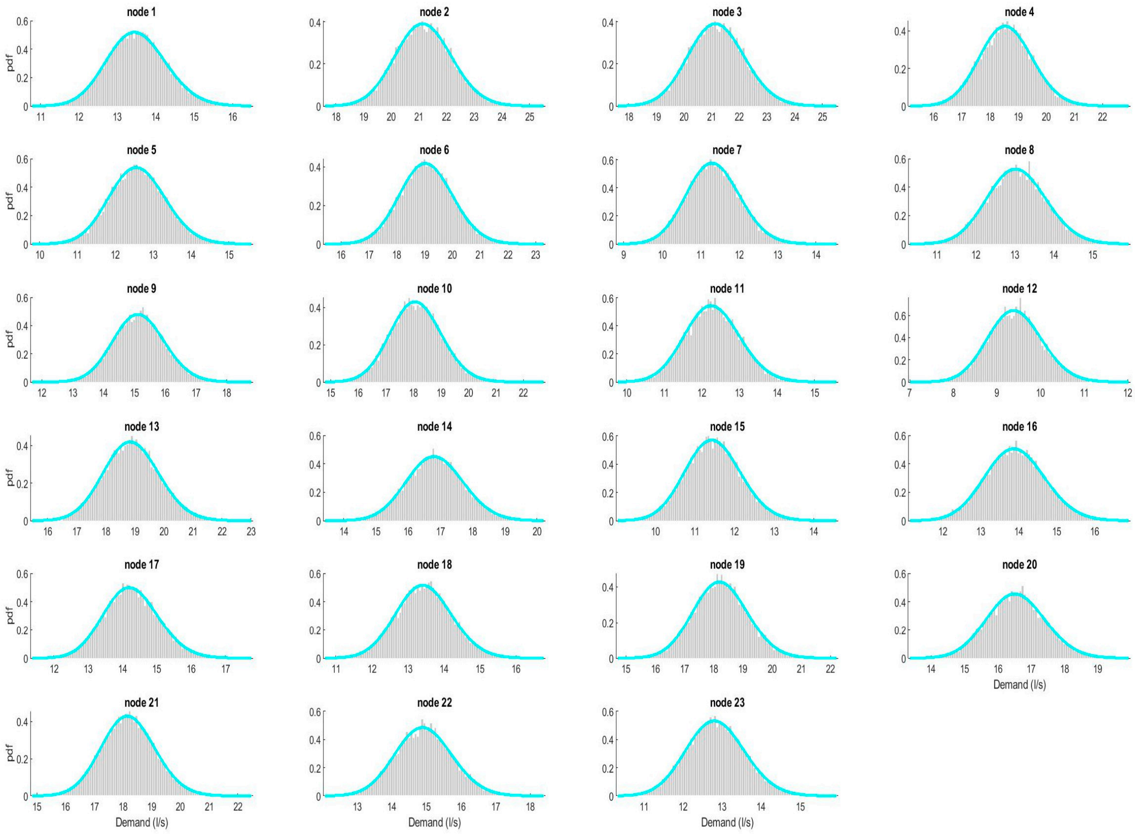


Figure 3. Gamma pdf distributions (cyan line) fitting demand data (grey bar), *DemandA*.

Table 3. Parameters *a* and *b* of Gamma distributions of input and output data, *DemandA*

	INPUT		OUTPUT			INPUT		OUTPUT	
Node ID	*a*	*b*	*a*	*b*	Node ID	*a*	*b*	*a*	*b*
1	298.74	0.0452	306.93	0.044	**13**	386.68	0.0489	393.31	0.048
2	422.3	0.0502	429.7	0.0493	**14**	354.04	0.0476	362.01	0.0465
3	381.87	0.0487	391.91	0.0474	**15**	263.29	0.0436	267.61	0.0429
4	368.5	0.0482	373.56	0.0475	**16**	305.63	0.0455	310.37	0.0448
5	283.07	0.0445	287.49	0.0438	**17**	311.49	0.0458	317.8	0.0449
6	389.65	0.049	396.84	0.0481	**18**	297.75	0.0452	302.28	0.0445
7	260.98	0.0434	266.95	0.0425	**19**	376.35	0.0485	382.78	0.0477
8	291.06	0.0449	296.87	0.044	**20**	349.34	0.0474	356.56	0.0464
9	326.26	0.0464	330.6	0.0459	**21**	375.66	0.0484	381.25	0.0477
10	374.28	0.0484	381.03	0.0475	**22**	322.65	0.0463	328.8	0.0454
11	277.78	0.0443	280.73	0.0438	**23**	287.32	0.0447	292.85	0.0439
12	226.35	0.0416	230.29	0.0409	**-**	-	-	-	-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 January 2019 doi:10.20944/preprints201901.0007.v2

