Water Science & Technology: Water Supply | 16.6 | 2016

A stochastic model for daily residential water demand

Rudy Gargano, Carla Tricarico, Giuseppe del Giudice
and Francesco Granata

ABSTRACT

Residential water demand is a random variable which influences greatly the performance of municipal water distribution systems (WDSs). The water request at network nodes reflects the behavior of the residential users, and a proper characterization of their water use habits is vital for the hydraulic system modeling. This study presents a stochastic approach for the characterization of the daily residential water use. The proposed methodology considers a unique probabilistic distribution – mixed distribution – for any time during the day, and thus for any entity of the water demanded by the users. This distribution is obtained by the merging of two cumulative distribution functions taking into account the spike of the cumulative frequencies for the null requests. The methodology has been tested on three real water distribution networks, where the water use habits are different. Experimental relations are given to estimate the parameters of the proposed stochastic model in relation to the users number and to the average daily trend. Numerical examples for a practical application have shown the effectiveness of the proposed approach in order to generate the time series for the residential water demand.

Key words | logistic distribution, mixed distribution, monitoring system, Monte Carlo method, residential water demand, stochastic approach, WDS

Rudy Gargano (corresponding author)
Carla Tricarico
Francesco Granata
Dipartimento di Ingegneria Civile e Meccanica,
University of Cassino and Southern Lazio,
Via G. Di Biasio 43,
03043 Cassino (FR),
Italy
E-mail: *gargano@unicas.it*

Giuseppe del Giudice
Dipartimento di Ingegneria Civile, Edile e Ambiente,
University of Naples Federico II,
via Claudio, 21,
80125 Napoli,
Italy

LIST OF SYMBOLS AND ACRONYMS

$C_D(t)$	[–] demand coefficient at time t
CDF	cumulative distribution function
CE	Castelfranco Emilia (one of three case studies)
CV(t)	[–] variation coefficient of the demand coefficient at time t
DMA	demand monitoring area
F_o	[–] probability of occurrence of null request
F^*	[–] CDF of the flow demand when it is different to zero
F_o^{max}	maximum value of the probability of null water demand
Fr	Franeker (one of three case studies)
L	logistic distribution
MD	mixed distribution
N	normal distribution
N_{us}	[–] number of users
PSG	Piedimonte San Germano (one of three case studies)
$Q(t)$	[L^3/T] demanded water flow at time t
$\mu_{C_D}(t)$	[–] mean demand coefficient at time t
μ_{C_m}	[–] minimum threshold value of the demand coefficient
μ_Q	[L^3/T] mean daily water demand
$\mu_Q(t)$	[L^3/T] mean water demand at time t
$\sigma_{C_D}(t)$	[–] deviation standard of demand coefficient at the time t

INTRODUCTION

Water distribution system (WDS) simulation models have been improved drastically with advances in computational algorithms, machine speed and data storage.

doi: 10.2166/ws.2016.102

For instance, hydraulic simulators have recently been developed by taking into account the functional relationship between delivered flows and nodal pressures. This is of particular importance when pressures are below the minimum level required to meet the desired water demand (e.g. Todini 2003; Giustolisi *et al.* 2008). In addition, some WDS models are able to calculate unsteady flow processes in WDSs (e.g. Vítkovský *et al.* 2006; Ferrante *et al.* 2014) or to detect pollution sources in the network (e.g. Preis & Ostfeld 2006; Liserra *et al.* 2014). Calibration techniques have been proposed for determining pipe roughness coefficient more realistically (e.g. Walski 1983; Greco & del Giudice 1999; Jung *et al.* 2014).

But, the performance of a WDS model is highly dependent on water demands. Hence, the value and utility of these WDS modeling advances is determined by the degree to which the WDS simulation is able to reproduce real water demands. For this aim the water demand models have to be capable of representing the random nature of the water requirements.

Bao & Mays (1990) used a Monte Carlo sample generator to model water demand uncertainty in a WDS reliability study. Subsequently, several approaches were proposed to incorporate water demand uncertainty in WDS hydraulic reliability (e.g. Gargano & Pianese 2000; Babayan *et al.* 2006; Chung *et al.* 2009), in sizing urban reservoirs (e.g. Nel & Haarhoff 1996; Van Zyl *et al.* 2008) and in the optimal rehabilitation/design of hydraulic networks (e.g. Kapelan *et al.* 2005; Tricarico *et al.* 2006; Tricarico *et al.* 2014).

A key challenge for practical application, however, is obtaining reliable estimates of the parameters of the corresponding cumulative distribution functions (CDFs) that need to be implemented for water demand generation.

Unfortunately, in the technical literature, there are only a few studies which reproduce water demand variability with CDF parameters that were estimated from experimental data (Alvisi *et al.* 2003; Blokker *et al.* 2010). This is probably due to the difficulty of getting experimental data, which can effectively represent the water requirements of a specific number of residential users.

The target of this paper is to define a practical stochastic approach which allows us to model realistically the residential water demand in relation to the users number and thus to generate daily time series. This is referred to an appreciable number of aggregated users (200–1,250), therefore the water demand can be handled as a unique random variable that is continuous positive.

The proposed model leads to an important simplification in respect to the approaches which are capable of reproducing the residential water demand of the end users such as the *Poisson Rectangular Pulse* (PRP – Buchberger & Wu 1995; Buchberger & Wells 1996), or the *Neyman-Scott Rectangular Pulse* (NSRP – Alvisi *et al.* 2003; Alcocer-Yamanaka *et al.* 2012) but without losing accuracy. In fact, the end user approaches need to model the random phenomena (usually at least three phenomena: frequency, duration and intensity of the residential request) which contribute to defining the water demand for each user. This issue arises also when the aim of the end user approach is to model the total request of the dwellings, neglecting the demands at the single taps, as the *Overall Pulse model* (Gargano *et al.* 2016).

In addition, in the end user models it is necessary to aggregate the water demand of single flats in order to represent the water demand for WDS nodes, i.e. for several users (*bottom-up approach*). This operation can be influenced by important scale effects, especially when the user number increases. Indeed, the estimation of the parameters for an aggregated number of users N_{us} on the basis of the parameters of a single consumer – at most a few – represents an arduous issue (e.g. Moughton *et al.* 2006; Filion *et al.* 2007; Magini *et al.* 2008), which can limit the practical applications of the end user approaches in modeling the residential water demand, especially when the inhabitant number is relevant.

This issue is avoided when the residential water demand is directly modeled as a clustered request of a significant number of users (e.g. Tricarico *et al.* 2007; Gato-Trinidad & Gan 2012). Indeed, the proposed stochastic model represents the aggregated water demand of users whose number ranges roughly between 200 and 1,250.

It is worth noting that the range of the considered number of users, even if restricted, is nevertheless of significance for characterizing the water demand at WDS nodes.

Instead of considering several probability distributions in relation to the time of the day, it is herein suggested a novel distribution which can be applied generally for any time of the day. This specific probabilistic distribution, called mixed distribution (MD), and the relationships for estimating its parameters are proposed in order to achieve an approach concretely applicable to generate synthetic time series of water demand during the day.

The proposed approach allows us to model the residential water demand in relation to the daily time, assuming the simplifying hypothesis to disregard the time correlation effects. However this assumption does not prejudice the effectiveness of the proposed approach, as is shown in the following sections by the comparison between the observed data and the generated time series.

The proposed relationships for the estimation of the MD parameters are in relation to the user number and the average water request during the day. Therefore, the implementation of the proposed model requires only the knowledge of the daily trend of the water demand, which is related to the considered users, but it is usually known by the water companies (e.g. by measuring the outflows of the municipal reservoirs).

In order to obtain results of general application, the actual time series of the water demand of users with different life styles were analyzed. In fact, a monitoring system for the WDS of a small town in Southern Italy (Piedimonte San Germano (PSG)) was carried out to obtain reliable data on residential water request. These data have been then compared with the time series of two other towns: Castelfranco Emilia (I) (Alvisi *et al.* 2003) and Franeker (NL) (Blokker *et al.* 2006).

The residential water request has been considered dimensionless by means of the daily mean water demand.

The flow request was measured [L/s] and it was averaged on a 1 min time interval. This interval allows us to obtain a detailed description of the water demand, and a 1 min interval is not so approximate as to induce appreciable scale effects caused by the time aggregation.

