water

MDPI

Article

Spatial Aggregation Effect on Water Demand Peak Factor

Giuseppe Del Giudice [1], Cristiana Di Cristo [1] and Roberta Padulano [2,*]

[1] Department of Civil, Environmental and Architectural Engineering, Università degli Studi di Napoli Federico II, Via Claudio 21, 80125 Naples, Italy; delgiudi@unina.it (G.D.G.); cristiana.dicristo@unina.it (C.D.C.)

[2] Regional Models and geo-Hydrological Impacts, Centro Euro-Mediterraneo sui Cambiamenti Climatici, Via Thomas Alva Edison, 81100 Caserta (CE), Italy

* Correspondence: roberta.padulano@cmcc.it

Received: 28 May 2020; Accepted: 14 July 2020; Published: 16 July 2020

Abstract: A methodological framework for the estimation of the expected value of hourly peak water demand factor and its dependence on the spatial aggregation level is presented. The proposed methodology is based on the analysis of volumetric water meter measurements with a 1-h time aggregation, preferred by water companies for monitoring purposes. Using a peculiar sampling design, both a theoretical and an empirical estimation of the expected value of the peak factor and of the related standard error (confidence bands) are obtained as a function of the number of aggregated households (or equivalently of the number of users). The proposed methodology accounts for the cross-correlation among consumption time series describing local water demand behaviours. The effects of considering a finite population is also discussed. The framework is tested on a pilot District Metering Area with more than 1000 households equipped with a telemetry system with 1-h time aggregation. Results show that the peak factor can be expressed as a power function tending to an asymptotic value greater than one for the increasing number of aggregated households. The obtained peak values, compared with several literature studies, provide useful indications for the design and management of secondary branched pipes of water distribution systems.

Keywords: cross-correlation; data spatial aggregation; finite population effect; metering; sample mean; sampling design; standard error; stochastic analysis; water demand peak factor; water distribution networks

1. Introduction

In the last decades, the understanding and prediction of water consumption have become a focal point of EU policies and directives, with the general aim of supporting safe access to drinking water and basic sanitation services to the people. In this context, the estimation of water demand in a distribution system is a key issue when applying management strategies to reduce costs and preserve the resource [1].

The water demand of a single user exhibits a random and pulsing behaviour; however, the aggregation of a large number of consumers is able to highlight trends, seasonal cycles, and the possible existence of peaks. Such quantities usually have different values and features according to the scales of the measured or aggregated data (hourly, daily, weekly, monthly, seasonally, yearly). The estimate of peak values is crucial to design drinking water distribution networks, in order to obtain reliable systems, able to provide a good level of service in terms of demands and pressures [2]. The knowledge of water consumptions and of the relative peak values is also required in many

applications where the simulation of the system functioning is needed, whose results are strongly affected by demand uncertainty [3–6].

The estimation of hourly or sub-hourly peak demand due to residential uses has been widely studied adopting different methods and techniques. Top-Down Deterministic Approaches (TDAs) provide empirical relationships based on the number of users for the estimation of the hourly or sub-hourly demand peak factor, defined as the ratio between the maximum and mean flow. TDAs usually focus on the whole network, analysing the water consumption of the total served population. The first relationships [7,8] estimated the dependence on the population of the instantaneous peak factor in sewer systems. Some research found that the hourly peak factor can be considered constant when the population is lower than a fixed threshold, while it decreases when the users exceed the threshold value [9]. Those empirical equations for wastewater peak factors tend to be restricted to a minimum population of one thousand and a maximum population of one million. More recently, a formula was proposed for characterizing the mean value of the peak water demand for small towns through statistical inferences on a large database [10], providing a lower estimate compared to the Babbitt's formula [7]. Moreover, investigated the effect of the data time sampling interval on the evaluation of the peak factor was investigated [10]. The dependence of peak factors on the number of users was also the subject of investigations [11], to provide empirical relationships for the estimation of the parameters of the Gumbel probability distribution, able to represent the stochastic behaviour of peak water demand.

Bottom-Up Approaches (BUAs) try to reconstruct nodal demands generating a large number of synthetic realizations of individual users' consumptions described by a stochastic variable. It has been proved that at the fine temporal scale the nodal demand takes the shape of a pulse [12]. In this context, temporal trends of instantaneous nodal consumptions are reconstructed aggregating demands produced by stochastic pulse generation methods, such as the Poisson Rectangular Pulse (PRP) (e.g., [13–18]) or the cluster Neyman-Scott Rectangular Pulse (NSRP) (e.g., [4,19,20]). A single pulse is associated with each demand event, whose arrival time is described through a Poisson process. In the proposed methods, pulse duration and intensity have been generated assigning different specific probability distributions: Normal [15], exponential [4,19–21], log-normal [12] for the duration; exponential [4,15,19,20], Weibull [21], log-normal [12] for the intensity. More recently, a method was proposed to account for the correlation between pulse duration and intensity, which led to some improvement in pulse consistency [22].

To apply these methods, model parameters need to be assigned. The parameters' values can be obtained using measured pulse features obtained by monitoring consumptions with an ultra-high time resolution [12,17,18] or reproducing statistical properties of aggregated consumptions, when they are known at a higher temporal step (1 min or larger) [19,23].

In this context, Blokker E.J.M., et. al., [24,25] proposed the SIMDEUM model for the reconstruction of water consumptions starting from the micro-components of water demand. The PRP model was used, but different distributions were adopted to generate the pulses produced by the different household fixtures and users. Then, for its parametrization, knowledge is required about the occupants' habits and about the end uses of the fixtures obtained from a survey of the considered households. This can be done, for example, by analysing the water end-users that drive peak daily demand and examining their diurnal demand patterns using data obtained from high resolution smart meters [26]. The PRP and SIMDEUM models have similar performances [27], with the former prevailing at the single household scale and the latter prevailing in case of multiple households. In all cases, BUAs require a significant computational effort and, for their parametrization, a detailed knowledge of the consumptions at a small spatial scale is required.

More recently, proposed a probabilistic approach was proposed for a reliable estimation of the maximum residential water demand represented by a single variable [28], showing the reliability of the log-normal and Gumbel distributions in representing peak water demand during the day. The authors

suggested practical equations for the estimation of the expected value and coefficient of variation of the daily peak factor and investigated time scaling effects.

Many studies investigated the influence of the acquisition time step in water demand modelling [23,29,30]. In this context, analysing water consumption data recorded at time intervals from 5 min to one hour, a significant effect of the sampling time step was observed [31] and new equations were derived for the evaluation of the peak value. A comparison of the instantaneous estimate of the maximum demand obtained at a 1 s time step through a BUA with the one computed from hourly average estimates using a TDA was also performed [32]. As expected, results showed that the latter gives small demand values, especially at small spatial aggregation scales, while at increasing aggregation levels the difference decreases, because the random fluctuations tend to be smoothed with consequently smaller peak values.

The effect of spatial aggregation is less studied. First attempts investigated the effect of both time step and spatial aggregation on the cross-correlation between nodal demands, however limiting the analysis to a group of five and ten houses [33]. Results highlighted an increase in correlation for increasing spatial aggregation, while a decrease of the standard deviation was observed.

In the last decades, the rising development of smart meters systems for household water consumption monitoring provided new modelling perspectives [34,35]. Smart metering can provide data recordings at different levels of accuracy, from 1 s to hours, depending on the characteristic of the system and on the objective of the investigations [36,37]. With a reasonable economic impact, water companies started with the installation of smart water meters, usually placed in a large number of households and collecting hourly measurements. In fact, water companies are mainly interested in controlling and understanding aggregated consumptions in order to make decisions on pricing strategies, on future interventions, and on consumption reduction. Some approaches have been recently proposed for modelling demand patterns using measurements at large time steps [38–41].

