

Multivariate statistical analysis for water demand modelling: implementation, performance analysis, and comparison with the PRP model

Chiara M. Fontanazza, Vincenza Notaro, Valeria Puleo and Gabriele Freni

ABSTRACT

Water demand is the driving force behind hydraulic dynamics in water distribution systems. Consequently, it is crucial to accurately estimate the actual water use to develop reliable simulation models. In this study, copula-based multivariate analysis was proposed and used for demand prediction for a given return period. The analysis was applied to water consumption data collected in the water distribution network of Palermo (Italy). The approach produced consistent demand patterns and could be a powerful tool when coupled with water distribution network models for design or analysis problems. The results were compared with those obtained using a classical water demand model, the Poisson rectangular pulse (PRP) model. The multivariate consumption data statistical analysis results were always higher than those of the PRP model but the copula-based method maintained the daily water volume of actual consumptions and provided maximum daily consumption that increased with the return period.

Key words | multivariate analysis, Poisson rectangular pulse model, vine copula, water demand modelling

Chiara M. Fontanazza (corresponding author)
Vincenza Notaro
Valeria Puleo
Gabriele Freni
Faculty of Engineering and Architecture,
University of Enna 'Kore',
Enna,
Italy
E-mail: chiara.fontanazza@gmail.com

INTRODUCTION

Urban development has created new water distribution system management problems. Water supply networks are required to respond to growing drinking water demands, which are highly variable in terms of space and time. The general goal for any water utility is to constantly supply good quality water under sufficient pressure to all customers (Zhou *et al.* 2002; Herrera *et al.* 2010). The reliability of a utility's water distribution system depends on a combination of the following factors that play important roles in the system design and management: the water demand variability, size and maintenance of pipes, and volumes of urban reservoirs. The development of powerful computers has enabled hydraulic engineers to simulate the behaviour of water supply systems for most scenarios. However, the accurate prediction of pressures, flows and water quality parameters depends strongly on the quality of the input data. The data required to simulate the behaviour of a water distribution

network, such as pipe friction coefficients and nodal demands, along with their temporal variations, contain uncertainty, which consequently affects our confidence in the simulation outcome. There has been agreement in the literature that uncertainties in nodal demands and their variation with time is one of the primary sources of error responsible for discrepancies between measured and model-simulated flows and pressures.

Residential water demand is one of the most difficult parameters to determine when modelling drinking water distribution networks. The simulation of a water supply system is often performed by assuming averaged values (in space and in time) of the water demands. The average spatial values are obtained by clustering the water consumption of users towards each node of the network, and the time-averaged values are obtained as the mean of the instantaneous values of the nodal demands. The simulation

results obtained using these simplifications may not be reliable for the hydraulically disadvantageous zones of the network. Therefore, water demand modelling has been an active field of study. Researchers have been primarily interested in household domestic water consumption, which is the principal rate of the total volume supplied by the water distribution system in urban areas, often equal to 75%. Water demand can be predicted on different time scales. Short- and long-term forecasting of the municipal water demand is essential to water utilities for system planning, design and asset management. Short-term forecasting is useful for operating and managing existing water supply systems within a specific period, whereas long-term forecasting is particularly important for system planning and design.

Analysing and forecasting urban water demand is a complex but imperative task. Unfortunately, accurately predicting demands and simulating the short- and long-term pattern of demands are difficult propositions for two reasons. First, it is impossible to predict the extent to which factors that influence the water demand (e.g., household income, urban population, number of service connections, water price and land uses) will change in the future. Peak daily use is estimated by applying peak factors to the average daily level, and design and expansion decisions are based on this value. The second consideration that complicates water demand simulation is its variability in time. By studying a hydrograph of the daily water demand in a system, it can be recognised nearly immediately that changes in water demand occur at many different time scales. For example, domestic water use typically changes rapidly, nearly in a matter of seconds and minutes, as people perform their domestic tasks of washing, cooking, cleaning, bathing, etc.

A detailed model of the hydraulic behaviour of drinking water distribution systems could be obtained by implementing a domestic water demand model in one of several software programs that have been recently developed. To date, stochastic models for instantaneous residential water demand (e.g., impulse models, see the subsequent sections for references) have been used to obtain realistic demand patterns for the hydraulic distribution network solvers. Several basic parameters related to the residential water usage are required to apply these models, such as the frequency, duration and intensity statistics of a single consumption

event. This methodology, which maintains its validity, reveals some limitations, in our opinion. Impulse models (e.g., the Poisson rectangular pulse (PRP) model) can reliably simulate continuous patterns of water demand for a specific user only when continuous series of their recorded consumption are available to calibrate them. The reliability of the simulated demand is thus strictly dependent on the amount of available data. These models demonstrate good performance, as well as arising from hydrology to define rainfall series, even if the generated demand patterns will have the same temporal and spatial aggregation as the recorded patterns. If the temporal aggregation of available data is too coarse, model parameters (strictly related to single consumption pulses) lose their physical sense.

In actual water distribution network management, it is unlikely that continuous series of recorded consumption data are available for more than a few users. Furthermore, performing this kind of monitoring campaign at large scale is usually hardly viable by means of common systems of automatic meter reading (mainly due to the high costs of data transmission and management). The large amount of data that should be stored and transferred to a central database require high effort both in terms of memory storage and energy consumption which greatly reduces the battery life of the monitoring equipment.

As reported in [Alvisi *et al.* \(2007\)](#), the water demand process is not entirely stochastic, but a rate can be linked to deterministic relationships between the variables characterising the process: seasonal and weekly periodicities in daily consumption and daily periodicities in hourly water demand patterns can usually be identified. Impulse models neglect the statistical dependence of the parameters characterising the consumption process considering pulse intensity, duration and frequency as independent variables. Nevertheless, the statistical proprieties of water demand (e.g., the mean, total daily volume, maximum daily value) are few, easily monitored and processed. Therefore, a statistical methodology able to assess and analyse the multivariate relationships between these statistical properties involved in the demand process could lead to the reduction of the amount of data useful to describe the water consumption process. As an advantage, the management of the monitoring equipment at large scale could be feasible and cost effective.

In the present study, a multivariate statistical method based on the copula functions recently introduced in hydrology and proposed by the authors in Fontanazza *et al.* (2014) has been further developed. The main peculiarity of the copula function is that it permits separate investigation of the marginal properties and interdependence structures of the variables. It synthesises the dependence structure of the variables in the purest and most essential form (Bárdossy & Pegram 2009) without assuming that the variables are normal or have the same marginal distribution.

Starting from the above-mentioned considerations, this study has the following two objectives: (1) to propose a procedure based on a multivariate statistical analysis of the primary features of the water consumption process at the domestic level and (2) to obtain a practical use water consumption model that is not characterised by high resolution data requirements. The proposed method was validated by comparing its results with those obtained using the PRP model. This paper is organised as follows. The multivariate and PRP analyses of the consumption process are described next. Then, the case study to which the procedure is applied is presented. This is followed by presentation and comparison of the resulting demand patterns, according to the multivariate analysis and the PRP model. Finally, the study conclusions are presented.

MATERIALS AND METHODS

Multivariate consumption data statistical analysis based on copula function

The copula function is a new multivariate statistical analysis method that is frequently used in hydrology to perform multivariate process simulations, extreme value analysis and dependence structure modelling (De Michele & Salvadori 2003; Favre *et al.* 2004; Salvadori & De Michele 2004a, b; Grimaldi & Serinaldi 2006; Bárdossy & Pegram 2009). While there is a multitude of bivariate copula, building higher dimensional copulas is generally recognised as a difficult problem.

The idea of constructing a multivariate dependence model from bivariate copulas as building blocks (i.e., pair-copulas) dates back to Joe (1996). The author detailed the construction of the first pair-copula in terms of distribution

functions. Bedford & Cooke (2001, 2002) realised that there were a significant number of possible pair-copulas constructions (PCC); thus, they organised them in a graphical manner by sequentially designing trees that identify the bivariate copula densities needed to make up a d-dimensional density. It involves only the products of bivariate copulas. As the trees are intrinsically related, the authors called these distributions regular vines (R-vines). Vine copulas are flexible functions for multivariate dependencies, which specify a factorisation of the copula density into a product of conditional bivariate copulas. The class of regular vines is still general, and it embraces a large number of possible pair-copula decompositions (Aas *et al.* 2009), including two simple tree structures, such as line trees and star trees. The former corresponds to D-vines, while the latter corresponds to C-vines. Recently, vine copula functions have demonstrated great potential in several hydrological applications (Graler *et al.* 2013).

Several procedures analysing and modelling user water consumption, such as the Poisson, Neyman-Scott and Bartlett-Lewis models, have been widely adopted in the literature. These procedures have been primarily used to study the stochastic process inherent in rainfall events (Rodríguez-Iturbe 1986; Buchberger & Wu (1995); Cowpertwait *et al.* 1996; Alvisi *et al.* 2003; Alcocer-Yamanaka *et al.* 2012). Similarly, the present paper proposes to model the daily pattern for user water consumption through a multivariate statistical approach first used to generate synthetic rainfalls by employing copula functions that simulated the multivariate relationships between the main variables characterising the temporal rainfall pattern, such as the rainfall total depth, duration and maximum intensity (Fontanazza *et al.* 2011).

