To be cited as: Vertommen, I., Magini, R., Cunha, M.C., "Scaling Water Consumption Statistics", Journal of Water Resources Planning and Management, 141 (5), 04014072, 2015. doi: 10.1061/(ASCE)WR.1943-5452.0000467

SCALING WATER CONSUMPTION STATISTICS

Ina Vertommen[1], Roberto Magini[2] and Maria da Conceição Cunha[3]

Abstract

Water consumption is perhaps the main process governing Water Distribution Systems. Due to its uncertain nature, water consumption should be modeled as a stochastic process or characterized using statistical tools. This paper presents a description of water consumption using statistics as mean, variance, and correlation. The analytical equations expressing the dependency of these statistics on the number of served users, the observation time and the sampling rate, namely the scaling laws, are theoretically derived and discussed. Real residential water consumption data are used to assess the validity of these theoretical scaling laws. Results show a good agreement between the scaling laws and the scaling behavior of real data statistics. The scaling laws represent an innovative and powerful tool, allowing to infer the statistical features of overall water

[1] Ph.D. Student, Department of Civil Engineering, Faculdade de Ciências e Tecnologia da Universidade de Coimbra, Rua Luís Reis Santos – Pólo II da Universidade, 3030-788 Coimbra, Portugal (corresponding author). E-mail: *ivertommen@dec.uc.pt.*

[2] Professor, Department of Civil, Building and Environmental Engineering, La Sapienza University of Rome, Via Eudossiana 18, 00184 Roma, Italia. E-mail: *roberto.magini@uniroma1.it*.

[3] Associate Professor, Department of Civil Engineering, Faculdade de Ciências e Tecnologia da Universidade de Coimbra, Rua Luís Reis Santos – Pólo II da Universidade, 3030-788 Coimbra, Portugal. E-mail: *mccunha@dec.uc.pt.*

consumption at each node of a network, from the process that describes the demand of a user unit without loss of information about its variability and correlation structure. This will further allow the accurate simulation of overall nodal consumptions, reducing the computational time when modeling networks.

Subject Headings

Water distribution systems, water use, statistics, correlation, scale effects, time series analysis.

Introduction

Optimal design and management solutions for Water Distribution Systems (WDS) can only be obtained when using accurate and realistic values of nodal consumptions. With the increasing computational capacity, consumption uncertainty and networks' reliability have become increasingly important in design practices. Residential use represents a significant proportion of the total consumption and is characterized by high variability, since it depends on many factors, known as explanatory variables, like climate, urban density, household size, water use policies, price and income (Polebitsky and Palmer, 2010). Moreover, even users belonging to the same type do not exhibit the same behavior every day. The conventional modeling of WDS considers deterministic consumptions at all nodes of the system. However, from the aforementioned reasons it seems evident that consumption is not deterministic, and its variability represents a great source of uncertainty when modeling WDS: uncertainty inherent to consumption propagates into uncertain pressure heads and flows, affecting the reliability of the system. A realistic approach for modeling WDS emerges from the explicit consideration of consumption uncertainty through its statistical characterization. In a probabilistic hydraulic analysis, nodal consumptions are assumed to be random variables, and their deterministic values are replaced by statistical information about them, such as the mean, variance and probability distributions, which express the uncertainty about the real value of the consumptions. A thorough statistical description of water consumption also requires the definition of the correlation between consumptions. Statistical correlation between residential indoor water consumptions was proved to be not negligible and to affect the hydraulic performance of a WDS (Filion *et al.*, 2007; Filion *et al.*, 2008). The probabilistic characterization of the performance of the network is thus essential for reliability purposes, but is difficult to solve. A considerable effort has been invested in developing methods and algorithms to solve this problem. However, the comprehension of the uncertainty itself has been overlooked. Quantities for the variance and correlation between nodal consumptions are always assumed; for instance, variance is mostly assumed to be 10% of the mean value (Kapelan *et al.,* 2005; Babayan *et al.*, 2004). Taking into account more realistic values for the uncertainty inherent to water consumption could significantly improve the optimization models.

Buchberger and Wu (1995) developed the first stochastic model for indoor water consumption, using three parameters: frequency, intensity and duration, characterized through a Poisson rectangular pulse process (PRP). Alvisi *et al.* (2003) proposed the alternative cluster Neyman-Scott rectangular pulse model (NSRP), resembling the PRP model, but differing in the means in which the total consumption and frequency of pulses are calculated, and better reflecting the daily variability of water consumption. A closer look to the arrival rate function of a PRP process intended to model automated meter reading demand data at different spatial and temporal scales is presented in Arandia-Perez *et al.* (2014). A predictive end-use model was developed more recently by Blokker and Vreeburg (2005), in which end-uses are simulated as rectangular pulses with specific probability distributions for the frequency, intensity and duration, attained from field surveys in the Netherlands. In Huang *et al*. (2014) annual urban water demand time series are forecasted, recognizing and embracing their non-stationary nature, and based on explanatory variables and the sensitivity of demand to them. The wavelet transform is used to decompose the non-stationary series, and then the kernel partial least squares and autoregressive moving average models are used to model the stationary sub-series. Another promising predictive model was developed by Aksela and Aksela (2011) and consists in the estimation of demand patterns at property level (single-family households). Estimation of nodal consumptions is taken a step further in Kang (2011) by combining the estimation of uncertain consumptions and pipe roughness coefficients with the prediction of pipe flows and pressure heads. The uncertainties in the estimated variables and pipe flows and pressure heads predictions are quantified in terms of confidence intervals using a first order second moment method. After verifying non trivial scaling of the variance of real consumption data with spatial aggregation, Magini *et al.* (2008) developed simple scaling laws relating the mean, variance and covariance of water consumption series with the number of aggregated users. The expected value for the mean consumption was found to increase linearly. The expected value for the variance and lag1 covariance of consumption was found to increase according to an exponent between one and two. The subject was further investigated in Vertommen *et al*. (2012). While the scaling laws were derived considering different time steps, the effect of the time window of observation on the statistics, due to the auto-correlation in the consumption series, was not completely established in this first approximation. The scaling laws were developed neglecting the space-time covariance function, an assumption made for the sake of simplicity at the time. Here, the spatial and temporal correlations will both be explicitly considered. The development of scaling laws for the cross-covariance and cross-correlation coefficient, between two groups with different characteristics is also an innovative and challenging task. To validate and calibrate the theoretically developed scaling laws, real residential consumption data are used.

