Five variants of a procedure for spatial aggregation of synthetic water demand time series

Stefano Alvisi, Nicola Ansaloni and Marco Franchini

ABSTRACT

A comparison of five variants of a procedure for spatial aggregation of synthetic water demand time series is presented. This procedure allows for the spatial aggregation of hourly synthetic water demand time series preserving mean and variance at user level in such a way that the statistics of the spatially aggregated time series are reproduced (mean, variance, lag-1 temporal correlation, lag-1 temporal covariance). Five different ways of application of the methodology are considered and compared. Application to a case study consisting of the water demands of 21 users highlights that the variants considered show different levels of effectiveness in reproducing the statistics of interest and different computational burden, but overall represent a valid tool for the bottom-up generation of synthetic water demand time series.

Key words | spatial aggregation, time series, water demand

Stefano Alvisi (corresponding author)
Nicola Ansaloni
Marco Franchini
Department of Engineering,
University of Ferrara,
Via Saragat 1,
44122 Ferrara,
Italy
E-mail: *stefano.alvisi@unife.it*

INTRODUCTION

Users' water demands are the main driver of water distribution systems (WDSs). Their proper characterization thus represents a fundamental prerequisite in order to set up a robust and accurate hydraulic model of a WDS. To this end, several models for the generation of synthetic user water demand time series have been proposed in the last decades; these models allow for the characterization of the user water demand even with fine time steps and are typically based on stochastic processes, like Poisson rectangular pulses (e.g., Buchberger & Wu 1995; Guercio *et al.* 2001), Neyman-Scott rectangular pulse (e.g., Alvisi *et al.* 2003; Alcocer *et al.* 2006) or based on the characterization of the uses of individual devices such as washing machines, toilets, etc., typically used in houses (e.g., Blokker *et al.* 2010). These models were shown to be very effective in statistically reproducing the observed data at the level of single users and at low temporal aggregation levels, but it is worth remembering that in WDS simulation models the water demands are typically allocated at the nodes of the model aggregating the consumption of many users adjacent to the node considered, and hourly or semi-hourly time steps are generally used. Thus, it is necessary to transfer the information from one level of spatial-temporal aggregation to another in order to set up synthetic water demand time series to be effectively used within hydraulic models of real/complex WDSs.

Indeed, several studies (e.g., Alvisi *et al.* 2003; Moughton *et al.* 2006) showed that both the simple temporal aggregation of the synthetic series, from, for example, 1-minute to 1-hour time step, and the spatial aggregation from single users to a group of users, performed through a simple sum, can lead to time series which do not reproduce the corresponding observed statistics (mainly variance and temporal covariances at different time-lags). In fact, users' water demand time series can be temporally and spatially correlated since they depend on the users' habits and lifestyle; thus, for example, a hot summer day can lead to very high water consumption of several users during evening hours due to irrigation of gardens, showers, etc. (Bakker *et al.* 2013). In this respect, Alvisi *et al.* (2003) highlighted that reproducing the covariances within and among the series to be aggregated (i.e., both temporally and spatially) is a prerequisite in order to properly reproduce the variances

doi: 10.2166/aqua.2015.041

of the spatially aggregated series. In fact, the aggregation procedure consists of adding up a certain number of random variables (which represent the users' water demands), and the variance of a random variable that is a sum of random variables is equal to the sum of the variances of the individual random variables plus twice the sum of their covariances.

Summing up, it is important to be able to 'transfer' the time series generated at low temporal and spatial aggregation levels (e.g., at 1-minute time step for single users) to higher aggregation levels (e.g., at 1-hour time step for groups of users), since these latter series are those typically utilized within the simulation models for the design and management of WDSs. At the same time, the main statistics (mean, variance and temporal covariance) should be reproduced at these higher aggregation levels since they can have a significant impact on the hydraulic performances of the WDSs, as demonstrated by some recent studies (e.g., Babayan *et al.* 2005; Filion *et al.* 2007).

The temporal aggregation problem of time series from minute to hourly time step has been recently addressed by Alvisi *et al.* (2014b). Thus, in this paper, attention is focused only on the spatial aggregation of hourly water demand time series of single users in order to obtain hourly time series representative of a group of users to be used within a hydraulic simulation model, which can reproduce the main statistics observed at the same spatial aggregation level, i.e., hourly *means, variances, covariances and correlations at different time-lags*. To this end, taking our cue from Alvisi *et al.* (2014a), a spatial aggregation procedure based on the method proposed by Iman & Conover (1982) is presented, and five different ways of applying this method are compared. In the following, the underlying idea of the procedure and its five variants are described; the results obtained by their application to the case study consisting of the water demands of 21 users of the WDS of Milford (Ohio) (Buchberger *et al.* 2003) are discussed; finally, conclusions are presented.

THE PROCEDURE

The spatial aggregation procedure consists of summing up the water demands of an assigned number n_{us} of single users in order to obtain the water demand of the group of users considered. Let us assume that synthetic water demand time series of the single user at $\Delta t = 1$ hour time step are available; in particular, let us assume that these series have been generated *independently* of each other and in such a way that they reproduce, at hourly level, *mean* and *variance* of the corresponding observed time series (see, for example, Alvisi *et al.* (2014b) for a possible procedure for the generation of series with these characteristics). The total water demand of the n_{us} users in the generic hour h (with $h = 1{:}24$) is given by

$$Q_h = \sum_{j=1}^{n_{us}} q_h^j \tag{1}$$

where q_h^j is the random variable representing the water demand of the jth user (with $j = 1{:}n_{us}$) in the corresponding hour h. From a statistical point of view, the discharge Q corresponding to a given spatial aggregation is thus equivalent to the sum of a certain number of random variables, where each variable represents the water demand at the level of a single user. As is well known from the scientific literature (e.g., Kottegoda & Rosso 1997), the mean of a random variable which is the sum of several random variables is equal to the sum of the means of the individual random variables, whereas the variance of a random variable which is the sum of several random variables is equal to the sum of the variances of the individual random variables plus twice the sum of their covariances. Thus, the mean of the spatially aggregated water demand Q_h at the generic hour h of the day is given by

$$E\{Q_h\} = E\left\{\sum_{j=1}^{n_{us}} q_h^j\right\} = \sum_{j=1}^{n_{us}} E\{q_h^j\} \tag{2}$$