A first objective of the present study is to understand how water companies can obtain information about the estimate of the peak factor, starting from measurements realized for different purposes on large networks with an hourly temporal scale. The paper presents a methodology for performing a statistical analysis of hourly data in order to analyse the behaviour of the hourly peak demand values as a function of spatial data aggregation using a high number of measurements. The considered test-case is a large-size District Metering Area of the water distribution network of Naples (Italy) equipped with a smart metering system, which provided water demand measurements performed with a one hour time aggregation on more than 1000 households for one year [39,40]. The main novelty of the study lies in the complex sampling design adopted, which allows treating hourly peak factors as stochastic variables for each fixed number of aggregated meters, accounting for possible cross-correlation and finite population effects. In this way, the main statistics (including expected values and variability) of the peak factor can be obtained as a function of the size of the considered group of users, and compared with other literature indications adequately scaled to account for different time scales. The main goal of the research is to provide the operators with a procedure for understanding the reliability of the network in terms of demand and pressure at different levels of users' aggregation using available data. This information is particularly useful to analyse the behaviour of old water networks, where the operating conditions may differ from the ones considered at the design stage, or to design future measures to improve the system management, such as the creation of District Meter Areas.

The paper is structured as follows. Section 2 describes the District Metering Area under investigation and the collected measurements, the main objectives and features, and the methodological framework of the analysis. Section 3 reports the outcomes, while Section 4 provides a discussion about the applicability of results. Finally, significant conclusions are drawn in Section 5.

2. Materials and Methods

2.1. The District Metering Area

The area under investigation is "North Soccavo" (Figure 1), which is one of the administrative neighbourhoods of the Municipality of Naples (Italy), counting about 20,000 inhabitants. The area was selected by the local water company as the pilot case for the implementation of a District Metering Area. The reason for this particular choice lies in the observation that this area is connected by a single branch to the water network of the City, which makes it particularly prone to be distrectualized. In recent years, the local water company connected all the existing water meters with a telemetry system, allowing for the automatic radio collection of consumed water volumes at a 1-h time step. The connection involved water meters related to both single and multiple households, as well as commercial activities and public buildings.

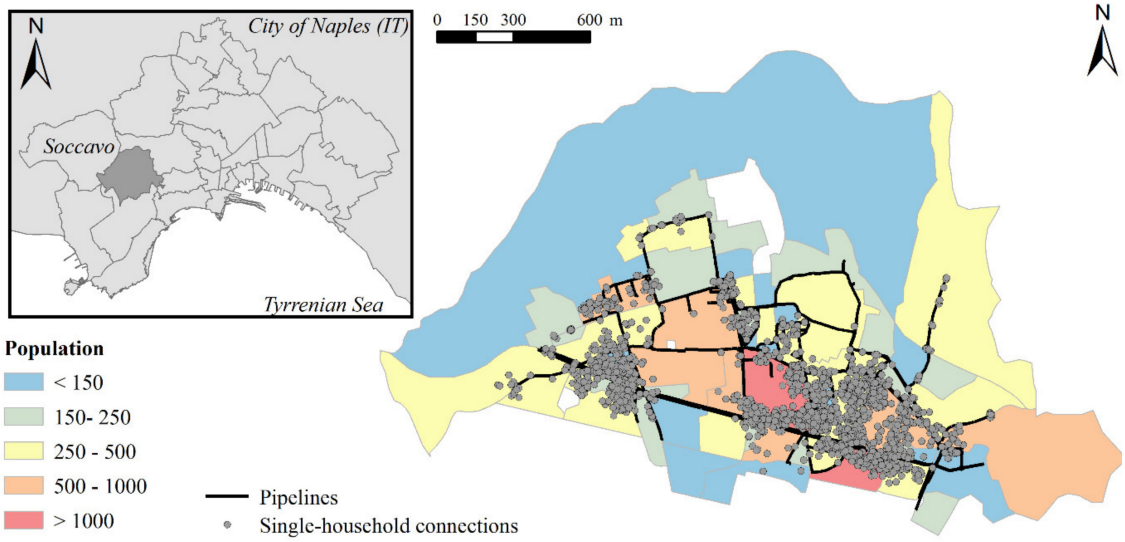

Figure 1. Pilot area: Water distribution network, water meters, and census particles.

Data referring to the year 2016 have recently been the subject of a multi-purpose research focusing on the prediction of water demand patterns, useful for the local water company to define management and leakage detection strategies [39,40,42]. The findings, based on hourly consumption data collected from 1 January to 31 December 2016, highlighted the following issues:

1. The neighbourhood is mainly residential. Of the total $K = 4253$ water meters, about 86% serve flats, apartments, and other inhabited buildings, whereas 14% have a non-residential purpose (commercial activities, public offices, and schools).
2. Not all the consumption time series collected by the telemetry system were suitable for the analyses. An anomaly/outlier detection procedure was applied, based on the use of the Completeness-Continuity Triangle (CCT) and on the application of the MAD criterion at different time aggregation levels. Such a strategy enabled the identification and subsequent removal of unreliable hourly data or the entire time series having a large number of outliers and/or missing values, as shown in [42]. Focusing on residential water meters serving single households, a number of 1162 passed the proposed anomaly/outlier detection procedure. Those data are used in the present research.
3. Focusing on the connections serving single households, a reduced number of significant patterns showing the annual cycle of water demand was detected [39]. Specifically, in the pilot area five patterns were identified representing different clusters of consumption behaviours. Figure 2 shows that those patterns are significantly similar at large aggregation levels (e.g., monthly),

only differing because of the consumption in August. This occurs because most of the people in Italy spend their holidays in August, and this produces a decrease in consumptions. However, the trends are different depending on the number of vacation days that is, in turn, proportional to the income level; a cluster without reduction, corresponding to users taking no summer holidays, can also be observed. Such a behaviour is expected to repeat cyclically every year. Removing the August consumption, no further seasonal cycle can be observed, and the five patterns in Figure 2 show no significant differences in the remaining months. As a consequence, if August data is discarded, daily discharges can be considered a random variable with no deterministic dependence. As far as the daily cycle is concerned, three different non-dimensional patterns were identified corresponding to Sundays, Saturdays, and Mondays–Fridays, respectively [40].

4. For residential connections, scaling laws were proposed [40] providing the mean hourly discharges and related standard deviations as a function of the number of aggregated households. The regression parameters depend on the characteristics of the specific dataset in terms of single-user behaviour and cross-correlation structure.

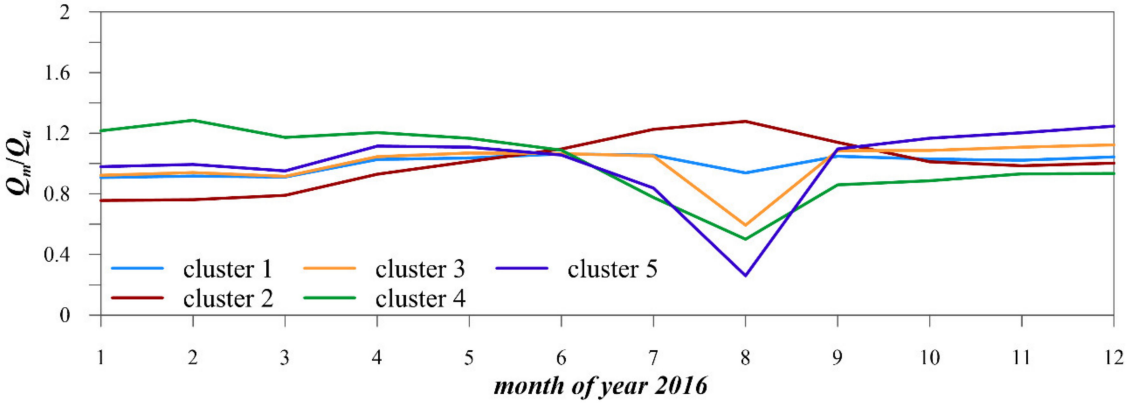

Figure 2. Average nondimensional patterns for the five main different clusters of consumption behaviours observed in the pilot area (Q_m and Q_a are the monthly and annual discharges, respectively).

The scaling laws proposed for the pilot area [40] are a function of the number of aggregated households, instead of the number of consumers, because the information about the number of people "hiding" behind each water meter was known only for a few cases. In the present paper, however, this information was derived by intersecting the spatial distribution of residential water meters (Figure 1), with the number of inhabitants at the census scale. As a result, the average number of users per connection ranges from 2.8 to 3. This uncertainty is caused by the delay between the date of the census survey (2011) and that of data collection (2016), and to a small number of water meters that still missed their connection with the telemetry system by the end of 2016.