According to this procedure, the multivariate consumption data statistical analysis (MCDSA) uses 3-d vine copulas for the main variables characterising the daily pattern of domestic water consumption: V_d , the total daily volume consumed by the user, which is analogous to the total rainfall volume; K_p , the daily peak coefficient, which is expressed as the ratio between the maximum consumption in a given time step, V_{max} , and the total daily volume, V_d , which is analogous to the maximum rainfall intensity; and the time to peak coefficient, $T_p/24$, which is expressed as the ratio between the time to peak and the total duration of the

demand pattern, 24 hours. Figure 1 shows a flowchart of the MCDSA used in the present study.

First, prior to performing the procedure, a triplet (V_d, K_p, T_p) is assessed for each daily pattern recorded for a monitored dwelling. Namely, MCDSA requires a sample of triplets (V_d, K_p, T_p) at least 1-year long to ensure reliability. Thus, the related dimensionless transformed variables $(V, K, T)_{historical}$, which are approximately uniformly distributed in $[0, 1]$, are evaluated together with the marginal distribution functions F_V, F_K and F_T (step 1).

Second, the Kendall's τ_k rank correlation of each couple of variables (i.e., V-K, V-T and K-T) is evaluated to estimate the statistical dependence between the variables. Hence, several 3-d vine copulas are built to fit the sample of the variables $(V, K, T)_{historical}$. In the three-dimensional case, there are no differences between a C- or a D-vine, only the ordering of variables can be changed. Figure 2 shows the possible schemes for composing a 3-d vine copula. In the second tree, the two conditional cumulative distribution function (CDF) values are calculated for all triplets $(V, K, T)_{historical}$. These 'conditioned observations', which are again approximately uniformly distributed in $[0, 1]$, are then used to fit

another bivariate copula (e.g., $C_{KT|V}, C_{VK|T}$ or $C_{VT|K}$). Considering the 3-d vine structure shown in Figure 2(a), the full density function c_{VKT} of the three-dimensional copula is given as follows:

$$c_{VKT}(v, k, t) = c_{KR|V}(F_{K|V}(k, v), F_{T|V}(t, v)) \cdot c_{KV}(k, v) \cdot c_{TV}(t, v) \tag{1}$$

The three-dimensional distribution function of the original variables sample (V_d, K_p, T_p) is obtained by combining the bivariate copulas, as shown in Equation (1), and substituting the marginal distribution functions F_{V_d}, F_{K_p} and F_{T_p} , which are defined as $v = F_{V_d}^{-1}(V_d), k = F_{K_p}^{-1}(K_p)$ and $t = F_{T_p}^{-1}(T_p)$. Therefore, the full density function $f_{V_dK_pT_p}$ is given as follows:

$$f_{V_dK_pT_p}(v_d, k_p, t_p) = c_{KT|V}(F_{K|V}(k, v), F_{T|V}(t, v)) \cdot c_{KV}(k, v) \cdot c_{TV}(t, v) \cdot f_{V_d}(v_d) \cdot f_{K_p}(k_p) \cdot f_{T_p}(t_p) \tag{2}$$

where f_{V_d}, f_{K_p} and f_{T_p} , represent the marginal density functions of (V_d, K_p, T_p) .

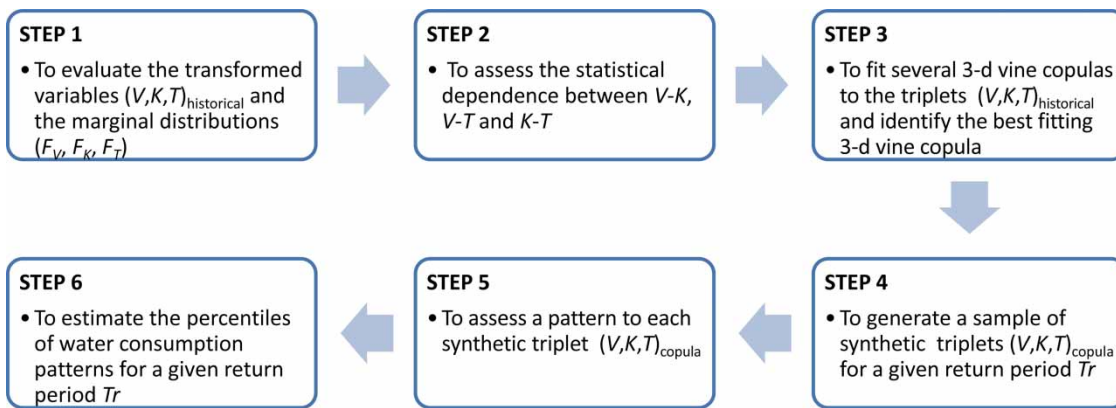


Figure 1 | A flowchart of the MCDSA performed in the present study.

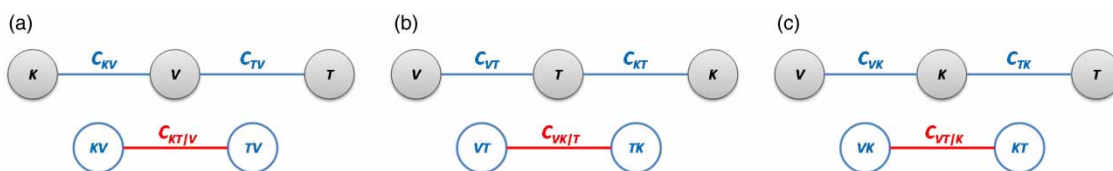


Figure 2 | Possible structures of a 3-d vine.

According to Equations (1) and (2), three bivariate copulas must be fitted to derive the building blocks of a 3-d vine copula (e.g., in Figure 2(a), the C_{KV} , C_{TV} and $C_{KT|V}$ bivariate copulas). The maximum likelihood estimation (MLE) method can be adopted to fit a copula from each family investigated for each pair of variables; the copula showing the highest log-likelihood value or the minimum Akaike's information criterion (AIC) value is selected as the best fitting vine copula model for the analysed data set. The methodology considers all possible schemes of the 3-d vine copula shown in Figure 2.

Using the best fitting 3-d vine copula model, which is identified in step 3, a sample of synthetic triplets $(V, K, T)_{copula}$ related to a given return period was generated (step 4). The multivariate return period of the triplets $(V, K, T)_{copula}$ was assessed following the procedure proposed by Salvadori et al. (2011) and based on the copula's Kendall distribution function $K_C(t)$. According to this procedure, the copula-based return period T_{KEN3} (expressed in days) is provided as follows:

$$T_{KEN3} = \frac{\mu}{1 - K_C(t)} \leftrightarrow t_{KEN3} = K_C^{-1}\left(1 - \frac{\mu}{1 - T_{KEN3}}\right) \quad (3)$$

where μ is the mean inter-arrival time expressed in days (in the case of a daily event, $\mu = 1$), and $K_C(t)$ is the copula's Kendall distribution function. Namely, $K_C(t): I \rightarrow I$ is defined as:

$$K_C(t) = P(C(v, k, r) \leq t) \quad (4)$$

with $t \in I$ defined as the probability level.

Thanks to Equation (4), after fixing the design return period T_{KEN3} , the corresponding probability level t_{KEN3} was assessed using the inverse of the copula's Kendall distribution function $K_C(t)$. In 3-d, this level corresponds to an iso-surface, i.e., all triplets $(V, K, T)_{copula}$ on this surface have the same copula value equal to t_{KEN3} . $K_C(t)$ allows the calculation of the probability that a random point $(V, K, T)_{copula}$ in the unit cubic space has a smaller or larger copula value than a given critical probability level ($t = t_{KEN3}$). The Kendall distribution function is a univariate representation of multivariate information because it is the CDF of the copula's iso-surface. Therefore, $K_C(t)$ is an

essential tool for calculating a copula-based return period for multivariate events (Graler et al. 2013).

In step 5 of the analysis, a pattern, derived from historical time series of consumption, was assigned to each synthetic triplet $(V, K, T)_{copula}$ using a similarity criterion. Namely, for each recorded daily consumption pattern, at first, the related mass curve (Huff curve), defined as the representation of the normalised time vs. the normalised cumulative water consumption from the beginning of the day, was evaluated. Then, each statistical triplet $(V, K, T)_{copula}$ was linked to the Huff curve of the historical event $(V, K, T)_{historical}$ that minimised the following objective function (Fontanazza et al. 2011):

$$S = w_V \cdot |[V_{copula} - V_{historical}]| + w_K \cdot |[K_{copula} - K_{historical}]| + w_T \cdot |[T_{copula} - T_{historical}]| \quad (5)$$

where w_V , w_K and w_T are weights with a sum equal to 1. They were evaluated using a Monte Carlo analysis aimed at maximising the modelling efficiency (valued by means of the Nash and Sutcliff method) between the 50th percentile of the synthetic and recorded patterns.

Finally, (as step 6) the synthetic patterns obtained for a given return period were statistically processed to estimate the related percentiles.

PRP model for consumption data analysis

Qi & Chang (2011), House-Peters & Chang (2011) and Donkor et al. (2012) have presented an overview of water demand prediction models on various time scales. The time scale is dictated by the purpose for which the prediction model is to be used (Bakker et al. 2013). Most past water consumption studies have started based on the requirements of quantifying global demand using long-term forecasting (Maidment et al. 1986; Dziegielewski & Boland 1989; Miao 1990; Zhou et al. 2000) and establishing a suitable rate structure (Rothstein 1992). New reasons to better characterise the domestic water consumption have recently emerged; among these reasons, the requirement to ensure water volumes demanded by customers and to supply them with sufficient pressure and good quality has been prominent (Clark et al. 1993; Buchberger & Wu 1995; Buchberger & Wells 1996; Guercio et al. 2001).

