

Article

Bottom-Up Generation of Peak Demand Scenarios in Water Distribution Networks

Enrico Creaco ^{1,2,*} , Giacomo Galuppini ¹ , Alberto Campisano ³ and Marco Franchini ⁴ 

¹ Dipartimento di Ingegneria Civile ed Architettura, Università degli Studi di Pavia, Via Ferrata 3, 27100 Pavia, Italy; giacomo.galuppini01@universitadipavia.it

² School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, SA 5005, Australia

³ Dipartimento di Ingegneria Civile ed Architettura, Università degli Studi di Catania, Viale Andrea Doria 6, 95125 Catania CT, Italy; alberto.campisano@unict.it

⁴ Dipartimento di Ingegneria, Via Saragat 1, 44100 Ferrara, Italy; marco.franchini@unife.it

* Correspondence: creaco@unipv.it; Tel.: +39-(0)382-98-5317

Abstract: This paper presents a two-step methodology for the stochastic generation of snapshot peak demand scenarios in water distribution networks (WDNs), each of which is based on a single combination of demand values at WDN nodes. The methodology describes the hourly demand at both nodal and WDN scales through a beta probabilistic model, which is flexible enough to suit both small and large demand aggregations in terms of mean, standard deviation, and skewness. The first step of the methodology enables generating separately the peak demand samples at WDN nodes. Then, in the second step, the nodal demand samples are consistently reordered to build snapshot demand scenarios for the WDN, while respecting the rank cross-correlations at lag 0. The applications concerned the one-year long dataset of about 1000 user demand values from the district of Soccavo, Naples (Italy). Best-fit scaling equations were constructed to express the main statistics of peak demand as a function of the average demand value on a long-time horizon, i.e., one year. The results of applications to four case studies proved the methodology effective and robust for various numbers and sizes of users.

Keywords: water distribution network; peak demand; scenarios; stochastic generation; beta probability distribution; correlation



Citation: Creaco, E.; Galuppini, G.; Campisano, A.; Franchini, M. Bottom-Up Generation of Peak Demand Scenarios in Water Distribution Networks. *Sustainability* **2021**, *13*, 31.

<https://dx.doi.org/10.3390/su13010031>

Received: 25 November 2020

Accepted: 21 December 2020

Published: 22 December 2020

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



Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The last three decades have seen the spread of physically based models, which are used by water utility staff for replicating the nonlinear behavior of water distribution networks (WDNs) with specific objectives, such as contingency planning, network optimization, and strategy planning [1,2]. Among the critical inputs of WDN modelling, nodal demands play a major role [3]. Therefore, numerous research efforts have recently been dedicated to the accurate assessment of nodal demands, while benefitting from increasingly large datasets provided by smart water metering technologies [4].

A viable option to reconstruct demand lies in the application of the bottom-up stochastic approach [5], which is the topic of the present paper. This approach aims to reconstruct the whole WDN demand starting from the single nodes upward, while representing the single nodal demands as stochastic variables.

If the focus is the representation of demand at very fine spatial and temporal scales, pulse generation models can be used (e.g., see [6] for a review of these models), which can reproduce the arrival time, duration, and intensity of demand pulses generated by each faucet present in any WDN node. The output of these models can be used as an input to unsteady flow models with large computational overhead, as was shown in work [3].

Other kinds of approaches can be used when larger time steps are considered. If the focus is to reproduce nodal demand time series with values aggregated at temporal step

ranging from minutes to hours, methodologies such as those proposed in the works [7–10] can be used. The output of these methodologies can be used as an input to the extended period simulation, which represents the behavior of the WDN as a succession of steady states [5]. The overall pulse model of Gargano et al. [8] aims to extend the concept of pulse from the representation of the instantaneous demand of the single faucet to the representation of the overall demand of one user or one node at a larger time step. In the methodology of Alvisi et al. [7], the demand associated with the generic time step is represented by the random variable through the multinomial probability distribution, with predetermined mean and standard deviation. Furthermore, cross-correlations between timeseries at lag 0 and lag 1 are considered. Compared to [7], the methodology of Creaco et al. [10], based on the beta probability distribution, preserves the skewness in demand time series and enforces correlations between time series at all lags. Furthermore, Creaco et al. [10] also presented a procedure to reconcile the generated demand time series with demand pulses generated at fine time step, thus enabling reconstruction of demand at any time step. The methodology of Kossieris et al. [9] is based on the mixed-type distribution for the representation of demand time series, while yielding similar performance to the methodology of Creaco et al. [10].

In various cases, the modelling objective has focused on exploring the performance and reliability of the WDN only in peak demand scenarios, with no need to investigate the temporal pattern of demand [5]. Starting from the early decades of the twentieth century [11], the peak demand has been the subject of numerous works, aimed at estimating the daily peak demand coefficient, as the ratio of maximum to daily average demand, as a function of the number of users. While the first works described this coefficient as a deterministic variable e.g., see [11–14], Tricarico et al. [15] were among the first to present the daily peak demand coefficient as a random variable. The probabilistic approach in [15] was later refined by Gargano et al. [16], who successfully described the daily peak demand coefficient using three probability distributions, namely the log-normal, Gumbel, and log-logistic distributions. Furthermore, Gargano et al. [16] used statistical analysis to express the parameters of these distributions as a function of the number of users and of the temporal step used for demand representation. Balacco et al. [17] used a different probabilistic framework, based on the exponential and Gumbel distributions, to develop a relationship between the above mentioned daily peak demand coefficient and the number of users. The applications of Balacco et al. [17] concerned several towns in Puglia. Finally, Del Giudice et al. [18] provided a methodological framework to investigate the main features of water demand hourly peak coefficients, with the objective to estimate the sample mean of hourly peak factors, the associated standard error (allowing definition of confidence bands), and its probability distribution. While considering the sample means normally distributed, they also provided empirical relations to express the sample mean and standard error as a function of the number of users. Though providing formulations and relationships that can be profitably used by practitioners to estimate the overall peak demand coefficient in a WDN, the methodologies in [15–18] fail to describe how the peak must be allocated to its nodes. In fact, in a generic peak scenario, the adoption of a single peak demand coefficient at all nodes, according to the deterministic top-down approach [5], may lead to an unrealistic representation of the peak demand in the WDN. One of the few works aimed at reconstructing the snapshot peak demand scenario in an entire WDN, as the combination of peak demands at all the nodes, is that of Magini et al. [19]. In the methodology presented in [19], combinations of simultaneous nodal demands are first generated through the gamma probability distribution with parameters expressed as a function of the number of nodal users by means of scaling laws. Then, nodal demand series are resorted to preserve correlation between nodes through the Iman–Canover method [20]. Though the methodology of Magini et al. [19] performs effectively at reconstructing peak demands through the bottom-up approach, the following aspects in [19] require further analysis and are therefore explored in the present paper:

- Being based on two parameters, the gamma probability distribution used in [19] enables respecting statistics of user demands only up to the second order, i.e., mean and standard deviation. Therefore, the question arises if more complex distributions, such as the beta distribution with tunable bounds [10], can be used to respect bounds and statistics of user demands up to the third order i.e., including the skewness;
- The scaling laws used in [19] for expressing the moments of the peak demand are based on the number of users as independent variable. However, demand may differ significantly from a user to another, being affected by the number and kind of household occupants, which is often unknown and/or variable in the year. Therefore, the average demand value on the long run, i.e., one year, is a more significant variable than the number of nodal users and can be considered as the independent variable for more well-founded scaling equations;
- In some cases, there may not be enough demand data to calibrate the scaling equations for estimating rank cross-correlations between nodal peak demands. Therefore, the question arises if an alternative procedure can be conceived to impose cross-correlations in the WDN.

The rest of the paper is organized as follows. Section 2 contains the methodology. Section 3 contains the applications, i.e., datasets from the district of Soccavo, Naples, Italy [10,18,21] and results. Sections 4 and 5 contain the discussions and conclusions, respectively.

2. Methodology

The methodology adopted in this paper consists of two steps, i.e., step 1 and step 2, which follow a preliminary step 0. As the flow chart in Figure 1 shows, step 0 operates on the smart metering data of measured consumption at the users of a WDN. It aims to create aggregated datasets of measured peak demands associated with growing values of the average demand on the long run. Step 1 concerns parameterization of a beta probability distribution model for the bottom-up generation of peak demand samples as a function of the average demand. Step 2 concerns the resorting of generated demand samples to preserve the expected degrees of rank cross-correlation in a WDN. In the following subsections, the preliminary step 0, and steps 1 and 2 of the methodology are described.