Table 4. Min, average, max values of input and output cross-correlation matrix, demand A

	ρ_1=0.0043	
	ρ input correlation matrix (scaling laws)	ρ output correlation matrix (scenarios)
min	0.7442	0.7542
average	0.8146	0.8147
max	0.8567	0.8568

3.2.2 DemandB

Also, in this case twelve-thousand scenarios were generated using the statistics estimated by the scaling laws, considering Gamma-distributed demands at each node, Figure 4. Differently from *DemandA* case, the value of generated data shows a great excursion between node and node due to the large variability in the number of users.

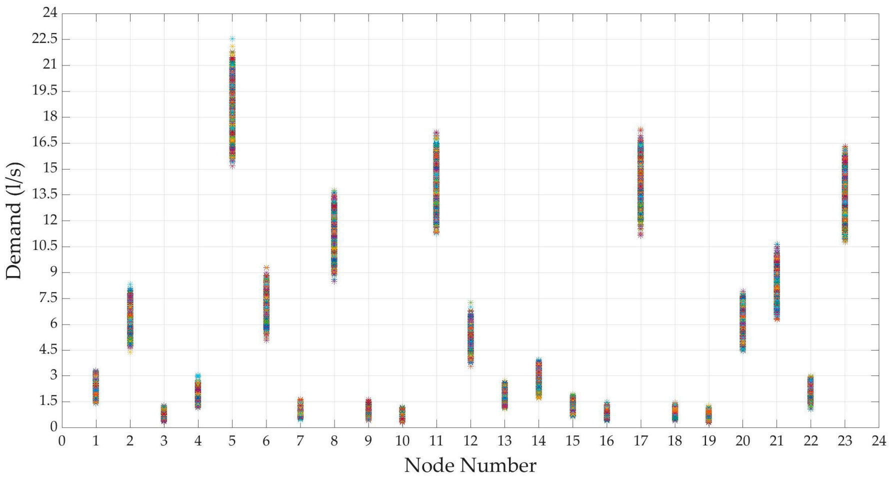


Figure 4. Generated demand scenarios, *DemandB.*

Similarly, to the previous case, Gamma distribution proves to be the best in fitting the generated data, Figure 5. Also, the input distributions parameters, estimated by the scaling laws, well agree with the corresponding parameters in output distributions, Table 5. In this case the correlation matrix of the generated scenarios, it almost perfectly matches with the input correlation matrix obtained from the scaling laws. Table 6 compares the minimum, average and maximum value of the input and output correlation matrices. The input and output correlation matrices are almost identical, but it should be noted that the correlation coefficient has very low values when considering node pairs with low number of users, intermediate values when one of the two has many users, higher values when the number of users is high for both. This responds to the fact that the correlation coefficient depends on the product of the number of users of the two considered nodes, see eq.3.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 January 2019 doi:10.20944/preprints201901.0007.v2

