

## Demand uncertainty in modelling WDS: scaling laws and scenario generation

R. Magini<sup>1</sup>, F. Capannolo<sup>2</sup>, E. Ridolfi<sup>3</sup> & R. Guercio<sup>1</sup>

<sup>1</sup>*Department of Civil, Constructional and Environmental Engineering, Sapienza University of Rome, Italy*

<sup>2</sup>*Engineer MSc, Italy*

<sup>3</sup>*Department of Civil and Environmental Engineering, University of Perugia, Italy*

### Abstract

Water distribution systems (WDS) are critical infrastructures that should be designed to work properly in different conditions. The design and management of WDS should take into account the uncertain nature of some system parameters affecting the overall reliability of these infrastructures. In this context, water demand represents the major source of uncertainty. Thus, uncertain demand should be either modelled as a stochastic process or characterized using statistical tools. In this paper, we extend to the 3rd and 4th order moments the analytical equations (namely scaling laws) expressing the dependency of the statistical moments of demand signals on the sampling time resolution and on the number of served users. Also, we describe how the probability density function (pdf) of the demand signal changes with both the increase of the user's number and the sampling rate variation. With this aim, synthetic data and real indoor water demand data are used. The scaling laws of the water demand statistics are a powerful tool which allows us to incorporate the demand uncertainty in the optimization models for a sustainable management of WDS. Specifically, in the stochastic/robust optimization, solutions close to the optimum in different working conditions should be considered. Obviously, the results of these optimization models are strongly dependent on the conditions that are taken into consideration (i.e. the *scenarios*). Among the approaches for the definition of demand scenarios and their probability-weight of occurrence, the moment-



matching method is based on matching a set of statistical properties, e.g. moments from the 1st (*mean*) to the 4th (*kurtosis*) order.

*Keywords: water distribution systems, water demand, scaling laws, scenario generation, stochastic/robust optimization.*

## 1 Introduction

The conventional modeling of water distribution systems, WDS, is usually based on a deterministic approach, not merely with regard to the geometrical and hydraulic features, but also with respect to the demand loadings. Further, it must be taken into account that many data, e.g. roughness, diameter and length of pipes, exact form and location of junction and valves, etc., are not perfectly known.

At the same time, great effort has been invested by researchers in the development and improvement of optimization models for the design and management of WDS, as well as techniques to solve these models. However, these optimization models are usually based on the assumption of perfectly defined working conditions, leading to solutions which are optimal for the considered inputs, but may be unreliable if reality turns out to be different.

Among all the cited factors, the variability of the demand represents the major source of uncertainty (Pierleoni *et al.* [1], Di Francesco *et al.* [2]). Uncertainty inherent to water demand propagates into uncertain pressure heads and pipe flows, affecting the overall reliability of any kind of model describing a WDS.

With this aim, after verifying nontrivial scaling of the variance of real consumption data with spatial aggregation, Magini *et al.* [3] developed simple scaling laws relating the mean, variance, and covariance of water consumption series with the number of aggregated users. The development of the scaling laws relies on the assumption that water flowing in a meter, corresponding to the water consumption of a unit user, is a random variable or realization of a stationary stochastic process. The expected value for the mean consumption was found to increase linearly and the expected value for the variance and lag1 covariance of consumption was found to increase accordingly to an exponent between 1 and 2. Following up this work, Vertommen *et al.* [4] theoretically derived the scaling laws for the cross-covariance, and cross-correlation coefficient and showed that the exponent of the variance scaling law depends on the spatial correlation between consumptions. Statistical correlation between residential water demands was already proven to be not negligible and to affect the hydraulic performance of a WDS (Filion *et al.* [5]). The probabilistic description of the water demand, through the definition of the scaling laws, provides the ability to insert the demand uncertainty in the modelling of WDS.

In general, the scaling laws are useful both for descriptive and decision models, thus, they can be a powerful tool not only in stochastic programming models but also in models that want predict how a WDS, which either already exists or will be constructed, will behave.

In particular, stochastic optimization methods (stochastic programming, robust optimization) tackles the issue of uncertainty in water demand by considering different scenarios and by obtaining a solution which stays close to the optimum



for all of them (da Cunha and de Sousa [6]). The outcome of stochastic optimization depends on the scenarios that are considered and on their probability of occurrence. 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.