2.2. Rationale and Structure of the Analysis

In the present paper the hourly water consumption database collected within the pilot area is used to obtain a comprehensive sample of hourly peak factors. Such a database can be potentially used to draw significant information allowing for the prediction of fundamental statistics of hourly peak demand such as central values, variability, and probability distribution. This research particularly focuses on the first issue, namely the sample mean of peak factors and related statistics.

As mentioned in the previous sections, peak factors in water networks can be deeply influenced by the amount and behaviour of consumers. Any statistical analysis should comply with the fact that the peak factors' values and the related statistics could be affected by the number and quality of the aggregated time series. For instance, if the network serves a small number of users, there is a large possibility that those consumers will highlight similar behaviours, resulting in higher peak factor values. On the contrary, if the network serves a large number of consumers, different behaviours are

expected and this translates in a global water demand more homogeneously distributed within the day, with smaller values of peak factors.

The peak factors evaluation in water networks usually consists of understanding how peak factors change under a progressive aggregation of the users, namely in finding a mathematical or statistical dependence of peak factors on the number of users N_u. Synchronicity of consumption behaviours is usually accounted for by means of the cross-correlation among consumption time series. Those considerations imply that, when N_u is small, a dependence can be found not only on *how many* but also on *which* time series is going to be aggregated. In other words, results may deeply vary according to the specific performed selection of consumers. On the contrary, when N_u is large, results are expected not to be significantly altered whichever time series is selected.

To overcome this issue and to investigate the statistical structure of peak factors in a way that is reliable, rigorous, and robust, the following sampling design is proposed. A discretized number N of households is set and, for each of them, the N time series (each corresponding to a water meter) with size D (corresponding to the considered monitored days) are extracted from the consumption database of the pilot area and aggregated. For each N, the operation is repeated M times, allowing the same water meters to be extracted in different samples, whereas, for each sample, extraction is performed without replacement. In this way, N artificial populations are obtained (one for each aggregation level) and M representative samples with size D are available. Finally, for each N, the main focus concerns the analysis of the following quantities assumed as the most important when using the concept of peak factors for the design or verification of water networks:

- Expected value of the sample mean of hourly peak factors;
- Standard error of the sample mean of hourly peak factors.

To correctly address the above-mentioned items, the usual sampling theory (e.g., [43]) cannot be adopted straightforwardly. The first reason lies in the observation that each random sample consists of a time series made up of a number D of independent realizations of the variable of interest (hourly peak factor), but there could be a non-negligible cross-correlation among the M samples that has to be taken into account. In this perspective, literature provides suggestions about including cross-correlation in the analyses [44].

The second reason is that the effect of a finite population must be taken into account. In this perspective, literature suggests that the classic sampling theory should be adopted when the population fraction ψ (namely the ratio of the amount of extracted data to the maximum number of available data, or, in other words, the ratio of sample size to finite population size) is small [45]. Indeed, in this condition, sample sizes comparable with the population size provide unnaturally small variabilities, since different samples will contain the same elements when $\psi \to 1$, with a progressive degeneration of the variance [45]. In turn, this could result in the need for very large and expensive databases to investigate large aggregation levels. For large ψ values, in case of sampling without replacement, suitable correction factors should be applied when estimating standard errors from the population variance, whereas the effect of a finite population on central values is usually considered negligible [45]. Especially concerning the variance of sample means, a correction factor, usually referred to as the Finite Population Correction Factor (*FPCF*) [45,46], a function of the population fraction, should be used when relating this quantity to the population variance. In the present research, the investigated population is characterized by two different dimensions, namely the number of monitored days D and the number of aggregated households N. For the adopted sampling design, D is the sample size, directly affecting computations, but the scientific interest mainly lies in understanding the effect of N, which, in turn, acts as a hidden variable with no explicit mathematical effect.

2.2.1. Parameter Definition

Let $q_{h,i}(d)$ be defined as a random variable which describes the water volume consumed by a single household i within a specific hour h of a specific day d; if D is the number of days with hourly

registrations, the recorded sample for the hour h is made up of a maximum of D data. M random samples of N households are drawn from the database of N_{max} households ($1 \leq N \leq N_{max}$) so that each household can belong to different samples, but every household can only be extracted once within each sample. The aggregated water demand for each day d at hour h of the random sample m is:

$$Q^m_{h,N}(d) = \sum_{i=1}^{N} q_{h,i}(d) \qquad h = 1, \ldots, 24 \tag{1}$$

For a group of N households, the hourly peak water demand $Q^m_{p,N}(d)$ of the random sample m for each day d is defined as:

$$Q^m_{p,N}(d) = \max_{h=1,\ldots,24} \left[Q^m_{h,N}(d) \right] \tag{2}$$

where $Q^m_{\mu,N}(d)$ is the daily mean water demand of the random sample m for a group of N households for each day d, expressed as:

$$Q^m_{\mu,N}(d) = \frac{\sum_{h=1}^{24} Q^m_{h,N}(d)}{24} \tag{3}$$

Then, for a group of N households, the dimensionless hourly peak water demand factor $CP_{m,N}(d)$ of the random sample m for each day d is defined as:

$$CP_{m,N}(d) = \frac{Q^m_{p,N}(d)}{Q^m_{\mu,N}(d)} \tag{4}$$

By the adopted notation, $CP_{m,N}(d)$ stands for a CP value belonging to the m-th sample of size N and referring to day d. According to the purpose of the analysis, it could be either seen as part of a sub-sample of size D made up of all the daily observations of CP within one specific sample m, or, alternatively, it can be considered as part of a sub-sample of size M made up of all the observations referring to one specific day d across all the extracted samples. In all cases, $CP_{m,N}(d)$ is a single realization drawn from the population of the random variable CP_N with expected value μ_N.

2.2.2. Expected Value, Variance, and Distribution of the Sample Mean

For a group of N households, the sample mean of the hourly peak water demand factor related to a sample m of size D is:

$$\overline{CP}_{m,N} = \frac{\sum_{d=1}^{D} CP_{m,N}(d)}{D} \tag{5}$$

If CP_N is an independent random variable, the mean (i.e., the expected value) of the sample means $\overline{CP}_{m,N}$ coincides with the population mean μ_N:

$$\mu_N = \frac{\sum_{m=1}^{M} \overline{CP}_{m,N}}{M} \tag{6}$$

Literature suggests an empirical relationship between μ_N and the number N of aggregated households in the following form [7,10,11,16,28]:

$$\mu_N = \frac{a}{N^b} + c \tag{7}$$

where, a, b, and c are the estimated regression coefficients. c is the horizontal asymptote of the function, representing the expected value of the sample mean of peak factor CP_N for a very large N.

According to the classic sampling theory, the standard deviation of the sample mean, usually referred to as, "standard error of the sample mean" [47], $ES_{D,N}$, is directly related to the population variance and to sample size D:

$$ES_{D,N}^2 = Var\left\{\overline{CP}_{m,N}\right\} = Var\left\{\frac{\sum_{d=1}^{D} CP_{m,N}(d)}{D}\right\} = \frac{1}{D^2}\sum_{d=1}^{D} Var\left\{CP_{m,N}(d)\right\} = \frac{\sigma_N^2}{D} \quad (8)$$

where σ_N^2 is the population variance of CP_N.

If the random variable CP_N is normally distributed, the sample mean will be normally distributed too, with $\overline{CP}_{m,N} \sim N(\mu_N, ES_{D,N})$ independently of sample size D. Otherwise, based on the central limit theorem, when the dimension of the random sample becomes sufficiently large ($D \geq 30$), the distribution of the sample mean can be approximated by a normal distribution independently of the specific distribution of the random variable CP_N. To verify the normality of the sample mean $\overline{CP}_{m,N}$, well-known statistical tests can be adopted such as the Kolmogorov-Smirnov (KS) test [48].