MONITORING SYSTEMS AND SAMPLE DATA

Nowadays the possibility of installing economic and reliable water demand monitoring systems allows us to have in-situ laboratories on real WDSs. But in order that the collected data be effectively representative of the water demand, the WDSs need to be redundant, i.e. with pressures at all nodes of the network higher than the minimum required level.

Moreover the number of users is a relevant parameter to describe the water demand (e.g. Mays 1999; Martinez-Solano *et al.* 2008; Gargano *et al.* 2012), therefore the measured flows were analyzed in relation to the relative served inhabitants. Hence, attention has to be focused on the measurement point locations, where the number of users is often difficult to estimate due to the looped structure of the WDSs.

Measurement points should be thus allocated along the pipes which supply a circumscribed demand monitoring area (DMA). DMAs are present when the topological WDS scheme – or part of it – is branched, or when in the network it is possible to detect areas which are connected with the rest of the system by means of a limited number of pipes (e.g. Buchberger & Nadimpalli 2004). The DMA characterization requires an accurate census in order to define the number of users requiring water.

Only when the pressure is adequate, can the flows collected be considered an expression of the residential water demand; therefore, the monitoring systems must record jointly the pressure and the flow rate.

A further preprocessing procedure has to be applied when the observed data are affected by leakage phenomena. In fact, in this circumstance, the time series have to be filtered from the water losses in order to represent the effectively demanded flows.

On the basis of these circumstances, a specific monitoring system which involves the real WDS of a small town – PSG – was realized by the Laboratorio di Ingegneria delle Acque of the Università di Cassino e del Lazio Meridionale.

The monitoring system consists of four measurement points, each with a pressure cell, and an electromagnetic flow meter, and all probes were connected to a data logger. The measurements are recorded continuously with an acquisition frequency of up to 1 Hz. The strategic location of the meters allowed measurement of the water demand for four DMAs with different numbers of users, N_{us}: 239, 777, 981 and 1,220.

The size of the water users has been measured by means of the number of the supplied inhabitants N_{us}, to which the relationships refer in the following sections.

The water demanded flow was made dimensionless by means of the ratio:

$$C_D(t) = \frac{Q(t)}{\mu_Q} \tag{1}$$

where μ_Q is the daily mean water demand, and $Q(t)$ the flow demand during the day. Hence, all the following statistics are referred to the dimensionless random variable of Equation (1).

In order to check the robustness of the results of this research, the proposed approach and the relative relationships have been tested with the water demand of users which present lifestyles totally different in respect to PSG inhabitants.

Indeed, the recorded time series of two further monitoring systems have been analyzed:

- Castelfranco Emilia (Italy) (CE) with a number of users equal to 596 (Alvisi et al. 2003);
- Franeker (The Netherlands) (Fr) with a number of users equal to 1,150 (Blokker et al. 2006).

These two monitoring systems were realized specifically for studying the water demand, hence they gave reliable time series for the water request. In addition, CE and Fr monitoring systems, as the field laboratory of PSG, provide a fine description of the water demand during the day because the time step Δt equals 1 min.

The assumption of a fine time step is precautionary in respect to the extreme water demands. Indeed, Tricarico et al. (2007) observed that when the time step is equal to 1 h, the peak demands are underestimated around the 20% for $\Delta T = 1$ min. Moreover Buchberger & Nadimpalli (2004) demonstrated that the probability of null request decreases exponentially with the increasing ΔT.

On this basis, all the following relations and considerations assume $\Delta t = 1$ min.

It is worth noting that in-situ laboratories with the above mentioned characteristics are rare, therefore it is difficult to obtain reliable time series of the flow demand.

Table 1 summarizes the principal characteristic of the field laboratories and of the observed time series, which were used for the statistical elaborations and inferences.

Because different consumptions were observed between the weekends and the weekdays, only the latter were analyzed.

For all the monitored urban areas it can be assumed that the users are exclusively residential and mainly characterized by an indoor water use. Users are mainly represented by middle class families where the major economical occupation is constituted by industrial workers.

Table 1 | Principal characteristics of the monitoring systems

	PSG (I)	Fr (NL)	CE (I)
N_{us}	239-777-981-1220	1,150	596
Monitored period	5 Jan.- 30 Mar. 2004 7 Jan.- 31 Mar. 2006	23 Jan. - 31 Mar. 2006	7 Jan. - 20 Mar. 2000
Type of counter	Flow meter	Flow meter	Volumetric meter
Frequency of sampling	1 Hz	1/60 Hz	1/60 Hz
Number of used days for the statistics	95	50	50

Figure 1 draws the trends of the water request for the three monitored users, where $\mu_{C_D}(t)$ was estimated considering the mean demanded flow for each minute of the day $\mu_Q(t)$:

$$\mu_{C_D}(t) = \frac{\mu_Q(t)}{\mu_Q} \tag{1'}$$

As the plots of Figure 1 show, the trends of the water demand for the three towns are completely different. Indeed, the request of PSG (Figure 1(a)) presents three decreasing peaks during the day, where the first peak demand in the morning is considerably more important (similar trends were observed for the other users of PSG). The Fr trend (Figure 1(b)) presents two daily peaks only, the sizes of which are very similar. Finally, the water demand of CE (Figure 1(c)) is a middle ground trend, where the evening peak is more significant than the morning one.

The significant differences among the daily trends of the considered users allowed the development of severe tests for the proposed stochastic model.

THE MIXED DISTRIBUTION

The mixed distribution allows us to model the random component of the residential water demand by means of a unique probabilistic CDF, whatever the daily time.

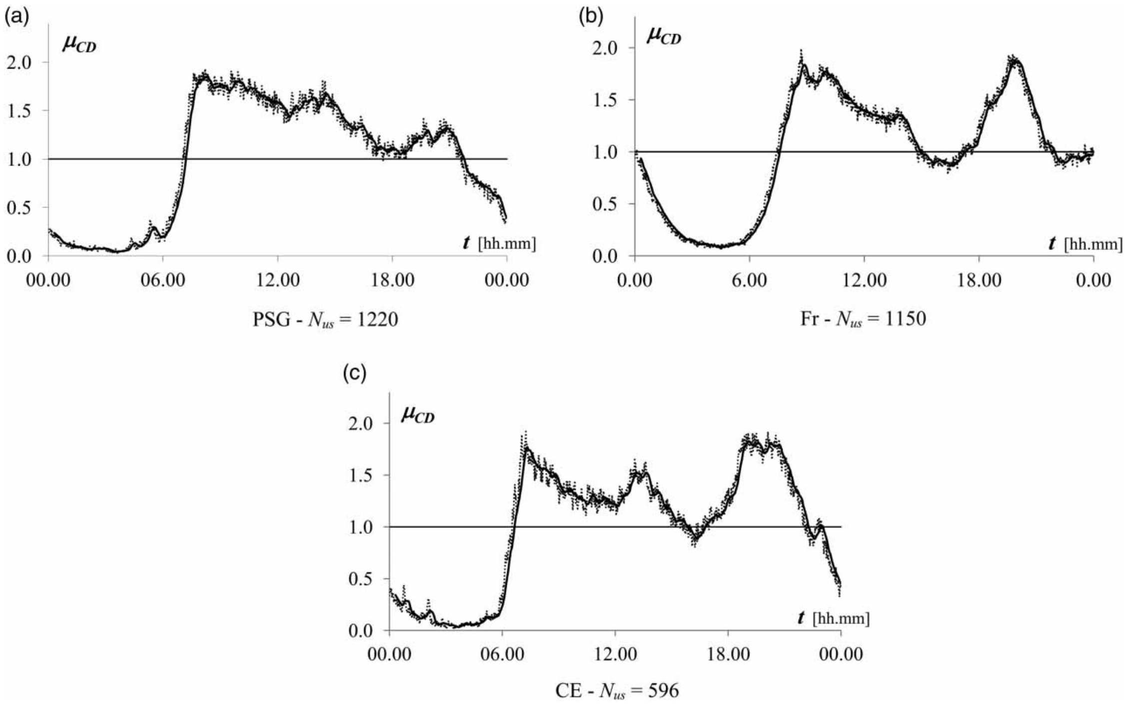

Figure 1 | Daily trend of the residential water demand for the three analyzed case studies.

