Available online at www.sciencedirect.com
ScienceDirect
Procedia Engineering 70 (2014) 51 – 60
Procedia Engineering
www.elsevier.com/locate/procedia

12th International Conference on Computing and Control for the Water Industry, CCWI2013

A procedure for spatial aggregation of synthetic water demand time series

S. Alvisi[a]*, N. Ansaloni[a], M. Franchini[a]

[a]*Department of Engineering, University of Ferrara, Via Saragat 1, 44122 Ferrara, Italy*

Abstract

A procedure for spatial aggregation of synthetic water demand time series is presented. Starting from synthetic water demand time series generated at user level and reproducing mean and variance of the corresponding observed series, the procedure allows the aggregated series to preserve the statistics of interest observed at the aggregation level considered. The procedure uses a method proposed by Iman and Conover (1982): the synthetic user water demand time series are reordered to preserve the observed spatial correlations of appropriate lag-time; then, the spatial aggregation of these series leads to a series representative of the user group reproducing the corresponding means and variances for the different hours of the generic day. Four different ways of performing this aggregation have been investigated and compared. Application to a case study consisting of the water demands of 21 users highlights that the approaches considered show different levels of effectiveness in reproducing the statistics, but overall the procedure proposed is a valid tool for the bottom-up generation of synthetic water demand time series.

Selection and peer-review under responsibility of the CCWI2013 Committee

Keywords: water demands; spatial aggregation; time series.

1. 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 robust and accurate hydraulic model of a WDS. To this

* Corresponding author. Tel.: +39-0532-974849; fax: +39-0532-974870.
E-mail address: stefano.alvisi@unife.it

1877-7058
Selection and peer-review under responsibility of the CCWI2013 Committee
doi:10.1016/j.proeng.2014.02.007

end several models for the generations 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 and 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 showed to be very effective in statistically reproducing the observed data at 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 consumptions 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 covariances at different time lags). More in details, Alvisi et al. (2003) highlighted that reproducing the covariances (and thus the cross-correlation) among the series to be aggregated (both temporally and spatially) is a prerequisite in order to proper reproduce the variances of the aggregated series.

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 level (e.g. at 1-hour time step for groups of users), since these latter are the series typically used within the simulation models for the design and management of the WDSs. At the same time, the main statistics (mean and variance) 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. (2013). 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. To this end, a spatial aggregation procedure based on the method proposed by Iman and Conover (1982) is presented; four different ways of applying this method are presented and compared. In the following, the proposed procedure is described and the results obtained by its application to the cases study consisting of the water demands of 21 users of the water distribution system of Milford (Ohio) (Buchberger et al. 2003) are discussed; finally, conclusions are presented.

2. 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 Δt = 1 hour time step are available; in particular, let us assume that these series have been generated independently each other and in such a way that they reproduce at hourly level mean, variance and temporal covariances (at the proper time-lag, as better explained in the subsequent sections) of the corresponding observed time series (see, for example, Alvisi et al., 2013 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 of the k-th day (with h=1:24, k=1:n_{die} where n_{die} is the number of days considered/generated) is given by:

$$Q_{h,k} = \sum_{j=1}^{n_{us}} q_{h,k}^{j} \tag{1}$$

where $q_{h,k}^{j}$ is the water demand of the *j*-th user (with j=1:n_{us}) in the corresponding hour *h* of *k*-th day. 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 lower spatial aggregation level (i.e. at single user level). As is well known from the scientific literature (e.g. Kottegoda and Rosso, 1999), 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:

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

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

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

It is worth noting that, according to eq. 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 eq. 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 order to impose the spatial cross-correlations at the single user water demands the method proposed by Iman and Conover (1982) (hereinafter indicated as *IC*) has been used. This method is based on the reordering the samples of water demands of the single users generated independently each other so as to impose the cross-correlations observed at the same level of spatial aggregation. The method can be schematized as follows.

Let **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 **R*** be the target (N_c x N_c) cross-correlation matrix calculated by using the observed data and **R** the (N_c x N_c) cross-correlation matrix calculated by using the generated samples: the generic element ρ_{ij} of **R** thus represent the Pearson correlation coefficient between the *i*-th column and the *j*-th column of **X** (with $i,j = 1,2,\ldots, N_c$). The main steps are:

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

In particular, for the spatial aggregation of the time series of the single user water demands *four* different ways of applying the *IC* procedure have been developed; these methods differ for the spatial and temporal cross-correlation coefficients imposed to the series to be spatially aggregated, as explained in the subsequent sections.