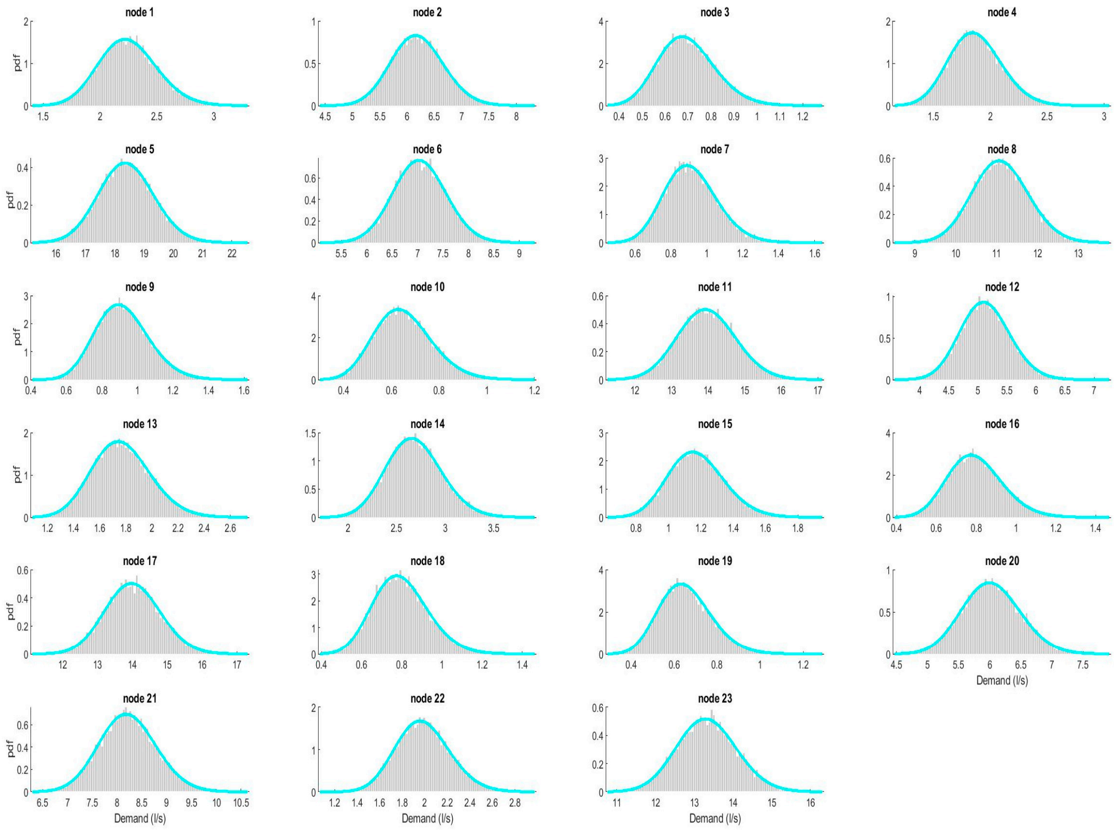


Table 5. Gamma PDFs parameters *a* and *b* input and output data at each node, *DemandB*.

	INPUT		OUTPUT			INPUT		OUTPUT	
Node ID	*a*	*b*	*a*	*b*	Node ID	*a*	*b*	*a*	*b*
1	75.06	0.0299	77.11	0.0291	**13**	62.42	0.0283	62.23	0.0284
2	163.78	0.0378	164.44	0.0376	**14**	86.01	0.0312	87.45	0.0307
3	30.44	0.0229	31.44	0.0221	**15**	45.54	0.0258	45.55	0.0258
4	65.16	0.0287	64.28	0.0291	**16**	33.80	0.0236	33.57	0.0237
	379.12	0.0486	381.21	0.0483	**17**	307.59	0.0456	310.45	0.0452
6	180.95	0.0389	181.88	0.0387	**18**	33.80	0.0236	33.56	0.0238
7	37.53	0.0243	38.28	0.0238	**19**	28.96	0.0225	28.81	0.0227
8	256.86	0.0432	259.31	0.0428	**20**	160.53	0.0376	162.17	0.0372
9	37.53	0.0243	36.95	0.0248	**21**	203.75	0.0404	203.55	0.0404
10	28.96	0.0225	28.92	0.0225	**22**	68.25	0.0291	68.70	0.0289
11	306.61	0.0456	306.60	0.0456	**23**	295.77	0.0451	298.27	0.0447
12	141.77	0.0362	142.72	0.0359	**-**	-	-	-	-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 January 2019 doi:10.20944/preprints201901.0007.v2

Table 6. Min, average, max values of input and output cross-correlation matrix, demand B

	ϱ1=0.0043	
	ϱ input correlation matrix (scaling laws)	ϱ output correlation matrix (scenarios)
min	0.1627	0.1627
average	0.4329	0.433
max	0.8261	0.8261

3.3 Reduction of demand scenarios

The final phase of the procedure consists in reducing the number of scenarios by aggregating them in relation to the distance of Kantorovich. For the application of the method the Euclidean norm was considered here. A reduced number of scenarios equal to 20 was chosen. In general, the choice of the number of scenarios should be based on the requirements of robust optimization problems and on the need for the reduced set to continue to describe the whole probability distribution of the demand at each node of WDN.

3.3.1 DemandA

In Figure 6 the reduced demand scenarios are represented. The 'mean scenario', i.e. the one defined by the average nodal values of nodal demand is plotted with black points. The red points indicate the most probable scenario, as reported in Figure 7. Mean and most probable scenarios, actually, do not coincide but are very close.