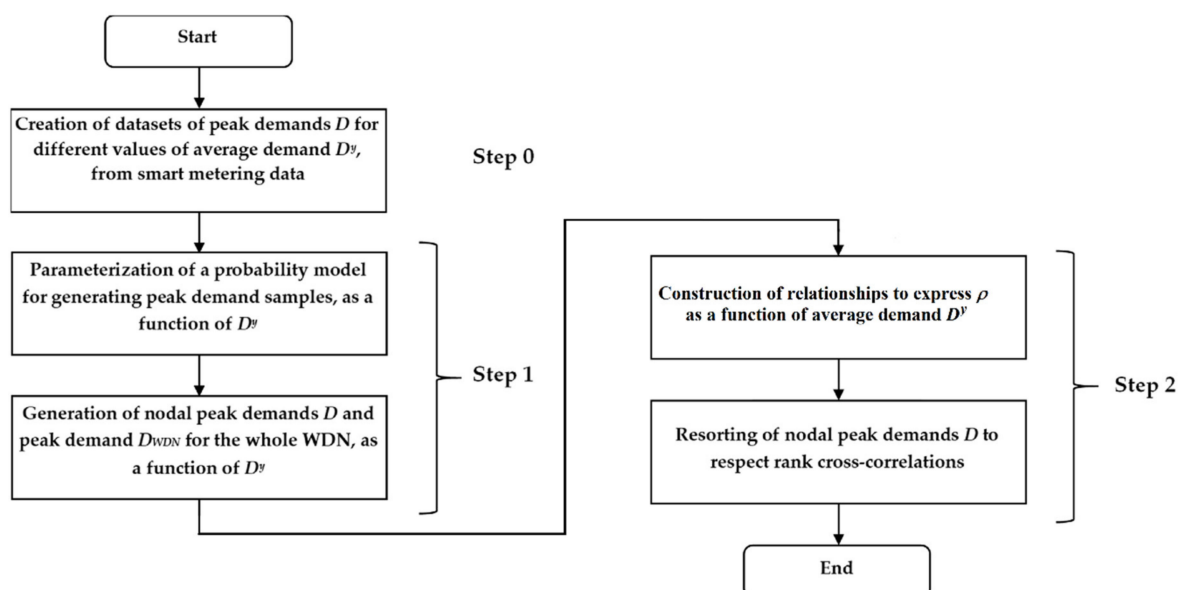


Figure 1. Flowchart of the methodology.

2.1. Step 0—Creation of the Dataset

The objective of Step 0 is to obtain samples of measured peak demand (for different aggregation levels) associated with various values of the average demand on the long run. As is explained below, these samples are obtained by aggregating the peak demand measurements at the users of a WDN. However, the level of aggregation at which the generic average demand value on the long run is obtained is not relevant for the statistical analyses of Steps 1 and 2.

Step 0 starts from the recorded timeseries of demand at a high number n_u of users of a WDN with the temporal resolution of one hour, which corresponds to the typical time step used for WDN modelling, covering a period of at least one year. This dataset will look like the demand data d_i^j , with subscript i and superscript j associated with the i -th of the n_u users and with the j -th of the 8760 h in the year ($8760 = 24 \text{ h} \times 365 \text{ days}$), reported as a matrix in Table 1. It must be noted that the number 8760 for the hours is used here for explicative reasons, without loss of generality. Longer durations of timeseries could also be adopted. For the timeseries recorded at the i -th user, the long-time average d_i^y can be easily calculated (final row of the matrix in Table 1). If the series duration is one year, as is assumed implicitly hereinafter, d_i^y represents the yearly average.

Table 1. Explicative dataset of recorded time series for n_u users in the whole year. See text.

Time (h)	User 1	User 2	User 3	User n_u-1	User n_u
1	d_1^1	d_2^1	d_3^1	$d_{n_u-1}^1$	$d_{n_u}^1$
2	d_1^2	d_2^2	d_3^2	$d_{n_u-1}^2$	$d_{n_u}^2$
8759	d_1^{8759}	d_2^{8759}	d_3^{8759}	$d_{n_u-1}^{8759}$	$d_{n_u}^{8759}$
8760	d_1^{8760}	d_2^{8760}	d_3^{8760}	$d_{n_u-1}^{8760}$	$d_{n_u}^{8760}$
Yearly Average	d_1^y	d_2^y	d_3^y	$d_{n_u-1}^y$	$d_{n_u}^y$

The first operation to carry out in Table 1 lies in the selection of the data to be used for the statistics, based on a certain criterion, e.g., the data of a certain hour of the day in a certain period of the year, e.g., the morning peak in working weekdays. In this context, let us assume that N_{tot} hours of the year (with $N_{tot} < 8760$), which are considered homogeneous from the statistical viewpoint, are selected leading to the dataset shown in Table 2. The yearly average value in Table 2 is kept the same as that in Table 1, since it does not depend on the criterion chosen to select the N_{tot} hours of the year.

Table 2. Explicative dataset of demands for n_u users in N_{tot} selected hours of the year. See text.

Selected Times	User 1	User 2	User 3	User n_u-1	User n_u
1	d_1^1	d_2^1	d_3^1	$d_{n_u-1}^1$	$d_{n_u}^1$
2	d_1^2	d_2^2	d_3^2	$d_{n_u-1}^2$	$d_{n_u}^2$
$N_{tot}-1$	$d_1^{N_{tot}-1}$	$d_2^{N_{tot}-1}$	$d_3^{N_{tot}-1}$	$d_{n_u-1}^{N_{tot}-1}$	$d_{n_u}^{N_{tot}-1}$
N_{tot}	$d_1^{N_{tot}}$	$d_2^{N_{tot}}$	$d_3^{N_{tot}}$	$d_{n_u-1}^{N_{tot}}$	$d_{n_u}^{N_{tot}}$
Yearly Average	d_1^y	d_2^y	d_3^y	$d_{n_u-1}^y$	$d_{n_u}^y$

Then, the spatial aggregation is performed by randomly summing the columns of Table 2 from the left to the right, to obtain the aggregated demands D_i^j and the aggregated yearly averages D_i^y (see Table 3), calculated by the two following formulas, respectively:

$$D_i^j = \sum_{k=1}^i d_k^j, \quad (1)$$

$$D_i^y = \sum_{k=1}^i d_k^y. \quad (2)$$

Table 3. Explicative dataset of n_u aggregated demands in N_{tot} selected hours of the year. See text.

Selected Times	Aggregation 1	Aggregation 2	Aggregation 3	Aggregation n_{u-1}	Aggregation n_u
1	D_1^1	D_2^1	D_3^1	D_{nu-1}^1	D_{nu}^1
2	D_1^2	D_2^2	D_3^2	D_{nu-1}^2	D_{nu}^2
$N_{tot}-1$	$D_1^{N_{tot}-1}$	$D_2^{N_{tot}-1}$	$D_3^{N_{tot}-1}$	$D_{nu-1}^{N_{tot}-1}$	$D_{nu}^{N_{tot}-1}$
N_{tot}	$D_1^{N_{tot}}$	$D_2^{N_{tot}}$	$D_3^{N_{tot}}$	$D_{nu-1}^{N_{tot}}$	$D_{nu}^{N_{tot}}$
Yearly Average	D_1^y	D_2^y	D_3^y	D_{nu-1}^y	D_{nu}^y

The spatial aggregation can be reiterated various times, considering a different (random) sequence of users at each time. This will result in a different sequence of aggregated demands. Of course, the last aggregations (for i slightly lower than n_u) will be potentially very similar, while the very last ones (for $i = n_u$) will be identical between each other. In fact, for $i = n_u$, all demands at time instant j are considered in all sequences within the aggregation. It is worth noting that reiterating n times Step 0 is likely to reproduce Table 3 n times, where in each column (i.e., for a given aggregation), different values are found related to the appropriate yearly average D^y . At the end of Step 0, the n tables obtained can be concatenated horizontally. The demand samples associated with the not aggregated measured demand (i.e., Table 2) are also concatenated at the end, to include numerous demand data associated with low values of D^y that would not be generated in the aggregation process. Finally, the columns of the final dataset are sorted in ascending order of D^y and the doublet columns potentially present are removed. As a result, the final dataset contains peak demand samples for growing values of D^y , independently of the aggregation level.

2.2. Step 1—Beta Probability Distribution with Tuneable Bounds

The beta probability distribution with tunable bounds was chosen in this work to represent the generic aggregated demand D_i in each column of Table 3. This choice was due to its capability of reproducing the statistics of demand series in terms of mean, standard deviation, and skewness, as was shown by Creaco et al. [10].