Scale effects have been identified in a wide variety of subjects and by many different researchers. In Ghosh and Hellweger (2012) a literature overview regarding spatial scaling in urban and rural hydrology can be found. Other scaling relations, such as the mean-variance scaling translated by Taylor’s power law, are well documented in many different systems: from the variability in population abundance (Ballantyne IV and Kerkhoff, 2007), to epidemiology, precipitation and river flows, stock markets, business firm growth rates (Eisler et al., 2008), car traffic, among others. By generically relating statistics of a stochastic process at different aggregation levels, these scaling laws are not restricted to water consumption modeling, and can be useful to different fields of science.

Being part of an ongoing research work, these scaling laws will be combined with optimization models for the design of WDS and scenario evaluations. Understanding the temporal and spatial variability of nodal consumptions is a fundamental pre-requisite for a risk-based approach in designing and managing WDS. At this aim, the scaling law approach will allow the development of more robust designs and management solutions for water distribution networks.

Theoretical Framework

The development of the scaling laws is 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 $Q_1(t)$. Herein, the water flow in a meter will be used to define the unit water consumption. This unit can refer, for instance, to one household. Hence, the spatial aggregation refers to the aggregation of meters with the same unitary consumption. Let there be n meters identified by m_i with $i = 1, 2, \dots, n$, let T denote the length of the observation time interval, and let $q_{m_i}(t)$, with $t \in [0, T]$, be different finite realizations of the stochastic process, representing the water consumption for the i^{th} meter. The mean and variance of water flow for the i^{th} meter, in the time interval T, are evaluated, respectively, by:

$$\mu_{m_i} = \frac{1}{T}\int_0^T q_{m_i}(t)dt \tag{1}$$

$$\sigma_{m_i}^2 = \frac{1}{T}\int_0^T \left[q_{m_i}(t) - \mu_{m_i}\right]^2 dt \tag{2}$$

The auto-covariance, $cov_{m_i}(\tau)$, and auto-correlation coefficient, $\rho_{m_i}(\tau)$ at a time lag τ are given by:

$$cov_{m_i}(\tau) = \frac{1}{T}\int_0^T \left(q_{m_i}(t+\tau) - \mu_{m_i}\right)\left(q_{m_i}(t) - \mu_{m_i}\right)dt \tag{3}$$

$$\rho_{m_i}(\tau) = \frac{cov_{m_i}(\tau)}{\sigma_{m_i}^2} \tag{4}$$

As aforementioned, to accurately describe stochastic consumption it is also necessary to determine the correlation between the signals in different meters, m_{i1} and m_{i2}. This correlation can be expressed through the cross-covariance, $cov_{m_{i1}m_{i2}}(\tau)$, and the cross-correlation coefficient, $\rho_{m_{i1}m_{i2}}(\tau)$, evaluated in time interval T, respectively, as followed:

$$cov_{m_{i1}m_{i2}}(\tau) = \frac{1}{T}\int_0^T \left(q_{m_{i1}}(t+\tau) - \mu_{m_{i1}}\right)\left(q_{m_{i2}}(t) - \mu_{m_{i2}}\right)dt \tag{5}$$

$$\rho_{m_{i1}m_{i2}}(\tau) = \frac{cov_{m_{i1}m_{i2}}(\tau)}{\sigma_{m_{i1}} \cdot \sigma_{m_{i2}}} \tag{6}$$

Where, $\sigma_{m_{i1}}$ and $\sigma_{m_{i2}}$ are the standard deviations of the consumption in m_{i1} and m_{i2}. If no lag is considered, these last two statistics become the lag-zero cross-covariance and lag-zero cross-correlation coefficient, given by the same expressions (5) and (6), but with $\tau = 0$.

Among the aforementioned statistics, the mean, the variance, the auto-covariance and the auto-correlation coefficient, coincide with the expected values of the stochastic process if the process is assumed to be ergodic and the observation time is long enough. The expected values assume different values depending on the 'spatial' aggregations in the discrete space of the positive integers associate with each meter (Magini *et al.*, 2008). The pooled water consumption, resulting from the aggregation of the n random variables is given by:

$$q_n(t)=\sum_{i=1}^{n} q_{m_i}(t) \qquad (7)$$

Where $q_n(t)$ is a finite realization of a pooled stochastic process $Q_n(t)$. The aim of this work is to determine the expected value of the above statistics for the pooled stochastic process, in a generic observation interval T, as function of the aggregation n, the length of T, assuming the expected values of the statistics for the stochastic process $Q_1(t)$, are known.

Scaling law for the variance

As aforementioned, Magini *et al.* (2008) developed the first equation for the expected value of the variance for n aggregated consumption series, $E[\sigma_n^2]$, which was further developed in Vertommen *et al.* (2012), neglecting the space-time correlation term. In order to solve the equation for $E[\sigma_n^2]$ without neglecting the referred term, the following equation obtained in Magini *et al.* (2008) is initially considered:

$$E[\sigma_n^2] = \frac{1}{T^2}\int_0^T \int_0^T \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \left[cov_{m_{i1}m_{i2}}(0) - cov_{m_{i1}m_{i2}}(\tau)\right] dt_1 dt_2 \qquad (8)$$

Where, $cov_{m_{i1}m_{i2}}(0)$ is the cross-covariance at lag $\tau = 0$ and $cov_{m_{i1}m_{i2}}(\tau)$ is the cross-covariance at lag $\tau = t_1 - t_2$. This expression can further be developed into:

$$E[\sigma_n^2] = \sum_{i=1}^{n} \sigma_{m_i}^2 + 2\sum_{i_1=1}^{n-1} \sum_{i_2=i_1+1}^{n} \sigma_{m_{i1}} \sigma_{m_{i2}} \rho_{m_{i1}m_{i2}}(0)$$
$$- \frac{1}{T^2}\int_0^T \int_0^T \left[\sum_{i=1}^{n} \sigma_{m_i}^2 \rho_{m_i}(\tau) + 2\sum_{i_1=1}^{n-1} \sum_{i_2=i_1+1}^{n} \sigma_{m_{i1}} \sigma_{m_{i2}} \rho_{m_{i1}m_{i2}}(\tau)\right] dt_1 dt_2 \qquad (9)$$

