Demand uncertainty in modelling water distribution systems

I. Vertommen[1,2], R. Magini[2] & M. da Conceição Cunha[1]
[1]Department of Civil Engineering, University of Coimbra, Portugal
[2]Department of Civil, Building and Environmental Engineering,
La Sapienza University of Rome, Italy

Abstract

Water demand, which is perhaps the main process governing water distribution systems (WDS), is affected by natural variability. The inherent uncertainty of demand is not negligible. Thus, uncertain demand should be modelled as a stochastic process or described using statistical tools. The stochastic modelling of water demand requires knowledge of the statistical features of the demand time series at different spatial and temporal scales. With this aim, this paper presents a stochastic description of demand and discusses in which measure its statistical properties depend on the level of spatial and temporal aggregation. The analytical equations, expressing the dependency of the statistical moments of demand signals on the sampling time resolution and on the number of served users, namely the 'scaling laws', are theoretically derived and discussed. These relationships have reference to the mean-variance scaling or Taylor's power law. The scaling laws are also validated using real water demand data of residential users. Through the scaling laws the statistical properties of the overall demand at each node of the WDS can be derived and the direct simulation of overall nodal demands can be done, reducing, among other things, the computational time in modelling or performing Monte Carlo sampling of these systems.
Keywords: water distribution systems, water demand, stochastic processes, correlation, variance, scaling.

1 Introduction

Recent studies on uncertainty in water distribution systems (WDS) indicate that nodal demands are the most significant inputs in hydraulic and water quality

models [1]. Thus, for a correct and realistic design and management, as well as simulation and performance assessment of WDS it is essential to have accurate values of water demand. The accurate knowledge of temporal and spatial variability of nodal demand is a fundamental pre-requisite for a risk-based approach in designing and managing WDS. The demand for water is influenced by various factors, from geographic, climatic and socioeconomic conditions, to cultural habits. Water demand uncertainty is made of both aleatory or inherent uncertainty, due to the natural and unpredictable variability of demand in space and time, and epistemic or internal uncertainty, due to a lack of knowledge about it. In particular, Hutton *et al.* [2] distinguish epistemic uncertainty in two types: the first one concerns the nature of the demand patterns, and what we do not know about this inherent variability when modelling WDN in both time and space. This uncertainty is defined as 'two-dimensional' uncertainty since it is composed by both aleatory and epistemic uncertainty. It may be constrained with extended and expensive spatial and temporal data collection or through the employment of descriptive/predictive water demand models. The second type of epistemic uncertainty takes the spatial allocation of water demand into account when modelling WDS [3].

Dealing with aleatory 'two-dimensional' uncertainty when modelling WDS, requires not only a complete statistical characterization of demand variability, but also the determination of the correlation among the different users and groups of users. In fact, statistical correlation of residential indoor water demands was proved to be not negligible [4] and to affect the hydraulic performance of WDS as well as its cost to achieve a desired level of reliability [5]. Li *et al.* [6] studied the spatial correlation of demand series generated by Poisson Rectangular Pulse (PRP) models. They verified that while time averaged demands that follow a homogeneous PRP process are uncorrelated, demands that follow a non-homogenous PRP process are correlated, and that this correlation increases with spatial and temporal aggregation. A similar conclusion about the correlation was achieved by Moughton *et al.* [7] from field measurements. After analysing the spatial and temporal features of water demand data for different spatial aggregation scales and sampled with different temporal resolutions, Magini *et al.* [8] summarized the aleatory uncertainty in mean-variance scaling relationships. The authors derived the 'scaling laws' for the first and second order moments, pointing out the role of the space–time correlation and how the level of spatial aggregation affects the statistical properties of the water demand series.

This work addresses the need to understand in which measure the statistical parameters of water demand signals depend on the number of aggregated users and on the temporal resolution in which they are estimated. It intends to describe these dependencies through the 'scaling laws', in order to derive the statistical properties of the total demand of a group of users from the features (mean, variance and correlation) of the demand process of a single-user. Special attention will be paid to the scaling of cross-correlation. Being part of the first author's PhD research, which aims the development of descriptive and predictive models for water demand that provide insight into peak demands, extreme events

and correlations at different spatial and temporal scales, these models will, in future stages, be incorporated in decision models for design purposes and scenario evaluations. Through this approach, we hope to develop more realistic and reliable WDS design and management solutions.