The variance of the spatially aggregated water demand Q_h at the generic hour h of the day is given by

$$\begin{aligned}\mathrm{var}\{Q_h\} &= \mathrm{var}\left\{\sum_{j=1}^{n_{us}} q_h^j\right\} \\ &= \sum_{j=1}^{n_{us}} \mathrm{var}\{q_h^j\} + 2\sum_{j=1}^{n_{us}}\sum_{l>j}^{n_{us}} \mathrm{cov}\{q_h^j, q_h^l\}\end{aligned} \tag{3}$$

It is worth noting that, according to Equation (2), if the synthetic time series of the single users reproduce the observed mean, this statistic, being purely additive, will be reproduced also at the spatially aggregated level; instead, from Equation (3), it follows that to reproduce the variance at the spatially aggregated level, it is necessary to ensure that the time series of the single users reproduce not only the variances, but also the covariances and thus the cross-correlations of the corresponding observed series. In fact, the covariance $\text{cov}\{q_h^j, q_h^l\}$ is related to the cross-correlations $\rho\{q_h^j, q_h^l\}$ and the variances $\text{var}\{q_h^j\}$ and $\text{var}\{q_h^l\}$ through the following equation:

$$\text{cov}\{q_h^j, q_h^l\} = \rho\{q_h^j, q_h^l\} \cdot \text{var}\{q_h^j\} \cdot \text{var}\{q_h^l\} \quad (4)$$

Since the two variances shown in Equation (4) are preserved at the level of single user (this is an assumption; see above), in order to reproduce the (spatial) covariances among the time series of the single users, the spatial cross-correlations have to be additionally preserved and can be obtained by using the method proposed by Iman & Conover (1982) (hereinafter indicated as IC). This method is based on the reordering of the samples of water demands of the single users generated independently of each other (this is an assumption; see above) so as to impose the cross-correlations among them. The method can be schematized as follows.

Let $\mathbf{X}$ be the matrix of the generated samples, made up of N_c columns, where each column represents one of the random variables considered (e.g., the water demands of the n_{us} users in the hour h of the day) and N_r rows, where each row represents one realization (e.g., the n_{die} generated days). Furthermore, let $\mathbf{R}^*$ be the target $(N_c \times N_c)$ cross-correlation matrix calculated by using the observed data and $\mathbf{R}$ the$(N_c \times N_c)$ cross-correlation matrix calculated by using the generated samples: the generic element ρ_{ij} of $\mathbf{R}$ thus represents the Pearson correlation coefficient between the ith column and the jth column of $\mathbf{X}$ (with $i,j = 1,2,\ldots,N_c$). The main steps are:

- the target cross-correlation matrix $\mathbf{R}^*$ is written as $\mathbf{R}^* = \mathbf{PP^T}$ by using Cholesky factorization (Press *et al.* 1990), $\mathbf{P}$ being a lower triangular matrix and $\mathbf{P^T}$ its transpose matrix;
- the cross-correlation matrix $\mathbf{R}$ is similarly written as $\mathbf{R} = \mathbf{SS^T}$ by using Cholesky factorization (Press *et al.* 1990);
- the matrix $\mathbf{X_1} = \mathbf{X(PS^{-1})^T}$ is computed; it is worth noting that the matrix $\mathbf{X_1}$ has the same dimensions as $\mathbf{X}$;
- the desired matrix $\mathbf{X}^*$ is created by reordering the columns of the matrix $\mathbf{X}$, so that the ranking of each column of $\mathbf{X}$ is equal to the ranking of each column of $\mathbf{X_1}$. In this way, the matrices $\mathbf{X}^*$ and $\mathbf{X_1}$ have the same rank correlation matrix, and, as a consequence, similar Pearson correlation matrices.

Summing up, imposing the spatial cross-correlations among the single user water demands through the IC method allows reproducing, according to Equations (3) and (4), the hourly *means* and *variances* of the spatially aggregated time series.

In order to preserve also the temporal covariances and correlations within the spatially aggregated time series, the IC method is again applied at this latter level by imposing the preservation of the temporal correlation coefficients from lag-1 to lag-24.

Five different ways of applying the IC procedure for the-*spatial* aggregation of the time series of the single user water demands are considered here; these variants differ for the spatial and temporal cross-correlation coefficients imposed simultaneously to the series to be spatially aggregated, as explained in the subsequent sections. Once the spatially aggregated time series is obtained, the IC method is applied again in order to preserve the temporal covariances and correlations at different time-lags.

The five variants of the procedure compared

In the first variant, called *V1*, the IC procedure is applied to the independently generated synthetic hourly time series of individual users which are expected to reproduce mean and variance for each hour h of the day; the target is the preservation of the lag-0 spatial cross-correlations among the water demands of individual users. The time series of each user are then summed up hour by hour producing the time series of the group of users. Operatively, the IC procedure is applied 24 times, one for each hour h of the day, in order to impose the lag-0 spatial cross-correlations among the water demands of individual users for each hour of the

day: the matrix **X** of the generated samples is thus made up of n_{us} columns, where each column represents the random variable water demand of the jth user (with $j = 1{:}n_{us}$) in the hth hour of the day, and the corresponding correlation matrix **R** has dimensions equal to $n_{us} \times n_{us}$.

Once the spatially aggregated time series is produced through this variant, the IC method is applied once more in order to directly impose the temporal correlations from lag-1 to lag-23. To this end, the matrix **X** is made up of 24 columns, where each column represents the random variable water demand of the spatially aggregated time series in each hour of the day. This last step will be hereinafter indicated as TC_{as} (Temporal Correlation of the aggregated series).

In the second variant, called *V2*, the IC procedure is preliminary applied to each of the synthetic hourly time series of the single users in order to reproduce the temporal correlations from lag-1 to lag-23, thus obtaining synthetic time series at user level which preserve mean, variance and temporal covariances at different time-lags. Operatively, the IC procedure is thus applied n_{us} times, one for each user: the matrix **X** of the generated samples is thus made up of 24 columns, where each column represents the random variable water demand of the considered user in each hour of the day, and the corresponding correlation matrix **R** has dimensions equal to 24×24. After that, the variant *V2* follows the same operating procedure as in *V1*, that is, the IC procedure is applied 24 times, one for each hour h of the day, to the synthetic hourly time series (which are now expected to reproduce mean, variance and covariance for each hour h of the day) in order to impose the lag-0 cross-correlations and the time series of the users are then summed up hour by hour producing the time series of the group of users. Finally the TC_{as} step is applied (see variant *V1*).