If the random variable CP_N is not independent (as will be demonstrated in the present paper), Equation (6) is still valid, whereas the standard error of the sample mean can be estimated according to the following Equation [45] that explicitly accounts for the covariance matrix:

$$\begin{aligned} ES_{D,N}^2 &= Var\left\{\frac{\sum_{d=1}^{D} CP_{m,N}(d)}{D}\right\} \\ &= \sum_{d=1}^{D} \frac{Var\left\{CP_{m,N}(d)\right\}}{D^2} + \sum_{i=1}^{D} \sum_{\substack{j=1 \\ j \neq i}}^{D} \frac{Cov\left\{CP_{m,N}(i), CP_{m,N}(j)\right\}}{D^2} \\ &= \frac{1}{D^2}\left[\sum_{d=1}^{D} Var\left\{CP_{m,N}(d)\right\} + \sum_{i=1}^{D} \sum_{\substack{j=1 \\ j \neq i}}^{D} Cov\left\{CP_{m,N}(i), CP_{m,N}(j)\right\}\right] \end{aligned} \quad (9)$$

where the first term sums up the cross-sample variance for each day d, and the second term sums up the cross-correlation among pairs of samples.

Equation (9) estimates the standard error of sample means. When sample data are extracted from a finite population, as in the present paper, the values of the standard error can be influenced and underestimated, because there is a high probability that the same elements are extracted from the total population. Indeed, for $N = N_{max}$ Equation (9) gives a null value for the standard error, which is a degeneration caused by the fact that the M samples are made up of exactly the same CP_N values. Instead, for an infinite population, a finite, although small, value for the standard error should be expected even for very high N values.

In case there is no spatial correlation among water demands, the covariance term in Equation (9) is null and the variance collapses back to Equation (8), with an inverse dependence on sample size D. In any other case, also accounting for the finite population effect (i.e., a null asymptotical value for ES) the dependence of ES on D can be formulated for each N in the generic form:

$$ES_D = \alpha_1 \times D^{\beta_1} \quad (10)$$

where the coefficients depend on the structure of the spatial correlation [30,41]. Since Equation (10) can be applied for each fixed N, the following general equation is proposed to consider the additional dependence of the variance on N:

$$ES_{D,N} = \frac{\alpha_1 \times D^{\beta_1}}{(\alpha_2 + N)^{\beta_2}} \quad (11)$$

When the distribution of the sample mean $\overline{CP}_{m,N}$ for a group of N households is (at least approximately) normal, the lower and upper limits $\left[\overline{CP}_{m,N}\right]_p$ of a confidence interval centered on the mean μ_N, for a predefined probability p, are:

$$\left[\overline{CP}_{m,N}\right]_p = \mu_N \pm \xi_p \times ES_{D,N} \quad (12)$$

where ξ_p is the normal p-th quantile; the standard error $ES_{D,N}$ can be estimated as the square root of either Equation (8) or Equation (9), or directly by its empirical approximation provided by Equation (11), based on the probability distribution of random variable CP_N. If the sample mean $\overline{CP}_{m,N}$ is normally distributed, substituting in Equation (12) the 2.5-th and 97.5-th normal percentile values $\xi_p = \pm 1.96$, the 95% confidence interval is obtained.

2.2.3. Variability of Peak Coefficient among Weekdays

The water demand can show different trends between working days and weekends, and this can affect the maximum daily water demand. For the investigated dataset, water consumption exhibits a significant weekly cycle, and water demand was clustered in three groups: weekdays, Saturdays, and Sundays [40].

To verify if the sample mean of hourly peak factors has a day-to-day variability, the ANOVA test is used, which is able to identify significant differences in the central values of different groups [49]. In the present study, seven groups, one for each day of the week, are defined and the related sample means are estimated by Equation (6). For each group, summation in Equation (6) is only extended to the days D_i with $i = 1, 2, \ldots, 7$ (1 = Mondays, 2 = Tuesdays, ... , 7 = Sundays). In other words, for each value of N, seven groups of M sample means are evaluated. It is worth noting that the mean of all the $7 \times M$ sample means coincides with the population mean μ_N.

As highlighted in the previous sections, the statistical behaviour of peak factors is influenced by the number N of aggregated households; thus, it is expected that the outcomes of ANOVA show a similar dependence. Specifically, for small N values any differences of the peak factor values among the days of the week could be covered by the high peak demand variability. In turn, those differences could become more evident for higher N values, when peak demand variability is lower due to the stabilization of the aggregated water demand pattern.

3. Results

The methodology proposed in the previous sections was applied to the water consumption database of the pilot area, made up of $N_{max} = 1162$ connections, each corresponding to a registered consumption time series. The analysis was initialized by setting 50 different values of N to be tested, ranging between 1 and 1162 (Table 1). For each N, $M = 150$ samples of N time series were randomly extracted from the consumption database and aggregated. Then, for each aggregated series, hourly peak demand factors $CP_{m,N}(d)$ were computed for the m-th sample by means of Equation (5). The total number of available monitored days D_{max} in the 2016 database is equal to 322; thus, for each day of the week, the maximum number of monitored days is 46 (Table 1).

The computation of sample means $\overline{CP}_{m,N}$ by means of Equation (6) was performed gathering $CP_{m,N}(d)$ values in seven groups according to the day of the week. Then, the ANOVA test was performed to highlight possible differences in the behaviour of peak factors during the week. Figure 3 shows the results of the ANOVA test as box-plots of peak factor sample means for two different values of aggregated households $N = 10$ and $N = 1000$. ANOVA outcomes highlight that there are significant differences in terms of expected values of sample means between the weekdays, the Saturdays, and the Sundays, so that three clusters can be identified, coherently with findings shown in [40]. Moreover, as expected, those differences are more and more evident the higher the N value and can be considered statistically significant starting from N = 5–10.

Table 1. Cluster definition and relevant parameters.

Cluster		1	2	3	4
Day		**Saturdays**	**Sundays**	**Weekdays**	**All Days**
D_{min}				30	
D_{max}		46	46	230	322
N_{min}				1	
N_{max}				1162	
Equation (7)	a			1.763	
	b			0.670	
	c	1.573	1.611	1.536	1.552
	R^2	0.998	0.997	0.997	0.998
Equation (11)	α_1	0.274	0.293	0.300	0.274
	β_1			0	
	α_2	−0.726	−0.632	−0.665	−0.695
	β_2			0.5	
	R^2	0.999	0.997	0.999	0.999

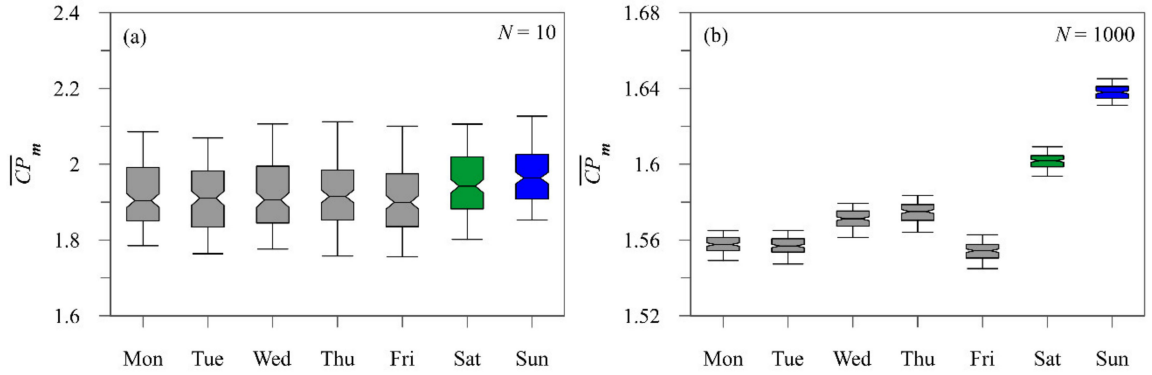

Figure 3. Box-plot of hourly peak factor sample means for the different days of the week and for two different numbers of aggregated households: (**a**) N = 10 and (**b**) N = 1000.

Figure 3 also shows that weekends are characterized by an expected value of peak factors higher than the weekdays. This could be explained considering that, during the weekend, people tend to adopt predictable schedules, translating in more homogeneous consumption behaviours, leading to more synchronous water uses and, therefore, producing more coherent water demand diurnal patterns. Finally, Figure 3 demonstrates that the differences among the three clusters are statistically significant and should be accounted for in further analyses. As a consequence, in the following sections four clusters will be investigated separately: Cluster 1, made up of Saturdays (D_{max} = 46); Cluster 2, made up of Sundays (D_{max} = 46); Cluster 3, made up of the remaining weekdays (D_{max} = 230); Cluster 4, made up of all the days of the week (D_{max} = 322). This last cluster is considered in order to better understand the significance of cluster separation in evaluating the statistics of interest.