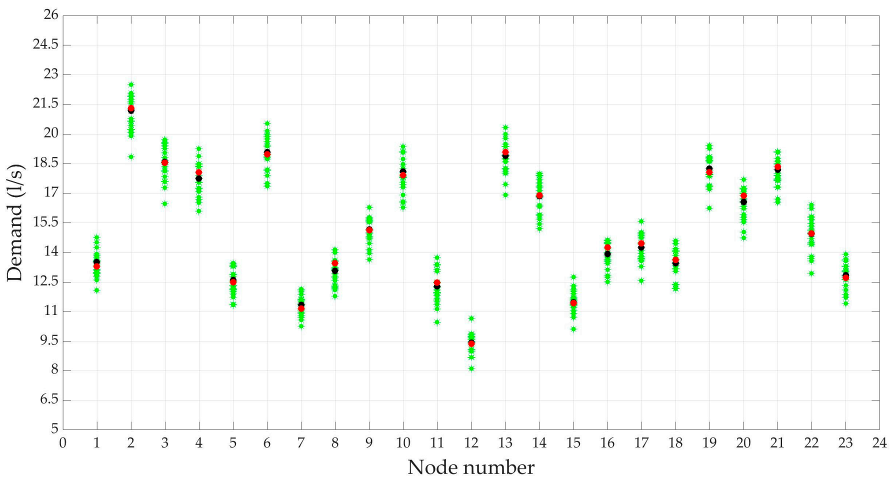


Figure 6. Reduced demand scenarios, *DemandA*.

- reduced nodal scenarios -> green dot
- average nodal values -> black dot
- scenario with max probability -> red dot

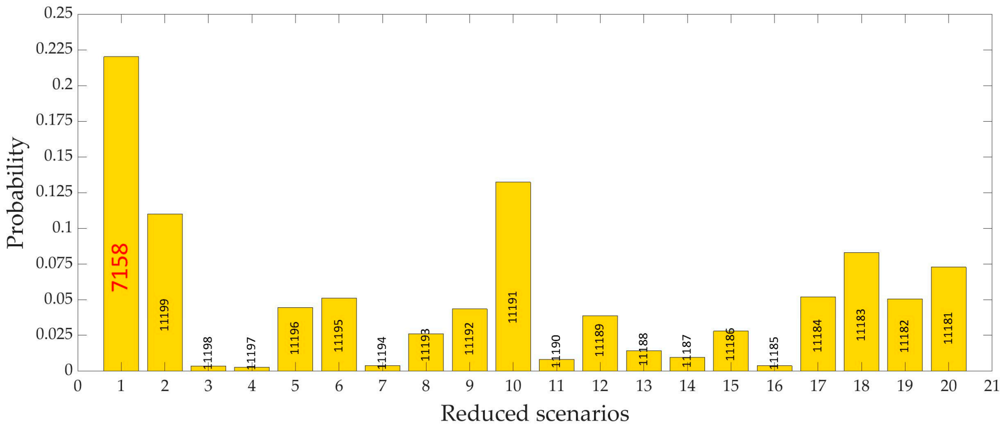


Figure 7. Reduced demand scenarios' probabilities, *DemandA*. Inside the bar is the number of the rank order of the generated scenario.

The most probable scenario, number 7158 in the set of generated scenarios, show a non-negligible value of its probability equal to 0.22. Only six scenarios have a probability value greater than 0.05.

3.3.2 DemandB

In Figure 7 the reduced demand scenarios are represented. They show a greater similarity compared to the previous case, perhaps owing to the very low number of users in some node and the great difference between some pairs of nodal mean water demand. Also, here there is no coincidence between the mean and the most probable scenario, which is number 3873 in the set of generated scenarios. In this case nine scenarios show a probability values greater than 0.05.