In the third variant, called *V3*, the spatial cross-correlations at lag-0 and the temporal cross-correlations among water demands at the level of individual users from lag-1 to lag-23 are simultaneously imposed. From a practical viewpoint, the IC procedure is applied in order to impose a cross-correlation matrix **R** of dimensions $24{\cdot}n_{us} \times 24{\cdot}n_{us}$ thus structured:

$$\begin{pmatrix} \begin{pmatrix} \rho_{1,1}^{1,1} & \rho_{1,2}^{1,1} & \cdots & \rho_{1,24}^{1,1} \\ \rho_{2,1}^{1,1} & \rho_{2,2}^{1,1} & \cdots & \rho_{2,24}^{1,1} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{24,1}^{1,1} & \rho_{24,2}^{1,1} & \cdots & \rho_{24,24}^{1,1} \end{pmatrix} & \begin{pmatrix} \rho_{1,1}^{1,2} & \rho_{1,2}^{1,2} & \cdots & \rho_{1,24}^{1,2} \\ \rho_{2,1}^{1,2} & \rho_{2,2}^{1,2} & \cdots & \rho_{2,24}^{1,2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{24,1}^{1,2} & \rho_{24,2}^{1,2} & \cdots & \rho_{24,24}^{1,2} \end{pmatrix} & \cdots & \begin{pmatrix} \rho_{1,1}^{1,n_{us}} & \rho_{1,2}^{1,n_{us}} & \cdots & \rho_{1,24}^{1,n_{us}} \\ \rho_{2,1}^{1,n_{us}} & \rho_{2,2}^{1,n_{us}} & \cdots & \rho_{2,24}^{1,n_{us}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{24,1}^{1,n_{us}} & \rho_{24,2}^{1,n_{us}} & \cdots & \rho_{24,24}^{1,n_{us}} \end{pmatrix} \\ \begin{pmatrix} \rho_{1,1}^{2,1} & \rho_{1,2}^{2,1} & \cdots & \rho_{1,24}^{2,1} \\ \rho_{2,1}^{2,1} & \rho_{2,2}^{2,1} & \cdots & \rho_{2,24}^{2,1} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{24,1}^{2,1} & \rho_{24,2}^{2,1} & \cdots & \rho_{24,24}^{2,1} \end{pmatrix} & \begin{pmatrix} \rho_{1,1}^{2,2} & \rho_{1,2}^{2,2} & \cdots & \rho_{1,24}^{2,2} \\ \rho_{2,1}^{2,2} & \rho_{2,2}^{2,2} & \cdots & \rho_{2,24}^{2,2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{24,1}^{2,2} & \rho_{24,2}^{2,2} & \cdots & \rho_{24,24}^{2,2} \end{pmatrix} & \cdots & \begin{pmatrix} \rho_{1,1}^{2,n_{us}} & \rho_{1,2}^{2,n_{us}} & \cdots & \rho_{1,24}^{2,n_{us}} \\ \rho_{2,1}^{2,n_{us}} & \rho_{2,2}^{2,n_{us}} & \cdots & \rho_{2,24}^{2,n_{us}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{24,1}^{2,n_{us}} & \rho_{24,2}^{2,n_{us}} & \cdots & \rho_{24,24}^{2,n_{us}} \end{pmatrix} \\ \vdots & \vdots & \ddots & \vdots \\ \begin{pmatrix} \rho_{1,1}^{n_{us},1} & \rho_{1,2}^{n_{us},1} & \cdots & \rho_{1,24}^{n_{us},1} \\ \rho_{2,1}^{n_{us},1} & \rho_{2,2}^{n_{us},1} & \cdots & \rho_{2,24}^{n_{us},1} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{24,1}^{n_{us},1} & \rho_{24,2}^{n_{us},1} & \cdots & \rho_{24,24}^{n_{us},1} \end{pmatrix} & \begin{pmatrix} \rho_{1,1}^{n_{us},2} & \rho_{1,2}^{n_{us},2} & \cdots & \rho_{1,24}^{n_{us},2} \\ \rho_{2,1}^{n_{us},2} & \rho_{2,2}^{n_{us},2} & \cdots & \rho_{2,24}^{n_{us},2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{24,1}^{n_{us},2} & \rho_{24,2}^{n_{us},2} & \cdots & \rho_{24,24}^{n_{us},2} \end{pmatrix} & \cdots & \begin{pmatrix} \rho_{1,1}^{n_{us},n_{us}} & \rho_{1,2}^{n_{us},n_{us}} & \cdots & \rho_{1,24}^{n_{us},n_{us}} \\ \rho_{2,1}^{n_{us},n_{us}} & \rho_{2,2}^{n_{us},n_{us}} & \cdots & \rho_{2,24}^{n_{us},n_{us}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{24,1}^{n_{us},n_{us}} & \rho_{24,2}^{n_{us},n_{us}} & \cdots & \rho_{24,24}^{n_{us},n_{us}} \end{pmatrix} \end{pmatrix} \tag{5}$$

where $\rho_{h,u}^{j,l}$ represents the coefficient of correlation between the water demands of the jth user and the lth user between the hth hour and the u-th hour. Once the spatially aggregated time series is obtained by summing up the reordered time series of the single users, the TC_{as} step is

applied (see variant *V1*) in order to preserve the temporal covariances.

It is worth noting that this method, depending on the number of users n_{us} involved, may imply the need to estimate and work with a very high number of correlation coefficients (very large correlation matrix). The numerical application will show that imposing such a high number of correlation coefficients can be difficult to achieve. For this reason, two other variants were considered as alternatives to *V3*.