In response to these considerations, we propose not arbitrary approaches for the establishment of different demand scenarios and the mathematical determination of their probability/weight of occurrence. Therefore, water demand will be modelled as a stochastic variable using the scaling laws for the statistical moments and cross-correlation of nodal demands. Also different marginal pdf functions will be considered at each node depending on the number of nodal users.

In stochastic optimization, it is necessary to transform the multivariate probability distributions in multiple scenarios, representative of situations of risk against which the performance of the system is evaluated. This process is known as precisely scenario generation (Kaut and Wallace [7]) and different methods for generating scenarios exist (Dupacová *et al.* [8] and Mitra [9]).

Most of these methods try to generate scenarios that match a given set of specifications for their marginal distributions and the dependence between them, which is almost always specified through the correlation (or variance-covariance) matrix. At this aim, in this work the scaling laws for the 1st and 2nd order moments are extended to the 3rd order moment (skewness) and 4th order moment (kurtosis) analyzing multi-year series of experimental data of water consumption. Moreover, the probability density functions, pdf, at different spatial and temporal aggregation level are derived.

## 2 Skewness and kurtosis

Dealing with the demand uncertainty when modelling WDS, requires not only a complete statistical characterization of demand variability, but also the determination of the correlation between the different users and groups of users. The natural variability of demand can be expressed using probability density functions, pdf. A pdf is characterized by its shape (e.g. normal, exponential, gamma, etc.) and by specific parameters like the population mean and variance. Thus, in order to represent uncertain water demand using a pdf, it is necessary to identify and estimate the values of these parameters. The consideration of different spatial and temporal aggregation levels induces changes in the pdf parameters, often leading to a reduction of the uncertainty. The auto-correlation and cross-correlation that characterize the water demand signals affect the extent to which the pdf parameters vary, and can introduce an additional sensitivity to the specific period of observation in question.

In order to understand the effects of spatial aggregation and sampling intervals on the statistical properties of demand, it is possible to develop analytical expressions for the moments (mean, variance, cross-covariance and cross-correlation coefficient) of demand time series, at a fixed time sampling frequency



$\Delta t$ , of  $n$  aggregated users as a function of the moments of the single-user series sampled in the observation period  $T$ . These expressions are referred to as *Scaling Laws*, and can be expressed as

$$E[m_{\Delta t, T}(n)] = E[m_{\Delta t, T}] \cdot n^\alpha \cdot f(\Delta t, T) \quad (1)$$

where  $E[m_{\Delta t, T}(n)]$  is the expected value of the moment  $m$  for  $n$  users for the time interval  $T$ ;  $E[m_{\Delta t, T}]$  is the expected value of the moment  $m$  for the single-user for the same time interval;  $\alpha$  is the exponent of the scaling law; and  $f(\Delta t, T)$  is a function that expresses the influence of both sampling rate and observation period.

The development of the scaling laws is based on the assumption that the demand can be described by a homogeneous and stationary process, which implies that the  $n$  aggregated users are of the same type (residential, commercial, industrial, etc.), and that the statistical properties of demand, mean and variance, can be assumed constant in time. Neglecting the bias that can arise when using small demand series (e.g. due to short observation periods; Vertommen *et al.* [10]) the expected value for the 2nd order moment, variance, can be expressed as:

$$E[\sigma_n^2] = \mathbf{1} \times \mathbf{M2} \times \mathbf{1}^T \quad (2)$$

where  $E[\sigma_n^2]$  is the expected variance of  $n$  aggregated demands,  $\mathbf{1}$  is the all-ones vector of size  $(1, n)$  and  $\mathbf{M2}$  the  $(n, n)$  covariance matrix of the  $n$  single-user demand signals. From eqn. (2), being the generic term of  $\mathbf{M2}$ ,  $c_{i,j} = \rho_{i,j} \sigma_i \sigma_j$ , where  $\rho_{i,j}$  is the Pearson correlation coefficient, Magini *et al.* [3] and Vertommen *et al.* [4] showed that the exponent  $\alpha$  of the scaling law for the variance varies between  $\alpha = 1$  (uncorrelated signals) and  $\alpha = 2$  (perfectly correlated signals).