The density function f of the beta distributed random variable D_i , representing the water demand of a generic aggregation (see Table 3), is [22]:

$$f(D_i) = \frac{(D_i - a)^{\alpha-1} (b - D_i)^{\beta-1}}{B(\alpha, \beta) (b - a)^{\alpha+\beta-1}}, \quad (3)$$

in which B , α and β are the beta function and the positive shape parameters, respectively. Finally, $a \geq 0$ and $b > a$ are the lower and upper bounds, respectively. Parameters α and β , and bounds a and b can be related to mean D_{mean} and standard deviation σ of D_i through the following Equations (4)–(7):

$$\alpha = \bar{\mu} \left[\frac{\bar{\mu}(1 - \bar{\mu})}{\bar{\sigma}^2} - 1 \right], \quad (4)$$

$$\beta = (1 - \bar{\mu}) \left[\frac{\bar{\mu}(1 - \bar{\mu})}{\bar{\sigma}^2} - 1 \right], \quad (5)$$

where:

$$\bar{\mu} = \frac{D_{mean} - a}{b - a}; \quad (6)$$

$$\bar{\sigma} = \frac{\sigma}{b-a}. \quad (7)$$

The skewness γ of the beta probability function can be evaluated as a function of α and β as:

$$\gamma = \frac{2(\beta - \alpha)\sqrt{\alpha + \beta + 1}}{(\alpha + \beta + 2)\sqrt{\alpha\beta}}. \quad (8)$$

For any pair of values of a ($< D_{mean}$) and b ($> D_{mean}$), α and β can be obtained from Equations (4) and (5) while respecting mean D_{mean} and standard deviation σ of demand D_i . This occurs if D_{mean} and σ , in Equations (6) and (7), respectively, are set equal to the mean and standard deviation of the measured demand variable D_i (method of the moments). However, each a, b pair determines different values of the shape parameters α and β , and subsequently of γ . In this work, a was set at the minimum observed value D_{min} of D_i (which was therefore an additional parameter of the probability distribution model), while b was suitably modified. In practice, the increase (or decrease) in b results in the lengthening (or shortening) of the right tail of the probability distribution, and therefore in the increase (or decrease) in the value of γ . The best value of b can be found by solving numerically an implicit equation obtained by setting γ at the observed value γ_{obs} in Equation (8), after replacing α , β , $\bar{\mu}$, and $\bar{\sigma}$, as they are expressed in Equations (4)–(7). In this work, a local optimization based on the active-set method [23] aimed at minimizing the absolute difference between γ in Equation (8) and γ_{obs} was applied with an initial value of b equal to the measured maximum value of D_i to find a solution of the implicit equation.

Summing up what is reported above, parameters α and β , and bounds a and b can be univocally identified starting from minimum D_{min} , mean D_{mean} , standard deviation σ , and skewness γ in the measured demand data. Best-fit scaling equations can be constructed to relate D_{min} , D_{mean} , σ , and γ of the random variable D_i , i.e., the demand in the peak hour in working weekdays, to the yearly average demand D_i^y . This enables obtaining a probability model to generate samples of peak demand values for any value of yearly average demand, when single or aggregated users have similar characteristics to those considered for model parameterization (available data). Specifically, the probability distribution model can be applied to generate a sample of N_g peak demand values (D_i) for the generic node i (with known value of the yearly average demand D_i^y) of a WDN. It can also be applied to generate a sample of N_g demand values (D_{WDN}) for the whole WDN, which has a known value of the yearly average demand D_{WDN}^y . However, if rank cross-correlations between nodes are neglected, the peak demand estimate D_{WDN}^* obtained for the whole WDN by summing all nodal demand samples, i.e., $D_{WDN}^* = \sum_{i=1}^{N_n} D_i$ with N_n equal to the number of nodes in the WDN, may not respect the expected statistics of standard deviation and skewness for the yearly average demand D_{WDN} of the whole WDN. The next subsection deals with the resorting of the demand samples generated for each WDN node to preserve the Spearman rank cross-correlations at lag 0.

2.3. Step 2—Re-Sorting of Generated Demand Samples

The algorithm used in this work to enforce rank cross-correlation is based on a modified version [24] of the Gaussian Copula [25]. Let us assume we have used Step 1 of the methodology to generate uncorrelated demand samples of N_g peak demands D_i^j for N_n nodes in a WDN, where subscript i and superscript j range from 1 to N_n and from 1 to N_g , respectively. This method consists of generating N_n samples of N_g values y_i^j , sampled from random variables with mean values to 0 and standard deviations equal to 1 through the multivariate normal distribution characterized by a correlation matrix with sizes $N_n \times N_n$, in which the generic element $\rho_{k,l}$ represents the expected rank cross-correlation between peak demand D_k for the generic node k and peak demand D_l for the generic node l . The rank cross-correlation matrix is symmetric with diagonal element equal to 1. To enforce the expected rank cross-correlations $\rho_{k,l}$, the values D_i^j must be reordered using the same sorting as y_i^j .

If datasets of peak demand are available, such as those obtained from step 0 of this methodology, the values of rank cross-correlation can be calculated by confronting a k -column of one dataset with the l -column of another. Then, a best-fit scaling equation can be constructed to relate rank cross-correlations $\rho_{k,l}$ to the yearly average demands of nodes/aggregations k and l . By applying this equation, the rank cross-correlation for two generic WDN nodes with similar characteristics of peak demand to the datasets can be evaluated.

If the peak demand datasets are not numerous enough to enable robust best-fit equations, the following simplified procedure can be applied. This procedure is based on a numerical optimization and considers a single value of correlation ρ in the WDN. In other words, $\rho_{k,l} = \rho$ for any pair of nodes k and l in the network. The optimal value of ρ is searched for by applying a local optimization based on the active-set method [23], to obtain a cumulative distribution of D_{WDN}^* close to that of D_{WDN} . Summing up, with reference to the uncorrelated demand samples of N_g peak demands D_i^j for the N_n nodes and to the sample of N_g peak demands D_{WDN}^j for the whole WDN, both available from Step 1, the following instructions are carried out for each value of ρ proposed by the optimizer:

1. Resort the N_n samples of N_g peak demands D_i^j following the ρ -based sort obtained through the Copula algorithm.
2. Estimate the 1-norm distance between the cumulative distribution of D_{WDN}^j and of D_{WDN}^j , i.e., the sum of the absolute difference between D_{WDN}^j and D_{WDN}^j evaluated for the different values of the cumulative frequency.

When the optimizer converges, the best value of ρ that minimizes the 1-norm distance between the two cumulative distributions (objective function) is obtained.

At the end of Step 2, N_g scenarios of peak demands D_i^j will be available. The generic j -th of the N_g scenarios contains the re-sorted peak demand values D_i^j , with i ranging from 1 to N_n nodes of the WDN.

3. Applications

3.1. Data

The applications of this work were carried out using the users' demand collected with a smart metering system from the Soccavo district of Naples, Italy. These data have already been used in the scientific literature [10,18,21]. They include the demand time series of $n_u = 1067$ residential users with one hour resolution, for a total time period of one year (8760 values for each user), from 20 March 2017 h 0–1, to 19 March 2018 h 23–24. The demand time series data available from the monitoring of the Soccavo district are characterized by yearly average demands d^y ranging from 3.47×10^{-4} to 0.014 L/s. Further details can be found in the work of Fiorillo et al. [21]. The application of the first part of the preliminary Step 0 of the methodology (see Tables 1 and 2) enabled selection of $N_{tot} = 228$ h in the year, corresponding to the morning peak from 8 am to 9 am in the working weekdays in all the months of the year except for August, which showed a very different behavior from the rest of the year. Then, in the second part of Step 0, two datasets were randomly generated for the Step 1 of the methodology. Each dataset was obtained by iterating the aggregation procedure based on Equations (1) and (2) (see Tables 2 and 3) for $n = 50$ times. As subsets of these two datasets, four smaller datasets (two with $n = 1$ and two with $n = 2$) were used for Step 2 of the methodology, that is for the characterization of the rank cross-correlation. The numerosity of these datasets was chosen to obtain a sufficiently representative group of datasets for assessing the parameters of the beta probability distribution and the rank cross-correlations, respectively. In the following subsections, first the results of Step 1 and Step 2 are described, followed by explicatory applications of the methodology for demand generation in WDNs. Only for clarity's sake, it is worth reminding at this stage that each dataset is organized like in Table 3, with each

column reporting peak demand values D associated with a certain value of yearly average demand D^y .

3.2. Results of Step 1

Since the parameters α and β , and bounds a and b of the beta probability distribution can be evaluated only if the minimum value D_{min} , mean D_{mean} , standard deviation σ , and skewness γ have been assigned, best-fit scaling equations were constructed to relate D_{min} , D_{mean} , σ , and γ of the peak demand D to the yearly average demand D^y . Therefore, the two datasets constructed in Step 0 were used for constructing/calibrating and for validating the best-fit equations, respectively. In the graphs in Figure 2, D_{min} , D_{mean} , σ , and γ are plotted against D^y for the peak demands D in the calibration datasets, independently of the aggregation level. These graphs show very regular increasing patterns of the first three parameters as a function of D^y . Though all peak demands have a clearly positive asymmetry, the pattern of γ shows instead some irregularities, maybe because D^y is unable to explain alone the variability in γ . Linear best-fit models were used to express D_{min} and D_{mean} . Due to the change in slope in the pattern $\sigma(D^y)$ at around $D^y = 0.02$ L/s, a combination of two linear models was chosen for σ . Finally, an exponentially decreasing model was chosen for γ . The best-fit scaling equations are reported in the following Table 4, pointing out very high-fitting performance in terms of R^2 for the first three parameters and a sufficient performance for γ (see also the lines in the graphs in Figure 2).