In the first of these, called *V4*, we consider only the temporal correlations among the hourly water demands from lag-1 to lag-23 *for each user*, and the lag-0 spatial cross-correlations among the n_{us} users for every hour, and disregard the spatial cross-correlations among the users from lag-1 to lag-23. In practical terms, within the framework of the IC reordering procedure, as in *V3*, we consider a single correlation matrix $\mathbf{R}$ having dimensions of $24{\cdot}n_{us} \times 24{\cdot}n_{\mathrm{us}}$, structured in this way:

$$\begin{pmatrix}
\begin{pmatrix} \rho_{1,1}^{1,1} & \rho_{1,2}^{1,1} & \dots & \rho_{1,24}^{1,1} \\ \rho_{2,1}^{1,1} & \rho_{2,2}^{1,1} & \dots & \rho_{2,24}^{1,1} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{24,1}^{1,1} & \rho_{24,2}^{1,1} & \dots & \rho_{24,24}^{1,1} \end{pmatrix} & \begin{pmatrix} \rho_{1,1}^{1,2} & 0 & \dots & 0 \\ 0 & \rho_{2,2}^{1,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \rho_{24,24}^{1,2} \end{pmatrix} & \cdots & \begin{pmatrix} \rho_{1,1}^{1,n_{us}} & 0 & \dots & 0 \\ 0 & \rho_{2,2}^{1,n_{us}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \rho_{24,24}^{1,n_{us}} \end{pmatrix} \\
\begin{pmatrix} \rho_{1,1}^{2,1} & 0 & \dots & 0 \\ 0 & \rho_{2,2}^{2,1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \rho_{24,24}^{2,1} \end{pmatrix} & \begin{pmatrix} \rho_{1,1}^{2,2} & \rho_{1,2}^{2,2} & \dots & \rho_{1,24}^{2,2} \\ \rho_{2,1}^{2,2} & \rho_{2,2}^{2,2} & \dots & \rho_{2,24}^{2,2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{24,1}^{2,2} & \rho_{24,2}^{2,2} & \dots & \rho_{24,24}^{2,2} \end{pmatrix} & \cdots & \begin{pmatrix} \rho_{1,1}^{2,n_{us}} & 0 & \dots & 0 \\ 0 & \rho_{2,2}^{2,n_{us}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \rho_{24,24}^{2,n_{us}} \end{pmatrix} \\
\vdots & \vdots & \ddots & \vdots \\
\begin{pmatrix} \rho_{1,1}^{n_{\mathrm{us}},1} & 0 & \dots & 0 \\ 0 & \rho_{2,2}^{n_{\mathrm{us}},1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \rho_{24,24}^{n_{us},1} \end{pmatrix} & \begin{pmatrix} \rho_{1,1}^{n_{us},2} & 0 & \dots & 0 \\ 0 & \rho_{2,2}^{n_{us},2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \rho_{24,24}^{n_{us},2} \end{pmatrix} & \cdots & \begin{pmatrix} \rho_{1,1}^{n_{us},n_{us}} & \rho_{1,2}^{n_{us},n_{us}} & \dots & \rho_{1,24}^{n_{us},n_{us}} \\ \rho_{2,1}^{n_{us},n_{us}} & \rho_{2,2}^{n_{us},n_{us}} & \dots & \rho_{2,24}^{n_{us},n_{us}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{24,1}^{n_{us},n_{us}} & \rho_{24,2}^{n_{us},n_{us}} & \dots & \rho_{24,24}^{n_{\mathrm{us}},n_{\mathrm{us}}} \end{pmatrix}
\end{pmatrix} \tag{6}$$

where it is evident that only the correlation coefficients $\rho_{h,u}^{j,l}$ in which either $j = l$ (independently of h and u, all the temporal correlations from lag-1 to lag-23 for every user) or $h = u$ (spatial correlations among users at lag-0) are considered, while all the other coefficients are assumed to be equal to 0. Finally, the TC_{as} step is applied (see variant *V1*) on the spatially aggregated time series.

In the second alternative (to *V3*), called *V5*, the water demands are spatially aggregated by applying the same operating procedures as in *V3* but in successive steps, each time considering subgroups of the total n_{us} users to be aggregated; as a result, the dimensions of the cross-correlation matrix will be smaller. Practically speaking, the n_{us} users are evenly divided into n_g subgroups and for each subgroup *V3* is applied in order to obtain the spatially aggregated time series of the subgroup; finally, *V3* is applied once again, in this case to the series of n_g spatially aggregated subgroups. For example, assuming a number of users $n_{\mathrm{us}} = 20$, these users could be divided into $n_g = 4$ subgroups containing five users each: by applying *V3* to each subgroup it is possible to obtain $n_g = 4$ series of 'partially' spatially aggregated water demands, which are in turn spatially aggregated, again by applying *V3*, so as to obtain the time series of the overall aggregate of $n_{\mathrm{us}} = 20$ users; this approach enables us to work with correlation matrices which are smaller than the one that would result if all users were considered together. In this case as well, the TC_{as} step is finally applied (see variant *V1*) to the spatially aggregated time series.

APPLICATION

Case study

The proposed procedure (featuring the five variants) was applied to the water demands of 21 users of the WDS of Milford (Ohio) (Buchberger *et al.* 2003); the observation period had a length of 31 days, from 11 May to 10 June 1997. The hourly time series for each of the $n_{us} = 21$ users to be spatially aggregated were generated by using the procedure described by Alvisi *et al.* (2014b); the length of each generated hourly time series representing the water demand of a user is equal to $n_{die} = 90$ days and the generation process, and consequently the spatial aggregation process, was repeated $n_{rep} = 500$ times. The generated synthetic series reproduce, for each user, the hourly mean and variance of the corresponding observed series. More details about the generations of the time series with these characteristics are provided in Alvisi *et al.* (2014b).

Application of the variants of the procedure and discussion of the results

The hourly water demands of each user were spatially aggregated by applying the five variants previously illustrated; moreover, in order to obtain a basis for comparing them, we also evaluated the results obtained by simply adding up the hourly water demands of the n_{us} users which preserve the hourly mean and variance. In all the cases, the TC_{as} step was systematically applied to the spatially aggregated time series. The efficacy of each variant was evaluated by using the coefficient of determination (CD) defined as

$$CD = 1 - \frac{\sum_{i=1}^{24} (\varepsilon_i)^2}{\sum_{i=1}^{24} (So_i - \mu_{So})^2} \quad (7)$$

where $\varepsilon_i = So_i - Ss_i$ is the ith difference between the value of the statistic of interest So_i (variance, lag-1 temporal correlation coefficient, etc.) obtained from the observed series at the ith hour of the day and the corresponding value Ss_i obtained as average of the $n_{rep} = 500$ synthetic series; μ_{So} is the daily average of the values of the statistic of interest obtained from the observed series; the closer the value of CD is to 1, the better the statistic of interest will be reproduced.