At a domestic service level, water demand is considered sporadic and is characterised by sudden demand pulses, and it tends to have a stochastic character (Buchberger & Wu 1995; Buchberger & Wells 1996), particularly when considering time scales on the order of seconds. The sporadic water demand can be characterised as a series of rectangular pulses with a set intensity, duration and frequency. Therefore, several stochastic models for domestic demand determination have been developed. These models include the PRP model (for reference, see the following) and the Neyman-Scott rectangular pulse model (Alvisi *et al.* 2003; Alcocer-Yamanaka *et al.* 2012), among others (Blokker *et al.* 2010).

The PRP model was initially proposed and applied to describe rainfall temporal patterns and to generate a synthetic series representing rainfall or storm events, according to the projected interval and duration (Rodríguez-Iturbe *et al.* 1984, 1987; Rodríguez-Iturbe 1986; Cowpertwait *et al.* 1996). Subsequently, the PRP model has been transferred to water demand modelling and has provided good results (Buchberger & Wu 1995; Buchberger & Wells 1996; Guercio *et al.* 2001; Buchberger *et al.* 2003; Freni *et al.* 2004; García *et al.* 2004), thereby confirming the adaptability of these models to describe and well-reproduce processes that are strongly time dependent and that arise from aggregation and the superposition of single events. The superposition is caused by each dwelling that takes water as a single pulse from the distribution network, independently of the other dwellings. Furthermore, Magini *et al.* (2008) and Vertommen *et al.* (2014) showed that the spatial and temporal aggregation strongly influences the water demand statistics and represents a key point for a correct stochastic model of water consumption.

The demand events are generally of short duration, followed by relatively long periods of no demand. When a user logs on to the water service, one or more appliances may be maintained as busy. Each water demand event may be composed of a random number of rectangular demand pulses with a height that represents the intensity and a width that represents the duration. The superposition of single pulses causes a complex water demand event. The Poisson process characteristics make it improbable that two demand pulses start at the same moment. However, two pulses that start at different moments are partially

superimposed due to the finite duration of each pulse. The intensity of the water consumption composed of more than one pulse is the sum of the intensities of the single pulses.

The basic parameters of the PRP model are the frequency of occurrence of the individual pulses (i.e., the number of pulses that occur in the selected time step) and the intensity and duration of the pulse. The pulse origins are independently displaced from the event origin. The duration and intensity of the pulses are assumed to be independent random variables. This assumption is expedient for the sake of operational efficiency and simplicity. The frequency, intensity and duration are formally described by their statistical properties of mean, variance and probability distribution. Since the water demand presents certain natural variations during the day, the arrival rate changes during the day (i.e., the Poisson process is non-homogeneous).

In this paper, the parameters of the PRP model are obtained by registering the instantaneous water demand every second using special equipment, separating the individual demand pulses, and statistically processing the resulting series of pulses. This approach, based on the direct measurement of the demand pulses, becomes costly because of the special equipment required to register the water demand, the required personnel for related field work, and the enormous amount of data to be processed.

A single elementary use is rectangular-shaped, and it is described by a random duration and a random steady intensity. More complex water demands that are caused by the superposition of multiple single pulses must be converted into more regular events, which are defined as single equivalent rectangular pulses (SERPs) by Buchberger *et al.* (2003). Rectangular-shaped ideal water consumption is described by the moment at which it begins and ends, constant and positive intensity and duration.

The recorded single-user water consumptions always show a wide variety of patterns that typically differ from the rectangular-shaped patterns, with an intensity fluctuation caused by network pressure variability and unsteady events due to the opening and closing of appliances. Each event was converted into one or more SERP to analyse the characteristics of the consumption process and to estimate the parameters of the PRP model. This process

involved two steps: signal smoothing and pulse separation. Each complex event was converted into constant rectangular-shaped signals by applying a signal smoothing based on the analysis of the difference between the progressive average (from the event beginning) and a five-point moving average. Each intensity variation for which the difference between the progressive average and the five-point moving average is higher than a set threshold value (changing with the duration and the total volume of each event) was considered to be significant and, thus, not negligible. Previous intensities were averaged to yield a constant intensity. This procedure was performed many times inside each event and for all events. The result was rectangular-shaped and more complex signals, defined as strings of random blocks (SORBs) by Buchberger *et al.* (2003), with more than one constant intensity. SORBs show that many appliances are contemporarily busy. In the pulse separation phase, SORBs were analysed to identify the actual superposition of single uses. The identification and separation of single pulses are based on two empirical rules: two pulses do not begin or end simultaneously, and the superposition of pulses causes an increase in intensity. A system of 'n' linear equations, with unknown values that are the mean intensities of the SERPs, was written for each SORB (where 'n' is the system dimension equal to the amount of positive change in intensity). The system solution provided the intensities, durations, beginning and ending of each single pulse. When the solution was not reached, the SORB was replaced by a single equivalent pulse with an intensity equal to an average value weighted by the duration of the different intensity levels.

The evaluation of the PRP model parameters depends on the aggregation time step for which the stochastic process may be considered to be homogeneous. In this study, the consumption process was assumed to be homogeneous at the daily scale for the intensity and duration and homogeneous at the hourly scale for the pulse occurrence (frequency). The daily consumption pattern appears to be primarily due to the occurrence of single events rather than their duration and intensity. This hypothesis reduced the number of parameters to evaluate for stochastic model identification.

THE CASE STUDY

The above-described MCDSA and PRP models were applied to water consumption data obtained from monitoring eight dwellings located in Palermo (Italy) throughout 2007 (Figure 3).

The customers that participated in the consumption monitoring program were selected according to the following characteristics: families with at least two members; family members ranging in age from 4 to 70 years; a minimum of one electric household appliance (dishwasher or washing machine); negligible outdoor consumptions; and cooperation. The selected eight families were the only families that agreed to participate in the consumption monitoring program. Instrument packs to monitor domestic water use were installed on the service line of each of the eight dwellings, downstream of the revenue water meter (Figure 4).

The instrument package included a multi-jet water meter and a data logger. The two devices were coupled by means of an impulse sensor. The water meter had a minimum flowrate, Q_1 , equal to 15 l/h, a transitional flowrate, Q_2 , equal to 22.5 l/h, a permanent flowrate, Q_3 , equal to 1,500 l/h, and an overload flowrate, Q_4 , equal to 5,000 l/h (ISO 4064-1 2005). When the cumulative volume consumed by the user was equal to or higher than 0.5 l in a given time period, the sensor transmitted a signal to the data logger for each 0.5 l; when the volume was lower than 0.5 l, the sensor did not transmit a signal but the volume was aggregated to the following consumption pulse until the cumulative volume was equal to 0.5 l. Considering that a common faucet is characterised by flows in the range 6–12 l/min, the sensor was able to transmit pulses longer or equal to 5 seconds (in the worst case) or equal to 2.5 seconds (in the best case). Cumulative volumes of more than 0.5 l were recorded in a text file containing six fields (i.e., day, month, year, hour, minute and second). The water demands were downloaded by connecting the data logger to a portable computer.

The monitoring period was approximately 1 year for five dwellings, shorter for dwellings 4 and 5, and longer (by two times) for dwelling 6 (Table 1).

A four-step process was used to transform the raw input signals into archived residential water consumptions (Figure 5). Step 1 involved data retrieval; step 2 involved data correction (removal of repetitive signals and placing



Figure 3 | Location of the monitored dwellings in the city of Palermo.

the water demand readings in chronological order) and water use separation; in step 3, leaks and ultra-low demands were censored; and in step 4, the volume of each pulse was uniformly distributed over the duration of the pulse. A sparse matrix collected the flow values (l/sec) using the seconds in a day as the number of columns and the days during which the consumption data were recorded (changing for each user) as the number of rows.

RESULTS AND DISCUSSION

MCDSA application to the case study

For each monitored dwelling, a triplet (V_d, K_p, T_p) was assigned to each recorded daily water demand pattern,

thus obtaining three samples of variables, with more than 365 data points for each one. Then, the related dimensionless transformed variables $(V, K, T)_{\text{historical}}$, approximately uniformly distributed in $[0, 1]$, were evaluated together with the marginal distribution functions (F_V, F_K, F_T) . These functions were identified by fitting several distribution functions to the empirical CDF of $(V, K, T)_{\text{historical}}$ and by performing a Kolmogorov–Smirnov (K-S) goodness-of-fit test to choose the best distribution. **Table 2** shows the parameters of the obtained marginal distributions (F_V, F_K, F_T) for all eight monitored dwellings. For dwelling 1, all three variables (V , K and T) show a good fit with the generalised extreme value (GEV) distribution. The GEV marginal distributions of the variables K and T were cut off in $[0, 1]$.