Since the consumption random variables have the same underlying stochastic process, $\sigma_{m_i} = \sigma_1$ and $\rho_{m_i}(\tau) = \rho_1(\tau)$, equation (9) can be simplified into:

$$E[\sigma_n^2] = n\sigma_1^2 + 2\sigma_1^2 \sum_{i_1=1}^{n-1} \sum_{i_2=i_1+1}^{n} \rho_{m_{i1}m_{i2}}(0) - \frac{n\sigma_1^2}{T^2}\int_0^T \int_0^T \rho_1(\tau) dt_1 dt_2$$
$$- \frac{2\sigma_1^2}{T^2} \sum_{i_1=1}^{n-1} \sum_{i_2=i_1+1}^{n} \int_0^T \int_0^T \rho_{m_{i1}m_{i2}}(\tau) dt_1 dt_2 = \qquad (10)$$
$$= n\sigma_1^2\left(1 - \gamma_1(T)\right) + 2\sigma_1^2 \left\{\sum_{i_1=1}^{n-1} \sum_{i_2=i_1+1}^{n} \left[\rho_{m_{i1}m_{i2}}(0) - \gamma_{m_{i1}m_{i2}}(T)\right]\right\}$$

Where, $\gamma_1(T)$ is the variance function for the consumption observed in the single meters, as defined by Vanmarcke (1983):

$$\gamma_1(T) = \frac{1}{T^2}\int_0^T \int_0^T \rho_1(\tau) dt_1 dt_2 \quad (11)$$

And similarly,

$$\gamma_{m_{i1}m_{i2}}(T) = \frac{1}{T^2}\int_0^T \int_0^T \rho_{m_{i1}m_{i2}}(\tau) dt_1 dt_2 \quad (12)$$

For the special case of spatial uncorrelated demands, expression (10) becomes:

$$E[\sigma_n^2] = n\sigma_1^2[1 - \gamma_1(T)] \quad (13)$$

And for the special case of spatial perfectly correlated demands, expression (10) becomes:

$$E[\sigma_n^2] = n^2\sigma_1^2[1 - \gamma_1(T)] \quad (14)$$

Since the spatial correlation between consumptions can assume values between 0 (uncorrelated consumptions) and 1 (perfectly correlated consumptions), equations (13) and (14) represent the minimum and maximum limits for the expected value of the variance of the pooled process $Q_n(t)$. The theoretical equation (10) relies on many different variables and can therefore be difficult to use in practical cases. An alternative and simplified generic equation is proposed through the following approximation:

$$E[\sigma_n^2] \cong n^\alpha\sigma_1^2[1 - \gamma_1(T)] \quad (15)$$

Where, the expected value of the variance of the pooled process $Q_n(t)$, is proportional to the variance of the process $Q_1(t)$, according to an exponent, which varies between 1 and 2. The value of the scaling exponent depends on n and on the existing spatial correlation: if consumption signals are uncorrelated in space, the variance increases linearly, if signals are perfectly correlated in space, the variance increases according to a quadratic order. The auto-correlation, or the correlation in time, of the consumption signals reduces the variance in a finite observation period T. This reduction is expressed through the variance function. When the observation period T is significantly larger than the scale of fluctuation, θ, the variance function is simplified into $\gamma_1(T) = \frac{\theta}{T}$ (VanMarcke, 1983), and its value will be much smaller than one, having therefore little influence on the expected value of the variance of the pooled process, $Q_n(t)$. In this case it seems reasonable to neglect the space-time covariance function, and the equation for $E[\sigma_n^2]$ becomes the equation derived in Vertommen *et al.* (2012). The approximation for the expected value of the variance of the pooled process, $Q_n(t)$, given by equation (15), disregards the fact that the scaling exponent could be a function of the number of aggregated meters. As a first approximation, the real demand data will be fitted to the power law, and a general and constant value of α will be estimated.

Scaling law for the cross-covariance

Let there now be two different types of consumption, A and B, each with a different underlying stationary stochastic process, $Q_{A,1}(t)$ and $Q_{B,1}(t)$, whose realizations are respectively, $q_{A,m_i}(t)$ and $q_{B,m_j}(t)$, with $i = 1, 2, \dots, n_A$ and $j = 1, 2, \dots, n_B$. The objective is now to derive the expected value for the lag zero cross-covariance between the pooled processes $Q_{n_A}(t)$ and $Q_{n_B}(t)$, whose realizations are respectively, $q_{n_A}(t)$ and $q_{n_B}(t)$, i.e., the n_A aggregated random variables with consumption type A, and the n_B aggregated variables with consumption type B. Following a similar approach as the one used to develop the scaling law for the variance, the expected value for the aforementioned cross-covariance, $E[cov_{n_A n_B}]$, considering an observation time T, is given by:

$$E[cov_{n_A n_B}] = \frac{1}{T^2}\int_0^T \int_0^T \sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\left[cov_{m_i m_j}(\,0) - cov_{m_i m_j}(\,\tau)\right] dt_1 dt_2 \tag{16}$$

Where $cov_{m_i m_j}(\,\tau)$, is the cross-covariance between the consumptions at m_i and m_j, at time lag τ, and $cov_{m_i m_j}(\,0)$, is the cross-covariance between m_i and m_j, at time lag $\tau = 0$. This expression shows that the expected value for the cross-covariance between the aggregated consumptions of two different groups depends on the spatio-temporal correlation between the unit consumption variables of the two groups. If the consumption variables of group A have no correlation with the consumption variables of group B, independently of the correlation that might exist between the variables within each group, then $cov_{m_i m_j}(0) = 0$ and $cov_{m_i m_j}(\,\tau) = 0$, for all pairs (m_i, m_j). In this case, equation (16) becomes null. Considering now a more generic case in which the consumptions of group A are at some level correlated with the consumptions of group B, then equation (16) becomes:

$$E\left[cov_{n_A n_B}\right] = \sum_{i=1}^{n_A}\sum_{j=1}^{n_B} cov_{m_i m_j}(\,0) - \sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\frac{1}{T^2}\int_0^T\int_0^T cov_{m_i m_j}(\tau)dt_1 dt_2 =$$

$$= \sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\sigma_{m_i}\sigma_{m_j}\rho_{m_i m_j}(0) - \sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\sigma_{m_i}\sigma_{m_j}\ \ \Phi_{m_i m_j}(T) \qquad (17)$$

$$= \sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\sigma_{m_i}\sigma_{m_j}\left[\rho_{m_i m_j}(0) - \Phi_{m_i m_j}(T)\right]$$