2 Theoretical framework on scaling laws

In order to understand the effects of spatial aggregation and sampling intervals on the statistical properties of demand, it is useful to develop analytical expressions for the moments (mean, variance, cross-covariance and cross-correlation coefficient) of demand time series, measured with a fixed sampling frequency, of n aggregated users as a function of the moments of the single-user series sampled at a different rate. These expressions are referred to as *"Scaling Laws"*, and can be expressed as:

$$E[m_T(n)] = E[m] \cdot n^\alpha \cdot f(T) \tag{1}$$

where $E[m_T(n)]$ is the expected value of the moment m for n users from signals measured with a sampling rate T; $E[m]$ is the expected value of the moment m for the single-user; α is the exponent of the scaling law; and $f(T)$ is a function that expresses the influence of the sampling rate.

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. The scaling laws for the mean, variance, and lag1 covariance were derived by Magini *et al.* [8] and validated with synthetic and real demand data. They can be also expressed in the form of the Taylor's power law [9], which is, as mean-variance/covariance relationships. Regarding the mean demand, results show a linear increase with the number of users, without any influence of the sampling time interval. Concerning the variance of demand, a non-linear increase with the number of users was verified. Theory shows that the exponent of the scaling law for the variance depends on the spatial correlation that exists between the single-user demands: if demands are uncorrelated in space, the scaling law is linear, if they are perfectly correlated in space, the scaling law is quadratic. In this case, the sampling time interval plays a significant role, and the expected value of the variance for n aggregated users, neglecting the narrowness of the sample, can be written as:

$$[\sigma_T^2(n)] = (n \cdot E[\sigma^2] + n \cdot (n-1) \cdot E[cov]) \cdot \gamma(T) \tag{2}$$

where $E[\sigma_T^2(n)]$ is the expected value of the variance for n users, for the sampling time interval T; $E[\sigma^2]$ is the expected variance of the single-user, for a certain sampling time interval; α is the exponent of the scaling law; $E[cov]$ the expected value of the cross-covariance between the single-user demands, for a certain sampling time interval; and $\gamma(t)$ is the variance function proposed by

VanMarcke [10] which expresses the influence of the sampling time interval. The variance function itself can be approximated by [10]:

$$\gamma(T) \cong \left[1 + \left(\frac{T}{\theta}\right)^m\right]^{-1/m} \tag{3}$$

where m is a model index parameter and θ is the scale of fluctuation of the stochastic process. The scale of fluctuation characterizes the correlation structure of the demand process, and can be considered constant when the process is stationary. Large values of the scale of fluctuation are related to long-memory processes, while small values of this parameter refer to no-memory processes.

The expected value of the cross-covariance between the demands of n_a aggregated users of group A and n_b aggregated users of group B can be given by:

$$E\big[Cov_{AB,T}(n_a, n_b)\big] = n_a \cdot n_b \cdot E\big[\rho_{ab,T}\big] \cdot E\big[\sigma_{a,T}\big] \cdot E\big[\sigma_{b,T}\big] \tag{4}$$

where, $E\big[\rho_{ab,T}\big]$ is the expected Pearson cross-correlation coefficient between the single-user demands of the two groups; and σ_a and σ_b are the standard deviations of the single-user demands of group A and B, respectively. The expected value of the cross-covariance increases accordingly to the product between the numbers of users of each group. In the particular case in which both groups have the same statistical properties, i.e., they belong to the same process, and assuming that $n_a = n_b$, the scaling law of the cross-covariance becomes quadratic.