Following the same procedure used in Magini *et al.* [3] and Vertommen *et al.* [4] to derive eqn. (2), similar equations can be obtained for the third and fourth moments, that are:

$$E[m_n^3] = \mathbf{1} \times \mathbf{M3} \times (\mathbf{1}^T \otimes \mathbf{1}^T) \quad (3)$$

$$E[m_n^4] = \mathbf{1} \times \mathbf{M4} \times (\mathbf{1}^T \otimes \mathbf{1}^T \otimes \mathbf{1}^T) \quad (4)$$

where  $E[m_n^3]$  is the expected 3rd order moment of  $n$  aggregated demands, skewness,  $\otimes$  the Kronecker product,  $\mathbf{M3}$  the  $(n, n^2)$  co-skewness matrix of the  $n$  single-user demand signals,  $E[m_n^4]$  the expected 4th order moment of  $n$  aggregated demands and  $\mathbf{M4}$  the  $(n, n^3)$  co-kurtosis matrix of the  $n$  single-user demand signals.

From eqns (3) and (4) it is not straightforward to derive the exponent  $\alpha$  of the scaling law (eqn. (1)), therefore, in the following we present the results obtained analyzing synthetically generated data and multi-year real data of water consumption.



## 2.1 Analysis of synthetically generated data

In order to verify the influence of the correlation among the demand data on the scaling of the 3rd and 4th order moments 30 single-user demand series are generated using a Monte Carlo simulation. At this aim, water demand is modelled as a random variable with a gamma pdf, mean = 0.50 l/min and variance = 2.44 l/min. For the cross-correlation coefficient three distinct values are considered, namely,  $\rho = 1 \times 10^{-9}$ ,  $\rho = 0.01$  and  $\rho = 1.00$ . The first value of the cross-correlation coefficient corresponds to virtually uncorrelated demands. The second value corresponds to a more realistic situation, in which some correlation between demand series exists, and the last one represents perfectly correlated demands. Using eqns (2), (3) and (4) the statistical moments at different aggregation levels are evaluated. The 3rd moment is a measure of the asymmetry of the probability distribution, whilst the 4th moment is a measure of whether the data are heavy-tailed or light-tailed.

Figure 1 presents the trends of the 3rd and 4th order moments changing the correlation coefficient  $\rho$ . Expressing these moments with the power laws

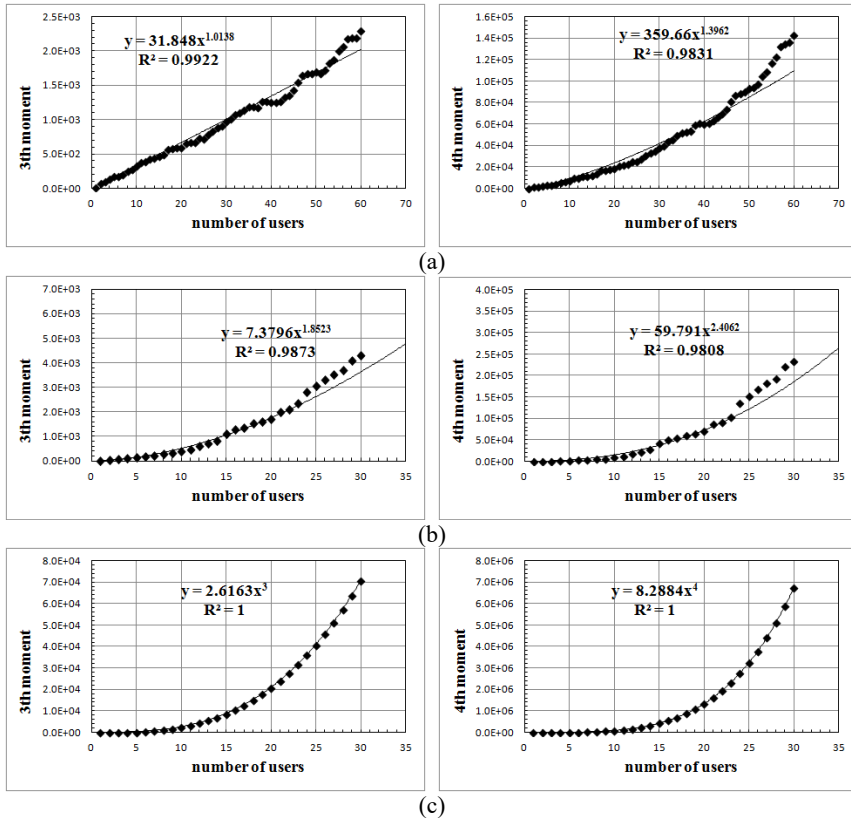


Figure 1: Scaling of the 3rd and 4th statistical moment: (a)  $\rho = 1 \times 10^{-9}$ , (b)  $\rho = 0.01$ , (c)  $\rho = 1.00$ .