Where, $\sigma_{m_i}, \sigma_{m_j}$ are the standard deviations of the consumption at m_i and m_j, respectively, $\rho_{m_i m_j}(0)$ is the cross-correlation function between m_i and m_j at time lag $\tau = 0$, $\rho_{m_i m_j}(\,\tau)$ is the cross-correlation function between m_i and m_j at time lag τ, and where:

$$\Phi_{m_i m_j}(T) = \frac{1}{T^2}\int_0^T\int_0^T \rho_{m_i m_j}(\,\tau)dt_1 dt_2 \qquad (18)$$

For practical purposes, if the consumption variables from each group have the same underlying process, then it is possible to assume a mean cross-correlation coefficient, denoted by $\bar{\rho}_{1A}(\,0)$, among the meters of group A, and a mean cross-correlation coefficient, denoted by $\bar{\rho}_{1B}(\,0)$, among the meters of group B. Consequently, a mean cross-correlation coefficient $\bar{\rho}_{1,AB}(\,0)$ between A and B, can also be assumed. In this case, the theoretical equation (17) can be approximated by:

$$E\left[cov_{n_A n_B}\right] = n_A n_B \sigma_{1A}\sigma_{1B}\left[\bar{\rho}_{1,AB}(0) - \bar{\phi}(T)\right] \qquad (19)$$

Where,

$$\bar{\phi}(T) = \sum_{i=1}^{n_A}\sum_{j=1}^{n_B} \frac{\Phi_{m_i m_j}(T)}{n_A n_B} \qquad (20)$$

The cross-covariance increases with the product of the spatial aggregation levels and is independent of the spatial correlation on the intern of each group. When the space-time covariance is neglected, the cross-covariance between water consumptions of two groups, scales with the product between the aggregation levels of both groups.

Scaling law for the cross-correlation coefficient

Finally, the objective is to derive the scaling law for the expected lag zero cross-correlation coefficient between the pooled processes $Q_{n_A}(t)$ and $Q_{n_B}(t)$, which is given by:

$$E\left[\rho_{n_A n_B}\right] = \frac{E\left[cov_{n_A n_B}\right]}{E\left[\sigma_{n_A}\right] \cdot E\left[\sigma_{n_B}\right]} \qquad (21)$$

Where, σ_{n_A} and σ_{n_B} are the standard deviations of the pooled processes $Q_{n_A}(t)$ and $Q_{n_B}(t)$, respectively. Using the more generic obtained scaling laws for the variance, equation (15), and the cross-covariance, equation (16), equation (21) becomes:

$$E\left[\rho_{n_A n_B}\right] = \frac{\sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\left[\rho_{m_i m_j}(0) - \Phi_{m_i m_j}(T)\right]}{\left[n_A^{\alpha_A}(1-\gamma_{1A}(T))\right]^{\frac{1}{2}} \cdot \left[n_B^{\alpha_B}(1-\gamma_{1B}(T))\right]^{\frac{1}{2}}} \qquad (22)$$

Where α_A, α_B and $\gamma_{1A}(T), \gamma_{1B}(T)$ are the exponents of the scaling law for the variance and the variance function specific to the meters in groups A and B, respectively. In parallel, the simplified equation (19) for the cross-covariance produces:

$$E\left[\rho_{n_A n_B}\right] = \frac{n_A n_B\left[\bar{\rho}_{1,AB}(0) - \bar{\phi}(T)\right]}{\left[n_A^{\alpha_A}(1-\gamma_{1A}(T))\right]^{\frac{1}{2}} \cdot \left[n_B^{\alpha_B}(1-\gamma_{1B}(T))\right]^{\frac{1}{2}}} \tag{23}$$

Also, if the number of aggregated in both groups is the same, i.e., $n_A = n_B$, equation (23) becomes:

$$E\left[\rho_{n_A n_B}\right] = n^{\beta} \frac{\bar{\rho}_{1,AB}(0) - \bar{\phi}(T)}{[(1-\gamma_{1A}(T))]^{\frac{1}{2}} \cdot [(1-\gamma_{1B}(T))]^{\frac{1}{2}}} \tag{24}$$

Where, $\beta = 2 - \frac{(\alpha_A+\alpha_B)}{2}$. In this case the cross-correlation coefficient increases according to an exponent that is equal to the difference between the exponents of the expected value for the cross-covariance between the pooled process $Q_{n_A}(t)$ and $Q_{n_B}(t)$ (expected to be equal to two) and the average between the exponents of the expected values of the standard deviation associated to each process $Q_{n_A}(t)$ and $Q_{n_B}(t)$, respectively. Since $1 \le \alpha_A, \alpha_B \le 2$, the exponent of the scaling law for the cross-correlation coefficient, β, will assume values between 0 and 1. These limits represent the possible extreme cases: perfectly correlated consumptions within each group, and uncorrelated consumptions within each group.

Equation (23) shows that the cross-correlation coefficient between the pooled processes $Q_{n_A}(t)$ and $Q_{n_B}(t)$, depends separately on the two aggregation levels, n_A and n_B, and not only on their product as happens with the cross-covariance. The cross-correlation coefficient between the pooled processes also depends on the cross-correlation coefficient between the realizations of each of the group individually, $q_{A,m_i}(t)$ and $q_{B,m_j}(t)$, i.e., $\rho_{m_{i1}m_{i2}}(0)$ and $\rho_{m_{j1}m_{j2}}(0)$, other than the cross-correlations existing between the realizations of both groups, i.e., $\rho_{m_i m_j}(0)$.