The expected value of the Pearson cross-correlation coefficient between the demands of n_a aggregated users of group A and n_b aggregated users of group B, can be given by:

$$\begin{aligned}
E\big[\rho_{AB,T}(n_a, n_b)\big] &= \frac{E\big[Cov_{AB,T}(n_a, n_b)\big]}{E\big[\sigma_{A,T}(n_a)\big] \cdot E\big[\sigma_{B,T}(n_B)\big]} \\
&= \frac{n_a \cdot n_b \cdot E\big[\rho_{ab,T}\big]}{\sqrt{n_a\big(1 + E\big[\rho_{a,T}\big] \cdot [n_a - 1]\big)} \cdot \sqrt{n_b\big(1 + E\big[\rho_{b,T}\big] \cdot [n_b - 1]\big)}}
\end{aligned} \tag{5}$$

This equation represents the scaling law for the Pearson cross-correlation coefficient. It shows that this coefficient depends separately on the spatial aggregation levels of each group, n_a and n_b, and not only on their product as happens for the covariance. In order to take the different sampling rates into account, a function $f(T)$ could be considered in equations (4) and (5). If $n_a = n_b = n$ and if there is no difference in the sampling time intervals equation (5) becomes:

$$E\big[\rho_{AB}(n)\big] = \frac{n \cdot E[\rho_{ab}]}{\sqrt{\big[1 + (n - 1) \cdot E[\rho_a]\big]\big[1 + (n - 1) \cdot E[\rho_b]\big]}} \tag{6}$$

From equation (5) it is possible to observe that the expected value $E[\rho_{AB}(n_a, n_b)]$ increases with the number of users, n_a and n_b, reaching the following limit value:

$$E[\rho_{AB,lim}(n_a, n_b)] = \lim_{\substack{n_a \to \infty \\ n_b \to \infty}} E[\rho_{AB}(n_a, n_b)] = \frac{E[\rho_{ab}]}{\sqrt{(E[\rho_a] \cdot E[\rho_b])}} \qquad (7)$$

Since by definition $E[\rho_{AB,lim}] \leq 1$, the maximum value that the expected value of the cross-correlation coefficient between the single-user demands of group A and B can assume is:

$$E[\rho_{ab,max}] = \sqrt{(E[\rho_a] \cdot E[\rho_b])} \qquad (8)$$

From equation (6) it is possible to observe that the cross-correlation coefficient between the n aggregated users of group A and the n aggregated users of group B depends on both the cross-correlations in each group and the cross-correlation between the groups. Therefore, it seems interesting to investigate the way in which these two aspects, one at a time, affect its expected value as the number of aggregated users increases. In order to do so let us first considered a fixed value of ρ_{ab} and the values of ρ_a and ρ_b vary. Figure 1 shows graphically the results for $\rho_{ab} = 0.1$ and different pairs of ρ_a and ρ_b.
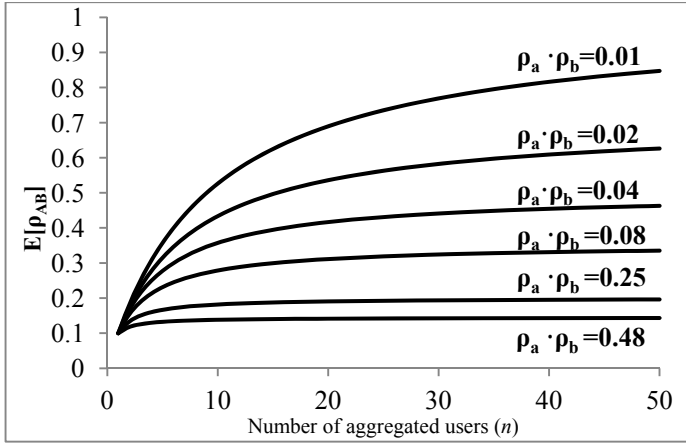

Figure 1:   Scaling laws of $E[\rho_{AB}(n)]$, for different values of $\rho_a \cdot \rho_b$.

As expected, all the curves have a common starting point, since ρ_{ab} is fixed. According to equation (7) a gradual flattening of the curves and a reduction of the shape ratio $\rho_{AB,lim}/\rho_{ab}$ can be noticed when the product $\rho_a \cdot \rho_b$ increases. Let us now consider a different case in which ρ_a and ρ_b are fixed and ρ_{ab} varies. The results are shown graphically in figure 2. The curves have now different starting points and equal shape ratios $\rho_{AB,lim}/\rho_{ab}$. Increasing ρ_{ab} produces only an upward shift of the curves, extending their transient.
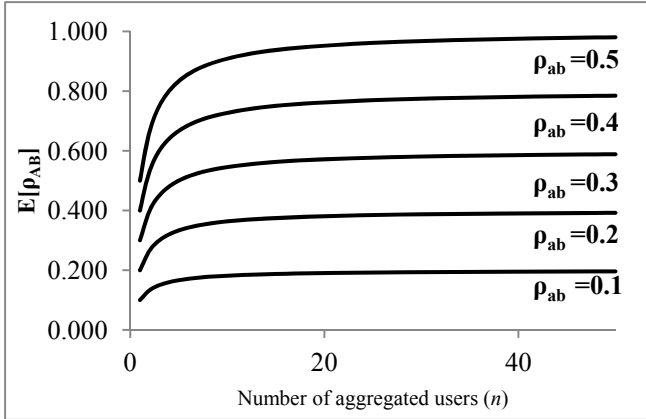
Figure 2:   Scaling laws of $E[\rho_{AB}(n)]$, for different values of $E[\rho_{ab}]$.