3.1. Sample Mean: Expected Value, Standard Error, and Scaling Laws

According to the proposed methodology, the analysis of hourly peak factor sample means consists of the estimation of the expected value, associated standard deviation, and confidence band.

For each N value, M sample means $\overline{CP}_{m,N}$ were computed by means of Equation (5) and the corresponding expected values μ_N were estimated by means of Equation (6); then, the empirical relation between N and μ_N was found by calibrating parameters in Equation (7). Sample means and expected values are shown, for each Cluster, in Figure 4 as a function of the number of aggregated households. Table 1 shows the estimated values of the regression coefficients a, b, and c and the value for the coefficient of determination, which is very high for all Clusters. Figure 5a shows the comparison

between the observed expected values, computed by means of Equation (6), and the predicted expected values, obtained from Equation (7), for all Clusters. It is evident that points gather almost perfectly along the 1:1 line, showing a high accordance between the observed and the predicted values, with just a slight deviation for the highest mean values, corresponding to $N = 1$.

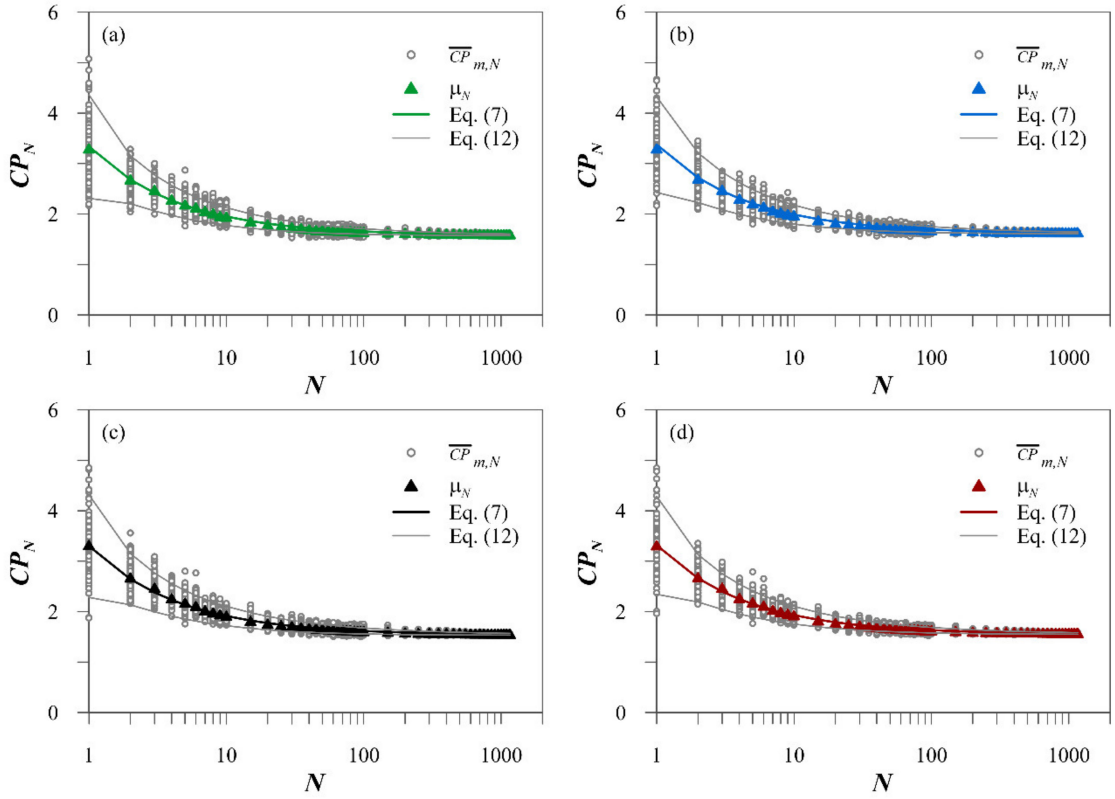

Figure 4. Sample means, expected values, and confidence bands as a function of the number of aggregated households for: (**a**) Cluster 1; (**b**) Cluster 2; (**c**) Cluster 3; (**d**) Cluster 4.

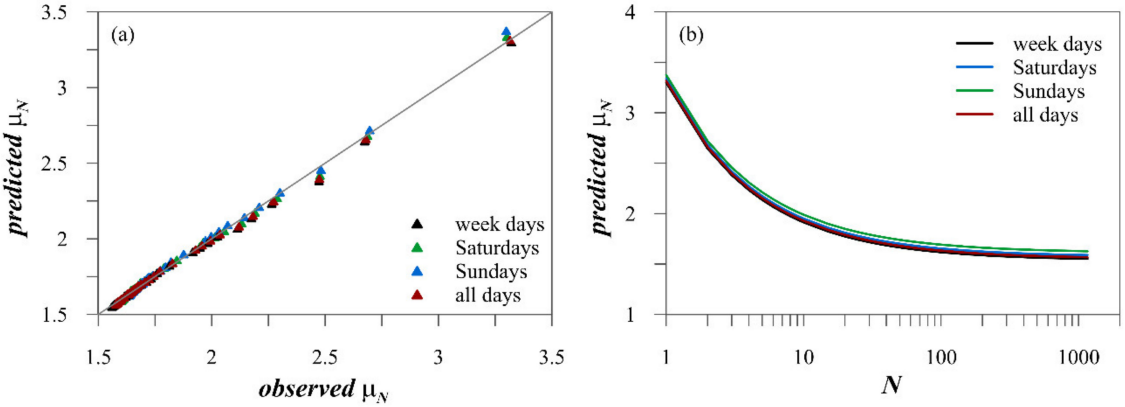

Figure 5. (**a**) Accordance between expected values of sample means estimated by Equations (6) and (7) for all the Clusters. (**b**) Comparison among calibrations of Equation (7) performed on the different Clusters.

Table 1 and Figure 5b show that the regression curves of the four Clusters are very similar, with only a different value for the c coefficient, which represents the expected value of hourly peak demand factor for a large number of households. As Figure 4 shows, this asymptotic value can be considered attained for $N > 100$–200 for every Cluster. Figure 5b and Table 1 also show that the highest asymptotic

expected value is observed for the Sundays Cluster, followed by the Saturdays, and the Weekdays Clusters. Cluster 4 shows intermediated values.

As Figure 4 shows, for a fixed N, the M sample means $\overline{CP}_{m,N}$ show a non-negligible variability, which can be quantified by means of the standard error $ES_{D,N}$. In order to compute standard errors, the regression coefficients in Equation (11) were calibrated for each Cluster by using the estimate of $ES_{D,N}$ provided by Equation (9), and their values are shown in Table 1 along with the very high coefficient of determination. To capture the dependence of $ES_{D,N}$ on both N and D, different values of D were tested in the range D_{min}–D_{max}, where D_{min} = 30 was set to ensure normality, as previously mentioned. However, in all cases the dependence on D resulted to be negligible with respect to the aggregation level, with very small values for the exponent β_1, that was approximated to zero for all Clusters (Table 1). Moreover, Table 1 shows that for all the Clusters β_2 resulted equal to 0.5, with a simplification in the proposed regression equation.

Figure 6 shows, for each Cluster, the regression curve provided by Equation (11) as well as the standard errors estimated as the square root of Equation (9) for three different values of D (D_{min}, D_{max}, and intermediate value depending on D_{max}). Coherently with the approximation $\beta_1 = 0$, no effect of the number of recording days can be observed, with all the points gathering along the regression curve, with just a slight deviation for $N = 1$.