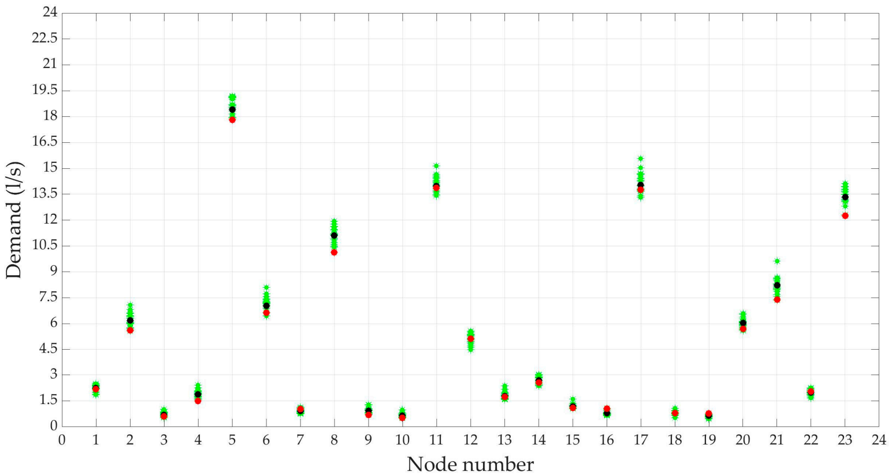


Figure 8. Reduced demand scenarios, *DemandB*.

- reduced nodal scenarios -> green dot
- average nodal values -> black dot
- scenario with max probability -> red dot

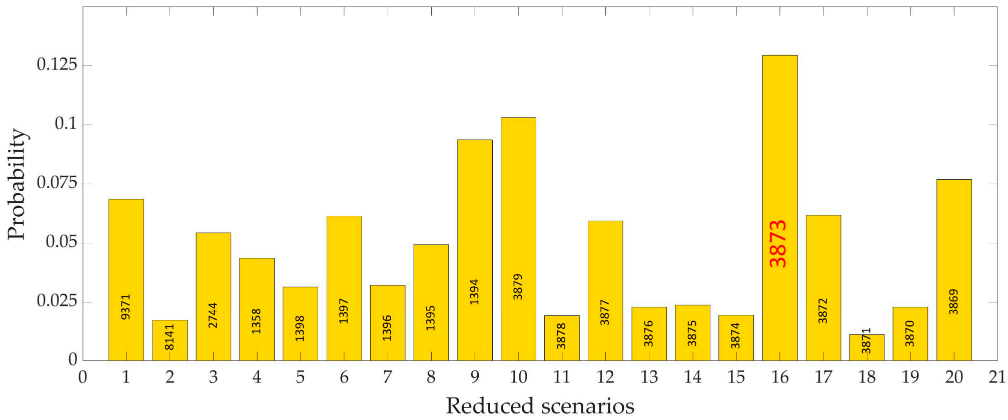
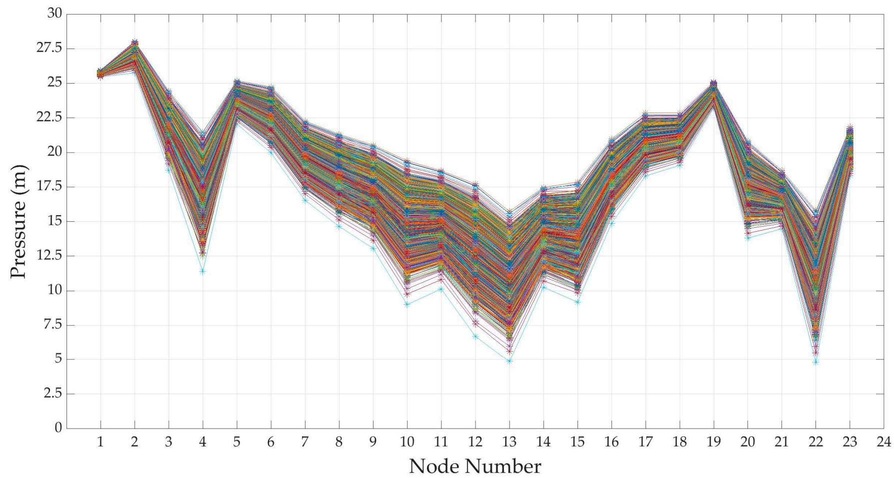


Figure 9. Reduced demand scenarios' probabilities, demand B. Inside the bar is the number of the rank order of the generated scenario.