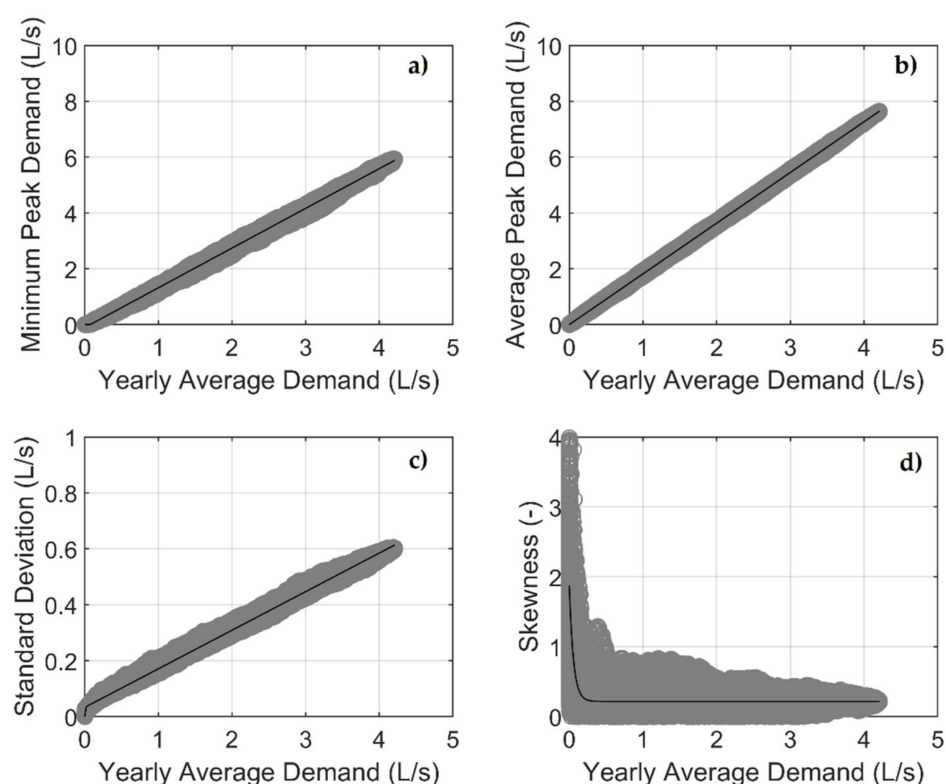


Figure 2. (a) Minimum value D_{min} , (b) mean value D_{mean} , (c) standard deviation σ , and (d) skewness γ in peak demand D as a function of the yearly average demand D^y , as obtained from calibration dataset (grey dots) and from best-fit equations (black lines).

Table 4. Best-fit equation obtained for each of the four parameters of the probability model and for the rank cross-correlation. R^2 in calibration and validation.

Equation	R^2 (-) in Calibration	R^2 (-) in Validation
$D_{min} = \max(0, 1.423D_y - 0.105)$	0.998	0.998
$D_{mean} = 1.817D_y$	1.000	1.000
$\sigma = 1.78D_y$ for $D_y \leq 0.02$ L/s $\sigma = 0.138D_y + 0.033$ for $D_y > 0.02$ L/s	0.995	0.994
$\gamma = 0.216 + (1.886 - 0.216) \exp(-17.781 D_y)$	0.428	0.459
$\rho = [-110.73 \text{ prod}^2 + 9.199 \text{ prod} + 0.002] \cdot \text{ratio}^{0.15}$ for $(\text{prod}) \leq 0.04$ L ² /s ² $\rho = [0.130 \ln(\text{prod}) + 0.624] \cdot \text{ratio}^{0.15}$ for $(\text{prod}) > 0.04$ L ² /s ²	0.904	0.921

The goodness of fit of the best-fit equations persisted in the validation dataset (see graphs in Figure 3 and the last column in Table 4).

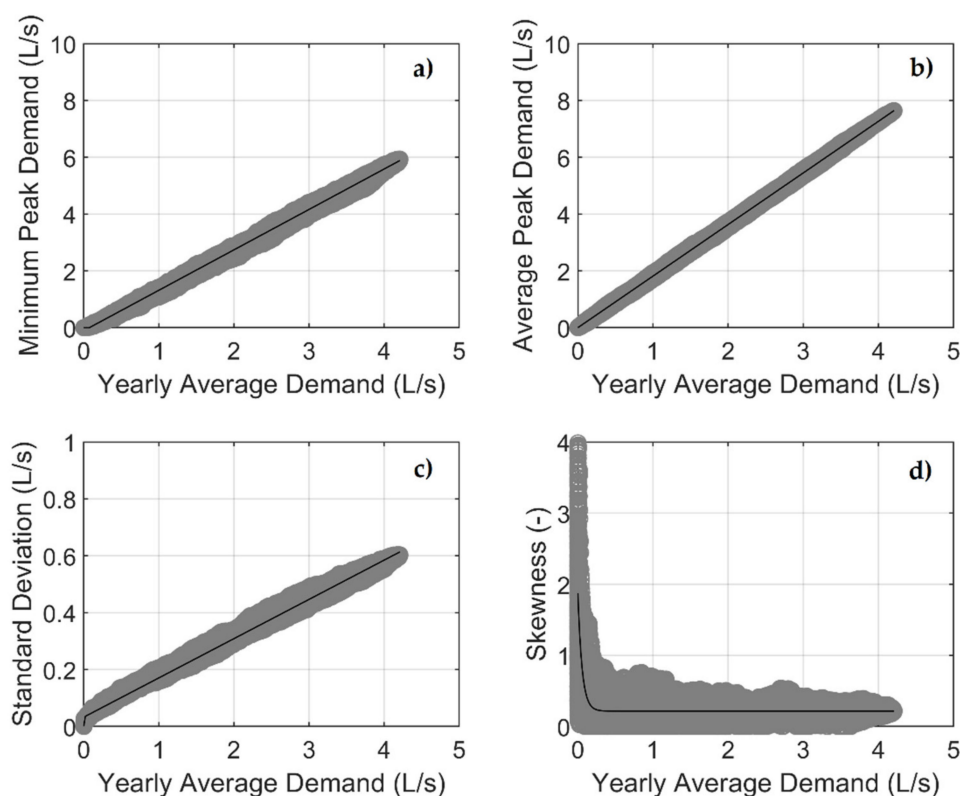


Figure 3. (a) Minimum value D_{min} , (b) mean value D_{mean} , (c) standard deviation σ , and (d) skewness γ in peak demand D as a function of the yearly average demand D_y , as obtained from validation dataset (grey dots) and from best-fit equations (black lines).

A last verification of the beta probability model with parameters estimated through the best-fit scaling equations for D_{min} , D_{mean} , σ , and γ is shown in the graphs in Figure 4. Each graph allows comparing the histogram of normalized frequencies of the measured peak demand for a certain value of yearly average demand D_y , with the beta probability density function associated with same value of D_y . Globally, Figure 4 points out a satisfactory agreement for both low (graphs a and b) and high (graphs c and d) values of D_y . Incidentally, it must be remarked that the low and high values of yearly average demand D_y determined different patterns for the probability density function of peak demand D , since the probability of values of D close to 0 is high for the former and null for the latter. Furthermore, it must be remarked that the beta probability model enabled preservation

of the skewness values obtained with the best-fit equation reported in Table 4. In fact, in the four case studies, it provided γ values equal to 1.7595, 1.6148, 0.2156, and 0.2156, respectively. If applied in the same context and parameterized only based on D_{mean} and σ , the gamma probability model described in the work [19] would provide γ values equal to 1.9560, 1.9560, 0.1760, and 0.1605. Therefore, it would overestimate or underestimate the skewness for the low or high values of D^y , respectively. Another benefit of the beta probability model is that it fits the minimum value D_{min} of D better than the gamma probability model, for which $D_{min}=0$ for any value of D^y .

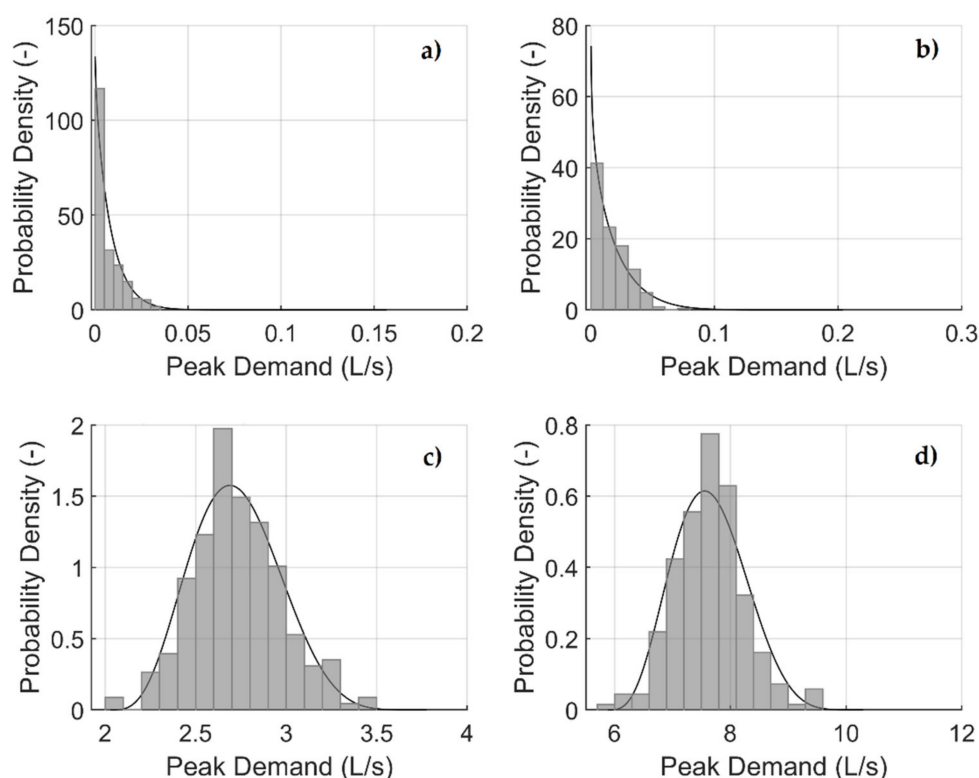


Figure 4. For different values of the yearly average demand D^y [(a) 0.0044 L/s, (b) 0.010 L/s, (c) 1.50 L/s, and (d) 4.21 L/s] probability density p of the peak demand (black line) obtained with the beta probability distribution model compared with the histogram of normalized frequencies in the measured series.