Therefore this distribution has to be able to represent the extreme variability of the user daily habits in respect of their water requirements. This implies that with a unique probabilistic distribution it is possible to reproduce both a binary random variable (on/off of water flow demand), and a positive continuous random variable (water flow demand). Indeed, the night flow requests, and in general the minimum demands, lead to the necessity of a probabilistic model capable of representing not only the water demand entity, but also the event of null water demand (C_D-$(t) = 0$). Indeed, observed cumulative frequencies of minimum water demand (Figure 2) have pointed out the presence of a vertical mass spike for $C_D = 0$.

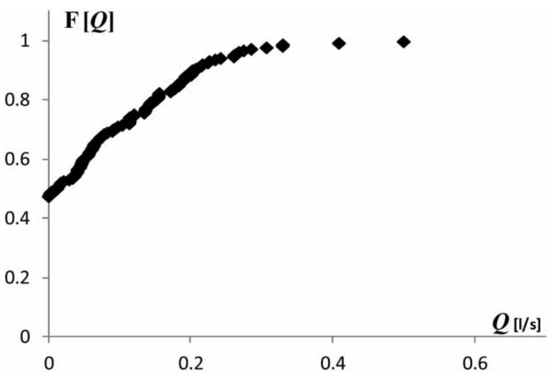

Figure 2 | Example of observed cumulative frequencies of the residential demanded flow (PSG users: $N_{us} = 777$ and $t = 6.00$ a.m.).

Obviously, the need to consider the spike for $C_D = 0$ is more evident when the number of users considered is negative. The null water requirement is quite probable for a reduced number of users, and this is not just during the night hours.

The water demand model can be thus considered as a combination of two distributions: the first one referred to as a random discrete variable, the second one as a continuous random variable (Gargano *et al.* 2014). In particular, the first distribution describes the event of null demanded flow, while the second distribution represents the circumstance for which the flow demand is different from zero.

The mixed distributions are effective when the analyzed random events are the summation of two phenomena. For instance, in the hydrological realm the two-component extreme value distribution is effective to model the flood peaks, as a product of the mixture of two types of storms (Rossi *et al.* 1984).

By applying thus the total probability theorem for the random variable C_D, the resulting CDF is:

$$F[C_D] = \Pr[C_D = 0] \cdot \Pr[C_D < c_D | C_D = 0] + \Pr[C_D > 0] \cdot \Pr[C_D < c_D | C_D > 0] \quad (2)$$

If F_o is the occurrence probability that the water demand is null ($F_o = \Pr[C_D = 0]$), and F^* represents the CDF of the flow demand when it is not null ($F^* = \Pr[C_D < c_D | C_D > 0]$), Equation (2) can be rewritten as:

$$F[C_D] = F_o + (1 - F_o)F^* \quad (3)$$

Equation (3) has been obtained by observing that the $\Pr[C_D \leq c_D | C_D = 0]$ is the probability of a certain event, and that the two conditions of flow demand $[C_D = 0]$ and $[C_D > 0]$ are incompatible and exhaustive events.

Therefore, Equation (3) requires the definition of the two component distributions F_o and F^*, the value of which depends on the daily time:

$$F[C_D(t)] = F_o(t) + (1 - F_o(t))F^*(t) \quad (3')$$

Equation (3′) represents the application of the MD for the context of a daily stochastic model.

Probability of null water demand, $F_o(t)$

The event for which the water demand in a WDS node is equal to zero $[C_D(t) = 0]$ or different by zero $[C_D(t) > 0]$ might be modeled by means of the simple Bernoulli distribution, where the probabilities of the two incompatible and exhaustive events are respectively $F_o(t) = \Pr[C_D(t) = 0]$ and the complement $1 - F_o(t) = \Pr[C_D(t) > 0]$.

On the basis of the field data recorded the $F_o(t)$ probability was estimated. It is appropriate, nevertheless, to highlight that as the peculiarity of the users' habits affects significantly the daily trend $\mu_{C_D}(t)$, so it affects also the probability $C_D(t) = 0$ estimation, as demonstrated in the following sections.

Moreover, it was shown (Buchberger & Nadimpalli 2004) that the $F_o(t)$ value depends significantly on the interval time Δt with which the day has been discretizated and on the number of users supplied.

The following analysis of F_o value has been developed for $\Delta t = 1$ minute and for a number of users ranging between 200 and 1,250 N_{us}.

Figure 3 shows the $F_o(t)$ values estimated by means of the observed time series (Table 1), for which it is evident – as it was expected – that the probability of null water request assumes significant values during the night time, and the $F_o(t)$ trend depends significantly on the users' habits.

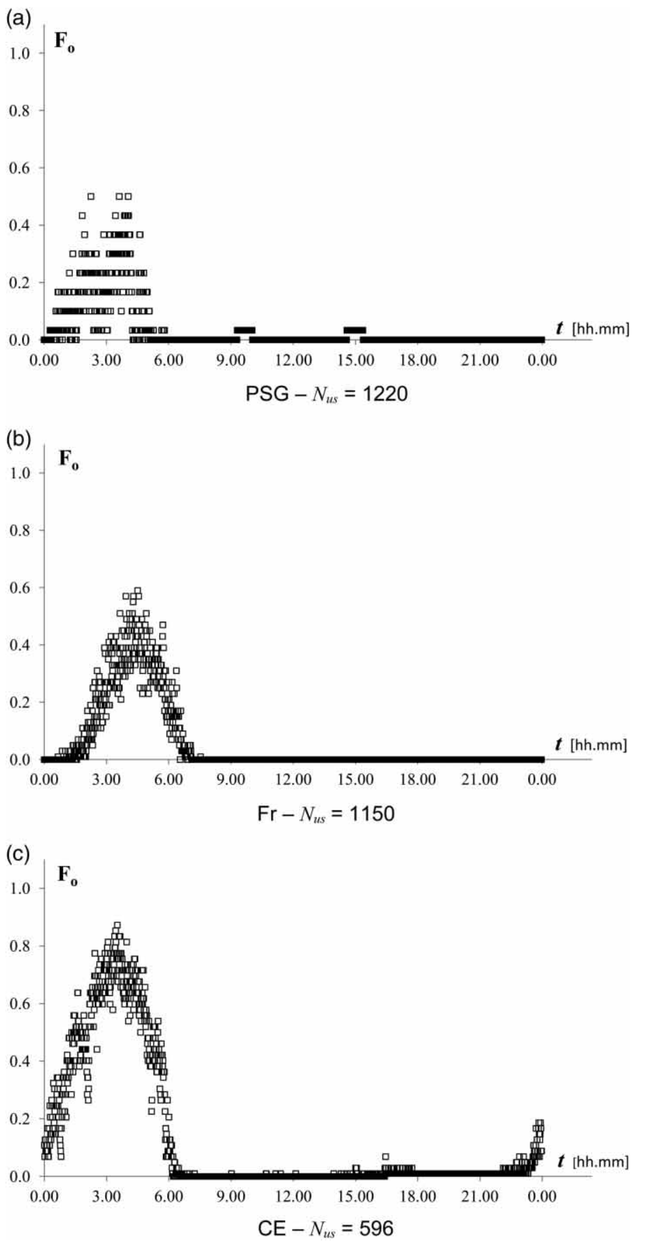

Figure 3 | Probability of null water request during the day, experimental data for different N_{us}.

In addition, Figure 3 shows that the F_o increases when the number of users decreases.

The F_o values during the rest of the day are substantially null and this depends on considered N_{us}. Hence, for a minus number of users minus respect to that investigated, the $F_o(t)$ should present non null values also in other daily time.

Therefore, the maximum value of the probability of null water demand F_o^{max}, decreases with the increasing the number of supplied users.

This trend has been pointed out in the plot of Figure 4, where the maximum value of F_o could be estimated in relation to the users number by means of:

$$F_o^{max} = 1 - 0.25\left(\frac{N_{us}}{1000}\right)^{2.5} \tag{4}$$

Equation (4), valid for $N_{us} = 200 \div 1{,}250$, shows that F_o^{max} tends rapidly to 1 for small user numbers. In this condition, the null water request defined during the night hours becomes practically a certain event.