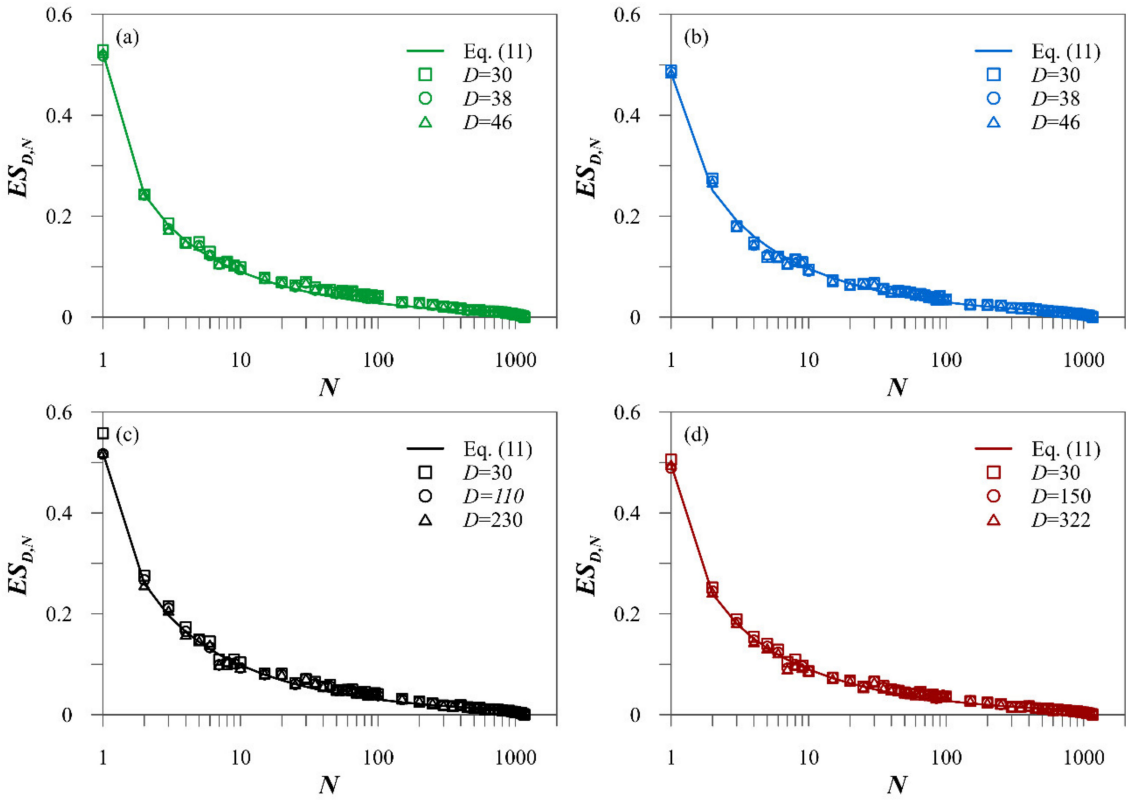

Figure 6. Standard errors of the sample mean estimated by Equation (9) and predicted by Equation (11) for three different D for: (**a**) Cluster 1; (**b**) Cluster 2; (**c**) Cluster 3; (**d**) Cluster 4.

As a goodness-of-fit measure, Figure 7a shows a comparison between the squared standard error estimated by means of Equation (9) and the values predicted by Equation (11) for all the Clusters, with regression coefficients shown in Table 1. The points in Figure 7a gather almost perfectly along the 1:1 line, ensuring an extremely satisfying prediction of the sample mean standard deviation by Equation (11). Figure 7 shows a comparison among the prediction curves of the standard error as a function of N for the different Clusters for $D = D_{max}$. It can be observed that the four curves show small differences for small values of N, which become negligible for N > 100–200. Coherently, in the same

range of N, the prediction curve for μ_N reaches its asymptotic value for all the Clusters, which suggests an extreme accuracy in the estimation of the expected value of the sample mean for $N > 100–200$. On the other hand, this can be regarded as an effect of investigating a finite population. Indeed, if the same N values were analysed based on a more extended database (i.e., if a higher number of recorded households were monitored), higher values for the standard error would possibly be expected. Moreover, for $N = N_{max}$, each of the M samples is made up of the same elements, so that the M estimates of the sample mean are equal, and the standard error of the sample mean is equal to zero.

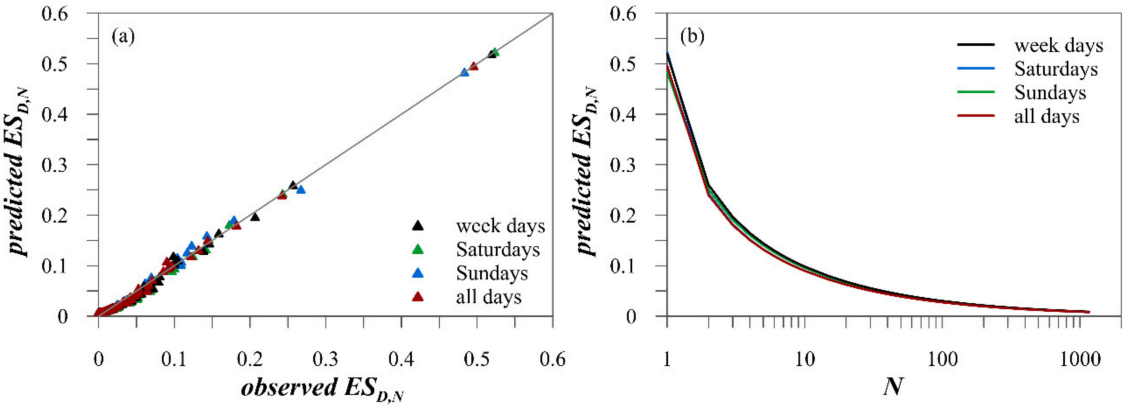

Figure 7. (**a**) Accordance between the standard error estimated for $D = D_{max}$ by Equations (9) and (11) for all the Clusters. (**b**) Comparison among calibrations of Equation (11) performed on the different Clusters for $D = D_{max}$.

The empirical estimates of the standard error were adopted in Equation (12) to obtain the 95% confidence band centred on the expected value of the sample mean, as shown in Figure 4. Confirming the previous evidence, the confidence band reduces as N increases, with an amplitude that can be considered negligible for $N > 100–200$.

3.2. *Sample Mean: Probability Distribution and Final Considerations*

The assumption of normality was verified for sample sizes $D \geq 30$ independently on the distribution of the original sample variable CP_N. However, in order to highlight the possible effect of a finite population, the normality assumption was checked for each Cluster and each value $N < N_{max}$ by means of the Kolmogorov-Smirnov (KS) test [48]. For $N = N_{max}$ no probability distribution can be defined since the variance is null.

The KS test was run under two different assumptions for the distribution of the sample mean:

- normal distribution with unknown mean and variance parameters m and s ("assumption 1");
- normal distribution with $m = \mu_N$, estimated by means of Equation (6), and $s = ES_{D,N}$, estimated as the square root of Equation (9) ("assumption 2").

Figure 8 shows the results of the Kolmogorov-Smirnov test for the two considered assumptions, in terms of percentage of samples passing/not passing the KS test for the four Clusters. It can be observed that under assumptions 1 and 2 the KS test is passed for all the Clusters for all the tested N values. This proves that the sample means are rigorously distributed by means of a normal model with the mean and variance correctly estimated by Equations (6) and (9), respectively. This also confirms that the estimation of the probability distribution of the sample mean is not affected by any finite population effect.

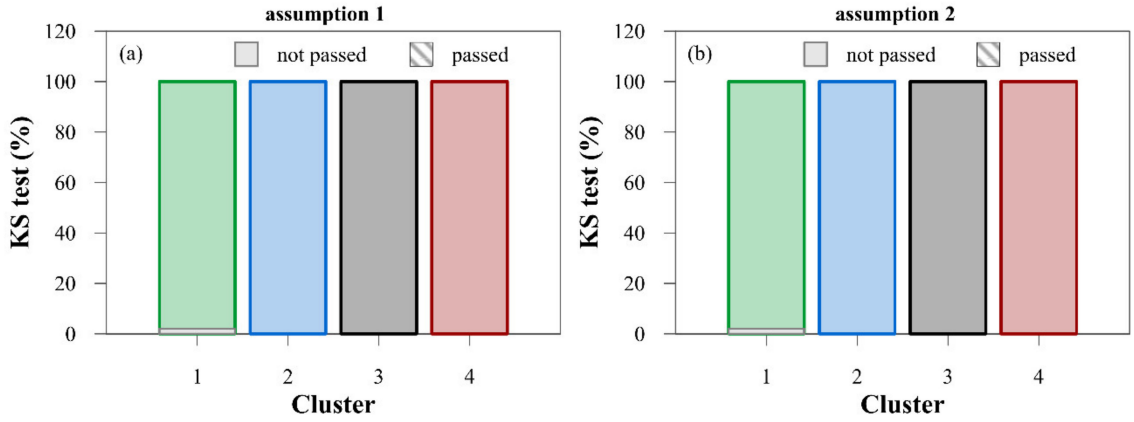

Figure 8. Percentage of samples passing/not passing the Kolmogorov-Smirnov test for the four tested Clusters under (**a**) assumption 1 and (**b**) assumption 2 for the underlying normal distribution of sample means.