3.4 Hydraulic simulation with scenarios from demand A

Water demand is the forcing parameter of WDN and its natural variability is reflected on the variability of the quantities describing the hydraulic behavior of the whole system. The following question arises: how uncertainty of water demand determines uncertainty of nodal pressure-heads and pipe flow-rate in a WDN? Specifically: the most probable water demand scenario coincides with the most probable pressure-head scenario? At this aim a set of pressure scenarios was derived from the set of generated demand scenarios in the *DemandA* case using a demand-driven hydraulic model based on the Global-Gradient algorithm proposed by Todini and Pilati [30]. Twelve-thousand pressure scenario have been obtained for Apulian network, Figure 10. The reduction algorithm was also carried on and twenty reduced scenarios and corresponding probabilities were derived, Figures 11 and 12.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 January 2019 doi:10.20944/preprints201901.0007.v2

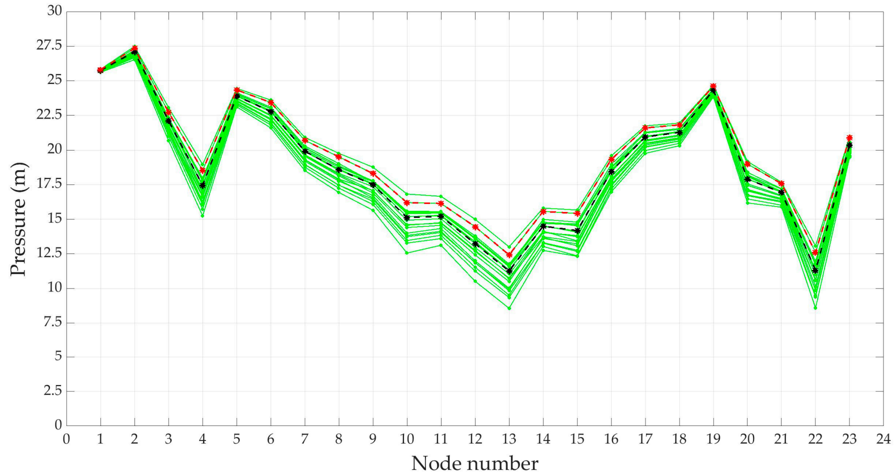


Figure 11. Reduced pressure headlines scenarios, demand A.

- average nodal values -> black dotted line
- scenario with max probability -> red dotted line

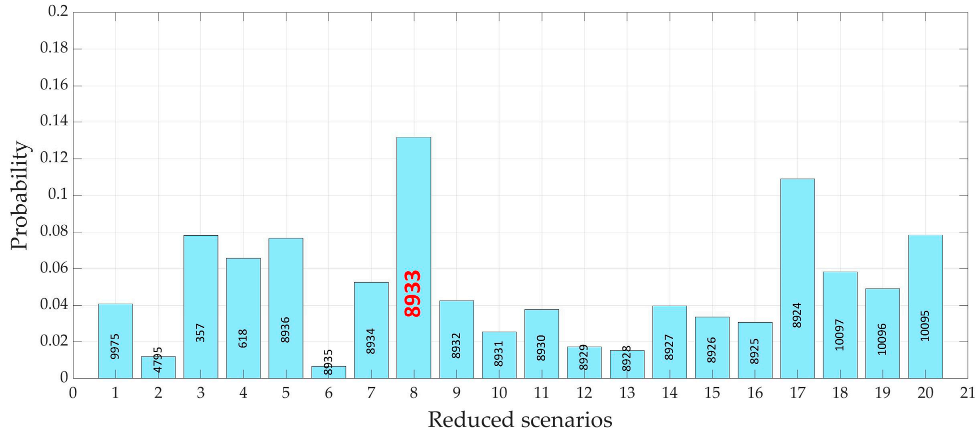


Figure 12. Reduced pressure scenarios' probabilities, demand A. Inside the bar is the number of the rank order of the generated demand scenario.

The results show how the twelve-thousand scenarios generated, unlike the corresponding ones of the water demand, show a low variability in their 'shape'. Their trend is strongly influenced by the geometrical and hydraulic characteristics of WDN. From the results it is also possible to identify the nodes most affected by flow variability and this could be important for designing a proper pressure-head monitoring system. Moreover, the most probable pressure headline is, in this case, higher than that produced by average water demand. Its number in the set of generated demand scenarios is 8933, that is, it does not coincide with the most probable in this set, Figure 13. Otherwise, it is possible to observe the coincidence between the average pressures in each node and the pressure values produced by the average demand scenario. The greatest value of probability is about 0.14 and nine scenarios exceed the probability value 0.05. The filtering effect on water demand by the WDN determines greater uniformity in the probability values of the reduced scenarios leaving uncertainty in estimating the most probable one.