The experimental data reported in Figure 5 shows that F_o can be estimated in relation to the mean values of the demand coefficients, μ_{C_D}. Indeed, the experimental data can be represented by the exponential relationship:

$$F_o = \exp\left(-5\frac{N_{us}}{1000}\mu_{C_D}\right) \tag{5}$$

The F_o probability value promptly decreases towards the null value already for the flows close to the mean daily value of the demand coefficient (Figure 5). This trend is also more evident for the increase of the user number.

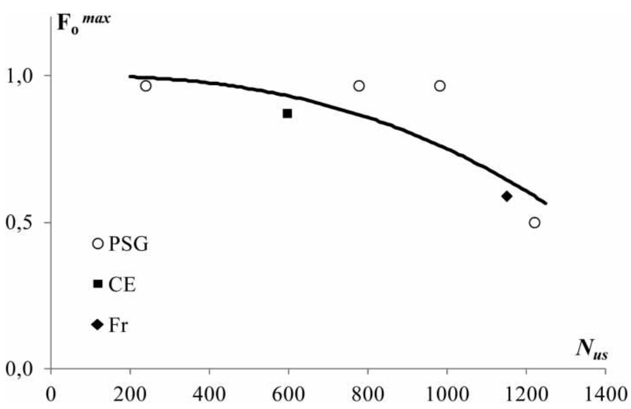

Figure 4 | Equation (4) (—) for the estimation of F_o^{max} versus N_{us} and observed data.

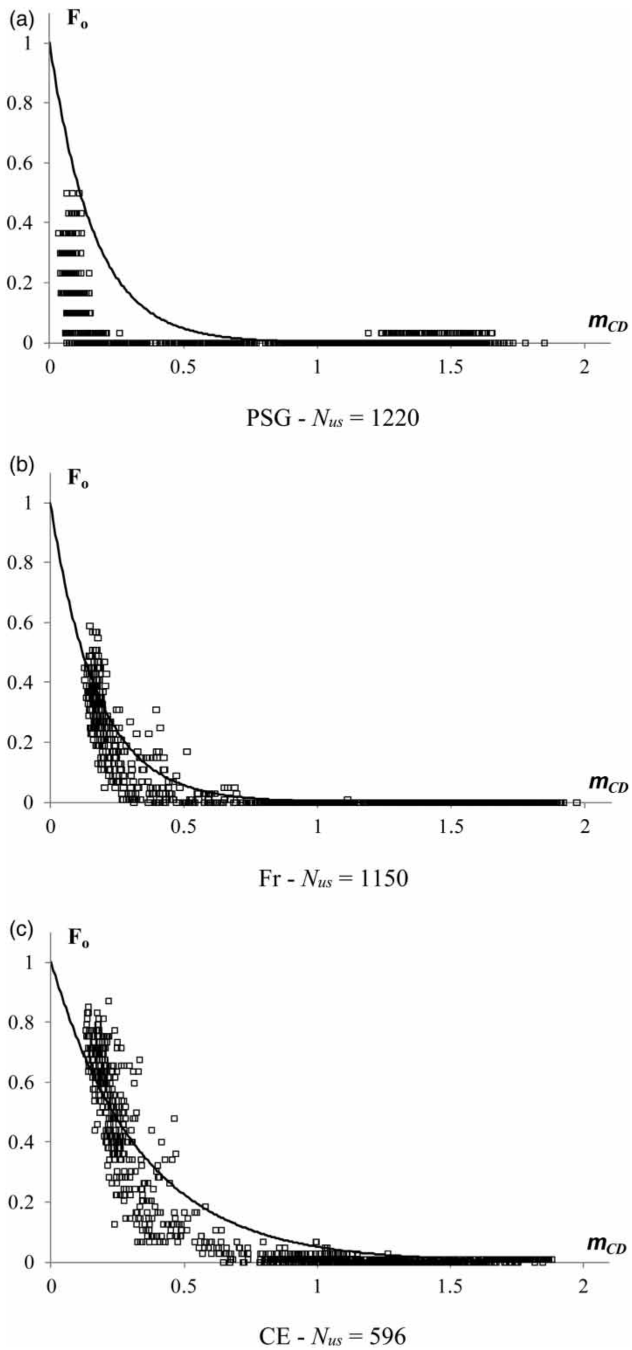

Figure 5 | Equation (5) and observed data of F_o versus μ_{C_D}.

It is worth noting that the condition $F_o = 1$ (which implies $\mu_{C_D} = 0$) represents a theoretical limit, because the event of null water request cannot be a certain event, even during the night time. Therefore, the Equation (5) becomes unfounded when the value of $\mu_{C_D}(t)$ leads to a probability of null water request close to 1. Therefore, from the observed data it is clear that a threshold value (μ_{C_m}) of μ_{C_D} should be defined in such a way as to limit inferiorly the Equation (5), hence Equation (5) is valid for $\mu_{C_D} \geq \mu_{C_m}$.

If it assumes that μ_{C_m} occurs at the same time of the maximum value $F_o^{\max}$ – condition however close to the reality – then the μ_{C_m} value can be obtained by equaling Equations (4) and (5):

$$\mu_{C_m} = -\frac{200}{N_{us}} \ln\left[1 - 0.25\left(\frac{N_{us}}{1000}\right)^{2.5}\right] \tag{6}$$

where Equation (6) is valid for $N_{us} = 200 \div 1{,}250$.

Equation (6) gives an estimation of the minimum admissible value of the water demand in relation to the user number. Hence, when μ_{C_D} is equal to μ_{C_m} (Equation (6)) $F_o(t)$ should be assumed equal to the maximum value obtained by means of Equation (4).

The application of Equations (4)–(6) jointly with the μ_{C_D} pattern allows us to estimate the daily trend of the probability $F_o(t)$, as it will be shown in the following sections.

The distributions for not null water demand, *F**(*t*)

The identification of the distributions suitable to model the not null water demand $F^*(t)$ was carried out by statistical inferences that have considered the data of the actual users (CE, Fr and PSG), eliding the null water requests from the experimental samples.

In this way it was observed that the normal (N) distribution is quite robust to represent the F^* for the whole day. Furthermore, it has been considered the possibility in representing the $F^*(t)$ also by means of the logistic (L) distribution that presents the following equation:

$$F_{C_D}[C_D(t)] = \left[exp\left(-\frac{\pi}{\sqrt{3}}\frac{C_D(t) - \mu_{C_D}(t)}{\sigma_{C_D}(t)}\right) + 1\right]^{-1} \tag{7}$$

The L distribution, in respect to the N distribution, has the advantage that the probability density function can be integrated, although it has the same trend of the Normal model (Swamee 2002; Ashkar & Aucoin 2012).

Both the distributions (N and L) need to be truncated being $C_D > 0$. In fact for low values of μ_{C_D} the probability of a negative value of C_D by Equation (7) is not negligible.

The choice of taking into account bi-parametric models has been guided by the need to find models that permit reliable estimation of few parameters and thus lend themselves to a practical application.

As an example, plots in Figure 6 report the comparison – for Fr $N_{us} = 1{,}150$ (Figure 6(a)) and PSG $N_{us} = 777$ (Figure 6(b)) users – between the N and L distributions and the observed cumulative frequencies at 5.00 a.m., the time at which the F_o probability is significant for both PSG and Fr users.

Table 2 reports the parameters of Figure 6 of data filtered with respect to the null request.

Tests of the MD

The comparison between the field data and the mixed distribution (Equation (3′)) have demonstrated the effectiveness of the proposed distribution in representing the flow demand during the whole day.

Indeed, Figure 7 for different daily time (six examples) shows that Equation (3′) fits well with the observed cumulative frequencies in different times of the day. The data of PSG

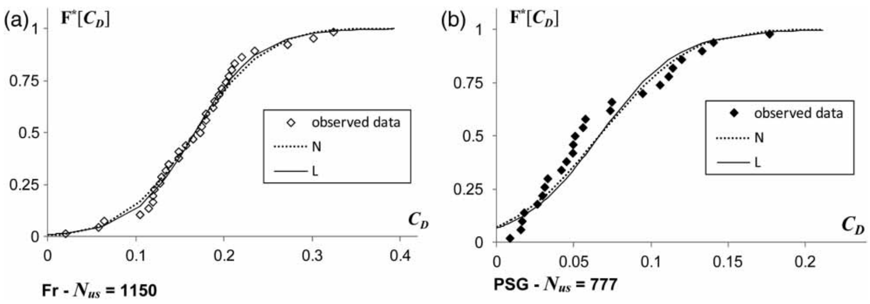

Figure 6 | Normal and logistic CDFs and observed cumulative frequencies for filtered data – $t = 5.00$ a.m.