3.3. Results of Step 2

The four datasets constructed in Step 0 to explore rank cross-correlation were used to construct/calibrate and validate best-fit scaling equations. Calibration was carried out using one dataset with $n = 1$ and one dataset with $n = 2$. Validation was carried out using the other dataset with $n = 1$ and the other dataset with $n = 2$. In either phase, rank cross-correlations (different combinations of k and l columns) were calculated between the peak demand samples in the dataset with $n = 1$ and those in the dataset with $n = 2$. Values of the rank cross-correlations equal to 1 due to the presence of identical peak demand samples were discarded from the analysis. The rank cross-correlations between the very much aggregated peak demands were also removed because the random sequential aggregation leads to overly similar peak demand samples at the last iterations and, therefore, to misleading rank cross-correlations very close to 1, as it was said above. The analysis of the rank cross-correlation ρ between two samples of peak demand showed that this variable tends to increase when the product $prod$ of the yearly average demands grows. Under the same value of $prod$, ρ tends to increase when the ratio of the minimum (low) to the maximum (high) yearly average demand grows from 0 to 1. The relationship between ρ and $prod$ was expressed by a polynomial model and a logarithmic model, valid for low and high

values of $prod$, respectively. Both models were multiplied by a power of $ratio$ to introduce the dependency of ρ on $ratio$. The calibrated best-fit equations are reported in Table 4, as well as the high-fitting performance in terms of R^2 in both calibration and validation. The graphs in Figure 5 compare the ρ values obtained from the datasets with those obtained through the best-fit scaling equations as a function of $prod$, attesting to the goodness of fit of the latter equations. In fact, in both graphs, the clouds of dots associated with the measured and calculated values of ρ as a function of $prod$ are satisfactorily overlapped.

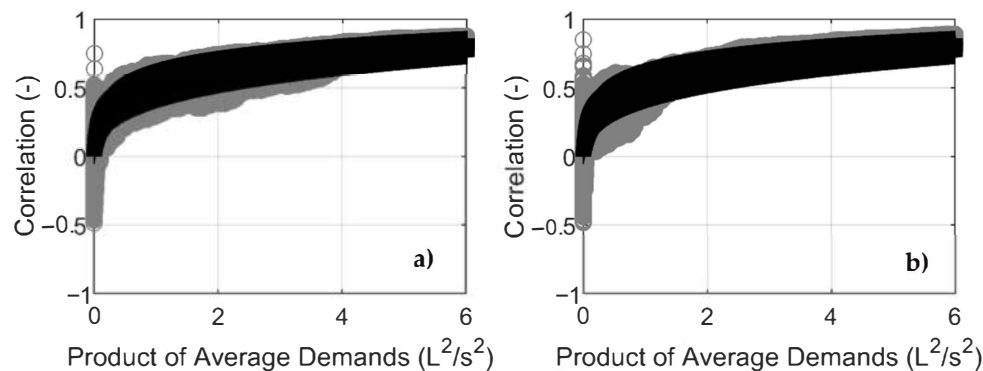


Figure 5. Rank cross-correlations ρ between peak demands D as a function of the product $prod$ of yearly average demands D^y , as obtained from datasets (grey dots) and from best-fit equations (black crosses). Calibration and validation in graphs (a) and (b), respectively.

Then, some tests were carried out on the numerical optimization procedure for sample re-sorting. In fact, for various pairs of values of D^y , peak demand samples were generated and re-sorted by using the optimization procedure described above and by enforcing the correlation degree obtained through the best-fit formula. In all cases, the optimization-based re-sorting yielded similar rank cross-correlations to those suggested by the formula. Other tests of step 2 of the methodology can be found in the following subsection.

3.4. Peak Demand Generation in WDNs

Four case studies, numbered as 1a, 1b, 2a, and 2b, were considered for peak demand generation, to test the methodology under different demand conditions. All case studies were assumed to have similar peak demand features to the Soccavo district, thus legitimizing the applicability of the relationships in Table 4. Obviously, the relationships would need to be recalibrated under different peak demand conditions.

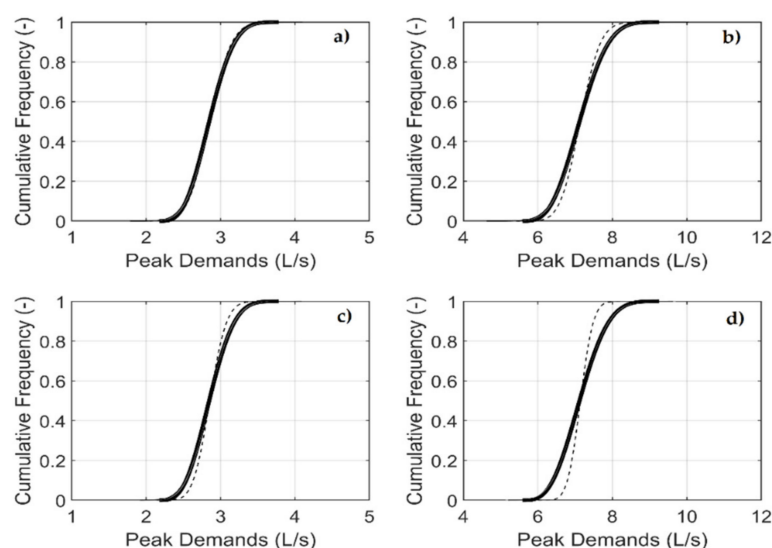
Case study 1a was represented by the WDN serving a small fictitious town quarter in Italy. Further details can be found in the work [26]. The quarter was assumed to host 850 inhabitants, linked to 183 nodes, 170 of which have known nonzero average demand. The yearly average demand D^y of the whole WDN was about 1.57 L/s. Case study 1b differed from case study 1a due to the amplification of each node demand through a factor of 2.5, resulting in a yearly average demand D^y of the whole WDN of about 3.93 L/s.

Case study 2a considered in this work was a skeletonized version of case study 1a. Further details can be found in the work [26]. In particular, demanding people's allocation in the skeletonized WDN was carried out considering the criterion of connection proximity to the node. To this end, the demanding people placed along the upstream part of the generic pipe were allocated to the upstream node in the skeletonized layout, and vice-versa. This resulted in 12 nodes associated with known nonzero demand values. Case study 2b differed from the case study 2a due to the amplification of demand through a factor of 2.5. As expected, the values of D^y for the whole WDN in case studies 2a and 2b were equal to those in case studies 1a and 1b, respectively. The D^y values of each demanding node in case studies 2a and 2b are reported in Table 5.

Table 5. Yearly average demands D^y (L/s) at the various nodes and in the whole WDN in case studies 2a and 2b.

Demanding Node	Case Study 2a	Case Study 2b
1	0.11470	0.2868
2	0.11285	0.2821
3	0.13135	0.3284
4	0.09250	0.2313
5	0.14615	0.3654
6	0.16835	0.4209
7	0.21645	0.5411
8	0.15540	0.3885
9	0.09065	0.2266
10	0.13505	0.3376
11	0.13135	0.3284
12	0.07770	0.1943
Total	1.57250	3.9313

By applying step 1 of the methodology with the best-fit equations for D_{min} , D_{mean} , σ , and γ reported in Table 4, (starting from the D^y values in the nodes) samples of peak demand D were generated for the WDN nodes in all case studies using the corresponding beta probability distributions. Specifically, one sample with $N_g = 100,000$ peak demand values was generated for each node of given yearly average D^y . Then, the WDN peak demand D_{WDN}^* was obtained by summing the nodal peak demands without resorting. Obviously, it was made up of a sample of $N_g = 100,000$ values. As the graphs in Figure 6 show, the cumulative distribution of D_{WDN}^* may be far from the theoretical cumulative distribution D_{WDN} obtained by applying the beta probability model related to the whole yearly average demand D_{WDN}^y , especially for larger nodal values of the yearly average demand (case studies 1b, 2a, and 2b).

**Figure 6.** In case studies (a) 1a, (b) 1b, (c) 2a, and (d) 2b, cumulative frequency of the peak demand in the whole network: theoretical frequency (—), frequency obtained with no re-sorting (---), frequency obtained with the equation-based re-sorting (.....), frequency obtained with the optimization-based re-sorting (-.-.-).