To estimate the statistical dependence between the three variables $(V, K, T)_{\text{historical}}$, Kendall's τ_k rank correlation of

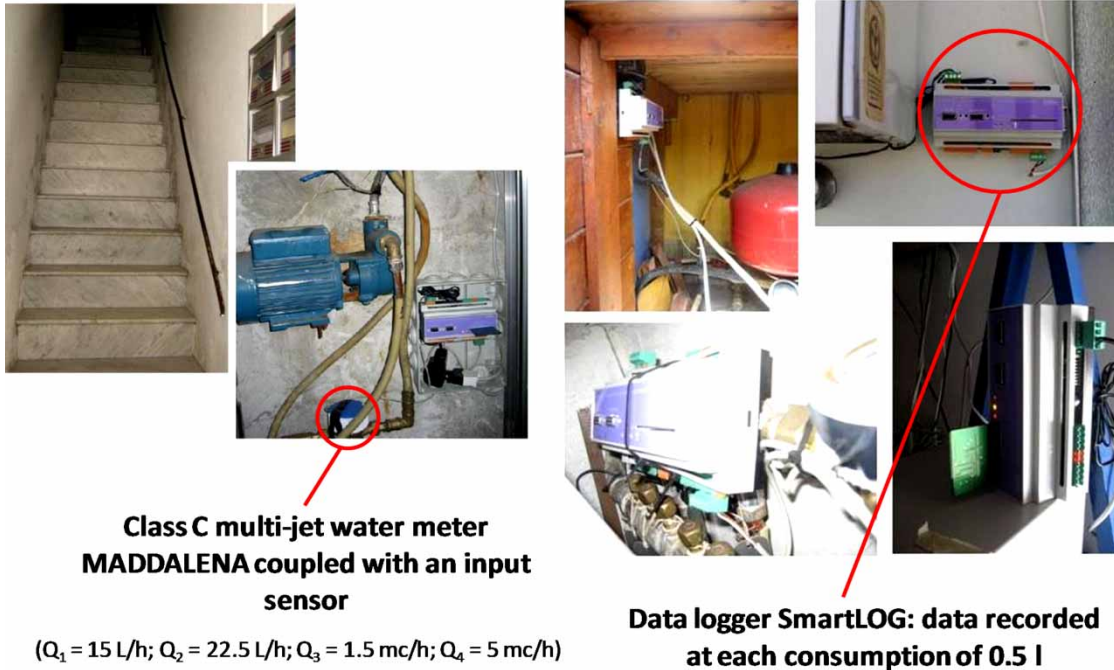


Figure 4 | Instrument packs and some installations.

Table 1 | Duration of the monitoring period

| Dwelling | Monitoring period [days] |
|----------|--------------------------|
| 1 | 334 |
| 2 | 359 |
| 3 | 317 |
| 4 | 237 |
| 5 | 212 |
| 6 | 637 |
| 7 | 345 |
| 8 | 320 |

each pairwise V - K , V - T and K - T was evaluated (Table 3). For dwelling 1, the pairwise V - K and V - T showed a negative correlation, with Kendall's τ_k values equal to -0.42 and -0.12 , respectively; only the K - T pair had a positive correlation, with τ_k equal to 0.07 . For the other seven dwellings, the correlation was always negative for V - K and positive for V - T . Furthermore, the correlation was higher for V - K and V - T and lower for K - T .

Then, the MLE method was adopted to fit several 3-d vine copulas to the variables $(V, K, T)_{historical}$ related to each monitored dwelling. All possible schemes of the 3-d

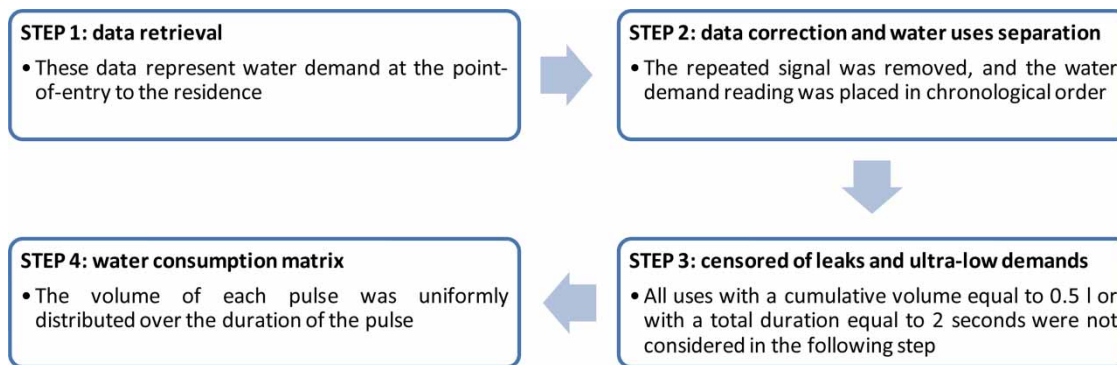


Figure 5 | Steps in the process for transforming the raw input signals into archived water consumptions.

Table 2 | Marginal distributions parameters of the $(V, K, T)_{historical}$: a_1, a_2, a_3 are the statistic parameters of the marginal distributions and the K-S is provided with a confidence level equal to 0.01

| Dwelling | | Distribution | a_1 | a_2 | a_3 | K-S test $D_{0.01}$ |
|--|-------|------------------|-----------|-------------|-----------------|---------------------|
| Dwelling 1 | F_V | GEV | -0.23 | 0.17 | 0.37 | 0.044 |
| | F_K | GEV | 0.37 | 0.04 | 0.13 | 0.063 |
| | F_T | GEV | 0.45 | 0.08 | 0.37 | 0.028 |
| Dwelling 2 | F_V | Logistic | - | 0.50 | 0.10 | 0.079 |
| | F_K | GEV | 0.2189 | 0.14 | 0.04 | 0.078 |
| | F_T | Inverse Gaussian | - | 0.53 | 2.99 | 0.015 |
| Dwelling 3 | F_V | Logistic | - | 0.37 | 0.09 | 0.077 |
| | F_K | GEV | 0.34 | 0.17 | 0.07 | 0.071 |
| | F_T | Log-logistic | - | -0.69 | 0.22 | 0.017 |
| Dwelling 4 | F_V | Weibull | 0.3711 | 2.009 | - | 0.096 |
| | F_K | GEV | 0.2078 | 0.1775 | 0.066 | 0.088 |
| | F_T | Beta | 31.145 | 19.776 | - | 0.001 |
| Dwelling 5 | F_V | Log-logistic | - | 0.8975 | 0.1916 | 0.104 |
| | F_K | GEV | 0.1304 | 0.1392 | 0.0409 | 0.099 |
| | F_T | Log-logistic | - | -0.8532 | 0.1989 | 0.010 |
| Dwelling 6 | F_V | Logistic | - | 0.3395 | 0.0751 | 0.023 |
| | F_K | GEV | 0.1690 | 0.1012 | 0.0132 | 0.020 |
| | F_T | Log-logistic | - | -0.9322 | 0.3664 | 0.004 |
| Dwelling 7 | F_V | GEV | -0.133 | 0.28 | 0.1589 | 0.076 |
| | F_K | GEV | 0.3356 | 0.1514 | 0.0597 | 0.069 |
| | F_T | Log-logistic | - | -1.038 | 0.206 | 0.001 |
| Dwelling 8 | F_V | GEV | -0.16 | 0.3972 | 0.1249 | 0.071 |
| | F_K | GEV | 0.0626 | 0.1045 | 0.027 | 0.088 |
| | F_T | Logistic | - | 0.4982 | 0.1123 | 0.024 |
| GEV | | | $a_1 = k$ | $a_2 = \mu$ | $a_3 = \sigma$ | |
| Logistic and log-logistic distribution | | | | $a_2 = \mu$ | $a_3 = \sigma$ | |
| Inverse-Gaussian distribution | | | | $a_2 = \mu$ | $a_3 = \lambda$ | |
| Weibull and beta distribution | | | $a_1 = a$ | $a_2 = b$ | | |

vine copula shown in Figure 2 were built. The investigated copula families include Normal, Student, Gumbel, Frank, Clayton, BB1, BB6, BB7, BB8 and their rotated version.

Table 3 | Kendall's τ_k rank correlation of each couple of variables V - K , V - T and K - T

| Dwelling | τ_k vk | τ_k vt | τ_k kt |
|----------|-------------|-------------|-------------|
| 1 | -0.420 | -0.118 | 0.068 |
| 2 | -0.322 | 0.087 | 0.077 |
| 3 | -0.385 | 0.062 | -0.036 |
| 4 | -0.170 | 0.001 | 0.113 |
| 5 | -0.348 | 0.100 | 0.020 |
| 6 | -0.499 | 0.161 | -0.161 |
| 7 | -0.558 | 0.073 | -0.101 |
| 8 | -0.157 | 0.023 | 0.019 |

The best fitting 3-d vine copula model for the analysed data set was the 3-d vine copula showing the highest log-likelihood value or the minimum AIC values. Table 4 shows the bivariate copula families, parameters and Kendall's τ_k of the building blocks of the 3-d vine copula built for dwelling 1. The 3-d vine copula structure best fitting the analysed data is the C-vine (Figure 2(a)). In the Appendix (available in the online version of this paper), the corresponding tables related to the other seven monitored dwellings are shown.

After identifying the best fitting 3-d vine copula model, the analysis focused on the generation of synthetic triplets $(V, K, T)_{copula}$ related to a given return period.

Two return period were set, $T_{KEN3} = 100$ days and $T_{KEN5} = 365$ days, the typical value for water distribution system management.