Table 2 | Parameters of filtered data of Figure 6

	Fr (NL)	PSG (I)
$\mu(t = 5.00)$	0.167	0.067
$\sigma(t = 5.00)$	0.065	0.046

($N_{us} = 777$) and CE ($N_{us} = 596$) have been reported in Figure 7 as an example of the satisfying results that have been obtained also in reference to the other monitored users number. The less number of the experimental points in the CE plots depends on the degree of accuracy of the water meters used for Castelfranco Emilia monitoring system (volumetric metres).

In Figure 7 the observed cumulative frequencies are compared with the mixed distribution (continuous line), for which the not null water request was described by means of the L distribution (Equation (7)). The parameters of the distribution were estimated on the basis of experimental data and the relative values are reported in Table 3 (columns $F_o(t)$, $\mu_{C_D}(t)$ and $\sigma_{C_D}(t)$). The plots with dot lines

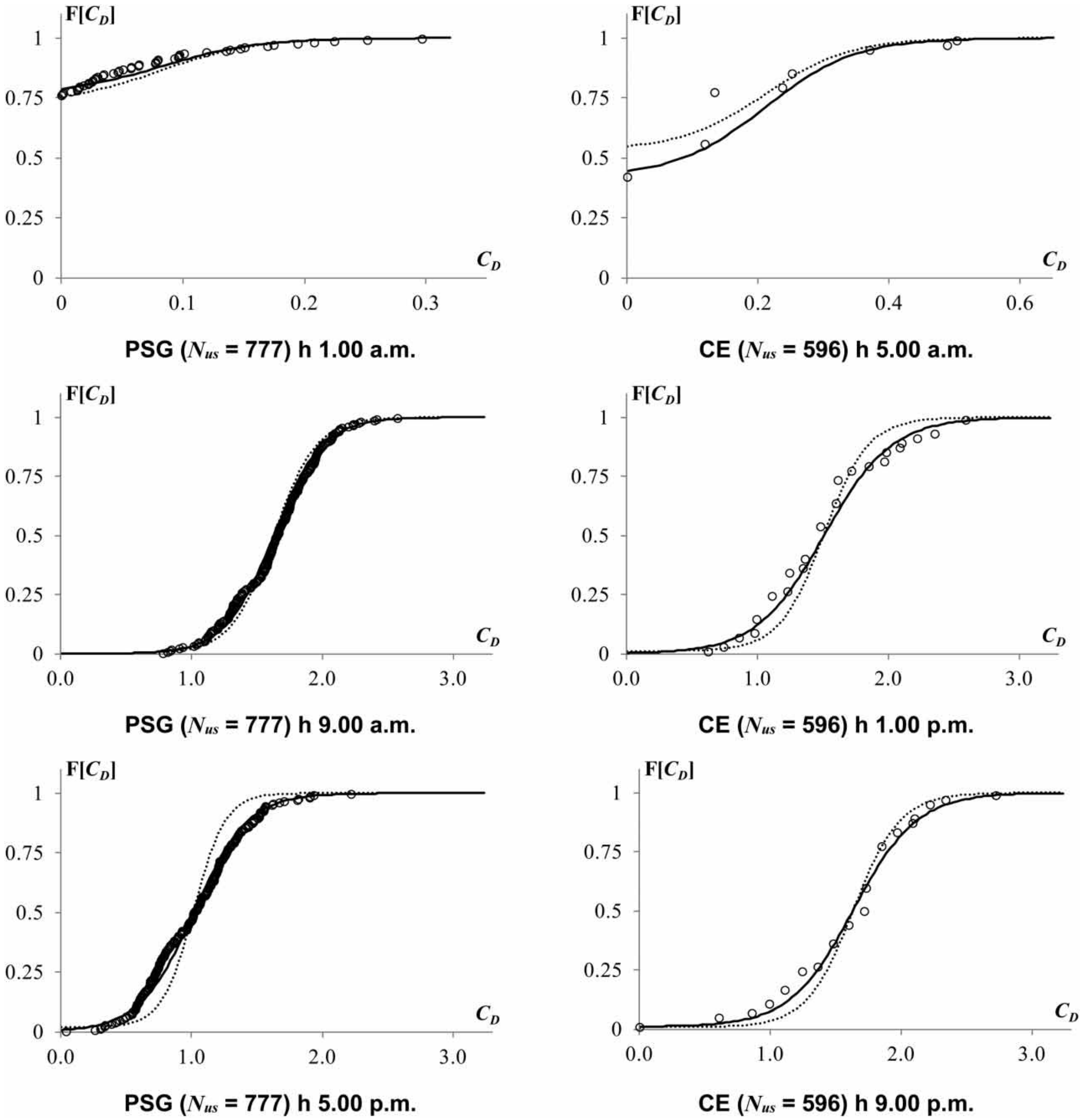

Figure 7 | CDFs of MD and observed cumulative distribution frequencies (O) for different daily time and for users PSG $N_{us} = 777$ and CE $N_{us} = 596$.

Table 3 | Parameters of the plots in Figure 7

Users	t [hh.mm]	$F_o(t)$ [%]	$\overline{F_o(t)}$ [%]	[–]	$\sigma_{C_D}(t)$ [–]	$\overline{\sigma_{C_D}}$ [–]
PSG $N_{us} = 777$	1.00 a.m.	76.2	72.8	0.080	0.072	0.070
PSG $N_{us} = 777$	9.00 a.m.	0.0	0.2	1.630	0.336	0.293
PSG $N_{us} = 777$	5.00 p.m.	0.0	1.9	1.016	0.372	0.214
CE $N_{us} = 596$	5.00 a.m.	69.9	53.3	0.211	0.123	0.117
CE $N_{us} = 596$	1.00 p.m.	0.0	1.1	1.510	0.467	0.307
CE $N_{us} = 596$	9.00 p.m.	1.0	0.8	1.637	0.436	0.323

in Figure 7 furthermore represent the MD CDFs, where Equation (5) gave the probability of null request (in Table 3, column $\overline{F_o(t)}$) and the standard deviations (in Table 3, column $\overline{\sigma_{C_D}}$) has been estimated by means of the relation (Equation (8)) described in the following section.

The MD is able to manage the spike of cumulative frequencies of the water demand that assumes remarkable values during the night (see in particular plots of Figure 7 for h1.00–5.00 am).

Finally, Figure 8 summarizes the results of the statistical inferences in order to analyze the effectiveness of the MD throughout the whole day.

In particular, the robustness of the proposed approach was checked by means the Kolmogorov–Smirnov test reported in Figure 8 as an example for Fr and CE, considering the N distribution to describe the $F^*(t)$ component of Equation (3′). The tests were done every half hour of the day, and Figure 8 shows the Kolmogorov–Smirnov parameter D, which exceeds the threshold-significant level = 5% – for only a few points.

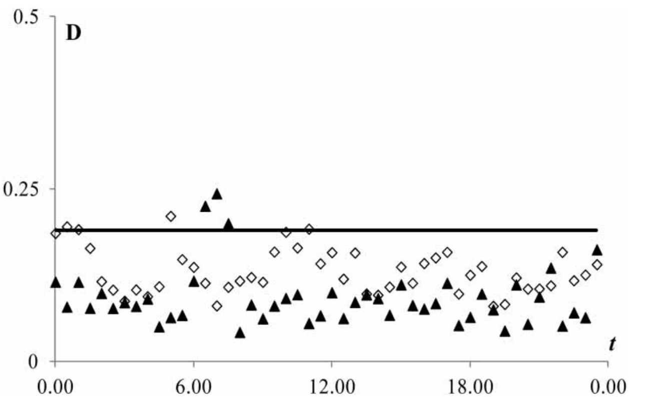

Figure 8 | Kolmogorov–Smirnov test for MD throughout the whole day – data of CE (◇) and Fr(▲).