In Figure 1, the variances, the lag-1 correlation coefficients before and after the TC_{as} step, and the lag-1 temporal covariances after the TC_{as} step, of the spatially aggregated synthetic series obtained following the application of the five variants are presented and compared with the corresponding observed statistics; in particular, as regards the synthetic series, the statistics obtained for each of the $n_{rep} = 500$ series are shown, along with the values averaged over the n_{rep} synthetic series.

First, as illustrated in Figure 1(a), the simple sum of the individual user water demands results in an underestimation of the (hourly) variance of the aggregated historical time series, returning a CD of 0.27. This is due to the fact that no spatial cross-correlations have been imposed and preserved, and in particular no spatial cross-correlations at lag-0 for each of the 24 hours; therefore, the sum of the covariances of Equation (3) is close to zero, resulting in an underestimation of the variance of the aggregated data. Clearly, also the lag-1 temporal correlations of the spatially aggregated time series before the TC_{as} step are not reproduced, the users' time series being temporally uncorrelated, whereas following the TC_{as} step, the lag-1 correlations of the aggregated time series are improved; overall, however, the lag-1 temporal covariances of the aggregated time series are not well reproduced (CD = 0.11) due to the low effectiveness in reproducing the variances (see Equation (4)).

Variant *V1* involves reordering the users' water demands so as to impose lag-0 spatial cross-correlations among the water demands of the individual users. For example, Figure 2 shows the target cross-correlation matrix $\mathbf{R}^*$ calculated by using the observed data for hour $h = 12$ a.m. and the cross-correlation matrix $\mathbf{R}$ calculated by using the generated samples before and after applying the Iman & Conover (1982) approach. As can be observed, the approach allows for obtaining generated water demands whose cross-correlation matrix approximates very well the observed one.

Consequently, as shown in Figure 1(b), *V1* enables excellent reproduction of the variance of the aggregated historical time series, with a CD of 0.94; as for the previous case, the lag-1 temporal correlations of the aggregated time series before the TC_{as} step are not reproduced, the aggregated users' time series being temporally uncorrelated, whereas

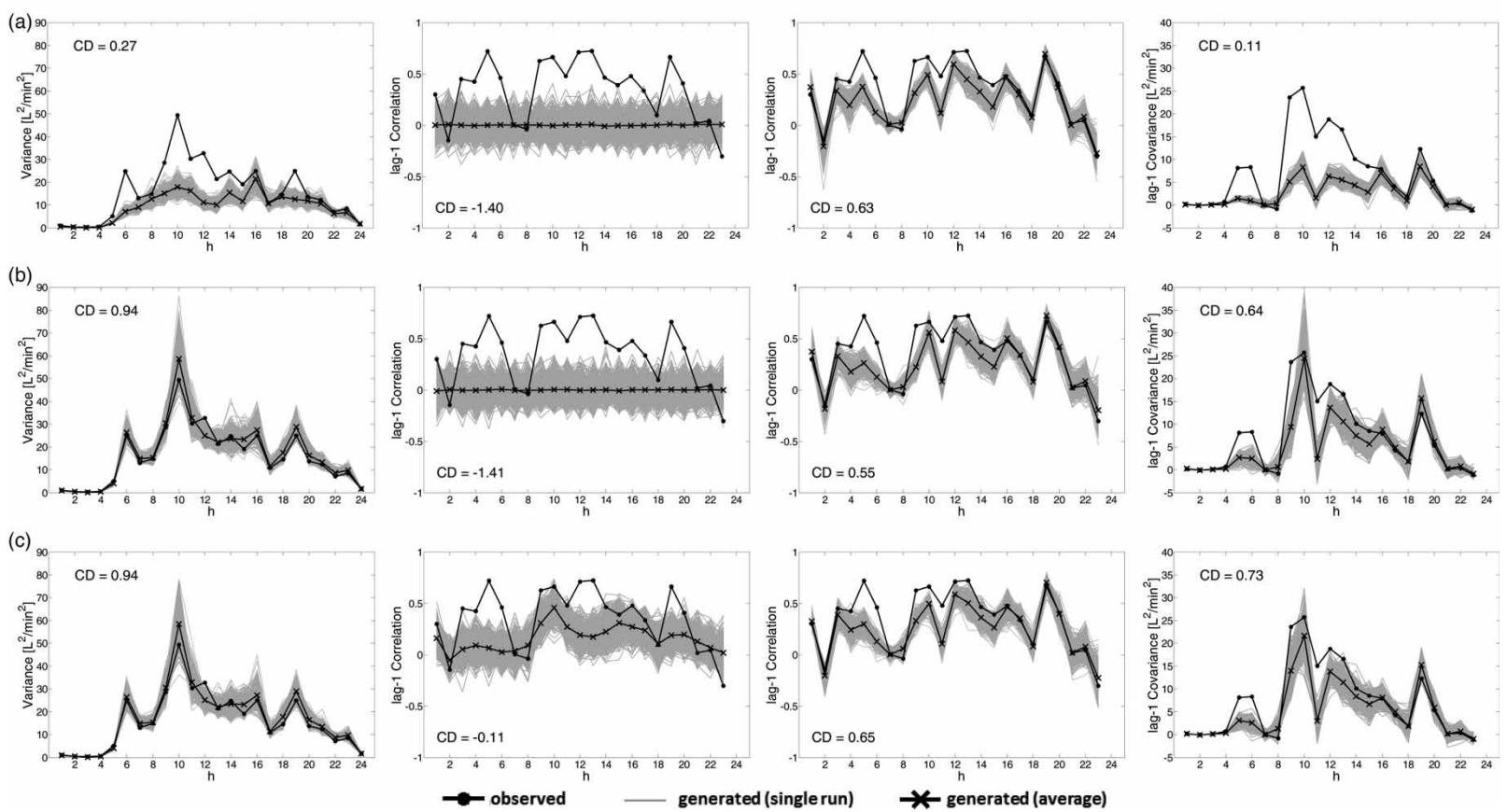

Figure 1 | Comparison of the variances (column I), lag-1 temporal correlation coefficients before (column II) and after (column III) the TC_{as} step, and lag-1 temporal covariances after the TC_{as} step (column IV), associated with the statistics of the observed time series and those of 500 synthetic series, as well as their average values, obtained through (a) calculation of the simple sum of water demands, and application of the variants (b) *V1*, (c) *V2*, (d) *V3*, (e) *V4* and (f) *V5*. *(Continued.)*