$E[\sigma_n^3] = \sigma_1^3 n^\beta$  and  $E[\sigma_n^4] = \sigma_1^4 n^\gamma$ , the exponents  $\beta$  and  $\gamma$  respectively vary in the range 1.0–3.0 and 1.0–4.0, going from completely uncorrelated demands to perfectly correlated demands.

More frequently, instead of the third and fourth moment the corresponding dimensionless coefficients of skewness and kurtosis are used. Data series with positive skewness generally tend to have heavy tails on the right side. On the contrary, data with negative skewness generally tend to have heavy tails on the left side. Data series with high kurtosis tend to have heavy tails, or outliers. Whilst, data with low kurtosis tend to lack of outliers: the uniform distribution is the extreme case. The exponent  $\beta'$  of power laws for skewness ranges in between  $-0.5-0.0$  and the exponent  $\gamma'$  for kurtosis ranges in between  $-1.0-0.0$ , going from completely uncorrelated demands to perfectly correlated demands (Figure 2).

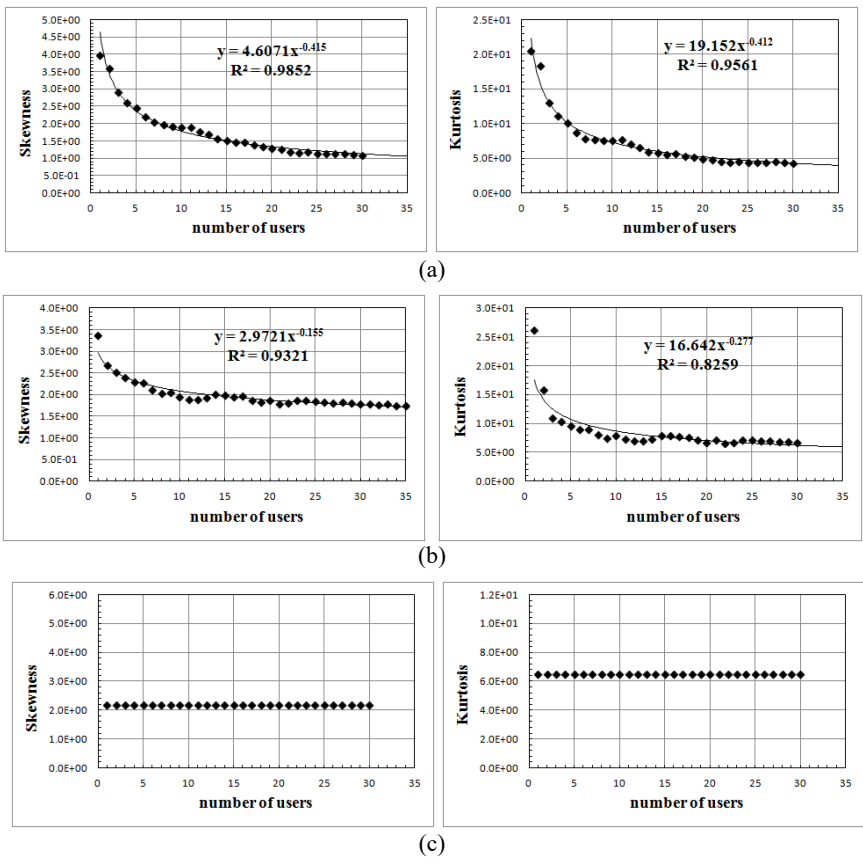


Figure 2: Scaling of skewness and kurtosis coefficients: (a)  $\rho = 1 \times 10^{-9}$ , (b)  $\rho = 0.01$ , (c)  $\rho = 1.00$ .