ESTIMATION OF THE PARAMETERS OF THE MD

The use of bi-parametric models for representing the F^* component implies the need to define practical relations for estimating the CDF parameter, i.e. mean and standard deviation.

As described above, the mean value $\mu_{C_D}(t)$ represents the daily water demand trend and it is a mirror of the users habits. Therefore, its pattern traces the lifestyle of the inhabitants (i.e. the start and the end time of the prevailing productive activities of the town; the average duration of the lunch break, etc.). Moreover, the following relations help to obtain a reliable daily trend:

- the ratio between the daily integral of the $\mu_{C_D}(t)$ (summation for the time discretization) and the daily time is equal to 1;
- several equations in the technical literature allow us to estimate the peak phenomenon in relation to the user number (e.g. Molino et al. 1991; Tricarico et al. 2007; Martínez-Solano et al. 2008);
- the minimum peak of the water demand can be estimated by means of Equation (6).

Otherwise, the daily trend can be deduced by means of in-situ measurements in a few points of the WDS. For instance, the water companies often have available flow meters on the outlet pipe of the urban reservoirs, which produce useful time series.

The standard deviation parameter can be obtained by means of a practical relationship herein suggested which allows us to estimate the variation coefficient CV(t) in relation to the user number.

The experimental data of the monitored users proved that the CV value decreases with the μ_{C_D} (Figure 9), as reported in the following equation:

$$\mathrm{CV}(t) = 0.1 + \frac{6}{\left(\frac{1}{4}\mu_{C_D}(t)\cdot N_{us}\right)^{3/4}} \tag{8}$$

According to Equation (8) CV also decreases with N_{us}, in agreement with other experimental evidences (e.g. Moughton et al. 2006; Gato-Trinitade & Gan 2012).

It is worth noting that at the increase of the number of users and of the demand coefficient the asymptotic value of Equation (8) is CV = 0.1. This result is in line with other

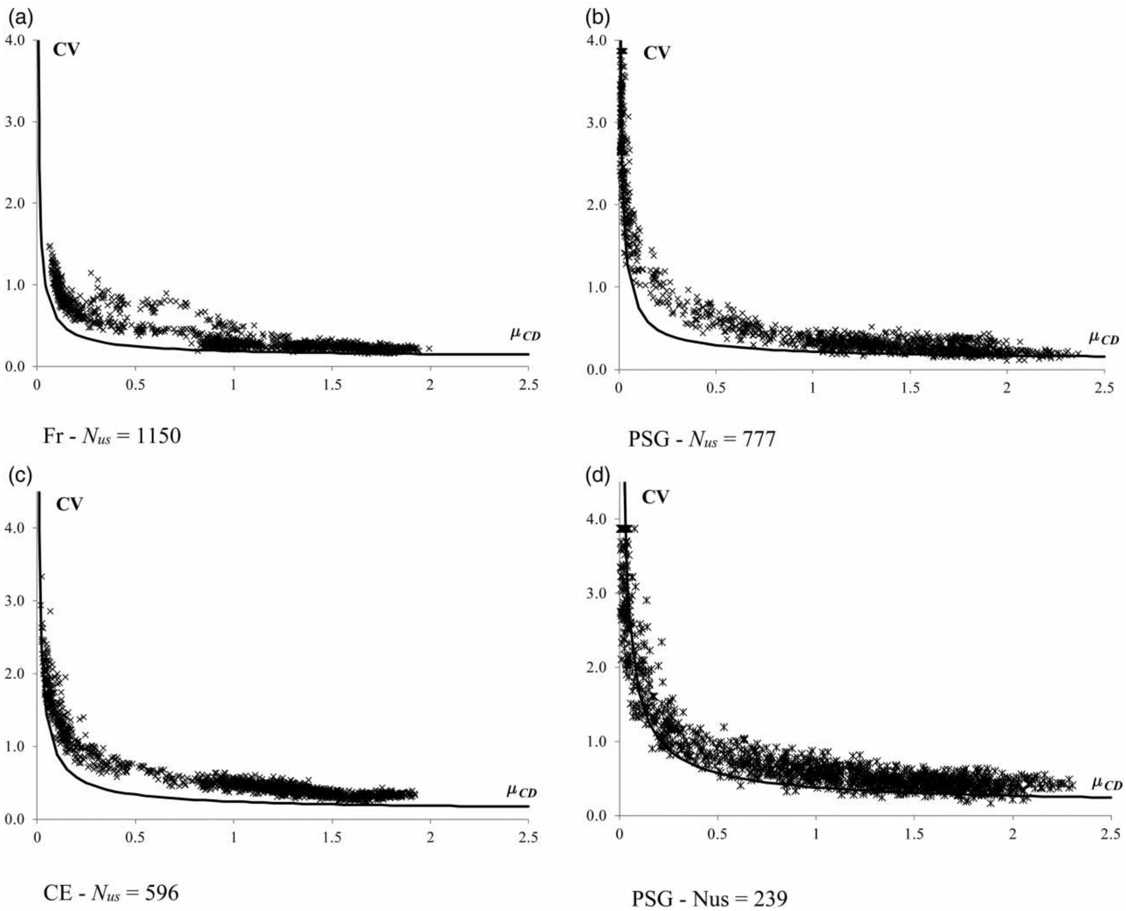

Figure 9 | CV of the demand coefficient, comparison between the observed data and Equation (8).

experimental studies relative to the peak phenomenon (Gargano & Pianese 2000; Tricarico *et al.* 2007).

The dot line and the continuous line of Figure 7 represent the CDFs of the MD that was obtained respectively by estimating the CV(t) by means of Equation (8) and on the basis of experimental data. Hence, the plots of Figure 7 show that Equation (8) implies admissible errors in the estimations of the coefficient of variation.

NUMERICAL EXAMPLE

The proposed approach, together with the suggested relationships and criteria to estimate the parameters, allows effortlessly the generation of the synthetic time series of the water demand applying the Monte Carlo method (Metropolis & Ulam 1949).

The residential water request of the different case studies examined has been taken into consideration for reporting a numerical example. Hence synthetic data of these users were generated by means of the proposed approach and the relative relationships to estimate the parameters.

Solely as an example, the numerical results obtained for Fr ($N_{us} = 1{,}150$) and CE ($N_{us} = 596$) have been herein reported.

Knowing the mean trend of the $\mu_{C_D}(t)$ during the day (Figure 1(b)–1(c)), by means of Equations (5) and (8) the trends of $F_o(t)$ and CV(t) were estimated for a time interval of 24 h (Figure 10) with a time interval $\Delta t = 1$ min. Moreover, the probability of the occurrence of the maximum null water demand was estimated by Equation (4) (the peaks of the plots in Figure 10(a)–10(c)).

The plots of Figure 10 show that F_o and CV increase with reducing number of users, in particular during the night.

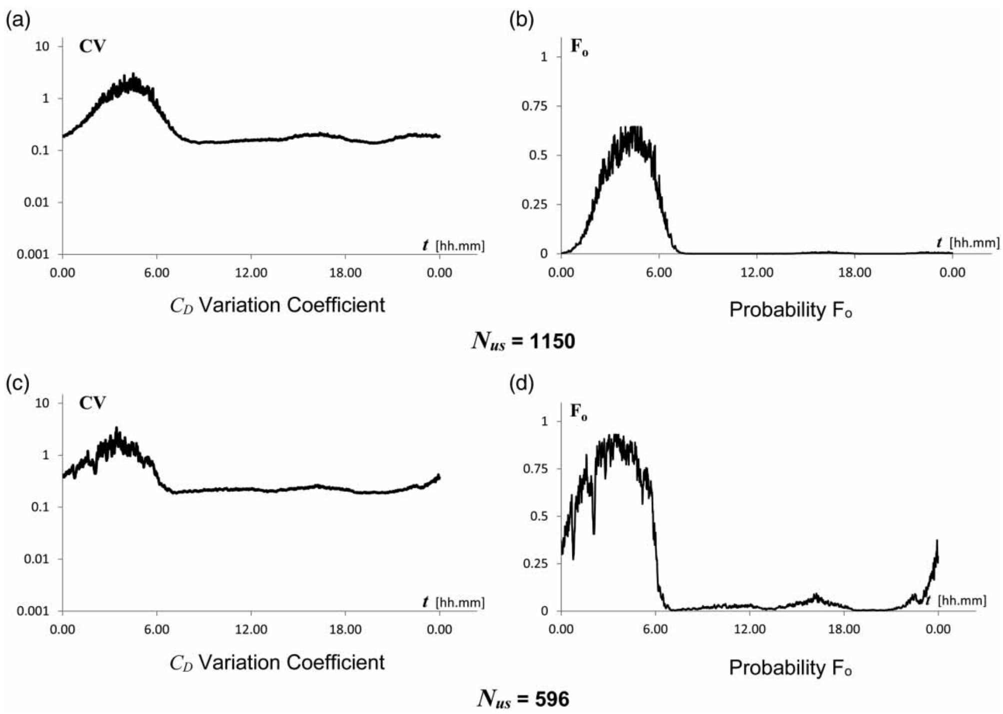

Figure 10 | Estimated daily trend of the variation coefficient and F_o probability for $N_{us} = 1{,}150$ and 596.