In the first method, called method A, the *IC* procedure is applied to the synthetic hourly time series which are expected to reproduce mean and variance for each hour h of the day and the temporal co-variances from lag-1 to lag-23,; 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 thus 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 *j-th* user (with $j = 1{:}n_{us}$) in the h-th hour of the day, and the corresponding correlation matrix **R** has dimensions equal to n_{us} x n_{us} . It is worth noting that imposing the lag-0 spatial cross-correlations ensures that the hourly variances of the spatially aggregated series are reproduced, but, from a theoretical point of view, it cannot be ensured that also the temporal co-variances at different time lags are reproduced at the aggregated level since the method is applied hour by hour, each independently of the other.

In the second method, called method B, 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. Given its very nature, this method can thus be applied starting from synthetic hourly water demand time series at level of single users which reproduce the mean and variance of the historical hourly demand time series, whereas these series are not required to reproduce also the temporal covariances (for the different lags), since these latter are imposed within the method simultaneously with the spatial cross-correlations at lag-0. From a practical viewpoint, the *IC* procedure is applied in order to impose a fit to 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{4}$$

where $\rho_{h,u}^{j,l}$ represents the coefficient of correlation between the water demands of the j-th user and the l-th user between the h-th hour and the u-th hour. It is worth noting that this method, since it imposes the preservation of the temporal correlation *within* each single user and the spatial correlation for the different lags *among* the single users, theoretically allows to preserve mean, variance and temporal correlation for different lags of the spatially aggregated series. On the other hand, depending on the number of users n_{us} involved, defining the correlation matrix 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 methods were considered as alternatives to method B.

In the first of these, defined method C, 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 case B, we consider a single correlation matrix **R** having dimensions of $24 \cdot n_{us}$ x $24 \cdot n_{us}$, structured in this way:

$$\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} & 0 & \cdots & 0 \\ 0 & \rho_{2,2}^{1,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_{24,24}^{1,2} \end{pmatrix} & \cdots &
\begin{pmatrix} \rho_{1,1}^{1,n_{us}} & 0 & \cdots & 0 \\ 0 & \rho_{2,2}^{1,n_{us}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_{24,24}^{1,n_{us}} \end{pmatrix} \\
\begin{pmatrix} \rho_{1,1}^{2,1} & 0 & \cdots & 0 \\ 0 & \rho_{2,2}^{2,1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \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}} & 0 & \cdots & 0 \\ 0 & \rho_{2,2}^{2,n_{us}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_{24,24}^{2,n_{us}} \end{pmatrix} \\
\vdots & \vdots & \ddots & \vdots \\
\begin{pmatrix} \rho_{1,1}^{n_{us},1} & 0 & \cdots & 0 \\ 0 & \rho_{2,2}^{n_{us},1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_{24,24}^{n_{us},1} \end{pmatrix} &
\begin{pmatrix} \rho_{1,1}^{n_{us},2} & 0 & \cdots & 0 \\ 0 & \rho_{2,2}^{n_{us},2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \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} \quad (5)$$

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.

In the second alternative method (to method B), defined method D, the water demands are spatially aggregated by applying the same operating procedures as in method B 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 method B is applied in order to obtain the spatially aggregated time series of the subgroup; finally, method B is applied once again, in this case to the series of n_g spatially aggregated subgroups. For example, assuming a number of users $n_{us} = 20$, these users could be divided into $n_g = 4$ subgroups containing 5 users each: by applying procedure B 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 method B, so as to obtain the time series of the overall aggregate of $n_{us} = 20$ 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.

3. Application

3.1. Case study

The proposed procedure was applied to the water demands of 21 users of the water distribution system 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. (2013); 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 and, for the application of the method A, also the lag-1 to lag-23 temporal covariances. More details about the generations of the time series with these characteristics are provided in Alvisi et al. (2013).

3.2. Application of the procedure and discussion of the results

The hourly water demands of each user were spatially aggregated by applying the 4 different methods previously illustrated; moreover, in order to obtain a basis for comparing the four spatial aggregation methods developed (A, B, C and D), we also evaluated the results obtained by simply adding up the hourly water demands of the n_{us} users which preserve the hourly mean, variance and covariance at different time lags for each user considered, but not the spatial cross-correlations at different lags (i.e. the starting series used for the application of method A).

In Fig.1 the variance and the lag-1 temporal covariance of the observed spatially aggregated hourly time series and of the synthetic series obtained following the application of the different spatial aggregation methods proposed are compared; 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.

Firstly, as illustrated in Fig.1a, the sum of the temporally correlated individual user water demands results in an underestimation of the variance of the aggregated historical time series, returning a coefficient of determination *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 eq.(3) is close to zero, resulting in an underestimation of the variance of the aggregated data. There is also marked underestimation of the lag-1 temporal covariance of the aggregated historical time series.