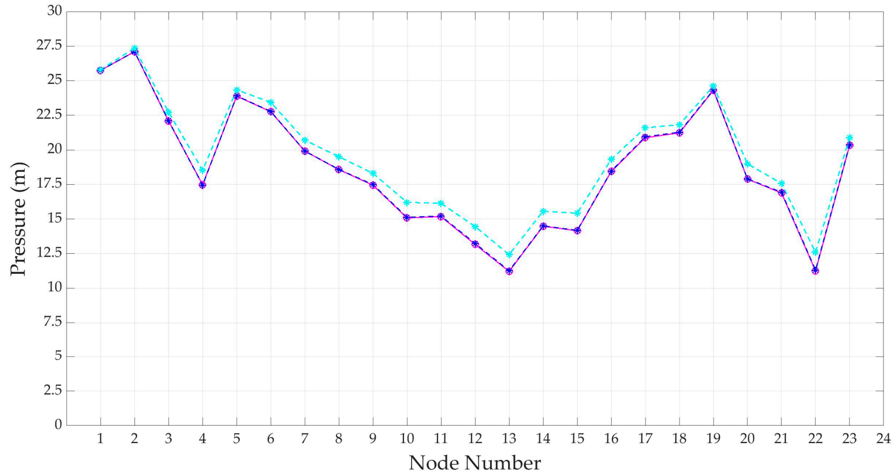


Figure 13. Pressure headlines:
- most probable demand scenario -> magenta line
- most probable pressure headline scenario -> continuous cyan line
- average pressure headline -> dotted blue line

5. Conclusions

This document proposes a complete procedure to generate and reduce the number of water demand scenarios able to adequately represent the uncertainty due to the variability of demand in WDN. Furthermore, with this procedure, an objective measure of their probability is associated to each of the reduced scenarios. Being able to determine demand scenarios for an entire distribution network and their corresponding probabilities of occurrence is a relevant step not only in the robust optimization problems but more generally in modelling WDN, both for their design and control. The proposed approach is based on the definition of the main statistical parameters and probability distributions of the water demand in each node of the network. In this regard, we highlight the importance of the scaling laws and the need to extend them to different types of use and user through further measurement campaigns or through the development of descriptive models of demand.

Acknowledgments: This research was supported by 'Progetti Ateneo Sapienza, small projects', protocol RP116154FD8D5BF3.

References

1. Magini, R., Pallavicini, I., Guercio, R., *Spatial and Temporal Scaling Properties of Water Demand*, Journal of Water Resources Planning and Management, 2008. **134**(5): 276–284, doi: 10.1061/(ASCE)0733-9496.
2. Bargiela, A., Sterling, M., *Adaptive forecasting of daily water demand*, in Comparative model for electrical load forecasting. D.W.Bunn and E.D.Farmer, eds.,1985. John Wiley and Sons, New York, N.Y.
3. Perelman, L., M. Housh, and A. Ostfeld, *Robust optimization for water distribution systems least cost design.* Water Resources Research, 2013. **49**(10): p. 6795-6809.
4. Xu, C.C. and I.C. Goulter, *Reliability-based optimal design of water distribution networks.* Journal of Water Resources Planning and Management 1999. **125**(6).
5. Steffelbauer, D., Neumayer, M., Günther, M., Fuchs-Hanusch, D., *Sensor Placement and Leakage Localization considering Demand Uncertainties,* 2014. Procedia Engineering **89**: p. 1160 – 1167.
6. Pallavicini, I., Magini, R., Guercio, R., *Assessing the spatial distribution of pressure head in municipal water networks,* 2005, CCWI 2005: Proceedings of the Eighth International Conference on Computing and Control for the Water Industry, Exeter, UK