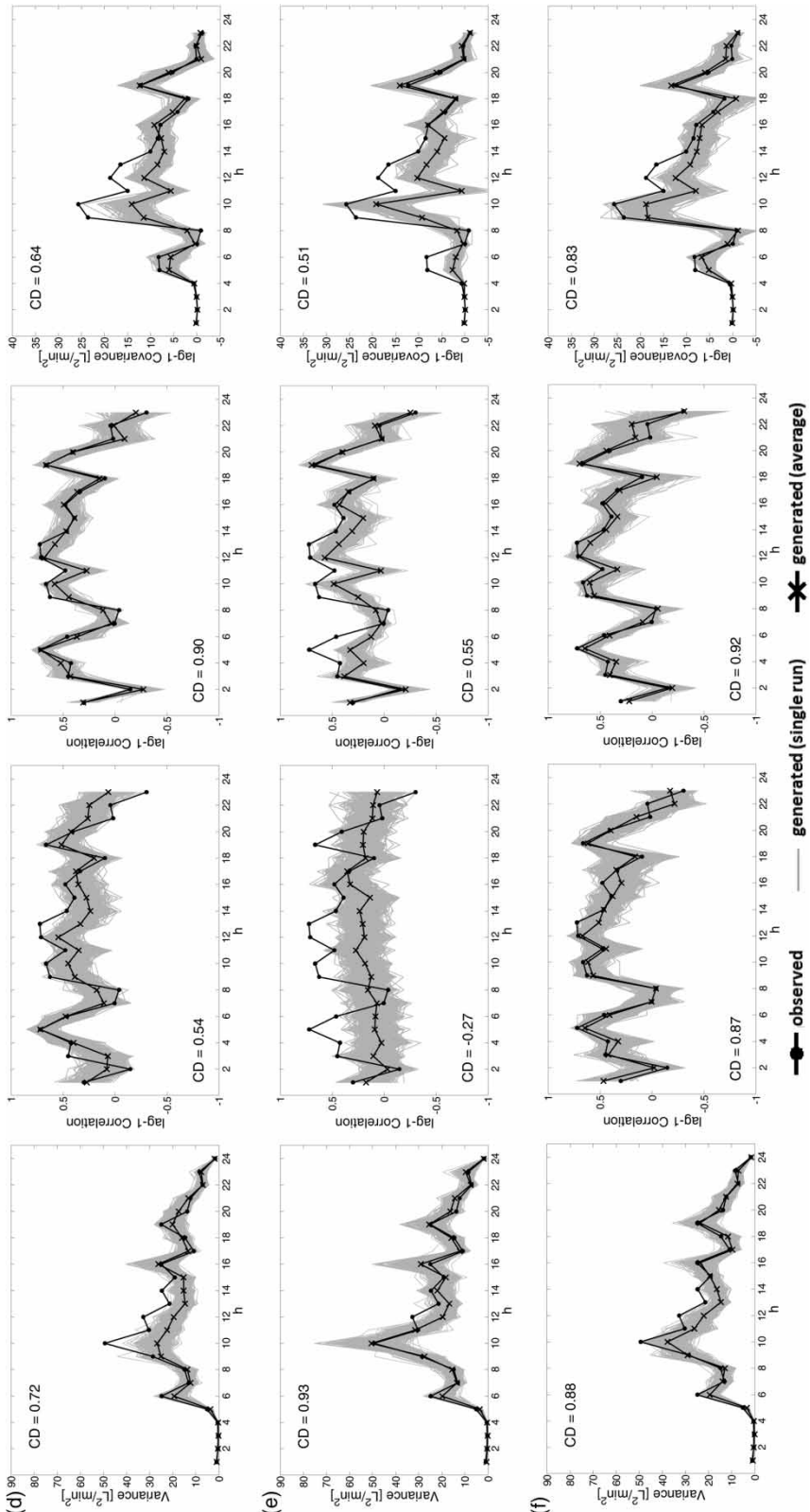

Figure 1 | Continued.

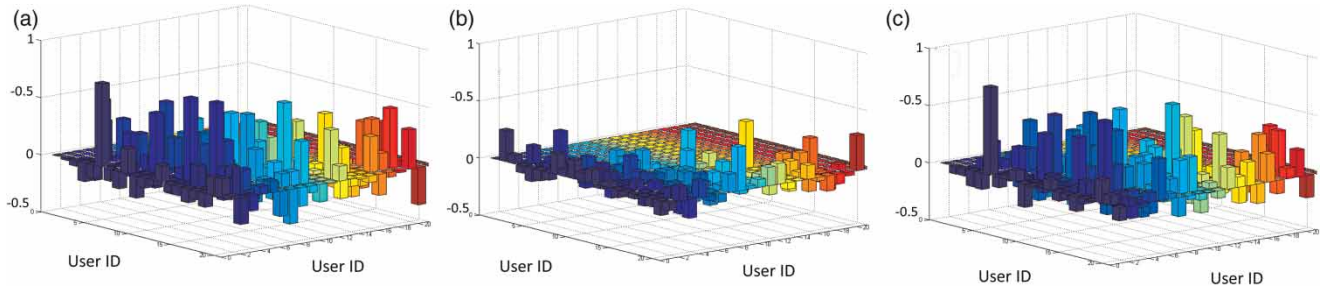

Figure 2 | Variant *V1*: comparison of (a) the target cross-correlation matrix **R*** calculated by using the observed data and the cross-correlation matrix **R** calculated by using the generated samples (b) before and (c) after applying the Iman & Conover (1982) approach.

following the TC_{as} step, the lag-1 correlations of the aggregated time series are improved, leading, together with the excellent reproduction of the variance (see Equation (4)), to a rather good reproduction of the temporal lag-1 covariances (CD = 0.64).