## 2.2 Analysis of observed data

The scaling of the skewness and kurtosis coefficient is also investigated for a set of observed demand data. The indoor water uses demand series of 82 single-family homes, with a total of 177 inhabitants, in a building belonging to the IIACP (Italian Association of Council Houses) in the town of Latina are considered (Guercio *et al.* [11]; Pallavicini and Magini [12], Vertommen *et al.* [4], Ridolfi *et al.* [13]). The apartments are inhabited by single-income families, belonging to the same low socioeconomic class. This circumstance enabled to consider a group of users substantially homogeneous in terms of lifestyle and water consumption habits, thereby making the assumption of a homogeneous process acceptable. The results are shown in Figure 3. It is evident that when the sampling time increases both the skewness and the kurtosis gets the typical trends of strongly correlated data.

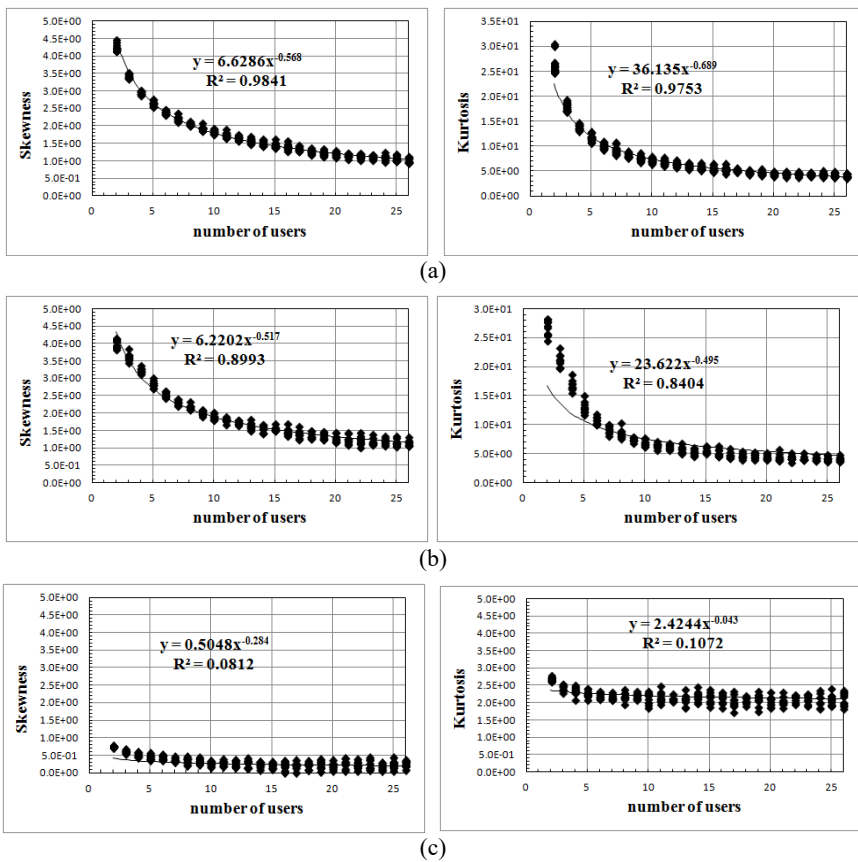


Figure 3: Scaling of skewness and kurtosis coefficients for real data for different sampling time  $dt$ : (a)  $dt = 20s$ , (b)  $dt = 60s$ , (c)  $dt = 600s$ .

In Table 1, the hourly average values of the coefficient and the exponent of the scaling law obtained from the analysis of three-week data are reported. Figure 4 shows the pdf of the demand signal measured at 6 am considering different numbers of users. It shows that when the number of users grow the exponential distribution is replaced, first, by the GEV distribution and then, by the gamma distribution.

Table 1: Hourly average values of the coefficient and exponent of the scaling law at different sampling times for hourly demand.