7. Kapelan, Z.S.,Savic, D.A., and Walters, G.A., *Multiobjective design of water distribution systems under uncertainty*, Water Resources Research, **41**(11), W11407, doi:10.1029/2004WR003787, 2005.
8. Iman, R. L., and Conover, W. J., *A distribution-free approach to inducing rank correlation among input variables*, Commununications in Statistics, 1982, **11**: p. 311-334.
9. Tolson, B. A., H. R. Maier, A. R. Simpson, and B. J. Lence, *Genetic algorithms for reliability-based optimisation of water distribution systems*, Journal of Water Resources Planning and Management, 2004. **130**(1), 141(5): p. 63-72.
10. Vertommen, I., R. Magini, and M.d.C. Cunha, *Generating Water Demand Scenarios Using Scaling Laws.* Procedia Engineering, 2014. **70**: p. 1697-1706.
11. Eck B., Fusco F., Taheri N., *Scenario Generation for Network Optimization with Uncertain Demands*. Proceedings of World Environmental and Water Resources Congress, 2015. 17th Water Distribution Systems Analysis Symposium(**3**): p. 844 - 852.
12. Ridolfi, E., Vertommen, I., Magini, R., *Joint probabilities of demands on a water distribution network: A non-parametric approach.* AIP Conference Proceedings, 2013. **1558**(1): p. 1681-1684.
13. Vertommen, I., Magini R., Cunha M.C., Guercio R., *Water demand uncertainty: the scaling law approach*, in Water supply systems analysis: selected topics, Ostfeld, D.A., Editor, 2012, ISBN: 978-953-51-0889-4, InTech: p. 1-25.
14. Vertommen, I., Magini, R., and Cunha M., *Scaling Water Consumption Statistics*. Journal of Water Resources Planning and Management, 2015. **141**(5): p. 04014072.
15. Giustolisi, O., Kapelan, Z.,Savic,D., *Algorithm for Automatic Detection of Topological Changes in Water Distribution Networks.* Journal of Hydraulic Eng., 2008. **134**(4):435-446, doi: 10.1061/(ASCE)0733-9429(2008)134:4(435).
16. Vershik, A., *Kantorovich metric: Initial history and little-known applications*. Journal of Mathematical Sciences, 2006. **133**(4):1410–1417.
17. Blokker, E.J.M., J.H.G. Vreeburg, and J.C.v. Dijk, *Simulating Residential Water Demand with a Stochastic End-Use Model.* Journal of Water Resources Planning and Management, 2010. **136**(1): p. 19-26.
18. Creaco, E., Alvisi, S., Farmani, R., Vamvakeridou-Lyroudia, L., Franchini, M., Kapelan, Z.,Savic, D., *Preserving Duration-intensity Correlation on Synthetically Generated Water Demand Pulses.* Procedia Engineering, 2015, **119**, p. 1463-1472
19. Filion, Y.R., et al., *Cross Correlation Analysis of Residential Demand in the City of Milford, Ohio*, in *Water Distribution Systems Analysis Symposium 2006*. 2008, ASCE.
20. Filion, Y., B. Adams, and B. Karney, *Cross Correlation of Demands in Water Distribution Network Design.* Journal of Water Resources Planning and Management, 2007. **133**(2): p. 137-144.
21. Kossieris, P., Makropoulos, C., *Exploring the Statistical and Distributional Properties of Residential Water Demand at Fine Time Scales*, Water, 2018, **10**, 1481, doi:10.3390/w10101481.
22. Ponomareva, K., Roman, D., Date, P., *An algorithm for moment-matching scenario generation with application to financial portfolio optimization,* European Journal of Operational Research, 2015, **240** (3): p.678-687.
23. Magini, R., Capannolo, F., Ridolfi, E., and Guercio R., *Demand uncertainty in modelling WDS: scaling laws and scenario generation*, WIT Transactions on Ecology and The Environment, 2016, **210**, WIT Press, doi:10.2495/SDP160621.
24. Mitra, S., *Scenario Generation for Stochastic Programming*, 2006, White Paper Series, Optirisk Systems, Domain: Finance Reference Number: OPT 004.

 doi:10.20944/preprints201901.0007.v2

25. Ekstrom, P. A., *A simulation toolbox for sensitivity analysis*, Masters Degree Project, 2005, Faculty of Science and Technology, Uppsala Universitet.
26. Dupacová, J., N. Gröwe-kuska, and W. Römisch, *Scenario reduction in stochastic programming, An approach using probability metrics.* Math. Program. , 2002. **Ser. A 95**: p. 493-511.
27. Morales J.M., Pineda, S., Conejo, A.J., , Carrión, M., *Scenario Reduction for Future Market Trading in Electricity Markets*. IEEE Trasactions on Power Systems, 2009, **24**(2): p. 878-888.
28. Heitsch, H., Romisch W., M., *Scenario reduction in stochastic programming*, Computational Optimization and Applications, 2003, **24**: p.187-206.
29. Giustolisi, O., D. Laucelli, and A.F. Colombo, *Deterministic versus stochastic design of water distribution networks.* Journal of Water Resources Planning and Management, 2009. **135**(2).
30. Todini, E., Pilati, S. , *A gradient algorithm for the analysis of pipe networks.* Computer Applications in Water Supply, Research Studies Press, Letchworth, Hertfordshire, UK, 1988.