V2 is similar to *V1*, but the single users' water demand time series are reordered in order to impose their own temporal correlations preliminary to the imposition of the lag-0 cross-correlations. As can be observed in Figure 1(c), this variant shows the same effectiveness of the previous one as far as the variance concerns; on the other hand, thanks to the impositions of the temporal correlations on each user time series, the lag-1 temporal correlation of the aggregated series before the TC_{as} step are better reproduced, and slightly better performances (with respect to *V1*) in reproducing both the lag-1 correlations (CD = 0.65 vs. CD = 0.55) and covariances (CD = 0.73 vs. CD = 0.64) can be observed also after the application of the TC_{as} step.

V3 simultaneously imposes *all* the temporal and spatial cross-correlations from lag-0 to lag-23 (see Equation (5)) within and among the different users. As shown in Figure 1(d), *V3* enables good reproduction of the variance of the aggregated historical time series, with a CD of 0.72, although this value is lower than the one obtained with *V1* and *V2*; the effectiveness in reproducing the lag-1 cross-correlation of the aggregated series is good before the TC_{as} step (CD = 0.54) and it significantly increases after the TC_{as} step (CD = 0.90), leading to a CD = 0.64 in reproducing the lag-1 covariances, as for *V1*. *V3* is less effective than *V1* and *V2* in reproducing the variance due to the fact that (1) the variance in each hour depends on the lag-0 spatial cross-correlations and (2) a larger number of correlation coefficients is imposed in *V3* than in *V1* and *V2*: in *V1* and *V2* only spatial cross-correlations at lag-0 are imposed whereas in *V3* all spatial-temporal cross-correlations from lag-0 to lag-23 are imposed. In practical terms, when we seek to impose so many correlation coefficients simultaneously, we obtain less accuracy in reproducing the lag-0 spatial cross-correlations compared to *V1* and *V2*, and thus the variances of the aggregated historical time series are not so well preserved.

V4 simultaneously imposes temporal correlations for every user and lag-0 spatial cross-correlations through the construction of a single banded correlation matrix (see Equation (6)). As shown in Figure 1(e), *V4* enables excellent reproduction of the variance of the aggregated historical time series, similar to *V1* and *V2*, with a CD of 0.93; the improvement over *V3* in terms of reproducing the variance is understandable considering the smaller number of cross-correlations imposed. However, it performs worse than *V3* in reproducing the lag-1 temporal correlation of the aggregated series both before and after the TC_{as} step resulting in a CD = 0.51 for the lag-1 temporal covariance of the aggregated time series, lower than the corresponding ones of *V1*, *V2* and *V3*.

V5, based on the aggregation of water demands in successive steps, was applied by dividing the $n_{us} = 21$ users into $n_g = 3$ groups of seven users each. Compared to *V3* it involves estimating a smaller number of correlation coefficients for each matrix. As shown in Figure 1(f), *V5* enables good reproduction of the variance of the aggregated historical time series, with a CD of 0.88; it is, moreover, highly effective in reproducing the lag-1 temporal correlation of the aggregated series both before and after the TC_{as} step resulting in a CD = 0.83 for the lag-1 temporal covariance of the aggregated time series, the highest value among all the variants.

Table 1 | Ranking of the five variants according to the effectiveness in reproducing the variance, lag-1 temporal correlation and lag-1 temporal covariance and the number of time the IC method is applied

	Variance		Lag-1 temporal correlation		Lag-1 temporal covariance		*IC* method application			
	CD	Rank	CD	Rank	CD	Rank	n_{IC}	Rank	Σrank	Total rank
V1	*0.94*	1	*0.55*	4	*0.64*	3	*24*	4	12	4
V2	*0.94*	1	*0.65*	3	*0.73*	2	*55*	5	11	2
V3	*0.72*	5	*0.90*	2	*0.64*	3	*1*	1	11	3
V4	*0.93*	3	*0.55*	4	*0.51*	5	*1*	1	13	5
V5	*0.88*	4	*0.92*	1	*0.83*	1	*4*	3	9	1

In short, as summarized in Table 1, where the five variants are ranked according to the effectiveness in reproducing each statistic, *V1* and *V2*, followed by *V4*, *V5* and *V3* showed better results in reproducing the hourly variance of the aggregated historical time series; *V5* in turn was shown to be the most effective in reproducing both the lag-1 temporal correlation and lag-1 temporal covariance of the aggregated historical time series after the TC_{as} step, followed by *V3* and *V2* as far as the lag-1 temporal correlation was concerned and by *V2* and *V1* for the lag-1 temporal covariance.

Finally, in terms of computational burden, it is worth noting that the number of times n_{IC}, the IC method, is applied (before the TC_{as} step, which is common to all the variants) can vary significantly from one variant to the other, as detailed in Table 1: in fact, in *V1* the IC method is applied 24 times, one for each hour h of the day, in order to impose the lag-0 spatial cross-correlations among the water demands of individual users, whereas in *V2*, remembering that also the temporal correlations within each user water demand time series are preliminarily imposed, the IC method is applied $n_{IC} = 21 + 24 = 55$ times; in *V3* and *V4* the IC method is applied only once, even if the correlation matrix is very large, and in *V5* the IC method is applied three times, considering each time a group of seven users, and once considering the three resulting partially aggregated series, for a total of $n_{IC} = 4$ times.

Overall, considering both the effectiveness in reproducing all the statistics and the computational burden, a total ranking of the five variants can be computed using the approach for ranking solutions according to different criteria proposed by Ostfeld *et al.* (2012), that is, summing up for each solution the rank values of the solution corresponding to each criteria, and identifying the solution which is characterized by the lowest total rank value. In this case, thus, the rank values obtained by each variant were summed up (see Table 1, column 'Σrank') and the variant characterized by the lowest total value identified. *V5* is the variant characterized by the best total rank, providing a good trade-off between effectiveness in reproducing the statistics and computational burden.