Knowledge of the daily trends of the parameters (Figure 1(b) and Figure 10) allowed us to generate synthetic daily time series by means of Equations (3′) and (7). Indeed, by knowing the parameters of the MD, the demand flow for a generic time step is obtained by means of the equation:

$$\begin{aligned} Q(t) &= \mu_Q C_D(t) \\ &= \mu_Q \left[1 - \frac{\sqrt{3}}{\pi} \ln \frac{1 - F(t)}{F(t) - F_o(t)} \mathrm{CV}(t)\right] \mu_{C_D}(t) \qquad (9) \\ &\text{for } F(t) > F_o(t) \end{aligned}$$

where F(t) is the generated probability for time t.

Figure 11 represents the comparison between the observed data and the generated data (Equation (9)) of a standard day. The comparison of the plots shows the effectiveness of the synthetic data in modeling the requested flows during the whole day.

The daily time series of the observed 50 days of water demand are compared with the simulated water demand of the same number of days in Figure 12. In detail, the plots of Figure 12 show the minimum, the maximum and the mean flow request for each minute of the day.

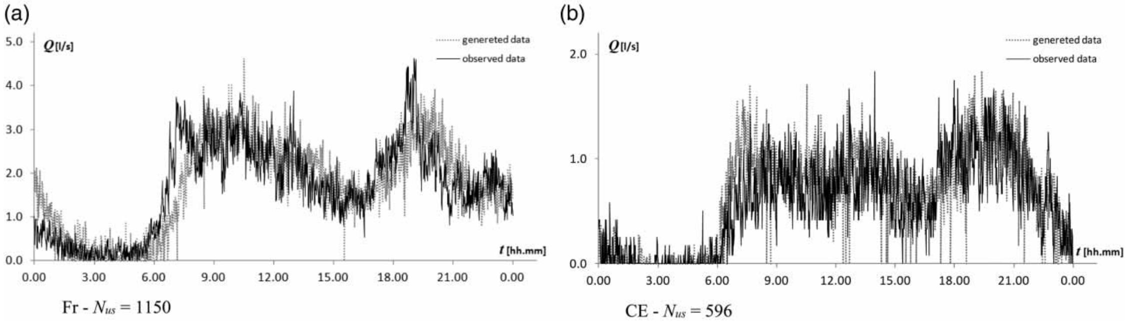

Figure 11 | Comparison between observed and generated demanded flows for a single day.

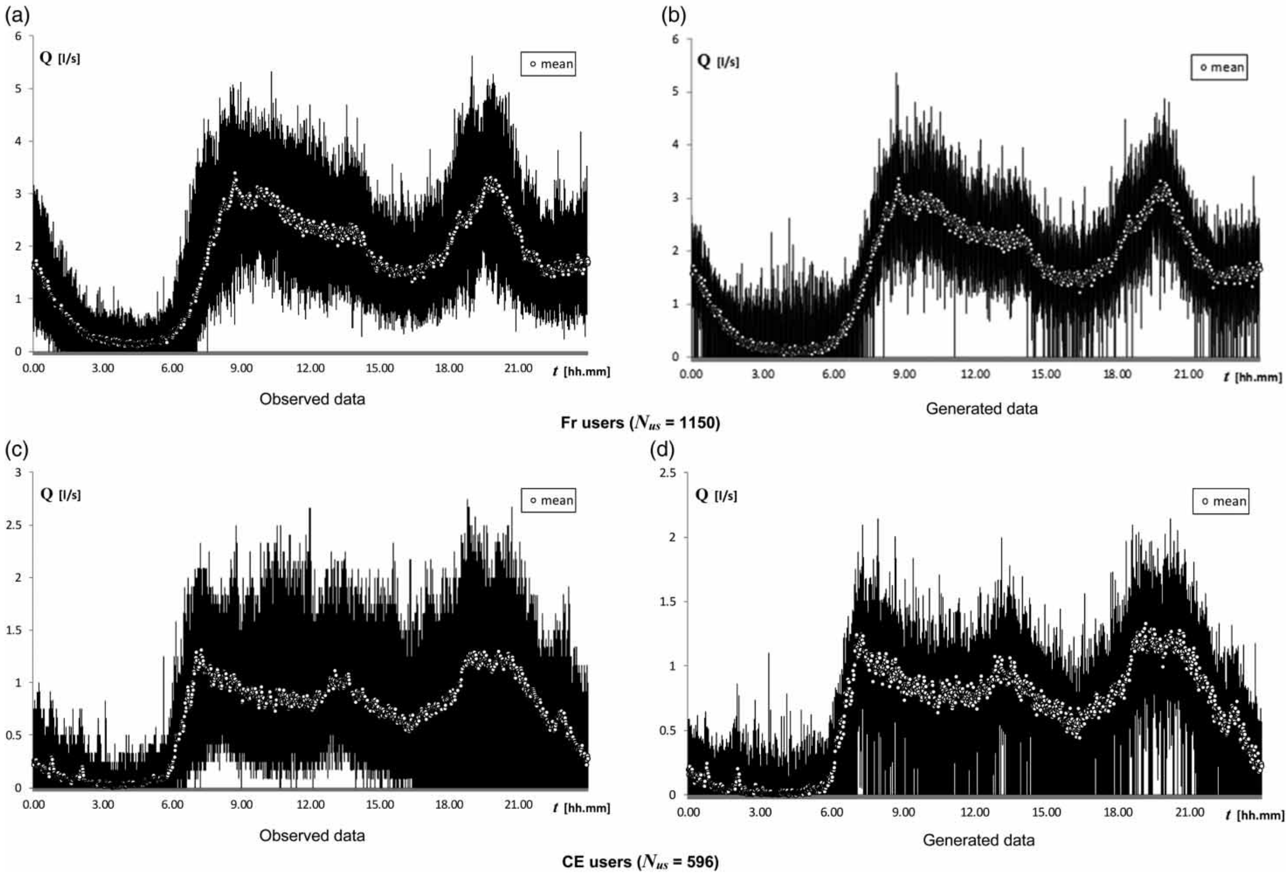

Figure 12 | Demanded flow (minimum, maximum and mean) of the observed (a) and (c) and generated (b) and (d) data.

CONCLUSION

A concrete contribution towards the generation of time series of the residential water demand during the day has been described in this paper. The fundamental importance of modelling the water demand accurately is widely recognized as an undeniable prerequisite for effective WDS analysis. The probabilistic approaches existing on the topic are not reliable, or they are too complex to be implemented in practice by the decision makers of the water company.

This is often due to the absence of a suitable experimental activity that allows to both test the effectiveness of the probabilistic models proposed, and provide indications for the estimation of the distribution parameters.

Hence, by means of statistical inferences on real residential demand data an independent stochastic approach and relative parameters able to represent the residential water requirements for 200–1,250 users have been herein suggested. This range represents the limits of the validity for the proposed stochastic model. However, the above mentioned validity extremes do not limit in a substantial way the effectiveness of the approach, which allows us to consider the whole water demand as a continuous and positive random variable.

With the aim of applying the obtained results to a wide range of cases, variegated sample data were examined allowing us also to study the incidence of the different life styles on water consumption. However, the obtained results are in line with other experimental analysis on water demand of real residential users.