In the particular case in which both groups of users have the same statistical properties, i.e., they belong to the same process, and assuming $n_a = n_b = n$, the scaling law for the cross-correlation coefficient, considering no differences in the sampling time intervals, is:

$$E[\rho_{AB}(n)] = \frac{n \cdot E[\rho]}{1 + (n-1) \cdot E[\rho]} \qquad (9)$$

From equation (9) it is clear that the cross-correlation coefficient $E[\rho_{AB}(n)]$ increases with the number of aggregated users, tending to one. This limit value is reached as sooner as the correlation $E[\rho]$ between the single-user demands is higher.

3   Validation of the analytical expressions

In order to have a practical understanding of the developed analytical expressions, the scaling laws were derived for samples of real residential demand data. The real demand data consist in measurements taken from 82 residences in Latina, during 2 years. For each residence the different days in which data were measured can be considered as different realizations of the same process, allowing the artificial extend of the number of users to 300, while preserving their homogeneity. The demand series were divided into 1 hour periods to guarantee the stationarity of the process. The single-user demands were aggregated in groups of $n = 10, 20, ... ,150$ users, for different averaging time intervals, ranging from 1 second to 1 hour. The scaling laws for the variance, cross-covariance and cross-correlation coefficient were obtained for different sampling time intervals.

Figure 3 represents the variance of the demand series at the peak hour (8-9h), for different levels of spatial aggregation, and obtained at different sampling intervals.

As expected the exponents of the scaling laws for the variance lie between 1 and 2. The variance decreases with the sampling interval. The exponent of the
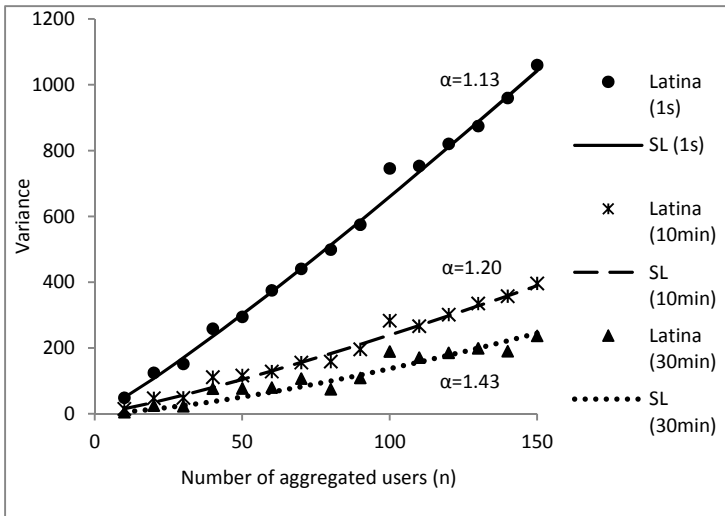
Figure 3:   Scaling Law for the variance for different sampling intervals, at peak hour.

scaling law increases with the sampling time, indicating that the correlation between demands increased.

The scaling laws obtained for the cross-covariance are close to quadratic.

The cross-correlation coefficient between demands increases with spatial aggregation and with the sampling interval. As an example, the scaling laws for the Pearson cross-correlation coefficient derived from the Latina data at different sampling intervals are illustrated in figure 4.

Figure 4:   Pearson cross-correlation coefficient of the data of Latina for different sampling intervals, at peak hour.

It is clear that the cross-correlation increases with the number of users and tends to one. The sampling interval influences the scaling law by shifting it upwards as the sampling time increases.