Table 4 | Copula families, parameters and Kendall's τ_k of the building blocks of the 3-d vine copulas built for dwelling 1

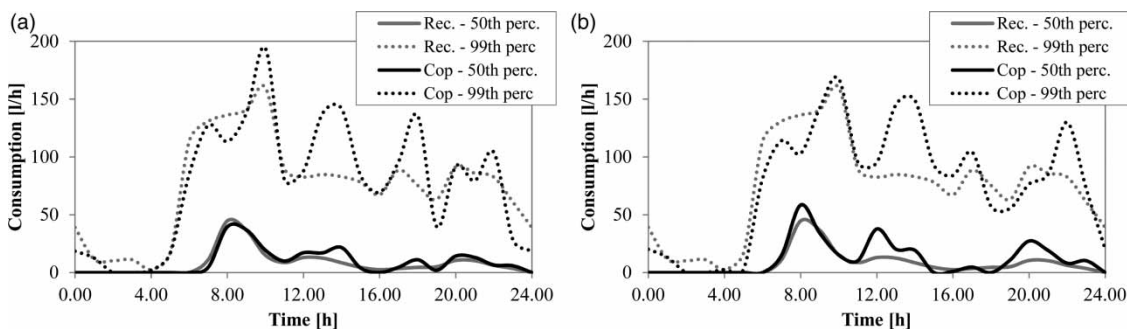
| | 3-d vine Figure 2(a) | | | 3-d vine Figure 2(b) | | | 3-d vine Figure 2(c) | | |
|-------------------------------|----------------------|----------|-----------|----------------------|----------|------------|----------------------|----------|------------|
| Log-likelihood | 106.31 | | | 79.22 | | | 105.60 | | |
| AIC | -204.63 | | | -148.44 | | | -201.20 | | |
| | C_{VK} | C_{VT} | C_{KTV} | C_{VK} | C_{KT} | $C_{VT K}$ | C_{VK} | C_{KT} | $C_{VT K}$ |
| Family copula* | 33 | 40 | 5 | 40 | 10 | 5 | 33 | 10 | 40 |
| par | 0.17 | -3.24 | 2.83 | -3.24 | 1.80 | 0.97 | -0.17 | 1.80 | -4.49 |
| par2 | 0.00 | 0.96 | 0.00 | 0.96 | 0.90 | 0.00 | 0.00 | 0.90 | -0.90 |
| Kendall's τ_k 2-d copula | -0.08 | -0.51 | 0.29 | -0.51 | 0.22 | 0.11 | 0.08 | 0.22 | -0.58 |

A numerical evaluation based on a sample of 2×10^7 points simulated by the best fitting 3-d vine copula was performed to calculate the inverse of $Kc(t)$, as no closed form exists for the CDF of the 3-d vine copula identified in this analysis (for more details, see Salvadori et al. 2011). According to Equation (4) and the numerical evaluation of $Kc(t)$, the related t_{KEN5} values were calculated, and the results equalled 0.99 and 0.997, respectively. Thus, 1,000 triplets $(V, K, R)_{copula}$ were sampled from the iso-surface related to $C(V, K, T) = t_{KEN5}$, and the corresponding 1,000 triplets $(V_d, K_p, T_p)_{copula}$ with iso-probability were obtained through the inverse marginal distribution. As the final step of the analysis, a pattern was statistically assigned to each triplet $(V_d, K_p, T_p)_{copula}$ after considering the historical time series of consumption. Namely, the Huff curve of the recorded daily consumption pattern that minimised the objective function of Equation (5) (Fontanazza et al. 2011) was assigned to each statistical triplet $(V_d, K_p, T_p)_{copula}$.

The 1,000 patterns were finally processed, and the percentiles for a given return period were estimated. Figure 6 shows the 50th and the 99th percentiles of the recorded and generated daily patterns for dwelling 1.

The curves for the 50th percentile of the recorded data and simulated consumption for the two return periods were similarly shaped: the peaks were preserved in the beginning of the morning, thus confirming that this user is typical of working families who are not often at home during the afternoon. The MCDSA provided good results, considering both the total daily volume, V_d , consumed by the user and the maximum daily consumption, V_{max} , that increased with the return period. The median of the total daily recorded volume was equal to 200 l, and the total daily volume simulated for a return period of 100 and 365 days was equal to 230 and 291 l, respectively. The maximum recorded water consumption was 44.67 l/h, and the maximum water consumption simulated for 100 and 365 days was 37.78 l/h and 56.75 l/h, respectively. The simulated pattern for $T_{KEN5} = 365$ days was higher than that for $T_{KEN5} = 100$ days.

The percentiles were consistent, demonstrating that the performed analysis can be efficiently used for water distribution network simulation. Even the curve for the 99th percentile of the recorded daily pattern, which is usually chosen for designing water supply loads to urban buildings,

**Figure 6** | Water demand percentiles of recorded data and generated daily pattern for dwelling 1: (a) 100-day return period; (b) 365-day return period.

showed an acceptable agreement with the corresponding percentile of the simulated patterns.

PRP model for the water demand simulation

The empirical CDF of intensity, duration and pulse frequency and the best fitting probability distribution function of the parameters of the PRP model, intensity, duration and pulse frequency were evaluated for each consumption time series. The intensity shows a good fit with a normal distribution that was cut off in $[0, 1]$, as demonstrated by [Guercio et al. \(2001\)](#), as does the duration with a log-normal distribution ([Table 5](#)) and the frequency with a Poisson distribution, as stated by the hypotheses of the PRP model.

The mean and the standard deviation of the duration shown in [Table 5](#) are comparable with those reported in [Buchberger & Wells \(1996\)](#) and [Buchberger et al. \(2003\)](#) which related that the duration of residential water demand showed a good fit with log-normal distribution as well ([Table 6](#)). The mean of the intensity in [Table 6](#) (that as well as the duration showed a good fit with log-normal distribution) was twice those in [Table 5](#), and the standard deviation of the intensity in [Table 6](#) exceeds those in [Table 5](#) by a factor 2–4.

As previously mentioned, the Poisson process is non-homogeneous and results in 24 distribution functions of the consumption frequency, one for each hour ([Table 7](#)), because the frequency with which customers log on to the water service is time dependent.

Table 5 | Parameters of the intensity and duration distribution functions

| Dwelling | Intensity | | Duration | |
|----------|--------------------------|---------------------------------------|----------------------|--------------------------------------|
| | Mean α [l/min] | Standard deviation β [l/min] | Mean τ [sec] | Standard deviation ω [sec] |
| 1 | 5.01 | 1.77 | 31 | 101 |
| 2 | 2.83 | 1.24 | 38 | 99 |
| 3 | 4.55 | 1.49 | 36 | 98 |
| 4 | 5.79 | 2.01 | 26 | 104 |
| 5 | 5.08 | 1.56 | 40 | 95 |
| 6 | 4.42 | 1.45 | 38 | 112 |
| 7 | 3.44 | 1.62 | 56 | 123 |
| 8 | 3.23 | 1.45 | 50 | 110 |

Table 6 | Parameters for PRP model taken from [Buchberger & Wells \(1996\)](#) and [Buchberger et al. \(2003\)](#)

| Intensity mean [l/min] | Standard deviation [l/min] | Duration mean [sec] | Standard deviation [sec] |
|---------------------------|-------------------------------|---------------------------|-----------------------------|
| 8.50 | 4.70 | 45 | 90 |

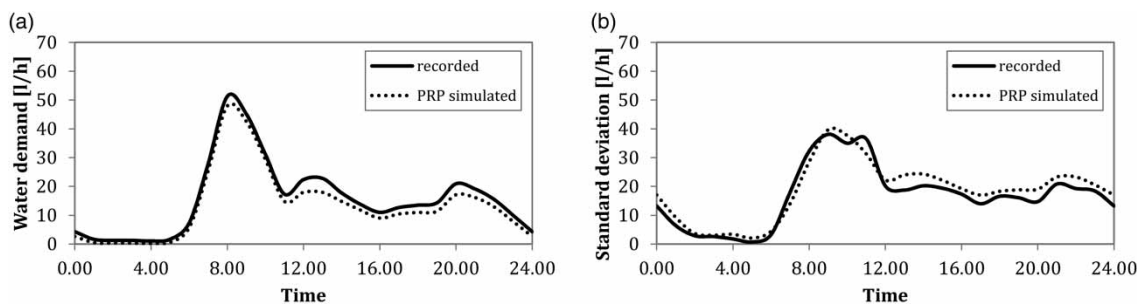
The synthetic series of water consumption was generated from the distribution functions of intensity (normal distribution), duration (log-normal distribution) and hourly frequency (Poisson distribution). A random probability was generated for each hour in a day, and the corresponding value of events/hour was defined using the distribution function of the frequency related to that hour. Two additional random probabilities were generated for each event along with the corresponding values of intensity and duration that arose from the distribution functions. The total water volume consumed by the user in that hour is the sum of the volume of each event (intensity by duration) that occurred simultaneously. Coupling this procedure with the Monte Carlo analysis, 100,000 synthetic water consumption series (with a length equal to those recorded) were generated, and the average water consumption was calculated. [Figure 7](#) shows that the average hourly synthetic consumption preserves the shape and volume of the recorded consumption. The comparison of the two curves shows that the PRP model is good at preserving the average demand value, and it is promising for the applications in real time management. To evaluate the forecasting ability of the PRP model, in the next section, the simulated patterns for different return periods are analysed.