CONCLUSIONS

This paper proposes and compares five different variants of a procedure for the spatial aggregation of synthetic water demand time series all based on the IC method. The variants differ for the spatial and temporal cross-correlation coefficients imposed to the series to be spatially aggregated and thus for the way the IC method is applied. Application of these variants to a real case involving the water demands of 21 users showed that all of them allowed the statistics of interest of the spatially aggregated historical time series to be reproduced. In particular, all of them provide better results than the simple sum of spatially uncorrelated synthetic time series in terms of ability to preserve the hourly variance of the aggregated data. This confirms the importance of preserving the spatial correlations of individual user water demands in the spatial aggregation procedure. On the other hand, comparing the variants among themselves, it was observed that the degree of efficiency in reproducing the statistics, and the computational burden, varied from one variant to another. In the specific case concerned, it was observed that the procedure of aggregation

based on user subgroups (variant *V5*) is the one that shows the best trade-off between computational burden and effectiveness in reproducing the main statistics, means, variances, lag-1 temporal correlations and lag-1 temporal covariances of the aggregated historical time series. In conclusion, the results obtained showed that the procedure in general, and the variant *V5* in particular, represent a valid tool for 'transferring' the time series generated at low levels of spatial (individual user) aggregation to series relating to groups of users, thereby enabling the main statistics of the corresponding historical time series to be preserved at these higher levels of spatial aggregation.

ACKNOWLEDGEMENTS

The authors wish to thank Prof. Steven Buchberger for providing the Milford data that were used in the numerical applications. This study was carried out as part of the PRIN 2012 project 'Tools and procedures for an advanced and sustainable management of water distribution systems' and under the framework of Terra & Acqua Tech Laboratory, Axis I activity 1.1 of the POR FESR 2007–2013, project funded by the Emilia-Romagna Regional Council (Italy) (http://fesr.regione.emiliaromagna.it/allegati/comunicazione/la-brochure-dei-tecnopoli).

REFERENCES

Alcocer, V. H., Buchberger, S. G. & Tzatchkov, V. 2006 Instantaneous water demand parameter estimation from coarse meter readings. In: *ASCE, Proceedings of 8th Annual International Symposium on Water Distribution Systems Analysis*. University of Cincinnati, Cincinnati, OH, available on CD-Rom.

Alvisi, S., Franchini, M. & Marinelli, A. 2003 A stochastic model for representing drinking water demand at residential level. *Water Resour. Manage* **17** (3), 197–222.

Alvisi, S., Ansaloni, N. & Franchini, M. 2014a A procedure for spatial aggregation of synthetic water demand time series. In: *Proc. 12th Int. Conf. on Computing and Control for the Water Industry – CCWI2013*. Procedia Engineering, Elsevier, Perugia, 70, pp. 51–60.

Alvisi, S., Ansaloni, N. & Franchini, M. 2014b Generation of synthetic water demand time series at different temporal and spatial aggregation levels. *Urban Water* **11** (4), 297–310.

Babayan, A., Kapelan, Z., Savic, D. & Walters, G. 2005 Least-cost design of water distribution networks under demand uncertainty. *J. Water Resour. Plann. Manage.* **131** (5), 19–26.

Bakker, M., Vreeburg, J. H. G., van Schagen, K. M. & Rietveld, L. C. 2013 A fully adaptive forecasting model for short-term drinking water demand. *Environ. Modell. Softw.* **48**, 141–151.

Blokker, E. J. M., Vreeburg, J. H. G. & van Dijk, J. C. 2010 Simulating residential water demand with a stochastic end-use model. *J. Water Resour. Plann. Manage.* **136** (1), 375–382.

Buchberger, S. G. & Wu, L. 1995 Model for instantaneous residential water demands. *J. Hydraul. Eng.* **121** (3), 232–246.

Buchberger, S. G., Carter, J. T., Lee, Y. & Schade, T. G. 2003 *Random Demands, Travel Times, and Water Quality in Deadends*. AWWA Research Foundation, Denver, CO, USA.

Filion, Y. R., Adams, B. & Karney, B. 2007 Cross correlation of demands in water distribution network design. *J. Water Resour. Plann. Manage.* **133** (2), 137–144.

Guercio, R., Magini, R. & Pallavicini, I. 2001 Instantaneous residential water demand as stochastic point process. In: *Water Resources Management* (C. A. Brebbia, P. Anagnostopoulos, K. L. Katsifarakis & A. H.-D. Cheng, eds). WIT Press, Southampton, UK, pp. 129–138.

Iman, R. L. & Conover, W. J. 1982 A distribution-free approach to inducing rank correlation among input variables. *Commun. Statist.* **11**, 311–334.

Kottegoda, N. T. & Rosso, R. 1997 *Statistic, Probability and Reliability for Civil and Environmental Engineers*. McGraw-Hill, New York, USA.

Moughton, L. J., Buchberger, S. G., Boccelli, D. L., Filion, Y. R. & Karney, B. W. 2006 Effect of time step and data aggregation on cross correlation of residential demands. In: *ASCE, Proceedings of 8th Annual International Symposium on Water Distribution Systems Analysis*. University of Cincinnati, Cincinnati, OH, available on CD-Rom.

Ostfeld, A., Salomons, E., Ormsbee, L., Uber, J. G., Bros, C. M., Kalungi, P., Burd, R., Zazula-Coetzee, B., Belrain, T., Kang, D., Lansey, K., Shen, H., McBean, E., Wu, Z. Y., Walski, T., Alvisi, S., Franchini, M., Johnson, J. P., Ghimire, S. R., Barkdoll, B. D., Koppel, T., Vassiljev, A., Kim, J. H., Chung, G., Yoo, D. G., Diao, K., Zhou, Y., Li, J., Liu, Z., Chang, K., Gao, J., Qu, S., Yuan, Y., Prasad, T. D., Laucelli, D., Vamvakeridou Lyroudia, L. S., Kapelan, Z., Savic, D., Berardi, L., Barbaro, G., Giustolisi, O., Asadzadeh, M., Tolson, B. A. & McKillop, R. 2012 The Battle of the Water Calibration Networks (BWCN). *J. Water Resour. Plann. Manage.* **138** (5), 523–532.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. 1990 *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, New York, USA.

First received 19 March 2014; accepted in revised form 15 January 2015. Available online 19 February 2015