Time step

Another important aspect when modeling WDS is the choice of the adequate time step to assess water consumption. The adequate time step for design purposes is obviously not the same as for operation planning purposes. Even for the same purpose, it might be necessary to consider different temporal resolutions for feeders and peripheral pipes of a system, since the temporal variation of consumption significantly increases from the first to the last one. Considering longer time steps results in loss of information about the consumption signals, which in turn results in lower estimates of the variance (Rodriguez-Iturbe *et al.*, 1984; Buchberger and Nadimpalli, 2004). At peripheral pipes this aspect is particularly relevant since the choice of the wrong time step will not reflect accurately the large consumption fluctuations that are, as aforementioned, characteristic of these parts of the network. It has been verified that the consumption variability deriving from different temporal aggregations specially affects flow rates and water quality at the peripheral pipes (Yang and Boccelli, 2013).

Water consumptions variables can be analyzed considering different time steps; for instance, a one second time step, a one minute time step, and so on. The realizations of the stochastic process observed at a smaller time step, can be aggregated in broader time steps. This is, a temporal aggregated water consumption variable, considering a time step Δt, is given by:

$$q_{mi,\Delta t}(\varphi)=\frac{1}{\Delta t}\int_{(\varphi-\Delta t/2}^{\varphi+\Delta t/2} q_{mi}(t)\,dt \qquad (25)$$

Where $q_{mi,\Delta t}(\varphi)$ is a realization of the time aggregated stochastic process $Q_{1,\Delta t}(\varphi)$. The temporal aggregated variable is divided by Δt to maintain the flow units. Some of the statistics of the temporal aggregated process $Q_{1,\Delta t}(\varphi)$, differ from the statistics of the original process $Q_1(t)$. The reduction of the variance of an instantaneous signal with the time step can be measured through the aforementioned variance function proposed by VanMarcke (1983). Making use of the variance function it is possible to obtain the variance at any desired time step from the variance of the instantaneous signal. Taking this into account the scaling law for the variance, in equation (15) becomes:

$$E\left[\sigma_{n,\Delta t}^2\right] = n^{\alpha}\sigma_1^2(1-\gamma_1(T))\gamma_1(\Delta t) \qquad (26)$$

Where, Δt is the desired time step, and $\gamma_1(\Delta t)$ is the variance function relating the variance of the original process $Q_1(t)$ and the variance of the temporal aggregated process $Q_{1,\Delta t}(\varphi)$.

Similarly for the cross-covariance, in equation (19) the following is obtained:

$$E\left[cov_{n_A n_B,\Delta t}\right] = n_A n_B \sigma_{1A}\sigma_{1B}[\bar{\rho}_{1AB}(0)-\bar{\phi}(T)]\bar{\phi}(\Delta t) \qquad (27)$$

Where, $\bar{\phi}(\Delta t)$ is the function relating the cross-covariance of the temporal aggregated process and the cross-covariance of the original process.

Validation of the analytical expressions using real consumption data

Effect of the spatial aggregation

The collected data consist in indoor water uses of 82 single-family residences, with a total of 177 inhabitants, from the town of Latina, Italy (Guercio *et al.*, 2003; Pallavicini and Magini, 2007). The 82 users were monitored in four different days (4 consecutive Mondays). For each user the different days of consumptions were assumed to be different realizations of the same stochastic process. In this way the number of variables was artificially extended to about 320, preserving at the same time the homogeneity of the sample. The temporal resolution of each time series is one second. The data series were divided into one hour periods to assure a stationary underlying process. The series were then temporally aggregated considering time steps ranging from one second to 30 minutes. To assess the scaling of the variance, all the consumption series were assumed to have the same underlying process and were aggregated in groups of $n = 10, 20, 30, \dots, 150$. It has to be noted at this point that the data series correspond to discrete and finite sequences of demand values (and are therefore called time series), while the theoretical developments were made for continuous variables. The statistics of the real demand data are thus obtained through the appropriate and well-known estimators. This might introduce some minor bias to the estimations. Bias corrections can be made (Koutsoyiannis, 2013), but fall out of the scope of this work. The variance of each group was estimated, obtaining real value pairs (σ_n^2, n) for all the considered time steps. To assess the scaling of the cross-covariance and cross-correlation coefficient, the time series were first randomly divided into two groups, A and B, assuming to have two distinct underlying processes, and then aggregated in groups of $n_A = n_B = 10, \dots, 150$. The cross-covariance and cross-correlation coefficient were estimated between all groups, obtaining real value pairs $(cov_{n_A n_B},\ n_A = n_B)$ and $(\rho_{n_A n_B}, n_A = n_B)$ for all the considered time steps. These pairs were used to validate the theoretical expressions for the scaling laws previously obtained and to calibrate them. For each parameter the value of the exponent α or β was obtained by adjusting the theoretical expression for the scaling law to the real value pairs. The least squares method was used for this adjustment. This process is repeated for all considered time steps, in order to verify its influence on the exponents of the scaling laws. The value of the variance function $\gamma_1(T)$ was estimated by numerically solving $\frac{1}{T^2}\int_0^T \int_0^T \rho_{mi}(\tau) dt_1 dt_2$, from the single consumption signals. The value of the function $\bar{\phi}(T)$ was estimated by numerically solving $\frac{1}{T^2}\int_0^T \int_0^T E\left[\rho_{m_i m_j}(\tau)\right] dt_1 dt_2$, from the single consumption signals. Results are summarized in Table 1.

Table 1 Values of $\boldsymbol{\gamma_1(T)}$ and $\boldsymbol{\Phi(T)}$ for single consumption values and considering the time steps $\Delta \boldsymbol{t} = \mathbf{1, 60\ and\ 600}$ seconds.

The obtained values for $\gamma_1(T)$ and $\bar{\phi}(T)$ show that for the considered consumptions series the effect of the temporal correlation cannot be neglected. Moreover, the values increase with the considered time step. The variance function assumes average values ranging between 0.195, for the instantaneous signal, and 0.274 for a time step of ten minutes. Being connected to the scale of fluctuation of the process, these values are indicative of a significant memory between consumption signals. The values obtained for $\bar{\phi}(T)$ range from an average of 0.062, for the instantaneous signal, to 0.180 for a ten minute time step, also indicating a considerable memory between consumption signals observed in different meters. Table 2 summarizes the obtained exponents of the scaling laws for the variance, cross-covariance and cross-correlation coefficients when considering time steps of one second, one minute and ten minutes.