MCDSA vs. PRP model for water demand simulation

[Figures 8](#) and [9](#) show the 50th percentiles (all monitored dwellings) of the water demand simulated by the PRP model and MCDSA for the two return periods chosen (100 and 365 days, respectively). In order to analyse the reliability of the generated patterns, the 50th percentiles of the recorded consumptions are plotted in [Figures 8](#) and [9](#) too; it must be highlighted that the median of the recorded

Table 7 | Parameters of the frequency distribution functions

| Dwelling | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|----------|------------------------------------|--------|-------|--------|-------|--------|-------|-------|
| Hour | Frequency (λ [uses/hour]) | | | | | | | |
| 0 | 0.933 | 0.997 | 0.650 | 0.045 | 1.151 | 3.370 | 0.299 | 2.651 |
| 1 | 0.169 | 0.532 | 0.107 | 0.032 | 0.618 | 2.352 | 0.110 | 2.760 |
| 2 | 0.081 | 0.164 | 0.035 | 0.023 | 0.344 | 1.693 | 0.136 | 1.895 |
| 3 | 0.092 | 0.139 | 0.019 | 0.027 | 0.151 | 0.653 | 0.200 | 1.704 |
| 4 | 0.067 | 0.270 | 0.054 | 0.113 | 0.193 | 2.496 | 0.371 | 1.368 |
| 5 | 0.130 | 0.329 | 0.142 | 0.275 | 0.198 | 7.553 | 1.629 | 1.701 |
| 6 | 1.088 | 2.741 | 0.719 | 2.599 | 0.410 | 12.401 | 4.948 | 2.592 |
| 7 | 4.729 | 12.095 | 2.596 | 9.230 | 2.241 | 11.699 | 6.794 | 4.355 |
| 8 | 9.852 | 8.501 | 4.716 | 14.032 | 5.571 | 5.215 | 4.365 | 6.572 |
| 9 | 9.264 | 5.020 | 3.568 | 8.770 | 4.113 | 3.645 | 2.904 | 8.428 |
| 10 | 6.141 | 3.657 | 3.259 | 5.469 | 4.637 | 4.175 | 2.954 | 8.234 |
| 11 | 3.532 | 2.786 | 2.274 | 4.464 | 5.179 | 6.771 | 1.684 | 6.964 |
| 12 | 4.613 | 4.059 | 2.416 | 2.752 | 3.830 | 9.416 | 0.945 | 6.984 |
| 13 | 5.190 | 6.393 | 2.625 | 2.793 | 2.863 | 8.980 | 1.957 | 6.141 |
| 14 | 4.370 | 4.332 | 3.041 | 5.401 | 2.142 | 7.957 | 1.696 | 3.770 |
| 15 | 2.732 | 5.894 | 2.675 | 5.257 | 1.877 | 8.029 | 1.209 | 4.319 |
| 16 | 2.236 | 6.886 | 1.899 | 2.973 | 1.458 | 6.954 | 1.548 | 3.240 |
| 17 | 2.430 | 4.875 | 0.890 | 2.568 | 2.491 | 7.140 | 1.690 | 2.520 |
| 18 | 2.521 | 3.911 | 0.959 | 3.455 | 2.184 | 7.762 | 2.273 | 3.497 |
| 19 | 2.901 | 4.596 | 2.063 | 2.707 | 2.561 | 7.063 | 2.919 | 3.342 |
| 20 | 4.718 | 3.933 | 2.757 | 2.189 | 3.330 | 5.287 | 2.417 | 2.556 |
| 21 | 4.338 | 2.660 | 2.101 | 1.396 | 2.491 | 5.066 | 1.397 | 1.434 |
| 22 | 3.796 | 2.811 | 2.139 | 0.766 | 1.472 | 4.321 | 0.991 | 1.875 |
| 23 | 2.574 | 2.150 | 1.656 | 0.257 | 1.259 | 3.974 | 0.765 | 2.016 |

**Figure 7** | Comparison between (a) the average hourly synthetic water demand and the recorded demand and (b) the standard deviation of measured and simulated water consumptions (dwelling 1).

data is strictly dependent on the length of the monitored consumption time series and cannot be linked to a specific return period. As shown in Table 1, this length is different for the eight monitored dwellings.

Regarding MCDSA model results, Figures 8 and 9 show that the time to peak (expressed by the coefficient K_p) was preserved for all dwellings and for the two return periods chosen.

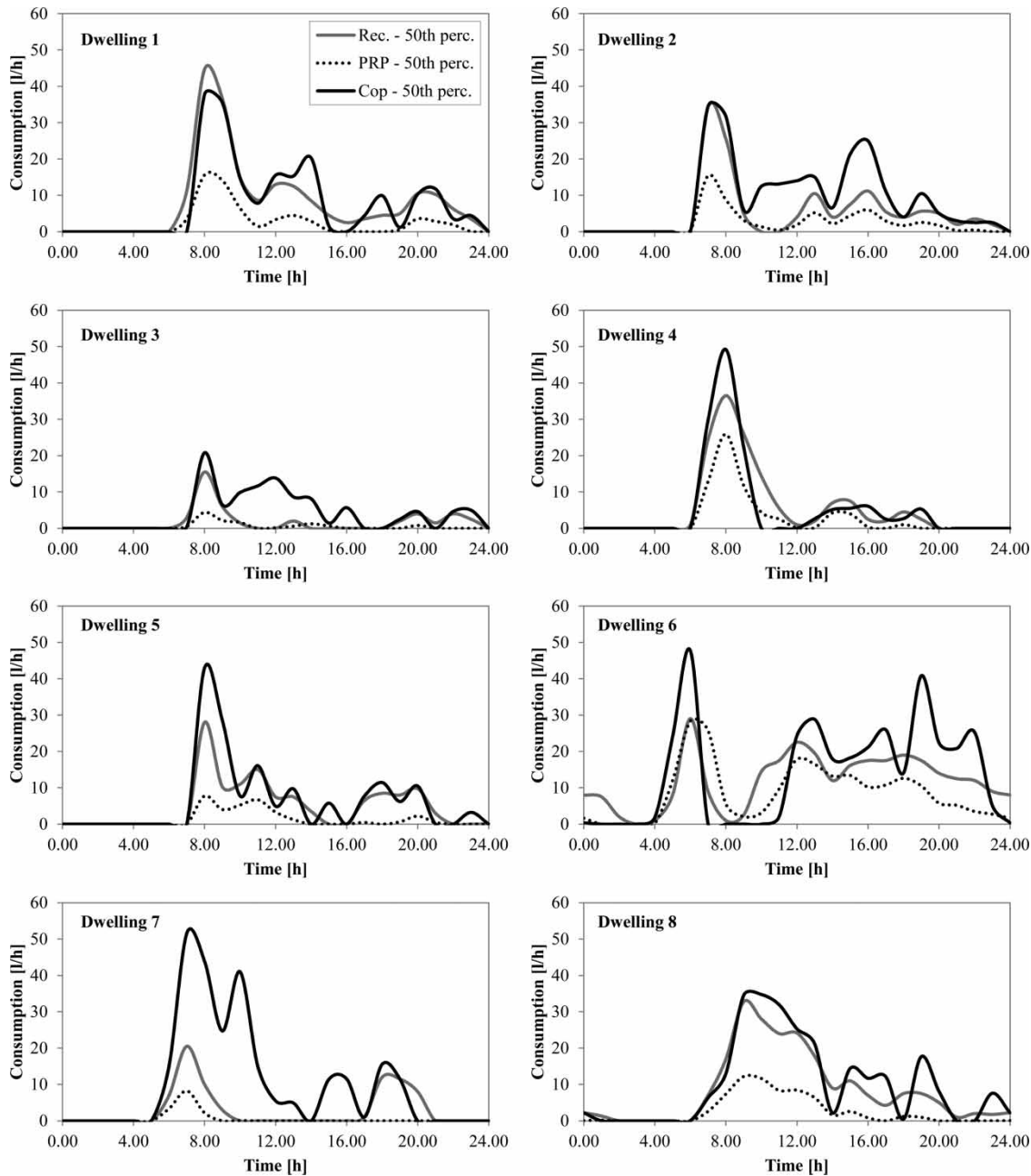


Figure 8 | Measured and synthetic average water demand (return period = 100 days).

For a return period equal to 100 days (Figure 8), making appropriate distinctions between the dwellings, the 50th percentiles of the MCDSA daily patterns reproduced quite accurately the median of the recorded daily patterns for six dwellings (1, 2, 3, 4, 5 and 8) while less accurately for dwellings 6 and 7. Namely, for dwellings 1, 4, 5, 6 and 8,

the total daily volume, V_d , showed quite acceptable values while for dwellings 2, 3 and 7, the related V_d values were much higher with relative error with respect to measured percentile, $\Delta V_d/V_d$ equal to -68% , -147% and -246% , respectively. The V_{max} characterising the generated pattern was acceptable for dwellings 1, 2, 3, 4 and 8; for the other