Method A involves reordering the temporally correlated user water demands from lag-1 to lag-23 so as to impose lag-0 spatial cross-correlations among the water demands of the individual users. As shown in Fig.1b, method A enables excellent reproduction of the variance of the aggregated historical time series, with a *CD* of 0.93; there is also an improvement in reproducing the lag-1 temporal covariance as compared to the simple sum of the water demands of the individual users, given that a *CD* of 0.46 is achieved.

Method B simultaneously imposes *all* the temporal and spatial cross-correlations from lag-0 to lag-23 (see eq. 4) within and among the different users. As shown in Fig.1c, method B enables good reproduction of the variance of the aggregated historical time series, with a *CD* of 0.73, though this value is lower than the one obtained with method A; the effectiveness in reproducing the lag-1 temporal covariance of the aggregated historical time series (*CD* = 0.47) is comparable to that of method A. Method B is less effective than method A in reproducing the variance due to the fact that, as observed previously, (a) the variance in each hour depends on the lag-0 spatial cross-correlations and (b) a larger number of correlation coefficients is imposed in method B than in method A: in method A only lag-0 spatial cross-correlations are imposed whereas in method B 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 method A, and thus the variances of the aggregated historical time series are not so well preserved.

Method C simultaneously imposes temporal correlations for every user and lag-0 spatial cross-correlations through the construction of a single banded correlation matrix (see eq. 5). As shown in Fig.1d, method C enables excellent reproduction of the variance of the aggregated historical time series, with a *CD* of 0.93; the improvement over method B in terms of reproducing the variance is understandable considering the smaller number of cross-correlations imposed. However, it performs worse than method B in reproducing the lag-1 temporal covariance of the aggregated historical time series, resulting in a *CD* of 0.04.

Method D, based on aggregating water demands in successive steps, was applied by dividing the $n_{us} = 21$ users into $n_g = 3$ groups of 7 users each. Compared to method B it involves estimating a smaller number of correlation coefficients for every matrix (42084 in the first aggregation and 2556 in the second, for a total of 44640, equivalent to about a third of the coefficients to be estimated in method B). As shown in Fig.1e, method D 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 temporal covariance at lag 1 (*CD* = 0.83).

In short, methods A, C and D showed better results in reproducing the hourly variance of the aggregated historical time series as compared both to the simple sum of hourly demands and to method B, and method D in turn demonstrated to be the most effective in reproducing the lag-1 temporal covariance of the aggregated historical time series.

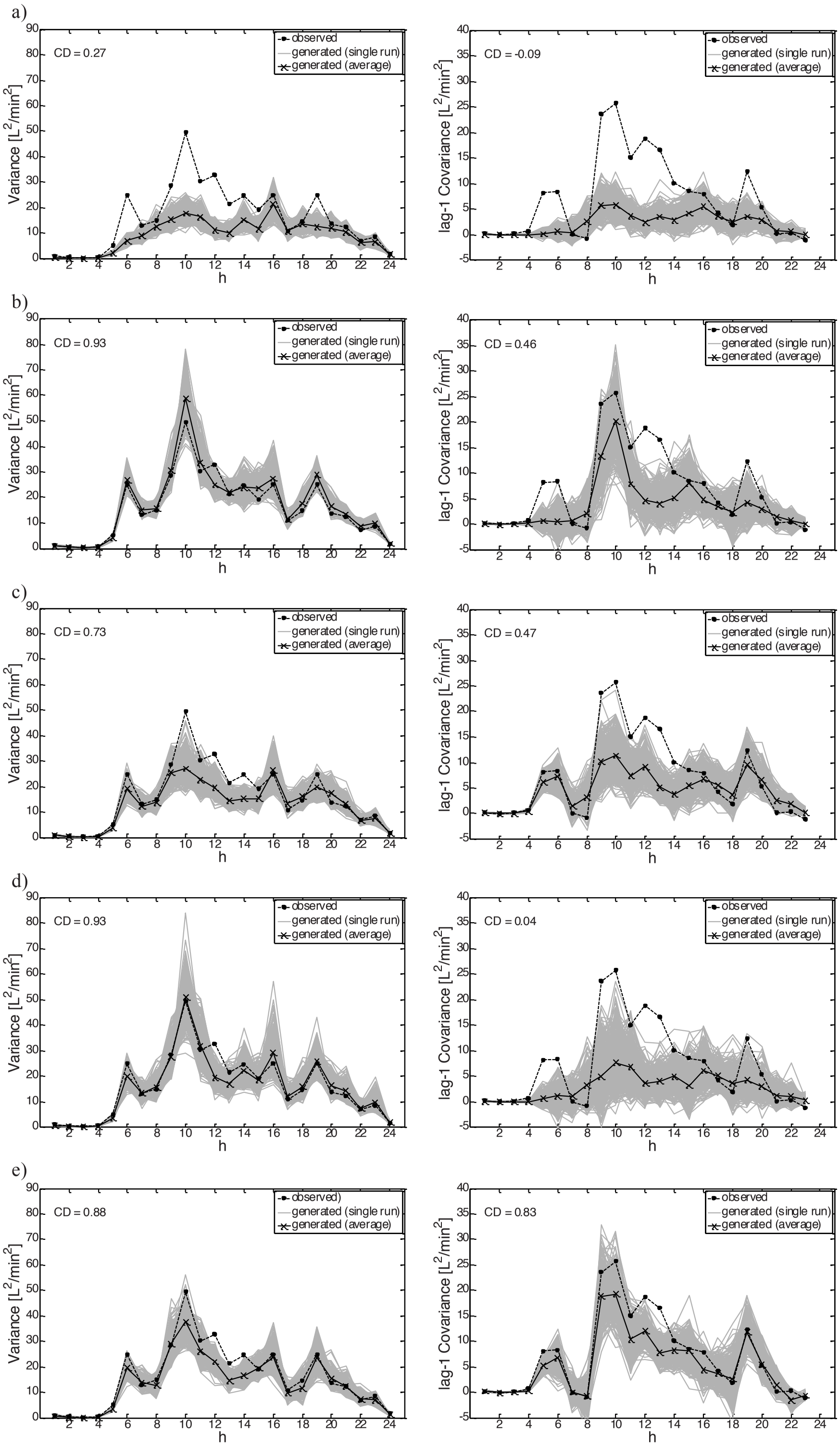
a)
b)
c)
d)
e)
Variance [L^2/min^2]
lag-1 Covariance [L^2/min^2]
h
observed
generated (single run)
generated (average)
CD = 0.27
CD = -0.09
CD = 0.93
CD = 0.46
CD = 0.73
CD = 0.47
CD = 0.93
CD = 0.04
CD = 0.88
CD = 0.83