Table 2 Exponents of the scaling laws for the variance, cross-covariance and cross-correlation coefficients at different time steps.

The variance of consumption increases slightly non-linearly with the aggregation, when considering time steps of one second and one minute. The average exponent of the scaling law for the variance is 1.033, considering a one second time step, and 1.063 considering a one minute time step. However, when considering a time step of ten minutes, the non-linearity of the scaling law for the variance becomes more evident, since in this case the exponent assumes an average value equal to 1.301. The assumption of linear scaling of the variance with the number of served users can lead to underestimated values of the variability of consumption at high spatial aggregation levels, especially when broader time steps are used. Being connected to the cross-correlation coefficient between consumptions, the results show that the consumption signals are slightly correlated and that this correlation increases when the time step increases. This observation can be explained by the fact that when considering longer time steps it is more likely to observe simultaneous water uses, than when very small time steps (ex.: one second) are considered. For a better understanding of these results, the scaling laws for the variance of consumption between hours 6 and 7 considering sampling times of one second, one minute and 10 minutes are graphically reported in figure1. The dots, plus sign and asterisk represent the average values of the variance of several different sets of n meters, for time steps of one second, one and ten minutes, respectively. "SL" stands for scaling law, and the relative error of the approximation is given by δ.

Figure 1 Scaling laws for the variance of consumption between 6 and 7 am, considering time steps of 1 second, 1 minute and 10 minutes.

Observing figure 1 it is clear that when broader time steps are considered, the variance decreases, but the exponent of the scaling law increases, due to the increase of the correlation. The relations between the variance and the exponent of the scaling law for the variance with the degree of correlation between consumptions are illustrated, one at a time, in figures 2 and 3. Figure 2 illustrates the relation between the variance and the cross-correlation coefficient. The values of the variance and cross-correlation are referring to the consumption series of 10 aggregated meters, between hours 6 and 7, evaluated at the time steps ranging from one second to 30 minutes. It is possible to observe that the variance decreases with the increase of the cross-correlation coefficient, directly related to the consideration of broader time steps, according to a power law.

Figure 2 Variance *versus* cross-correlation coefficient for $\boldsymbol{n = 10}$, between 6 and 7 am.

The exponents of the scaling laws for the variance, obtained for the different time steps can also be related to the degree of cross-correlation at each time step. Figure 3 shows this relation.

Figure 3 Exponents of the scaling laws for the variance *versus* the cross-correlation coefficient.

The exponent of the scaling law for the variance increases according to a power law with the degree of cross-correlation between the consumption series.

Regarding the cross-covariance, when considering the same number of aggregated meters in each group, it is expected to verify a quadratic increase of the parameter. The obtained results show that the value of the exponent of the adjusted scaling law is close to two at several hours of the day. The average value of the exponent decreases with the consideration of broader sampling rates, due to the effect of $\bar{\phi}(T)$. In order to obtain an exponent equal to two when considering broader sampling rates, a longer sampling time should be considered. Similarly to the variance, the cross-covariance itself decreases when broader sampling rates are considered. Figure 4 shows a graphical representation of the scaling laws for the cross-covariance of consumption between 6 and 7 am considering sampling times of one second, one minute and 10 minutes. The dots, plus sign and asterisk represent the average values of the cross-covariance between several different sets of n, for time steps of one second, one and ten minutes, respectively. "SL" stands for scaling law, and the relative error of the approximation is given by δ.

Figure 4 Scaling laws for the cross-covariance of consumption between 6 and 7 am, considering time steps of 1 second, 1 minute and 10 minutes.

The scaling law for the cross-correlation coefficient between consumption signals was also determined. The results show a significant increase of the correlation with n. As expected, the cross-correlation coefficient is higher when longer sampling rates are considered. The obtained results also show a flattening of the scaling curves when the sampling rate increases. This is expected to happen since in theory the exponent β is equal to the difference between the exponents of the cross-covariance and the average between the exponents of the standard deviation in each group, and the latter increase with the sampling rate. The value of the exponent is expected to assume values between zero and one, which is verified. The average exponent of the scaling law for the cross-correlation coefficient is 0.453, considering a one second sampling time, 0.363 considering a one minute sampling time, and 0.267 considering a 10 minute sampling time. A graphical representation of the scaling laws for the cross-correlation coefficient between 6 and 7 am, considering sampling times of one second, one minute and 10 minutes, can be found in figure 5. The dots, plus sign and asterisk represent the average values of the cross-correlation coefficient between several different sets of n, for time steps of one second, one and ten minutes, respectively. “SL” stands for scaling law, and the relative error of the approximation is given by δ.

Figure 5 Scaling laws for the cross-correlation coefficient of consumption between 6 and 7 am, considering time steps of 1 second, 1 minute and 10 minutes.

At some aggregation levels there seem to be some breaking points in the cross-correlation coefficient. These are due to the fact that the cross-covariance between groups and the standard deviation of each group do not increase in the same way, leading to a less smooth scaling of the cross-correlation coefficient. When different hours of the day are considered, these apparent breaking points can appear at different aggregation levels, or not be evident at all.

Effect of the time step

Let us know consider specifically the effect of the time step on the consumption statistics. For assessing the variance of consumption at any desired time step, one needs to know the value of the variance function at that time step. Vanmarcke (1983) suggested the following generic expression to estimate the variance function:

$$\gamma_1(\Delta t) \cong \left[1 + \left(\frac{\Delta t}{\theta}\right)^m\right]^{-1/m} \quad (28)$$

Where θ is the scale of fluctuation and m is a model index parameter. For assessing the cross-covariance at any desired time step, the function $\bar{\phi}(\Delta t)$ can be approximated by a similar expression as used for the variance function, in this case, being θ_{ab} the scale of fluctuation associated to the series of the two groups A and B. The values of the variance and cross-covariance of consumption at different time steps were used to calibrate the variance function and the function $\bar{\phi}(\Delta t)$ for the consumption data of Latina. Table 3 summarizes the obtained values for the scale of fluctuation and the index parameters m, when considering $n = 1, 10, 100$.