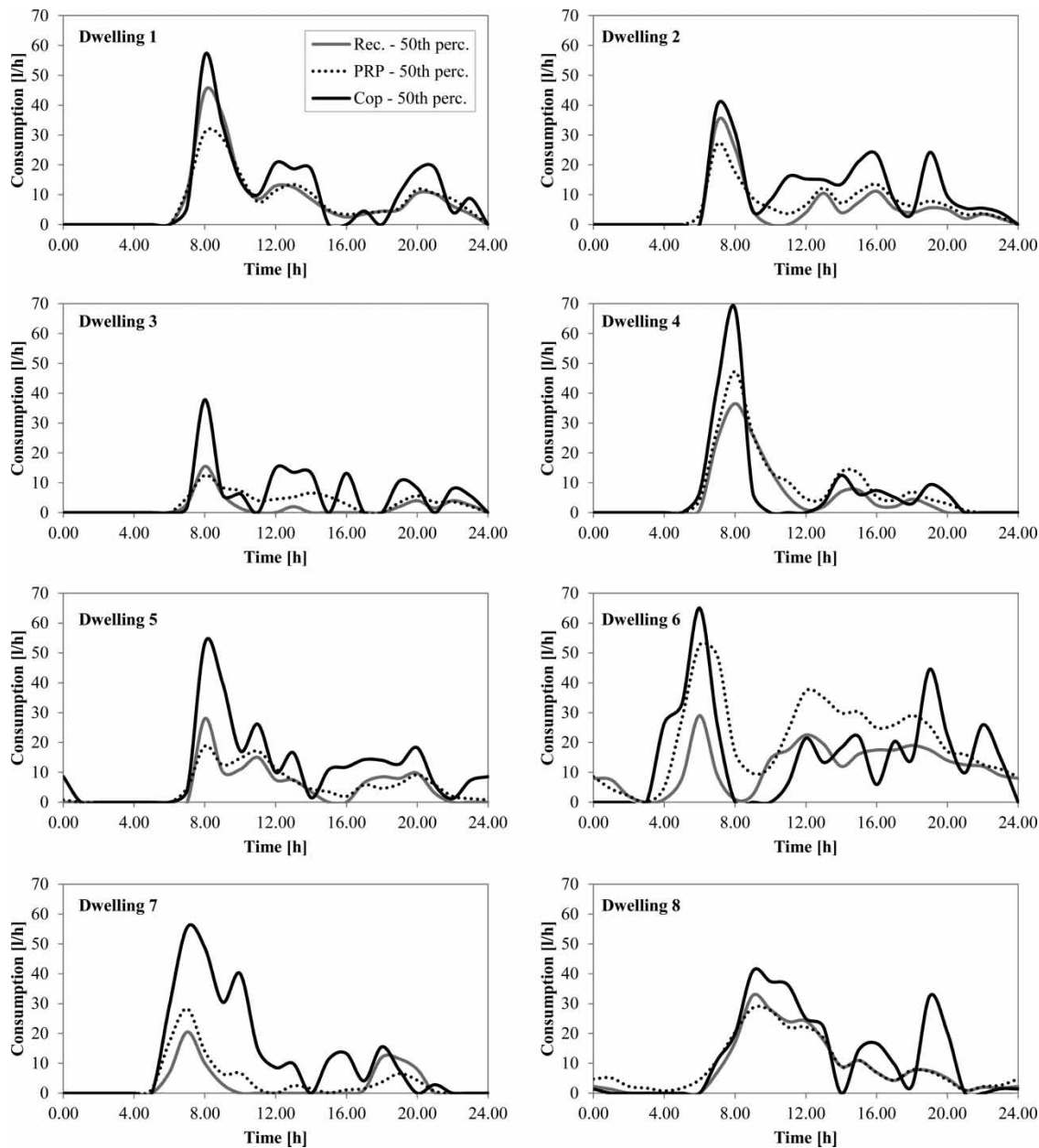


Figure 9 | Measured and synthetic average water demand (return period = 365 days).

dwellings (5, 6 and 7) V_{max} values were always higher than the recorded ones with a relative error, $\Delta V_{max}/V_{max}$, equal to -54% , -65% and -152% , respectively (Table 8).

For a return period equal to 365 days the V_d and V_{max} were higher than those for the shorter return period, as predicted. However, the daily patterns were well reproduced, except for dwellings 3, 6 and 7.

Summarising for the two return periods chosen, the 50th percentiles of the daily patterns generated by the MCDSA model usually tended to overestimate the consumptions with respect to the median of the recorded daily patterns (Table 8). The main reason can be found in the structure of the MCDSA model: the t_{KEN3} corresponding to the return periods chosen were high (0.99 and 0.997

Table 8 | Relative error of total daily volume (V_d) and maximum daily consumption (V_{max}) as ratio between the difference of the recorded value and the simulated value (50th percentile) and the recorded value, and error of peak coefficient (K_p) as difference between the recorded value and the simulated value (50th percentile)

| Dwelling | Recorded data | | | Copula method | | | PRP model | | |
|-----------------|---------------|-------|-----------|-------------------------------------|--------------|--------------------------|------------------|--------------|--------------------------|
| | V_d | K_p | V_{max} | $\Delta V_d/V_d$ $Tr = 100$ days | ΔK_p | $\Delta V_{max}/V_{max}$ | $\Delta V_d/V_d$ | ΔK_p | $\Delta V_{max}/V_{max}$ |
| 1 | 200.73 | 0.22 | 44.67 | 3% | -36% | 15% | 70% | 255% | 65% |
| 2 | 129.81 | 0.27 | 34.50 | -68% | 11% | 0% | 55% | 2% | 55% |
| 3 | 42.00 | 0.37 | 15.50 | -147% | 17% | -34% | 71% | 2% | 72% |
| 4 | 136.83 | 0.27 | 36.50 | 4% | -11% | -35% | 51% | -44% | 29% |
| 5 | 118.52 | 0.24 | 28.00 | -31% | -4% | -54% | 73% | -2% | 72% |
| 6 | 297.14 | 0.10 | 29.00 | -15% | -4% | -65% | 29% | -37% | 2% |
| 7 | 73.09 | 0.28 | 20.50 | -246% | 8% | -152% | 81% | -112% | 60% |
| 8 | 212.63 | 0.15 | 32.73 | -16% | 1% | -6% | 70% | -23% | 63% |
| $Tr = 365$ days | | | | | | | | | |
| 1 | 200.73 | 0.22 | 44.67 | -22% | -40% | -27% | 6% | -33% | 31% |
| 2 | 129.81 | 0.27 | 34.50 | -92% | 10% | -16% | -19% | 9% | 22% |
| 3 | 42.00 | 0.37 | 15.50 | -232% | 10% | -144% | -91% | 21% | 19% |
| 4 | 136.83 | 0.27 | 36.50 | -30% | -12% | -87% | -40% | 2% | -29% |
| 5 | 118.52 | 0.24 | 28.00 | -138% | 5% | -90% | -12% | 9% | 33% |
| 6 | 297.14 | 0.10 | 29.00 | -31% | -7% | -124% | -73% | 0% | -82% |
| 7 | 73.09 | 0.28 | 20.50 | -300% | 9% | -171% | -34% | -1% | -38% |
| 8 | 212.63 | 0.15 | 32.73 | -38% | 1% | -25% | -9% | 3% | 13% |

for 100 and 365 days, respectively), and the triplets statistically sampled $(V, K, T)_{copula}$ were characterised by infrequent values of the variables. As a result, the MCDSA model tended to overestimate water consumptions generating patterns characterised by high values of V_d and V_{max} .

In order to explain the different behaviour of dwellings 3, 6 and 7, some factors can be considered.

With regard to dwelling 6, the MCDSA model was not able to preserve V_{max} , probably due to the particular shape of the monitored daily patterns where more than one main peak of consumption occurred; dwelling 6 was the only dwelling that demonstrated secondary peaks compared to the main peak. Moreover, the length of the recorded data time series for dwelling 6 was longer than those of the other dwellings lasting approximately 2 years (Table 1). As a result, V_{max} was higher than the recorded one.

With regard to dwelling 3, the MCDSA model overestimated the consumptions between 9 a.m. and 5 p.m., and for dwelling 7, the MCDSA model was not able to accurately

forecast the consumption both in terms of V_d and V_{max} . The main reasons are that the median daily pattern of dwellings 3 and 7 showed no demand for several hours in the day (e.g., for dwelling 7 between 12 a.m. and 5 a.m., between 11 a.m. and 5 p.m., and after 9 p.m.), and all recorded patterns showed the same shape (and Huff curve), without any variation (e.g., weekday and weekend). Furthermore, as previously stated, the triplets $(V, K, T)_{copula}$ statistically sampled for the two return periods chosen were characterised by infrequent values of the variables (with t_{KEN3} values equal to 0.99 and 0.997 for 100 and 365 days, respectively). As result, the V_d and V_{max} were higher than the median of the recorded consumptions, and the generated pattern tended to overestimate water consumptions.

The procedure to define the hourly demand patterns and the percentiles for the given return period with the PRP model is similar to that previously described. Monte Carlo analysis was used to generate 100,000 series of 1,000 days of synthetic water consumption, from which the daily pattern water demand was derived for each set return period.

In the PRP model, the demand patterns showed the same shape (i.e., that of the historical consumption time series) for the two return periods: the 50th percentile of the simulated pattern was always lower than that of recorded consumption for the 100-day return period and was close to the median of the historical series for the 365-year return period (Table 8). The only exception was dwelling 6, for which the simulated pattern well reproduced the recorded pattern for the 100-day return period and was higher than the same curve for the 365-day return period (Table 8).

The length of the recorded consumption sample appears to influence the results of the two models; the daily MCDSA pattern preserved quite accurately all features of the recorded consumption for return periods that were shorter than the length of the historical time series, and the statistical analysis was less confident for the return period that was comparable with the length of the sample. The PRP model underestimated the water demand for return periods that were shorter than the historical series; the demand pattern was close to that recorded for the return periods that were comparable to the length of the consumption data.

CONCLUSIONS

Interest in domestic water demand modelling stems from the goal of achieving two primary objectives: to analyse the domestic consumption process to aid systems management and to define demand patterns at a given return period to aid systems design. From this viewpoint, the present study proposed a statistical methodology for the definition of water consumption patterns based on the return period and a multivariate probabilistic approach. The method is based on a multivariate statistical analysis in which a 3-d vine copula was constructed for the primary features of the consumption process at the domestic level. The water demand was predicted for the given return periods using the patterns that were statistically generated, considering the historical series of consumption to which the methodology was applied. The analysis of the percentiles of the water demand for the given return period show that the proposed approach produced consistent demand patterns and will be a powerful tool to couple with water distribution network models for design or analysis problems. The MCDSA forecasts water demand patterns by reproducing

quite accurately the main process properties: the time to peak, that is always preserved, the total daily volume, and the maximum daily consumption, increasing with the return period, for which appropriate comments are made. The total daily volume and the maximum daily consumption are in good agreement with those of recorded data only for some of the residences chosen for the case study; higher relative error occurred for others. The reason may be looked for in the shape of the monitored daily patterns: more than one main peak in water demand pattern and time series with the same shape without any variation among days results in less accurate outcomes.