Both variants of re-sorting in Step 2 of the methodology, based on the best-fit equation and on the optimization, respectively, brought the cumulative distribution of D_{WDN}^* very close to the theoretical cumulative distribution of D_{WDN} (see Figure 6).

The single values of rank cross-correlations ρ obtained through the optimization in the four case studies were equal to 0.002, 0.0063, 0.1025, and 0.3381, respectively. As expected, ρ grew when the average value of D^y at WDN nodes increased. In each case study, the $\rho_{l,k}$ values obtained through the best-fit equation changed from pair of nodes to pair of nodes, but lay around the single value obtained through the optimization. As further evidence of the effectiveness of the methodology, the graphs in Figure 7 compare, for case study 2a, the histogram of normalized frequencies of D_{WDN}^* with the theoretical beta probability density function of the peak demand D_{WDN} in the whole WDN, proving that only by resorting nodal demand samples can the sum of nodal demand represent the expected demand for the whole WDN.

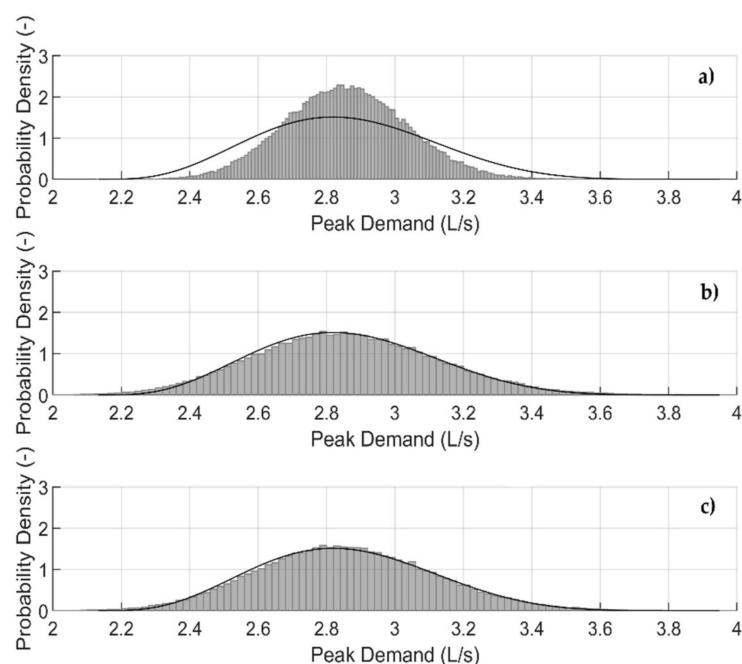


Figure 7. In case study 2a, comparison of theoretical probability density function of the whole peak demand with the histogram of normalized frequencies of the demand obtained by summing the single nodal demands and by applying (a) no re-sorting, (b) best-fit equation-based re-sorting, and (c) optimization-based resorting.

The theoretical mean, standard deviation, and skewness values and the values obtained by summing the generated peak nodal demand samples are reported in Table 6, showing that the only preserved statistic without data re-sorting was the mean. This is because the mean is simply additive. Best-fit equation-based and optimization-based resorting yielded significant improvements to preservation of standard deviation and skewness.

Table 6. Statistical consistency of the various methods in case study 2a.

	Theoretical	No Resorting	Equation	Optimisation
Mean (L/s)	2.86	2.86	2.86	2.86
Standard Deviation (L/s)	0.25	0.18	0.27	0.26
Skewness (-)	0.22	0.10	0.18	0.17

4. Discussion

The best-fit scaling equations obtained in the present work to estimate D_{mean} , σ , and ρ , shown in Table 4, can be compared with the scaling laws derived through a theoretical framework, e.g., [19,27]. These laws are based on the nodal number of consumers and the statistical features of water demand of the typical single user, the unitary user. To put them over the same ground as the best-fit equations of this work, the scaling laws were

reformulated through algebraic manipulations as a function of the yearly average demand, instead of the number of users (see Appendix A). The calibration of the scaling laws on the Soccavo dataset yielded the following equations:

$$D_{mean} = 1.817D^y. \quad (9)$$

$$\sigma = 0.182(D^y)^{0.822}. \quad (10)$$

$$\rho_{k,l} = \frac{\left(\frac{D_k^y}{0.0094}\right) \left(\frac{D_l^y}{0.0094}\right) 0.0145}{\sqrt{\frac{D_k^y}{0.0094} \left[1 + 0.0145 \left(\frac{D_k^y}{0.0094} - 1\right)\right]} \sqrt{\frac{D_l^y}{0.0094} \left[1 + 0.0145 \left(\frac{D_l^y}{0.0094} - 1\right)\right]}}. \quad (11)$$

Equation (9) for estimating D_{mean} is identical to the best-fit equation obtained in the present work (see Table 4). Equation (10) for estimating σ has a similar fit to the calibration data ($R^2 = 0.994$), in comparison with the best-fit equation obtained in the present work (see graphs a and c in Figure 8). However, it must be remarked that the calibration of the scaling law causes a significant underestimation of σ for the low values of D^y (see graphs b and d in Figure 8). Equation (11) for estimating rank cross-correlation has a globally lower goodness of fit ($R^2 = 0.897$, see Figure 9) and tends to underestimate the variability in the value of ρ .

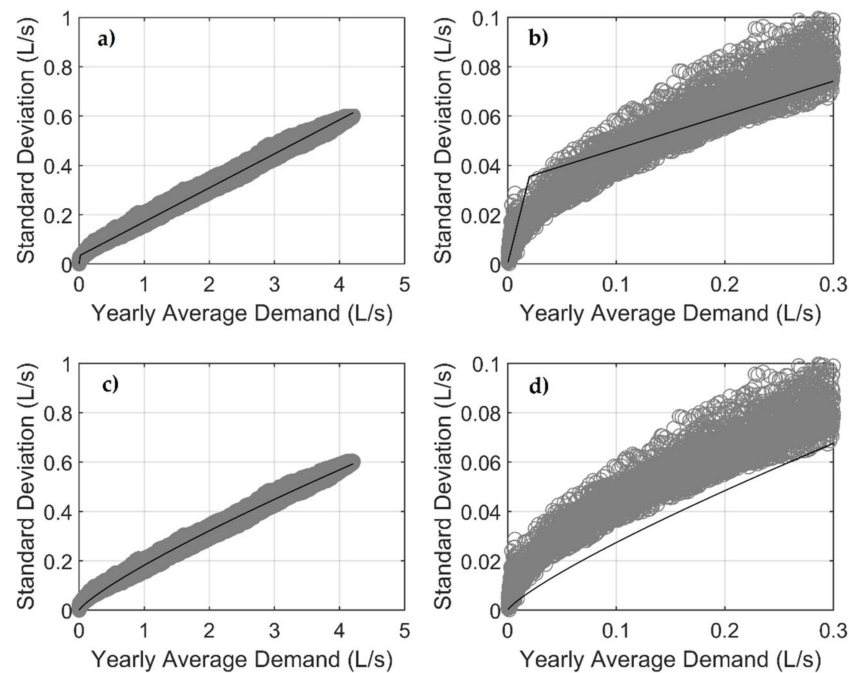


Figure 8. (a) Standard deviation σ in the peak demand D as a function of the yearly average demand D^y , as obtained from calibration dataset (grey dots) and from best-fit equations of this work (black lines); (b) zoom of (a) into the lower values of D^y and σ ; (c) standard deviation σ in the peak demand D as a function of the yearly average demand D^y , as obtained from calibration dataset (grey dots) and from calibrating the theoretical scaling law with the Soccavo data (black lines); (d) zoom of (c) into the lower values of D^y and σ .

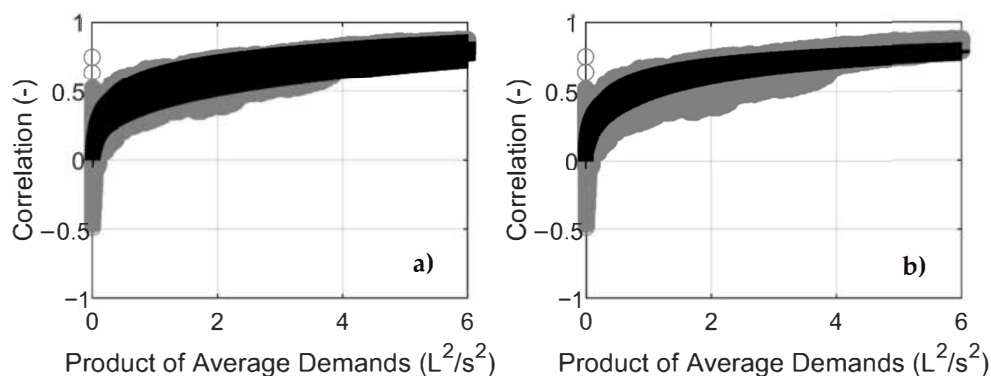


Figure 9. As a function of the product $prod$ of yearly average demands D^y , comparison of rank cross-correlations ρ between peak demands D obtained from calibration datasets (grey dots) with the values obtained through (a) the best-fit equations of this work (black crosses) and through (b) the calibration of the theoretical scaling law on the Soccavo district (black crosses).