Table 3 Average values for the scale of fluctuation and index parameters for the variance and cross-covariance, when considering $\boldsymbol{n = 1, 10, 100}$.

Regarding the variance, the scale of fluctuation assumes large values, enhancing the importance of considering the effect of the time step on the consumption statistics. The scale of fluctuation increases with the spatial aggregation, which indicates that the consumption signals stay correlated for a longer period in time when more meters are considered. The same is verified with the scale of fluctuation associated to the two groups of, A and B. The index parameters of the variance function are always smaller than one, and decrease with n. The index parameters associated to the cross-covariance also decrease with n, but are significantly larger than those obtained for the variance.

The value of the index parameter dictates the shape of the curve between the cross-covariance and the time step and the number of its inflection points. From the obtained results we observed that when the index parameter is smaller than one, the curve is convex. When the index parameter is greater than one there is at least one inflection point, and the larger its value the more evident becomes the S shape of the curve. Figure 6 shows a graphical representation of the evolution of the variance of the consumption with the time step, for $n = 100$.

Figure 6 Variance versus time step for $\boldsymbol{n = 100}$, between 6 and 7 am.

The variance of the real consumption data at different time steps is well estimated through the approximation for the variance function given by equation (4).

Conclusions

The accurate description of water consumption is as essential as challenging when dealing with the design and management of WDS. Understanding how, and in which measure, the statistics used to describe water consumption are affected by the spatial and temporal aggregation levels is therefore essential for an accurate description of stochastic consumption. Following up the work developed by Magini *et al.* (2008) and Vertommen *et al*. (2012), the scaling laws for the variance, cross-covariance and cross-correlation coefficient are theoretically derived. The correlation structure, both in space and time are explicitly considered. The variance is found to increase with the spatial aggregation, according to an exponent between one and two, depending on the spatial correlation between consumptions. The effect of the auto-correlation is measured through the variance function in the considered time interval and is responsible for a reduction of the overall variance. The development of scaling laws for the cross-covariance and cross-correlation between two different groups, with different characteristics, is innovative and will help understand the association between different signals which is crucial for a realistic assessment of water consumption in a network. The cross-covariance between two groups is found to increase according to the product between number of meters in each group and the correlation between the groups. An effect of the considered time step is also verified and is measured through the function $\bar{\phi}(T)$. The cross-correlation coefficient depends separately on the number of meters in each group, and on the correlation within each group, other than the correlation between groups. While the equations derived in Vertommen *et al.* (2012) were limited to cases in which it was guaranteed that $T >> \theta$, these new equations are not. We believe the main novelty of the paper is achieving the scaling laws that are valid for all cases, by fully developing the space time covariance function and attaining the correction term $1 - \gamma_1(T)$. These scaling laws might be a contribution to not only water consumption analysis, but also to other fields of science.

The theoretical scaling laws are found to well describe the scaling properties of the statistics of real residential consumption data of Latina, Italy. The values of $\gamma_1(T)$ and $\bar{\phi}(T)$, are obtained an found to be significant and to increase with the considered time step, indicating that when broader time steps are used, there is a higher auto-correlation between signals and there is a longer memory in the process. This finding highlights the importance of considering of the auto-correlation structure of water consumption series. The time step was also found to significantly affect the obtained exponents of the scaling laws: the exponents of the scaling law for the variance increase considerably with time step, the exponents of the scaling law for the cross-covariance and the cross-correlation decrease. Since the cross-correlation coefficient is closely related to the considered time step, it was possible to establish the relations between (1) the cross-correlation coefficient and the variance, and (2) the cross-correlation and the exponent of the scaling law for the variance. The variance was found to decrease with the degree of correlation between consumptions. On the contrary, the exponent of the scaling law for the variance increases with the correlation, according to a power function, meaning that for more correlated consumptions, their variability scales more rapidly. A more thorough relation between this exponent and the correlation could be an interesting topic to address in the future developments, besides verifying the existence of different regimes in the process of aggregation as a function of n. It could also be interesting to relate the correlation structure and scaling parameters to the factors that influence water consumption, such as temperature, precipitation, social habits economic conditions and price of water. We further believe it would be interesting to apply the scaling laws to a data set made up by a significantly larger number of unitary uses, in order to assess if significant errors might exist due to assumptions made, and also in order to validate the scaling laws for higher spatial aggregation levels, more common in real world water distribution systems.

The reduction of the variance and cross-covariance with the increase of the time steps was adequately approximated by the proposed variance function and $\bar{\phi}(T)$ function.

The obtained results clearly point out the importance of considering the scaling effects when describing or estimating nodal consumptions in a network for design or management purposes. The inclusion of uncertain consumptions in network design and management optimization problems is a challenging task, and we believe that the developed scaling laws are a step forward in unraveling it. From the developed laws it is possible to estimate the consumption statistics at any desired spatial or temporal scale. These parameters can then be used to generate consumption series for each node of the network. In this way, instead of generating random consumption series for all unitary uses at the network and aggregating them, the total consumptions at each node can be directly obtained through the scaling laws, achieving computational time savings. The network can then be simulated for all the consumption values from the series, obtaining a series of values for the pressure at each node. The approach provided by the scaling laws, which allows to take into consideration more accurate values of the consumption variability, can contribute to the design of networks capable of better enduring the stochastic nature of water consumption.

Acknowledgements

The participation of the first author in the study is supported by FCT - Fundação para a Ciência e Tecnologia under grant SFRH/BD/65842/2009.

References

Aksela, K., and Aksela, M. (2011). “Demand Estimation with Automated Data Reading in a Distribution Network.” *J. Water Resour. Plann. Manage.*, 137(5), 456-467, doi: 10.1061/(ASCE)WR.1943-5452.0000131.

Alvisi, S., Franchini, M., and Marinelli, A. (2003). “A Stochastic Model for Representing Drinking Water Demand at Residential Level.” *Water Resources Management*, 17 (3), 197-222, doi: 10.1023/A:1024100518186.