The theoretical scaling laws seem to provide a good fit for the Latina data, at all the considered sampling time intervals.

4 Variance function and scale of fluctuation

In order to verify the effect of the sampling time interval on the variance of demand, the variance was calculated at different sampling time intervals and for different groups of aggregated users. For each level of spatial aggregation, and each hour of the day, the index parameter m and the scale of fluctuation θ needed to be estimated. Two approaches were used to so. In the first approach the theoretical definition of the scale of fluctuation was used:

$$\theta = \int_{-\infty}^{+\infty} \rho(\tau) d\tau \tag{10}$$

where $\rho(\tau)$ is the auto-correlation function. An approximated value for the scale of fluctuation was obtained through the numerical integration of the auto-correlation function. Figure 5 represents the auto-correlation function for 10 aggregated users at the peak hour (8-9h).

In the second approach it was assumed that the index parameter m is constant and equal to one, and the scale of fluctuation was obtained from the variance function that provided the best fit to the real demand data.
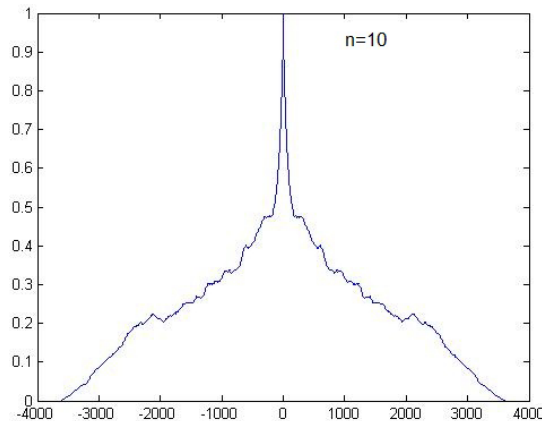

Figure 5:   Auto-correlation function for 10 aggregated users, at the peak hour.

The variance of the demand data from Latina at different sampling time intervals and for different numbers of aggregated users are represented in figures 6 and 7. In figure 6 the scale of fluctuation obtained through the auto-correlation function is used to obtain the scaling law. In figure 7 the scale of fluctuation obtained through the second approach is used instead.
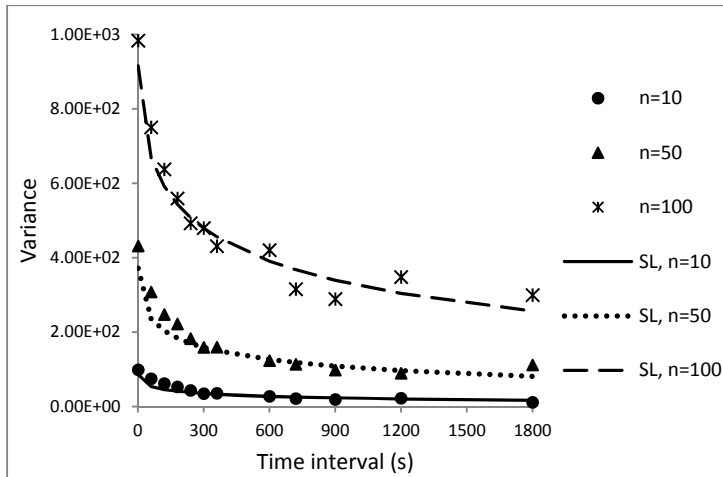
Figure 6:   Variance versus sampling time, with the scale of fluctuation obtained from the auto-correlation function, at peak hour.
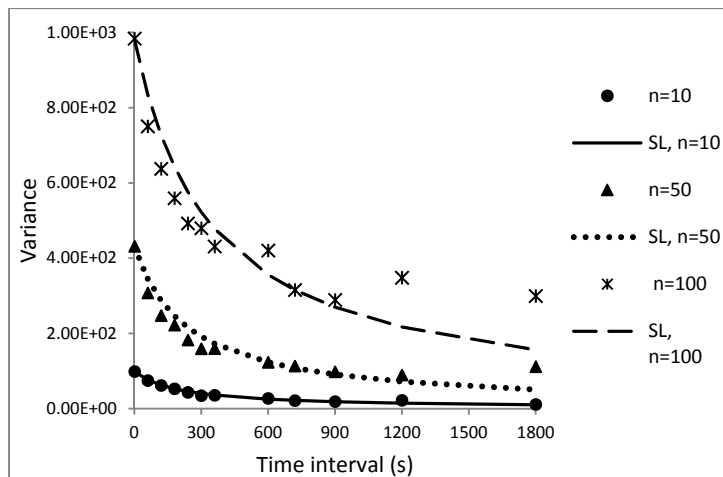

Figure 7:   Variance versus sampling time, with the scale of fluctuation obtained from the variance function, with $m = 1$, at peak hour.