Fig. 1. Spatial aggregation: comparison of the lag-1 temporal variances and covariances associated with the observed time series and each of the 500 synthetic series, as well as their average values, following a) calculation of the simple sum of water demands and b) application of method A, c) method B, d) method C and e) method D.

4. Conclusions

This paper proposes a procedure for the spatial aggregation of synthetic water demand time series based on the *IC* method. In particular, 4 different methods of application of the *IC* method have been investigated and compared. Application of these methods to a real case involving the water demands of 21 users showed that all these methods allow to reproduce the statistics of interest of the aggregated historical time series, though the degree of efficiency varied from one method to another. In particular, in terms of ability to preserve the hourly variance of the aggregated data, all 4 of the proposed spatial aggregation procedures based on the *IC* method provide better results than the simple sum of spatially uncorrelated synthetic series. This confirms the importance of preserving the spatial correlations of individual user water demands in the spatial aggregation procedure. In the specific case concerned, however, it was observed that the procedure of aggregation by user subgroups is the one that shows to be most effective in reproducing not only the variance but also the lag-1 temporal covariance of the aggregated historical time series. In conclusion, the results obtained showed that the procedure developed represents 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 (mean, variance and temporal correlations) 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 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. ASCE, Proceeding of 8th Annual International Symposium on Water Distribution Systems Analysis, Cincinnati, OH (USA).

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

Alvisi, S., Ansaloni, N., Franchini, M., 2013. Generation of Synthetic Water Demand Time Series at Different Temporal and Spatial Aggregation Levels. Urban Water (in press).

Babayan, A., Kapelan, Z., Savic, D., Walters, G., 2005. Least-cost design of water distribution networks under demand uncertainty. Journal of Water Resources Planning and Management 131(5), 19-26.

Blokker, E. J. M., Vreeburg, J. H. G., van Dijk, J. C., 2010. Simulating Residential Water Demand with a stochastic end-use model. Journal of Water Resources Planning and Management,136(1), 375-382.

Buchberger, S.G., Wu, L., 1995. Model for Instantaneous Residential Water Demands. Journal of Hydraulic Engineering, 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, Subject Area: High-Quality Water. Rep. 90963F, AwwaRF, Denver, CO.

Filion, Y. R., Adams, B., Karney, B., 2007. Cross correlation of demands in water distribution network design. Journal of Water Resources Planning and Management 133(2), 137-144.

Guercio, R., Magini, R., Pallavicini, I., 2001. Instantaneous residential water demand as stochastic point process. Water resources management, Eds. Brebbia et al., WIT Press, 129-138.

Iman, R. L., Conover, W. J., 1982. A distribution-free approach to inducing rank correlation among input variables, Communication in statistics, 11, 311-334.

Kottegoda, N. T., Rosso, R., 1998. Statistic, Probability and Reliability for Civil and Environmental Engineers, McGraw-Hill, New York.

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. ASCE, Proceeding of 8th Annual International Symposium on Water Distribution Systems Analysis, Cincinnati, OH (USA).

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