Arandia-Perez, E., Uber, J. G., Boccelli, D. L., Janke, R., Hartman D., and Lee, Y. (2014). “Modeling Automatic Meter Reading Water Demand as a Nonhomogeneous Point Process.” *J. Water Resour. Plann. Manage.*, 140(1), 55-64, doi: 10.1061/(ASCE)WR.1943-5452.0000318.

Babayan, A., Savic, D., Walters, G. (2004). “Multiobjective optimization of water distribution system design under uncertain demand and pipe roughness.” Presented at *IFAC Workshop on Modelling and Control of Water Resource Systems*, Venice, Italy, 29 September - 1 October.

Ballantyne IV, F., and Kerkhoff, A. (2007). “The observed range for temporal mean-variance scaling exponents can be explained by reproductive correlation.” *Oikos*, 116(1), 174-180 doi: 10.1111/j.2006.0030-1299.15383.x.

Blokker, E., and Vreeburg, J. (2005). “Monte Carlo Simulation of Residential Water Demand: A Stochastic End-Use Model.” In *Impacts of Global Climate Change*, 1-12, American Society of Civil Engineers, USA, doi: 10.1061/40792(173)34.

Buchberger, S., and Nadimpalli, G. (2004). “Leak estimation in water distribution systems by statistical analysis of flow reading.” *J. Water Resour. Plann. Manage.*, 130(4), 321–329, doi: 10.1061/(ASCE)0733-9496(2004)130:4(321).

Buchberger, S., and Wu, L. (1995). “Model for Instantaneous Residential Water Demands.” *J. Hydraul. Eng.*, 121(3), 232–246, doi: 10.1061/(ASCE)0733-9429(1995)121:3(232).

~~Cunha, M., Joaquim, S. (2001). “Hydraulic infrastructures design using simulated annealing.” *J. Infrastruct. Syst.*, 7(1), 32-39, doi: 10.1061/(ASCE)1076-0342(2001)7:1(32).~~

Eisler, Z., Bartos, I., and Kertész, J. (2008). “Fluctuation scaling in complex systems: Taylor's law and beyond.” *Advances in Physics*, 57(1), 89-142, doi: 10.1080/00018730801893043.

Filion Y., Adams, B., and Karney, A. (2007). “Cross Correlation of Demands in Water Distribution Network Design.” *J. Water Resour. Plann. Manage.*, 133(2), 137–144, doi: 10.1061/(ASCE)0733-9496(2007)133:2(137).

Filion, Y., Karney, B., Moughton, L., Buchberger, S., and Adams, B. (2008). “Cross Correlation Analysis of Residential Demand in the City of Milford, Ohio.” In *Water*

Distribution Systems Analysis Symposium 2006, 1-13, American Society of Civil Engineers, USA, doi: 10.1061/40941(247)43.

~~Fujiwara, O., Khang, D. (1990). “A two phase decomposition method for optimal design of looped water distribution networks.” *Water Resources Research*, 26(4), 539-49, doi: 10.1029/WR026i004p00539.~~

Ghosh, I., and Hellweger, F. L. (2012). “Effects of Spatial Resolution in Urban Hydrologic Simulations.” *Journal of Hydrologic Engineering*, 17(1), 129–137, doi:10.1061/(ASCE)HE.1943-5584.0000405.

Guercio, R., Magini R., and Pallavicini, I. (2003). “Temporal and spatial aggregation in modeling residential water demand.” In *Water Resources Management II*, 151-160, WIT Press, Southampton, UK, doi: 10.2495/WRM030151.

Huang, L., Zhang, C., Peng, Y., and Zhou, H. (2014). “Application of a Combination Model Based on Wavelet Transform and KPLS-ARMA for Urban Annual Water Demand Forecasting.” *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000397 , 04014013.

Kang, D., and Lansey, K. (2011). “Demand and Roughness Estimation in Water Distribution Systems.” *J. Water Resour. Plann. Manage.*, 137(1), 20–30, doi: 10.1061/(ASCE)WR.1943-5452.0000086.

Kapelan, Z., Savic, D., and Walters, G. (2005). “An efficient sampling-based approach for the robust rehabilitation of water distribution systems under correlated nodal demands." In *Impacts of Global Climate Change*, American Society of Civil Engineers, USA, 1-12, doi: 10.1061/40792(173)16.

Koutsoyannis, D. (2013). "Encolpion of stochastics: Fundamentals of stochastic processes". Department of Water Resources and Environmental Engineering – National Technical University of Athens, Athens.

Magini, R., Pallavicini, I. and Guercio, R. (2008). "Spatial and temporal scaling properties of water demand." *Water Resour. Plann. Manage.*, 134(3), 276–284, doi: 10.1061/(ASCE)0733-9496(2008)134:3(276).

Pallavicini, I., and Magini, R. (2007). "Experimental analysis of residential water demand data: probabilistic estimation of peak coefficients at small time scales." In *Water Management Challenges in Global Changes*, 379-384, Taylor & Francis, UK.

Polebitski, A. S., and Palmer, R. N. (2010). "Seasonal residential water demand forecasting for census tracts." *J. Water Resour. Plann. Manage.*, 136(1), 27–36, doi: 10.1061/(ASCE)WR.1943-5452.0000003.

Rodriguez-Iturbe, I., Gupta V.K., and Waymire E. (1984), Scale considerations in the modeling of temporal rainfall, *Water Resources Research* 20(11), 1611-1619, doi: 10.1029/WR020i011p01611.

VanMarcke, E. (1983). "Random Fields Analysis and Synthesis (Revised and Expanded New Edition)." *World Scientific Publishing Co. Pte. Ltd*, Singapore.

Vertommen, I., Magini, R., Cunha, M. C., and Guercio, R. (2012). “Water Demand Uncertainty: The Scaling Laws Approach.” *Water Supply System Analysis - Selected Topics*, Dr. Avi Ostfeld (Ed.), ISBN: 978-953-51-0889-4, InTech, DOI: 10.5772/51542.

Yang, X., and Boccelli, D. L. (2013). “A Simulation Study to Evaluate Temporal Aggregation and Variability of Stochastic Water Demands on Distribution System Hydraulics and Transport.” *J. Water Resour. Plann. Manage.*. doi:10.1061/(ASCE)WR.1943-5452.0000359.