Finally, for Cluster 2, Figure 9 shows a comparison between the empirical frequency and the normal probability models under the two assumptions, for the values $N = 60$ and $N = 500$; for these values, the KS test is passed under both assumptions. It can be noted that the CDF curves representing assumptions 1 and 2 are overlapped, highlighting the accuracy of the theoretical estimators adopted for the expected value and the standard deviation. Those results are shown for Cluster 2 but can be extended to all the Clusters.

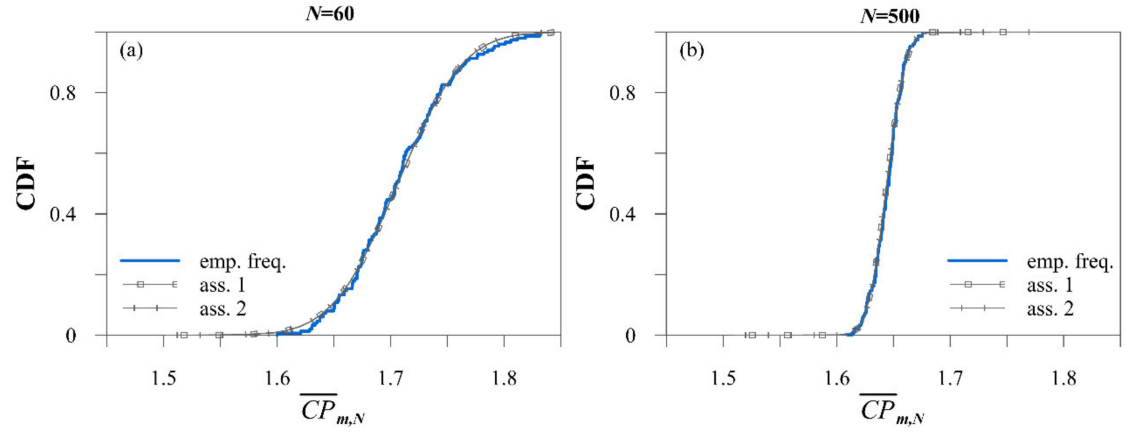

Figure 9. Empirical vs. theoretical probability distributions of sample means for Cluster 2 under two different assumptions for the underlying normal model: (**a**) $N = 60$ and (**b**) $N = 500$.

4. Discussion

4.1. Comparison with Literature

In this paragraph the results obtained from the presented analysis are compared with previous literature analyses. Different literature deterministic relationships for peak factor evaluation have been inspired by the following well known Babbitt's formula [7] deduced for domestic wastewater:

$$C_{P,B} = \frac{5}{\left(\frac{N_u}{1000}\right)^{0.2}} \tag{13}$$

where N_u is the number of users, usually ranging between one thousand and one million in a population, as previously mentioned. The Babbitt's relationship was successively reformulated [28] as:

$$\mu_{N_u}(\Delta t) = K_{CP}(\Delta t) \times \frac{10}{{N_u}^{0.2}} \tag{14}$$

Equation (14) is valid for $250 < N_u < 1250$, and was originally obtained analysing data measured with a 1-min frequency ($K_{CP} = 1$). K_{CP} is a reduction coefficient that takes into account the effect of the time aggregation scale for time steps higher than 1 min. The results of the above relationships are compared with Equation (7), herein rewritten in terms of number of users N_u assuming that each meter serves 2.9 inhabitants on average, and considering the parameters corresponding to all the days of the week (Cluster 4 in Table 1):

$$\mu_{N_u} = \frac{3.60}{{N_u}^{0.67}} + 1.552 \tag{15}$$

Differently from Equation (15), Equations (13) and (14) do not exhibit any asymptote.

For N_u = 500 and N_u = 1000, Table 2 reports: (i) The values of the peak factor estimated by Equation (15); (ii) the values obtained by adopting Equation (13), and (iii) the values obtained by adopting Equation (14). In Equation (14), considering the experimental field data reported in [28], K_{CP} is assumed to be equal to 0.65 for a sampling time step of 60 min. Table 2 highlights that the Babbitt's formula overestimates the peak factor [10], while the prediction obtained with the present analysis is comparable with the estimate of the formula proposed by [28]. In particular, the values obtained with Equation (14) are within the uncertainty range of Equation (7).

Table 2. Hourly peak factor values estimated with different relationships.

N_u	Equation (15)	Equation (13)	Equation (14)
500	1.61	6.60	1.87
1000	1.59	5.00	1.63

Forcing Equation (15) to assume a structure similar to Equation (14), it can be approximated by the following expression:

$$\mu_{N_u} = \frac{4.3}{{N_u}^{0.18}} \tag{16}$$

where the exponent for the number of users is very similar to the one in the empirical relationship in Equation (14) proposed by [28].

As previously mentioned, differently from the empirical literature relationships, the proposed Equation (7) for the evaluation of the hourly peak coefficient tends to an asymptotic value as the number of household increases. A similar result was also obtained by [16], who derived an estimation of the instantaneous peak factor using a probabilistic approach to describe the residential water use based on the Poisson Rectangular Pulse (PRP) model and adopting the Gumbel distribution for the extreme values. The asymptotic value can be assumed to be equal to the asymptotical hourly peak factor for a growing population [11]. Analyses performed on different towns in Italy showed that the asymptotic value ranges between 1.5 and 1.7 [11], similarly to the one deduced herein.

For a number of users varying between 3 and 3000, Equation (15) predicts a peak factor ranging between 3.3 and 1.55, which is the asymptotic value. Those values are also comparable with the range 1–5 reported in [26] considering the results of recent studies in different countries. The obtained values, smaller than the one provided by the empirical Babbitt's relationship, may be ascribed to a different kind of analysis and/or to a change in consumption behaviours compared to 30–50 years ago.

4.2. Applicability Example

The proposed procedure helps the operators in understanding the reliability of a network in terms of demand and pressure at a different level of the users' aggregation using hourly meter data. It can be adopted for understanding if peak values are changed with respect to the ones considered at the design

stage, for planning DMAs and for verifying the behavior of existing networks in case of problems in the branched pipes where a lower number of household is served.

As noted above, Equation (15) tends to an asymptotic value as the number of households increases (N > 100–200). This means that for looped networks, which serve more than about 600 inhabitants, the peak value can be considered equal to the asymptotic value. Conversely, when considering a single mainline serving different small groups of households, the variability of the peak factor should be accounted for. The synthetic following example shows an application of the proposed formulation in verifying a branched pipe serving different groups of households. Figure 10 shows the main line with six nodes and Table 3 reports the number of households (each represented by a water meter) assumed connected to each node, and the corresponding number of users under the assumption that each meter serves 2.9 inhabitants.

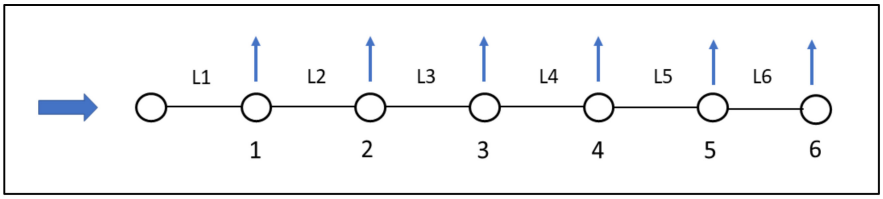

Figure 10. Sketch of a schematic mainline with six nodes serving different groups of households.

Table 3. Example data.