The variance of demand decreases significantly with the increase of the sampling time. The theoretical expression used to describe this effect, based on the variance function provides a good fit to the real demand data from Latina.

The scale of fluctuation varies through the day depending on the demand in each hour: the scale of fluctuation increases with the mean demand. Thus, at the peak hour longer memory signals are observed.

The scales of fluctuation obtained through both approaches differ considerably.

5   Conclusions

The scaling laws were derived analytically and for real demand data. A good agreement is found between the theoretical expressions and the real demand data. Results show that the variance increases with spatial aggregation according to an exponent that varies between 1 and 2. In theory, for spatially uncorrelated demands the scaling laws is linear and for perfectly correlated demands the scaling law is quadratic. The variance suffers a reduction in function of the sampling time interval, since for longer observation intervals peak values are flattened out. The exponent of the scaling law for the variance increases with the sampling time, reflecting the expected increase in the correlation. The reduction of the punctual variance with the sampling time can be expressed through the variance function once the scale of fluctuation is obtained.

The scaling law for the cross-covariance between two groups of users increases according to the product between the numbers of users in each group. The cross-correlation coefficient depends separately on the number of users in each group, and increases towards a limit value equal to one. This limit is achieved sooner for larger sampling intervals, since the starting cross-correlation coefficient between single-user demands itself is higher. This fact is easy to understand whereas for larger observation times there is a greater probability of having simultaneous uses.

Nodal demand loads are a key factor for an accurate simulation of WDS. A realistic description of demand that reflects its uncertainty can only be done in statistical terms. Moreover, the achieved results show that the assessment of the spatial and temporal scaling effects on the statistical properties of demand cannot be neglected.

Acknowledgements

References

[1]   Pasha, M.F.K.a.L., K., Analysis of uncertainty on water distribution hydraulics and water quality, Impacts of Global Climate Change, in 2005 World Water and Environmental Resources Congress. 2005: Anchorage, Alaska.

[2]   Hutton, C.J., Vamvakeridou-Lyrouda, L.S., Kapelan, Z., Savic, D.A., Uncertainty Quantification and Reduction in Urban Water Systems (UWS) Modelling: Evaluation Report. 2011.

[3] Giustolisi, O. and E. Todini, *Pipe hydraulic resistance correction in WDN analysis.* Urban Water Journal, 2009. **6**(1): p. 39-52.

[4] Filion, Y.R., Karney, B. W., Moughton, L., Buchberger, S. G. and Adams, B. J., *Cross Correlation Analysis of Residential Demand in the City of Milford, Ohio.* 43, 2006.

[5] Filion, Y., B. Adams, and B. Karney, *Cross correlation of demands in water distribution network design.* Journal of Water Resources Planning and Management (ASCE), 2007. **133**(2): p. 137-144.

[6] Li, Z., *et al.*, Spatial correlation analysis of stochastic residential water demands. Water Management Chalenges in Global Change, Taylor and Francis Group, London, 2007: p. 363-370.

[7] Moughton, L.J., et al., Effect of time step and data aggregation on cross correlation of residential demands, in 8th Annual Simposium on Water Distribution Systems Analysis. 2006: Cincinnati, Ohio, USA.

[8] Magini, R., I. Pallavicini, and R. Guercio, *Spatial and temporal scaling properties of water demand.* Journal of Water Resources Planning and Management (ASCE), 2008. **134**(3): p. 276-284.

[9] Taylor, L.R., *Aggregation, variance and the mean.* Nature, 1961. **189**: p. 732-735.

[10] VanMarcke, E., *Random Fields Analysis and Synthesis (Revised and Expanded New Edition).* 2010 ed. 1983: World Scientific Publishing Co. Pte. Ltd. 350.