As a result of the comparison shown in Figures 8 and 9, best-fit scaling equations with a better goodness of fit than the calibrated theoretical scaling laws can be obtained to relate the statistical parameters of consumption to the yearly average demand.

Another issue to be discussed lies in the comparison of the bottom-up approach for generating peak demand scenarios with the traditional deterministic top-down approach, based on applying amplificative peak coefficients on the yearly average demands D^y at WDN nodes. In case study 2a, in the case of nodal demands reconstructed using the optimization-based resorting, the peak demand coefficient was evaluated at the WDN nodes as the ratio of the peak demand on the yearly average demand. This was done in 20 demand scenarios featuring similar values of the cumulative frequency F close to 0.5 evaluated on the whole network demand (see graph c in Figure 6). As Table 7 shows, the peak demand coefficient in the whole WDN was about 1.81 in all scenarios. The nodal demand coefficients were, instead, quite variable around this value.

To better explore the difference of the bottom-up approach presented in this paper and the deterministic top-down approach, Figure 10 presents a comparison in terms of histogram of frequencies of the peak demand coefficient at node 1 in case study 2a. Graph (a) shows the histogram of frequencies obtained through the bottom-up approach, whereas graph (b) shows that obtainable through the deterministic top-down approach, by applying to all WDN nodes the same demand coefficient as the whole WDN. The comparison points out that the use of the deterministic top-down approach leads to the underestimation of the range of variation in the demand coefficient.

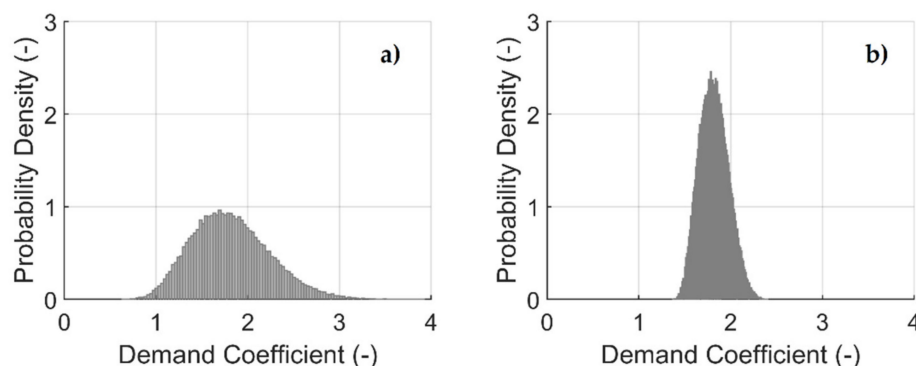


Figure 10. Histogram of frequencies of peak demand coefficient at node 1 in case study 2a, as was obtained (a) through the bottom-up stochastic approach presented in this paper and as would be obtained (b) through the deterministic top-down approach.

Table 7. In case study 2a with nodal demands reconstructed based on the optimization-based resorting, peak demand coefficient (-) at the nodes and in the whole network in various demand scenarios featuring similar values of the cumulative frequency $F = 0.5$ evaluated on the whole network demand.

Scenario	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Node 7	Node 8	Node 9	Node 10	Node 11	Node 12	Total
1	1.77	2.24	1.42	1.85	2.12	1.59	2.09	1.70	1.64	1.97	1.76	1.27	1.81
2	1.13	2.39	1.62	1.54	1.65	1.97	2.04	1.96	1.45	2.39	1.49	1.65	1.81
3	1.93	1.76	2.08	0.90	2.56	2.30	1.70	1.29	2.15	1.16	1.13	3.14	1.81
4	2.35	1.76	1.96	0.78	2.48	2.06	1.77	1.56	2.13	1.57	1.80	0.97	1.81
5	1.57	2.26	1.42	1.49	1.98	2.37	1.99	1.49	1.96	1.61	1.30	2.26	1.81
6	1.71	1.84	1.39	2.62	1.79	1.70	2.03	1.71	2.65	1.92	1.19	1.44	1.81
7	2.29	2.02	2.20	1.28	1.58	2.12	1.38	1.79	1.96	1.80	1.83	1.64	1.81
8	1.39	1.62	1.60	1.80	1.39	1.51	2.10	1.80	1.44	2.20	2.55	2.28	1.81
9	1.78	1.46	2.33	2.01	1.87	2.12	1.54	1.80	1.25	2.20	1.16	2.36	1.81
10	2.07	1.76	1.63	2.98	1.27	1.51	1.80	2.06	1.75	1.96	1.52	1.93	1.81
11	2.69	1.43	1.86	2.61	1.80	1.54	1.65	1.30	2.28	1.47	1.92	1.99	1.81
12	1.70	2.10	2.11	1.68	1.28	1.59	2.37	2.18	1.81	1.59	1.62	1.12	1.81
13	1.78	1.58	1.48	1.44	2.06	1.81	1.89	2.34	1.23	1.77	1.57	2.60	1.81
14	2.04	2.13	1.73	1.90	1.43	1.58	1.85	2.01	1.81	1.76	1.43	2.52	1.81
15	1.86	1.86	1.61	1.12	2.13	1.62	1.42	2.26	2.05	1.95	1.74	2.46	1.81
16	2.66	1.46	1.96	1.80	2.11	1.69	1.75	1.49	1.10	1.57	2.30	1.78	1.81
17	2.31	1.71	1.94	2.45	1.97	1.83	1.45	1.39	1.51	2.20	1.49	2.01	1.81
18	2.30	1.50	2.00	1.40	1.96	1.84	1.88	1.53	2.78	1.70	1.62	1.14	1.81
19	1.48	1.39	2.50	1.64	1.88	2.11	1.73	1.63	2.30	1.97	1.44	1.55	1.81
20	1.75	1.67	2.44	1.64	1.83	1.76	1.98	1.64	1.44	1.90	1.80	1.53	1.81
Min	1.13	1.39	1.39	0.78	1.27	1.51	1.38	1.29	1.10	1.16	1.13	0.97	1.81
Max	2.69	2.39	2.50	2.98	2.56	2.37	2.37	2.34	2.78	2.39	2.55	3.14	1.81

The results shown in Table 7 and in Figure 10 point out that the bottom-up generation presented in this work enables a much more realistic peak demand allocation than the standard top-down approach used in the engineering practice, which assigns the peak coefficient of the whole WDN to the single nodes. In this context, it must be noted that limitation is not inherent in the top-down approach but rather in the way practitioners apply it. In fact, there are notable examples in the scientific literature [28,29], where this limitation does not appear in the context of the top-down generation of demand time series. However, the top-down generation of peak demand scenarios has not been much explored so far.

5. Conclusions

The main conclusions of this work are the following:

1. The beta probability distribution model with tunable bounds is effective at representing the peak demand at different degrees of aggregation, while successfully preserving bounds and statistics up to the third order;
2. By making use of available data of measured consumption at WDN users, best-fit scaling equations can be constructed to relate mean, standard deviation, skewness, and rank cross-correlation of peak demands to the yearly average values;
3. An optimization-based re-sorting of data samples generated at each WDN node with the beta probability distribution can be used when there are not enough data to calibrate the scaling equations for estimating rank cross-correlations between nodal peak demands;
4. The bottom-up approach presented in this work yields more realistic demand scenarios than the deterministic top-down approach used in the engineering practice, based on applying single amplificative factors to all WDN nodes.

Author Contributions: Conceptualization, E.C. and M.F.; methodology, E.C. and G.G.; investigation, E.C. and G.G.; writing—original draft preparation, E.C.; writing—review and editing, G.G., A.C., and M.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: This research was conducted using the funds supplied by the Universities of Pavia, Catania and Ferrara.

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

Appendix A

The theoretical scaling laws are equations derived through a theoretical framework based on the assumption that water flow in a meter, corresponding to the water consumption of a unit user, is a random variable or realization of a stationary stochastic process e.g., see [19,27]. These laws relate the statistics D_{mean} , σ , and ρ of the aggregated users to the corresponding statistics $D_{mean,1}$, σ_1 , and ρ_1 of the typical unitary user, which must be calibrated based on monitored demands.