The random component of the stochastic approach is modeled by means of a novel distribution (mixed distribution) that is of effectiveness in describing the water demand for the entire day. The MD, obtained by merging two distributions, describes the residential water demand, regardless of its entity. In this way the proposed probabilistic model is able to manage at the same time a discrete random variable – necessary to take into account the null request during the night hours – and a continuous positive random variable, for the not null demanded flows.

In addition, some relationships are proposed in order to obtain robust estimations of the parameters of the mixed distribution, once the number of supplied users and the daily mean trend are known.

These equations, and the proposed stochastic approach, permit us to generate, by means of the Monte Carlo simulator, reliable sample data of the demanded water flows in a simple way.

The tests showed that the assumptions do not limit the effectiveness of the proposed stochastic approach.

ACKNOWLEDGEMENTS

The authors thank Prof. Marco Franchini and Dr Stefano Alvisi and Dr Mirjam Blokker for providing water demand data for Castelfranco Emilia and Franeker monitoring systems. The writers would also like to acknowledge Prof. Steven G. Buchberger for his fruitful suggestions.

REFERENCES

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** (7), 1779–1792.

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

Ashkar, F. & Aucoin, F. 2012 Discriminating between the lognormal and the log-logistic distributions for hydrological frequency analysis. *J. Hydrol. Eng.* **17** (1), 160–167.

Babayan, A. V., Kapelan, Z., Savic, D. A. & Walters, G. A. 2006 Comparison of two methods for the stochastic least cost design of water distribution systems. *Eng. Optimiz.* **38** (3), 281–297.

Bao, Y. & Mays, L. W. 1990 Model for water distribution system reliability. *J. Hydraul. Eng.* **116** (9), 1119–1137.

Blokker, E. J. M., Vreeburg, J. H. G. & Vogelaar, A. J. 2006 Combining the probabilistic demand model SIMDEUM with a network model. In: *Proceedings of the 8th Annual International Symposium on Water Distribution Systems Analysis*. Cincinnati, OH, USA [CD-Rom].

Blokker, E. J. M., Vreeburg, J. H. G. & van Dijk, J. C. 2010 Simulating residential water demand with a stochastic end-use model. *J. Water Resour. Plann. Manage.* **136** (1), 19–26.

Buchberger, S. G. & Nadimpalli, G. 2004 Leak estimation in water distribution systems by statistical analysis of flow readings. *J. Water Resour. Plann. Manage.* **130** (4), 321–329.

Buchberger, S. G. & Wells, G. J. 1996 Intensity, duration, and frequency of residential water demands. *J. Water Resour. Plann. Manage.* **122** (1), 11–19.

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

Chung, G., Lansey, K. & Bayraksin, G. 2009 Reliable water supply design under uncertainty. *Env. Model. Softw.* **24** (4), 449–462.

Ferrante, M., Brunone, B., Meniconi, S., Karney, B. W. & Massari, C. 2014 Leak size, detectability and test conditions in pressurized pipe systems: how leak size and system conditions affect the effectiveness of leak detection techniques. *Water Res. Manage.* **28** (13), 4583–4598.

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

Gargano, R. & Pianese, D. 2000 Reliability as a tool for hydraulic network planning. *J. Hydraul. Eng.* **126** (5), 354–364.

Gargano, R., Tricarico, C. & de Marinis, G. 2012 Residential water demand – daily trends. In: *Proceedings of the Annual International Symposium on Water Distribution Systems Analysis 2010*, Tucson, Arizona, pp. 1314–1323. DOI: 10.1061/41203(425)118.

Gargano, R., Tricarico, C., Buchberger, S. G., Del Giudice, G. & Di Palma, F. 2014 The mixed model for the residential flow demand of a small number of users. *Procedia Eng.* **89** (C), 975–981.

Gargano, R., Di Palma, F., de Marinis, G., Granata, F. & Greco, R. 2016 A stochastic approach for the water demand of residential end users. *Urban Water J.* **13** (6), 569–582.

Gato-Trinidad, S. & Gan, K. 2012 Characterizing maximum residential water demand. *Urban Water* – WIT Transactions on The Built Environment **122**, 15–24.

Giustolisi, O., Savic, D. A. & Kapelan, Z. 2008 Pressure-driven demand and leakage simulation for water distribution networks. *J. Hydraul. Eng.* **134** (5), 626–635.

Greco, M. & Del Giudice, G. 1999 New approach to water distribution network calibration. *J. Hydraul. Eng.* **125** (8), 849–854.

Jung, D., Kang, D., Kim, J. H. & Lansey, K. 2014 Robustness-based design of water distribution systems. *J. Water Resour. Plann. Manage.* **140** (11), 14.

Kapelan, Z., Dragan, S. & Walters, G. A. 2005 Multiobjective design of water distribution systems under uncertainty. *Water Resour. Res.* **41** (11), 1–15.

Liserra, T., Maglionico, M., Ciriello, V. & Di Federico, V. 2014 Evaluation of reliabilityindicatores for WDNs with demand-driven and pressure-driven models. *Water Resour. Manage.* **28** (5), 1201–1217.

Magini, R., Pallavicini, I. & Guercio, R. 2008 Spatial and temporal scaling properties of water demand. *J. Water Resour. Plann. Manage.* **134** (3), 276–284.

Martínez-Solano, J., Iglesias-Rey, P., Pérez-García, R. & López-Jiménez, P. 2008 Hydraulic analysis of peak demand in looped water distribution networks. *J. Water Resour. Plann. Manage.* **134** (6), 504–510.

Mays, L. W. 1999 *Water Distribution Systems Handbook*. McGraw-Hill Handbooks, New York, NY.

Metropolis, N. & Ulam, S. 1949 The Monte Carlo method. *J. Am. Stat. Assoc.* **44**, 335–341.

Molino, B., Rasulo, G. & Taglialatela, L. 1991 Peak coefficients of household potable water supply. *Water Resour. Manage.* **4** (4), 283–291.

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

Nel, D. & Haarhoff, J. 1996 Sizing municipal water storage tanks with Monte Carlo simulation. *J. Water Supply: Res. Technol. AQUA* **45** (4), 203–212.

Preis, A. & Ostfeld, A. 2006 Contamination source identification in water systems: a hybrid model trees-linear programming scheme. *J. Water Resour. Plann. Manage.* **132** (4), 263–273.

Rossi, F., Fiorentino, M. & Versace, P. 1984 Two-component extreme value distribution for flood frequency analysis. *Water Resour. Res.* **20** (7), 847–856.

Swamee, P. K. 2002 Near lognormal distribution. *J. Hydrol. Eng.* **7** (6), 441–444.

Todini, E. 2003 A more realistic approach to the 'extended period simulation' of water distribution networks. In: *Advances in Water Supply Management* (C. Maksimović, D. Butler & F. A. Memon, eds). A.A. Balkema Publishers, Abingdon, pp. 173–183.

Tricarico, C., Gargano, R., Kapelan, Z., Savić, D. A. & de Marinis, G. 2006 Economic level of reliability for the rehabilitation of hydraulic networks. *J. Civ. Eng. Env. Sys.* **23** (3), 191–207.

Tricarico, C., de Marinis, G., Gargano, R. & Leopardi, A. 2007 Peak residential water demand. *Water Manage. J.* **160** (WM2), 115–121.

Tricarico, C., Morley, M. S., Gargano, R., Kapelan, Z., de Marinis, G. & Savić, D. A. 2014 The Influence of the existing network layout on water distribution system rehabilitation analysis. *J. Hydroinformatics.* **16** (3), 1375–1389.

Van Zyl, J. E., Piller, O. & Legat, Y. 2008 Sizing municipal storage tanks based on reliability criteria. *J. Water Resour. Plann. Manage.* **134** (6), 548–555.

Vítkovský, J., Stephens, M., Bergant, A., Simpson, A. R. & Lambert, M. F. 2006 Numerical error in weighting function-based unsteady friction models for pipe transients. *J. Hydraul. Eng.* **132** (7), 709–721.

Walski, T. 1983 Technique for calibrating network models. *J. Water Resour. Plann. Manage.* **109** (4), 360–372.

First received 14 October 2015; accepted in revised form 2 June 2016. Available online 20 June 2016