	<i>dt</i> = 20s		<i>dt</i> = 60s		<i>dt</i> = 600s	
	<b>A</b>	<b><math>\alpha</math></b>	<b>A</b>	<b><math>\alpha</math></b>	<b>A</b>	<b><math>\alpha</math></b>
Variance	1.62	1.03	1.39	1.04	0.58	1.08
Skewness	5.17	-0.51	4.21	-0.50	0.85	-0.35
Kurtosis	20.82	-0.67	12.55	-0.30	2.40	-0.03

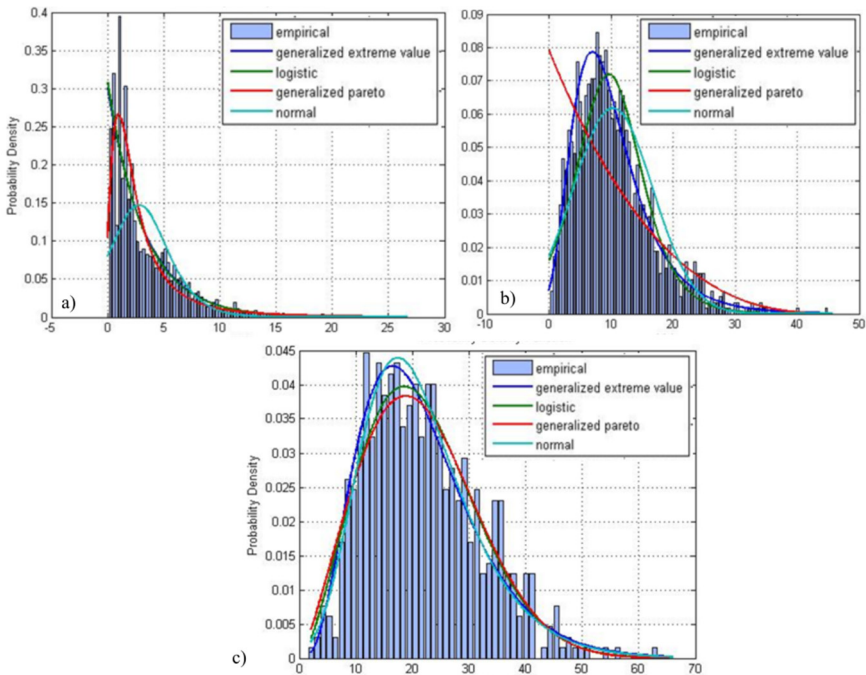


Figure 4: Probability density functions of real data for different number of aggregated users at 6 am: (a) 8 users, (b) 24 users, (c) 52 users.





### 3 Scenario generation

The knowledge of the scaling laws of the statistical moments and the shape of the probability distributions of water demand in relation to the number of users, is a useful tool to face the inherent uncertainty of demand in modeling the WDS and in particular to address the optimization problems for both design and management. Several methods are available for dealing with uncertain parameters in optimization problems: sensitivity analysis, stochastic programming and robust optimization. The choice of the scenarios is a critical step in obtaining meaningful solutions with the right compromise between system performance and robustness to variations in uncertain parameters.

If the statistical features of the uncertain parameter are known, numerical solutions can be obtained by approximating the pdf with discrete distributions having a finite number of outputs: scenarios. In this case scenarios can be generated following different methods. In one group of scenario generation methods, scenarios are derived by sampling from assumed marginal distributions. Techniques based on Monte-Carlo simulation and covariance matrix can be used (Ridolfi *et al.* [14], Vertommen *et al.* [15]). Another group of methods are based on matching a small set of statistical properties, e.g. moments. These methods can in turn follow two different paths (Ponomareva and Date [16] and Ponomareva *et al.* [17]). In the first path the statistical properties of the joint distribution are specified in terms of moments, usually including the covariance matrix. In the second group, specified (parametric) marginal distributions are sampled independently and the samples are then used along with Cholesky factorization of the covariance matrix to generate the necessary multivariate distribution (Lurie and Goldberg [18]).

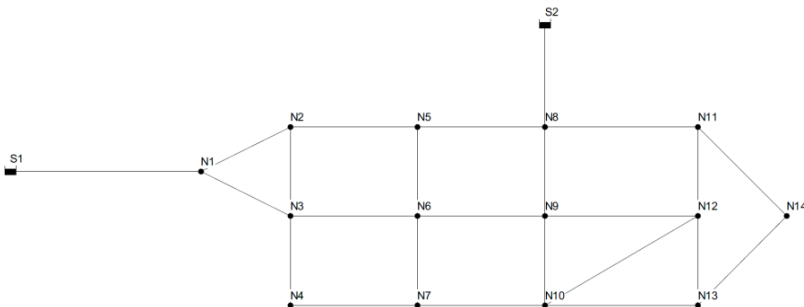


Figure 5: WDS considered for demands scenario generation.