The theoretical scaling law for the average demand D_{mean} is:

$$D_{mean} = n_u D_{mean,1}. \quad (12)$$

in which n_u is the number of aggregated users. The following manipulations involving the introduction of the yearly average demand D^y of the aggregated users lead to Equation (9):

$$D_{mean} = \frac{D^y}{n_u} n_u D_{mean,1} = \left(\frac{n_u}{D^y} D_{mean,1} \right) D^y = par_1 D^y. \quad (13)$$

The theoretical scaling law for the variance σ^2 , i.e., the square of the standard deviation σ , is:

$$\sigma^2 = n_u^\theta \sigma_1^2 \quad (14)$$

in which θ is an exponent ranging from 1 (fully uncorrelated demands) to 2 (fully correlated demands). The following manipulations involving the introduction of the yearly average demand D^y lead to Equation (10):

$$\sigma = n_u^{\theta/2} \sigma_1 = \left(\frac{D^y}{n_u} \right)^{\theta/2} n_u^{\theta/2} \sigma_1 = \left[\frac{n_u^{\theta/2}}{(D^y)^{\theta/2}} \sigma_1 \right] (D^y)^{\theta/2} = par_2 (D^y)^{\theta/2}. \quad (15)$$

The theoretical scaling law for the rank cross-correlation ρ between demands at nodes k and l is:

$$\sigma \rho_{k,l} = \frac{n_{u,k} n_{u,l} \rho_1}{\sqrt{n_{u,k} [1 + \rho_1 (n_{u,k} - 1)]} \sqrt{n_{u,l} [1 + \rho_1 (n_{u,l} - 1)]}} \quad (16)$$

in which $n_{u,k}$ and $n_{u,l}$ are the number of users at node k and the number of users at node l , respectively. The following manipulations involving the introduction of the yearly average demand D^y for both nodes k and l lead to Equation (11):

$$\rho_{k,l} = \frac{n_{u,k} n_{u,l} \rho_1}{\sqrt{n_{u,k} [1 + \rho_1 (n_{u,k} - 1)]} \sqrt{n_{u,l} [1 + \rho_1 (n_{u,l} - 1)]}} = \frac{\left(\frac{D_k^y}{D_1^y} \right) \left(\frac{D_l^y}{D_1^y} \right) \rho_1}{\sqrt{\left(\frac{D_k^y}{D_1^y} \right) \left[1 + \rho_1 \left(\frac{D_k^y}{D_1^y} - 1 \right) \right]} \sqrt{\left(\frac{D_l^y}{D_1^y} \right) \left[1 + \rho_1 \left(\frac{D_l^y}{D_1^y} - 1 \right) \right]}} = \frac{\left(\frac{D_k^y}{par_3} \right) \left(\frac{D_l^y}{par_3} \right) par_4}{\sqrt{\left(\frac{D_k^y}{par_3} \right) \left[1 + par_4 \left(\frac{D_k^y}{par_3} - 1 \right) \right]} \sqrt{\left(\frac{D_l^y}{par_3} \right) \left[1 + par_4 \left(\frac{D_l^y}{par_3} - 1 \right) \right]}}. \quad (17)$$

In the reformulated theoretical scaling laws, parameters par_1 , par_2 , par_3 , par_4 , and θ must be calibrated on the field based on monitored demands.

References

1. Machell, J.; Mounce, S.R.; Boxall, J.B. Online modelling of water distribution systems: A UK case study. *Drink. Water Eng. Sci.* **2010**, *3*, 21–27. [\[CrossRef\]](#)
2. Okeya, I.; Kapelan, Z.; Hutton, C.; Naga, D. Online modelling of water distribution system using data assimilation. *Procedia Eng.* **2014**, *70*, 1261–1270. [\[CrossRef\]](#)
3. Creaco, E.; Pezzinga, G.; Savic, D. On the choice of the demand and hydraulic modeling approach to WDN real-time simulation. *Water Resour. Res.* **2017**, *53*, 6159–6177. [\[CrossRef\]](#)
4. Cahn, A. An Overview of Smart Water Networks. *J. Am. Water Works Assoc.* **2014**, *106*, 68–74. [\[CrossRef\]](#)
5. Walski, T.M.; Chase, D.V.; Savic, D.A.; Grayman, W.; Beckwith, S.; Koelle, E. *Advanced Water Distribution Modeling and Management*; Haestad Press: Waterbury, CT, USA, 2003.
6. Creaco, E.; Blokker, M.; Buchberger, S. Models for generating household water demand pulses: Literature review and comparison. *J. Water Resour. Plan. Manag.* **2017**, *143*, 04017013. [\[CrossRef\]](#)
7. Alvisi, S.; Ansaloni, N.; Franchini, M. Generation of synthetic water demand time series at different temporal and spatial aggregation levels. *Urban Water J.* **2014**, *11*, 297–310. [\[CrossRef\]](#)
8. Gargano, R.; Di Palma, F.; de Marinis, G.; Granata, F.; Greco, R. A stochastic approach for the water demand of residential end users. *Urban Water J.* **2016**, *13*, 569–582. [\[CrossRef\]](#)
9. Kossieris, P.; Tsoukalas, I.; Makropoulos, C.; Savic, D. Simulating Marginal and Dependence Behaviour of Water Demand Processes at Any Fine Time Scale. *Water* **2019**, *11*, 885. [\[CrossRef\]](#)
10. Creaco, E.; De Paola, F.; Fiorillo, D.; Giugni, M. Bottom-Up Generation of Water Demands to Preserve Basic Statistics and Rank Cross-Correlations of Measured Time Series. *J. Water Resour. Plan. Manag.* **2020**, *146*, 06019011. [\[CrossRef\]](#)
11. Harmon, W.G. Forecasting sewage at Toledo under dry weather conditions. *Eng. News Rec.* **1918**, *80*, 1233.
12. Babbitt, H.E. *Sewerage and Sewage Treatment*, 3rd ed.; Wiley: New York, NY, USA, 1928; pp. 20–33.
13. Department of Environment and Conservation, Water Resources Management Division. *Guidelines for the Design, Construction, and Operation of Water and Sewerage Systems*; Newfoundland and Labrador Department: St. John's, NL, Canada, 2005.
14. Metcalf, L.; Eddy, H.P. *American Sewerage Practice, Volume III: Design of Sewers*, 3rd ed.; McGraw-Hill: New York, NY, USA, 1935; Volume I.
15. Tricarico, C.; De Marinis, G.; Gargano, R.; Leopardi, A. Peak residential water demand. Proceedings of the Institution of Civil Engineers 2007. *Water Manag.* **2007**, *160*, 115–121. [\[CrossRef\]](#)
16. Gargano, R.; Tricarico, C.; Granata, F.; Santopietro, S.; De Marinis, G. Probabilistic Models for the Peak Residential Water Demand. *Water* **2017**, *9*, 417. [\[CrossRef\]](#)
17. Balacco, G.; Carbonara, A.; Gioia, A.; Iacobellis, V.; Piccinni, A.F. Evaluation of Peak Water Demand Factors in Puglia (Southern Italy). *Water* **2017**, *9*, 96. [\[CrossRef\]](#)
18. Del Giudice, G.; Di Cristo, C.; Padulano, R. Spatial Aggregation Effect on Water Demand Peak Factor. *Water* **2020**, *12*, 2019. [\[CrossRef\]](#)
19. Magini, R.; Boniforti, M.A.; Guercio, R. Generating Scenarios of Cross-Correlated Demands for Modelling Water Distribution Networks. *Water* **2019**, *11*, 493. [\[CrossRef\]](#)
20. Iman, R.L.; Conover, W.J. A distribution-free approach to inducing rank correlation among input variables. *Commun. Stat. Simul. Comput.* **1982**, *11*, 311–334. [\[CrossRef\]](#)
21. Fiorillo, D.; Galuppini, G.; Creaco, E.; De Paola, F.; Giugni, M. Identification of influential user locations for smart meter installation to reconstruct the urban demand pattern. *J. Water Resour. Plan. Manag.* **2020**, *146*, 04020070. [\[CrossRef\]](#)
22. Johnson, N.L.; Kotz, S.; Balakrishnan, N. Chapter 21: Beta distributions. In *Continuous Univariate Distributions Vol. 2*, 2nd ed.; Wiley-Interscience: New York, NY, USA, 1995; ISBN 978-0-471-58494-0.
23. Nocedal, J.; Wright, S.J. *Numerical Optimization*, 2nd ed.; Springer: Berlin, Germany; New York, NY, USA, 2006; ISBN 978-0-387-30303-1.
24. Creaco, E.; Alvisi, S.; Farmani, R.; Vamvakieridou-Lyroudia, L.; Franchini, M.; Kapelan, Z.; Savic, D. Methods for preserving duration-intensity correlation on synthetically generated water-demand pulses. *J. Water Resour. Plan. Manag.* **2016**, *142*. [\[CrossRef\]](#)
25. Nelsen, R.B. *An Introduction to Copulas*; Springer: New York, NY, USA, 1999.
26. Creaco, E.; Signori, P.; Papiri, S.; Ciaponi, C. Peak Demand Assessment and Hydraulic Analysis in WDN Design. *J. Water Resour. Plan. Manag.* **2018**, *144*, 04018022. [\[CrossRef\]](#)
27. Vertommen, I.; Magini, R.; Cunha, M. Scaling Water Consumption Statistics. *J. Water Resour. Plan. Manag.* **2015**, *141*, 04014072. [\[CrossRef\]](#)
28. Alvisi, S.; Ansaloni, N.; Franchini, M. Comparison of parametric and nonparametric disaggregation models for the top-down generation of water demand time series. *Civ. Eng. Environ. Syst.* **2016**, *33*, 3–21. [\[CrossRef\]](#)
29. Fiorillo, D.; Creaco, E.; Paola, F.D.; Giugni, M. Comparison of bottom-up and top-down procedures for water demand reconstruction. *Water* **2020**, *12*, 922. [\[CrossRef\]](#)