The proposed methodology was validated by comparing the results with those obtained using the PRP model, as well. The PRP model calibrated on the consumption data of a single user was used to simulate the water demand during the same return periods.

The MCDSA results were always higher than those of the PRP model for both return periods; the MCDSA water demand pattern preserved and overestimated all recorded consumption features for return periods that were shorter and longer than the length of the historical series, respectively. For the two chosen return periods, the triplets values were characterised by high exceedance probability and, as a result, the total daily volume and the maximum daily consumption were higher and the generated pattern tended to overestimate water consumption.

However, MCDSA methodology even if less accurate than the PRP model is characterised by higher applicability because it needs inferior numbers of data compared to the PRP model, with the advantage that it can be easily performed at large scale by means of a common automatic meter reading system with restrained costs of data transmission and management.

As a future development of the study, also the pressure at the dwelling should be considered in the multivariate analysis in order to take into account the effect of the pressure on the water consumption phenomenon.

REFERENCES

- Aas, K., Czado, C., Frigessi, A. & Bakken, H. 2009 [Pair-copula construction of multiple dependence](#). *Insur. Math. Econ.* **44**, 182–198.

- Alcocer-Yamanaka, V. H., Tzatchkov, V. G. & Arreguin-Cortes, F. I. 2012 Modeling of drinking water distribution networks using stochastic demand. *Water Resour. Manage.* **26**, 1779–1792.
- Alvisi, S., Franchini, M. & Marinelli, A. 2003 A stochastic model for representing drinking water demand at residential level. *Water Resour. Manage.* **17**, 197–222.
- Alvisi, S., Franchini, M. & Marinelli, A. 2007 A short-term, pattern-based model for water-demand forecasting. *J. Hydroinform.* **9**, 39–50.
- 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.
- Bárdossy, A. & Pegram, G. G. S. 2009 Copula based multisite model for daily precipitation simulation. *Hydrol. Earth Syst. Sci.* **13**, 2299–2314.
- Bedford, T. & Cooke, R. M. 2001 Probability density decomposition for conditionally dependent random variables modeled by vines. *Ann. Math. Artif. Intel.* **32**, 245–268.
- Bedford, T. & Cooke, R. M. 2002 Vines – a new graphical model for dependent random variables. *Ann. Stat.* **30**, 1031–1068.
- 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 Res. Pl.-ASCE* **136**, 19–26.
- Buchberger, S. G. & Wells, G. J. 1996 Intensity, duration and frequency of residential water demands. *J. Water Res. Pl.-ASCE* **130**, 386–394.
- Buchberger, S. G. & Wu, L. 1995 A model for instantaneous residential water demands. *J. Hydraul. Eng.-ASCE* **121**, 232–246.
- Buchberger, S. G., Carter, J. T., Lee, Y. & Schade, T. G. 2003 *Random Demands, Travel Times, and Water Quality in Deadends*. IWA Publishing, London, UK.
- Clark, R., Grayman, W., Males, R. & Hess, A. 1993 Modeling contaminant propagation in drinking-water distribution systems. *J. Environ. Eng.* **119**, 349–364.
- Cowpervait, P. S. P., O'Connell, P. E., Metcalf, A. V. & Mawdsley, J. A. 1996 Stochastic point process modeling of rainfall. I. Single-site fitting and validation. *J. Hydrol.* **175**, 17–46.
- De Michele, C. & Salvadori, G. 2003 A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J. Geophys. Res.* **108**. Doi:10.1029/2002JD002534.
- Donkor, E., Mazzuchi, T. A., Soyer, R. & Roberson, J. A. 2012 Urban water demand forecasting: a review of methods and models. *J. Water Resour. Plann. Manage.* **140**, 146–159.
- Dziegielewska, B. & Boland, J. J. 1989 Forecasting urban water use: the IWR main model. *J. Am. Water Works Assoc.* **25**, 101–109.
- Favre, A. C., El Adlouni, S., Perreault, L., Thiémondge, N. & Bobée, B. 2004 Multivariate hydrological frequency analysis using copulas. *Water Resour. Res.* **40**. Doi:10.1029/2003WR002456.
- Fontanazza, C. M., Freni, G., La Loggia, G. & Notaro, V. 2011 Uncertainty evaluation of design rainfall for urban flood risk analysis. *Water Sci. Technol.* **63** (11), 2641–2650.
- Fontanazza, C. M., Notaro, V., Puleo, V. & Freni, G. 2014 Multivariate statistical analysis for water demand modeling. *Procedia Eng.* **89**, 901–908.
- Freni, G., La Loggia, G., Termini, D. & Vitellaro, P. 2004 A water demand model by means of the artificial neural networks method. In: *Proceedings of the 6th International Conference on Hydroinformatics*, Singapore, 21–27 June.
- García, V. J., García-Bartual, R., Cabrera, E., Arregui, F. & García-Serra, J. 2004 Stochastic model to evaluate residential water demands. *J. Water Res. Pl.-ASCE* **130**, 386–394.
- Graller, B., van de Berg, M. J., Vanderberghe, S., Petroselli, A., Grimaldi, S., De Baets, B. & Verhoest, N. E. C. 2013 Multivariate return period in hydrology: a critical and practical review focusing on synthetic design hyetograph estimation. *Hydrol. Earth Syst. Sci.* **17**, 1281–1296.
- Grimaldi, S. & Serinaldi, F. 2006 Design hyetograph analysis with 3-copula function. *Hydrolog. Sci. J.* **51**, 223–238.
- Guercio, R., Magini, R. & Pallavicini, I. 2001 Instantaneous residential water demand as stochastic point process. In: C. A. Brebbia, P. Anagnostopoulos & K. L. Katsifarakis (eds), *Water Resources Management*. WIT Press, Southampton, UK, pp. 129–138.
- Herrera, M., Torgo, L., Izquierdo, J. & Pérez-García, R. 2010 Predictive models for forecasting hourly urban water demand. *J. Hydrol.* **387**, 141–150.
- House-Peters, L. A. & Chang, H. 2011 Urban water demand modeling: review of concepts, methods, and organizing principles. *Water Resour. Res.* **47**, W05401. Doi:10.1029/2010WR009624.
- ISO 4064-1 2005 Measurement of water flow in fully charged closed conduits – Meters for cold potable water and hot water. Part 1: Specifications.
- Joe, H. 1996 Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In: L. Rüschendorf, B. Schweizer & M. D. Taylor (eds), *Distributions with Fixed Marginals and Related Topics*. Institute of Mathematical Statistics, Beachwood, OH, USA, pp. 120–141.
- Magini, R., Pallavicini, I. & Guercio, R. 2008 Spatial and temporal scaling properties of water demand. *J. Water Res. Pl.-ASCE* **134**, 276–284.
- Maidment, D. R., Miaou, S. P. & Crawford, M. M. 1986 Daily water use in nine cities. *Water Resour. Res.* **22**, 845–851.
- Miaou, S. P. 1990 A class of time series urban water demand models with non-linear climatic effects. *Water Resour. Res.* **26**, 169–178.
- Qi, C. & Chang, N. 2011 System dynamics modeling for municipal water demand estimation in an urban region under uncertain economic impacts. *J. Environ. Manage.* **92**, 1628–1641.
- Rodriguez-Iturbe, I. 1986 Scale of fluctuation of rainfall models. *Water Resour. Res.* **22**, 15S–37S.
- Rodriguez-Iturbe, I., Gupta, V. K. & Waymire, E. 1984 Scale considerations in the modeling of temporal rainfall. *Water Resour. Res.* **20**, 1611–1619.

- Rodriguez-Iturbe, I., Cox, D. R. & Isham, V. 1987 Some models for rainfall based on stochastic point process. *Proc. R. Soc. Lond. A* **410**, 269–288.
- Rothstein, E. 1992 Water demand monitoring in Austin, Texas. *J. Am. Water Works Assoc.* **84**, 52–58.
- Salvadori, G. & De Michele, C. 2004a Analytical calculation of storm volume statistics involving Pareto-like intensity duration marginals. *Geophys. Res. Lett.* **31**. Doi:10.1029/2003GL018767.
- Salvadori, G. & De Michele, C. 2004b Frequency analysis via copulas: theoretical aspects and applications to hydrological events. *Water Resour. Res.* **40**. Doi:10.1029/2004WR003133.
- Salvadori, G., De Michele, C. & Durante, F. 2011 On the return period and design in a multivariate framework. *Hydrol. Earth Syst. Sci.* **15**, 3293–3305.
- Vertommen, I., Magini, R. & da Conceição Cunha, M. 2014 Generating water demand scenarios using scaling laws. *Procedia Eng.* **70**, 1697–1706.
- Zhou, S. L., McMahon, T. A., Walton, A. & Lewis, J. 2000 Forecasting daily urban water demand: a case study of Melbourne. *J. Hydrol.* **236**, 153–164.
- Zhou, S. L., McMahon, T. A., Walton, A. & Lewis, J. 2002 Forecasting operational demand for an urban water supply zone. *J. Hydrol.* **259**, 189–202.

First received 31 December 2014; accepted in revised form 16 September 2015. Available online 17 November 2015