In the following the moment-matching method proposed by Date *et al.* [19] is applied for generating scenarios at the test WDS in Figure 5. This system is not used for simulation purposes, but only for setting the scene. The network comprises 14 nodes. Two tests are carried out: the following number of users per node is considered, in test (a):  $n = 40, 47, 28, 25, 43, 36, 35, 32, 40, 45, 26, 55, 38, 50$ ; in test (b):  $n = 400, 470, 280, 250, 430, 360, 350, 320, 400, 450, 260, 550, 380, 500$ .

500. The algorithm is based on convex optimization which matches exactly the mean, covariance matrix and marginal (zero) skewness of a symmetric distribution and also matches the marginal fourth moments approximately (by minimizing the worst case error between the achieved and the target marginal fourth moments (Date *et al.* [19]). Each scenario has its own weight, which constitutes a probability measure within the number of generated scenarios.

The scenarios generated with a low number of users show greater fluctuations around the average values with respect to those related to a higher number of users. In both cases, the statistical moments have been derived using the scaling laws.

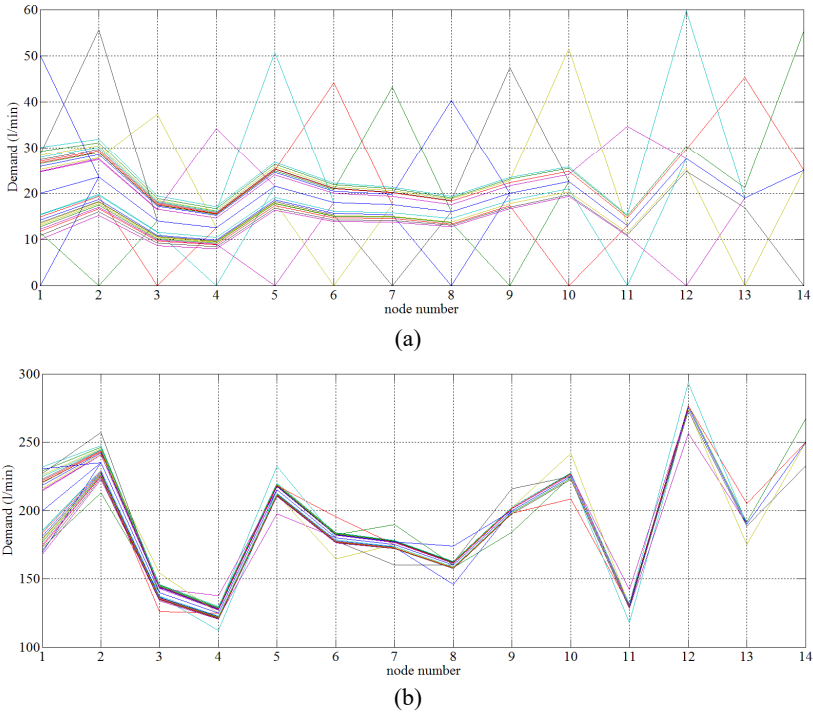


Figure 6: Generated scenarios: (a) low number of users per node; (b) high number of users per node.

## 4 Conclusions

In this paper, we have extended to the 3rd and 4th order moments the analytical equations, namely scaling laws, expressing the dependency of the statistical moments of demand signals on the sampling time resolution and on the number of served users. Analysis on synthetically generated and experimental data have shown that the scaling laws of the 3rd and 4th statistical moments depend on the Pearson correlation coefficient. In particular, the exponents of the power laws for the 3rd and 4th moment vary, respectively, in the range 1.0–3.0 and 1.0–4.0 and



for the skewness and kurtosis, respectively, in the range  $-0.5-0.0$  and  $-1.0-0.0$ , going from completely uncorrelated demands to perfectly correlated demands. Also, we have described how the pdf of the demand signal changes when the users number increases, ranging from the exponential to the gamma distribution. These results can be a useful tool in generating scenarios for WDS modelling and in particular in the moment-matching approach, as showed applying the method proposed by Date *et al.* [19].

## References

- [1] Pierleoni, A., Bellezza, M., Casadei, S., Manciola, P., Multipurpose water use in a system of reservoirs, in: IAHS-AISH Publ., 2007: pp. 107–116.
- [2] Di Francesco, S., Biscarini, C., Montesarchio, V., Manciola, P., On the role of hydrological processes on the water balance of Lake Bolsena, Italy, *Lakes and Reservoirs: Research and Management*, 2016, 21(1): 45–55, doi: 10.1111/lre.12120.
- [3] Magini, R., Pallavicini, I., Guercio, R., Spatial and Temporal Scaling Properties of Water Demand, *Journal of Water Resources Planning and Management*. 134: 276–284, doi: 10.1061/(ASCE)0733-9496(2008)134:3(276), 2008.
- [4] Vertommen, I., Magini, R., da Conceição Cunha, M., Scaling Water Consumption Statistics, *Journal of Water Resources Planning and Management*. 141: 04014072, doi: 10.1061/(ASCE)WR.1943-5452.0000467, 2015.
- [5] Filion, Y., Adams, B., Karney, B., Cross Correlation of Demands in Water Distribution Network Design, *Journal of Water Resources Planning and Management*. 133: 137–144, doi: 10.1061/(ASCE)0733-9496(2007)133:2(137), 2007.
- [6] Cunha, M. da C., Sousa, J.J. de O., Robust Design of Water Distribution Networks for a Proactive Risk Management, *Journal of Water Resources Planning and Management*. 136: 227–236, doi: 10.1061/(ASCE)WR.1943-5452.0000029, 2010.
- [7] Kaut, M., Wallace, S.W., Evaluation of scenario-generation methods for stochastic programming, *Stochastic*: 2003.
- [8] Dupacová, J., Consigli, G., Wallace, S.W., Scenarios for multistage stochastic programs, *Annals of Operations Research*. 100: 25–53, 2000.
- [9] Mitra, S., Scenario generation for stochastic programming, *OptiRisk S*: 2008.
- [10] Vertommen, I., Magini, R., da Conceição Cunha, M., Demand uncertainty in modelling water distribution systems, in: *WIT Trans. Built Environ.*, 2012: pp. 3–13.
- [11] Guercio, R., Magini, R., Pallavicini, I., Temporal And Spatial Aggregation In Modeling Residential Water Demand, in: *WIT Trans. Ecol. Environ.*, WIT Press, 2003.
- [12] Pallavicini, I., Magini, R., Experimental analysis of residential water demand data: Probabilistic estimation of peak coefficients at small time



- scales, in: Proc. Comb. Int. Conf. Comput. Control Water Ind. CCWI2007 Sustain. Urban Water Manag. SUWM2007, 2007: pp. 379–384.
- [13] Ridolfi, E., Servili, F., Magini, R., Napolitano, F., *et al.*, Artificial Neural Networks and Entropy-based Methods to Determine Pressure Distribution in Water Distribution Systems, *Procedia Engineering*. 89: 648–655, doi: 10.1016/j.proeng.2014.11.490, 2014.
- [14] Ridolfi, E., Vertommen, I., Magini, R., Joint probabilities of demands on a water distribution network: A non-parametric approach, in: AIP Publishing, 2013: pp. 1681–1684.
- [15] Vertommen, I., Magini, R., Cunha, M. da C., Generating Water Demand Scenarios Using Scaling Laws, *Procedia Engineering*. 70: 1697–1706, doi: 10.1016/j.proeng.2014.02.187, 2014.
- [16] Ponomareva, K., Date, P., Higher order sigma point filter: A new heuristic for nonlinear time series filtering, *Applied Mathematics and Computation*. 221: 662–671, doi: 10.1016/j.amc.2013.06.084, 2013.
- [17] Ponomareva, K., Roman, D., Date, P., An algorithm for moment-matching scenario generation with application to financial portfolio optimisation, *European Journal of Operational Research*. 240: 678–687, doi: <http://dx.doi.org/10.1016/j.ejor.2014.07.049>, 2015.
- [18] Lurie, P.M., Goldberg, M.S., An Approximate Method for Sampling Correlated Random Variables from Partially-Specified Distributions, *Management Science*. 44: 203–218, doi: 10.1287/mnsc.44.2.203, 1998.
- [19] Date, P., Mamon, R., Jalen, L., A new moment matching algorithm for sampling from partially specified symmetric distributions, *Operations Research Letters*. 36: 669–672, doi: 10.1016/j.orl.2008.07.004, 2008.