Node	N	N_u	Link	CP	Q_m (L/s)	Q_p (L/s)
1	60	174	L1	1.60	36	59
2	32	93	L2	1.62	24	40
3	40	116	L3	1.64	18	29
4	30	87	L4	1.68	9.7	16
5	15	43	L5	1.81	3.6	6.6
6	3	9	L6	2.41	0.61	1.5

For a total number of 180 served households, equivalent to 522 users, Equation (15) provides a peak factor (CP) equal to 1.60, which is the value that should be considered for designing the pipe L1. Indeed, while link L1 serves 180 households, L6 serves only three of them. Assuming a water supply of 0.07 L/s per inhabitant, Table 3 reports, for each link, the peak value obtained by means of Equation (15), μ_{N_u}, as well as the corresponding mean, Q_m, and peak, Q_p, discharge. A correct evaluation of the hourly peak factor is important for designing the trunks of branch pipes, where an underestimation of the discharge may produce situations of pressure deficit. Conversely, an overestimation of the pipe diameter may produce low velocity and an increase of the water age with a consequent decay of water quality [50]. Concluding, the performed study highlights that the peak factor changes drastically in the interval $1 < N < 100$, and this change has to be carefully considered for a correct design of branch pipes.

5. Conclusions

The proposed analysis provides a methodological framework to investigate the main features of water demand hourly peak factors based on hourly consumption data. The main objective is the estimation of the sample mean of hourly peak factors, the associated standard error (allowing for the definition of confidence bands), and its probability distribution. Those quantities are investigated in a perspective of spatial aggregation: For each considered aggregation level, artificial populations are created by aggregating multiple consumption time series and analysing the related statistics.

Theoretical expressions for the sample mean and for the standard error are provided (Equations (6) and (9), respectively), where the standard error expression accounts for the cross-correlation among samples. Moreover, empirical relations of the sample mean and standard error as a function of the number of aggregated households or meters (or users) are also provided (Equations (7) and (11),

respectively). Concerning the probability distribution, sample means can be considered normally distributed, with model parameters effectively estimated by Equations (6) and (9).

The outcomes of the research in terms of mean peak factor are consistent with previous literature analyses focusing on similar or higher-resolution consumption datasets. In addition, the confidence band suggests a high accuracy of its estimation. The structure of the dependence on the aggregation level suggests the presence of an asymptotic value for a high number of users, as also suggested by some recent literature works.

The research confirms the possibility of using 1 h-aggregation consumption datasets for the analysis of water demand peak factors and provides a general framework to perform the stochastic analysis for aggregated consumption data. The empirical relation for the estimation of the expected value of the hourly peak factor has a general validity, although regression parameters' values are a reflection of the specific consumptions of the pilot area. General validity can be also extended to Equation (11) for the estimation of the standard deviation if the effect of a finite population is neglected. Indeed, results showed that the finite population condition does not affect the probability distribution of sample means, which remains normal, but it may affect the amplitude of the confidence bands, which could be underestimated. The proposed methodology will be further applied on other distribution systems. Moreover, additional investigations about the effect of spatial correlation on the coefficient of variation of peak discharges, as well as the quantification of the peak factor variance, will be the object of future research.

As a final remark, the structure and the coefficients of the empirical relationship described by Equation (7) for the expected value of the hourly peak water demand factor allows formulating the following general considerations, that can be of significant aid in the design and verification of water distribution networks.

1. Several relationships provided by the literature asymptotically tend to zero as the number of households increase; conversely, in the present research the peak factor asymptotically tends to a constant value greater than one.
2. The asymptotic value is reached for values of the number of households N of about 100–200 (approximately corresponding to a number of users N_u of about 300–600); conversely, the increase in the peak factor mainly affects only the secondary pipe networks of the urban centres which serve a reduced number of users.
3. The secondary pipe networks generally consist of a branched pipe structure, which is more sensitive to the flow variation than looped networks; in this case, the peak factor must be adequately considered for design purposes.

Author Contributions: Conceptualization, G.D.G., C.D.C., and R.P.; data curation, G.D.G., C.D.C., and R.P.; formal analysis, G.D.G., C.D.C., and R.P.; investigation, G.D.G., C.D.C., and R.P.; methodology, G.D.G., C.D.C., and R.P.; software, G.D.G., C.D.C., and R.P.; supervision, G.D.G., C.D.C., and R.P.; validation, G.D.G., C.D.C., and R.P.; visualization, G.D.G., C.D.C., and R.P.; writing—original draft, G.D.G., C.D.C., and R.P.; writing—review and editing, G.D.G., C.D.C., and R.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cominola, A.; Spang, E.S.; Giuliani, M.; Castelletti, A.; Lund, J.R.; Loge, F.J. Segmentation analysis of residential water-electricity demand for customized demand-side management programs. *J. Clean. Prod.* **2018**, *172*, 1607–1609. [CrossRef]
2. National Research Council of the National Academies. *Drinking Water Distribution Systems: Assessing and Reducing Risks*; National Academies Press: Washington, DC, USA, 2007.

3. Babayan, A.; Kapelan, Z.; Savic, D.; Walters, G. Least-Cost design of water distribution networks under demand uncertainty. *J. Water Resour. Plan. Manag.* **2005**, *131*, 375–382. [CrossRef]
4. Alcocer-Yamanaka, V.H.; Tzatchkov, V.G.; Arreguin-Cortes, F.I. Modeling of drinking water distribution networks using stochastic demand. *Water Resour. Manag.* **2012**, *26*, 1779–1792.
5. Di Cristo, C.; Leopardi, A.; de Marinis, G. Assessing measurement uncertainty on trihalomethanes prediction through kinetic models in water supply systems. *J. Water Supply Res. Technol. AQUA* **2015**, *64*, 516–528. [CrossRef]
6. Marquez Calvo, O.O.; Quintiliani, C.; Alfonso, L.; Di Cristo, C.; Leopardi, A.; Solomatine, D.; de Marinis, G. Robust optimization of valve management to improve water quality in WDNs under demand uncertainty. *Urban Water J.* **2018**, *15*, 943–952. [CrossRef]
7. Babbitt, H.E. *Sewerage and Sewage Treatment*, 3rd ed.; Wiley: New York, NY, USA, 1928.
8. Johnson, C.F. Relation between average and extreme sewage flow rates. *Eng. News Rec.* **1942**, *129*, 500–501.
9. Metcalf, L.; Eddy, H.P. *American Sewerage Practice, Volume III: Design of Sewers*, 3rd ed.; McGraw-Hill: New York, NY, USA, 1935.
10. Tricarico, C.; de Marinis, G.; Gargano, R.; Leopardi, A. Peak residential water demand. *Proc. Inst. Civ. Eng. Water Manag.* **2007**, *160*, 115–121. [CrossRef]
11. Balacco, G.; Gioia, A.; Iacobellis, V.; Piccinni, A.F. At-site assessment of a regional design criterium for water-demand peak factor evaluation. *Water* **2019**, *11*, 24. [CrossRef]
12. Buchberger, S.G.; Carter, J.T.; Lee, Y.H.; Schade, T.G. *Random Demands, Travel Times and Water Quality in Dead-Ends*; AWWARF Report No. 294; American Water Works Association Research Foundation: Denver, CO, USA, 2003.
13. Buchberger, S.G.; Wu, L. Model for instantaneous residential water demands. *J. Hydraul. Eng.* **1995**, *121*, 232–246. [CrossRef]
14. Buchberger, S.G.; Wells, G.J. Intensity, duration and frequency of residential water demands. *J. Water Resour. Plan. Manag.* **1996**, *122*, 11–19. [CrossRef]
15. Guercio, R.; Magini, R.; Pallavicini, I. Instantaneous residential water demand as stochastic point process. *WIT Trans. Ecol. Environ.* **2001**, *48*, 129–138.
16. Zhang, X.; Buchberger, S.; Van Zyl, J. A theoretical explanation for peaking factors. In Proceedings of the ASCE EWRI Conferences, Anchorage, AK, USA, 15–19 May 2005.
17. Creaco, E.; Alvisi, S.; Farmani, R.; Vamvakeridou-Lyroudia, L.; Franchini, M.; Kapelan, Z.; Savic, D. Preserving duration–intensity correlation on synthetically generated water-demand pulses. *Procedia Eng.* **2015**, *119*, 1463–1472.
18. Creaco, E.; Farmani, R.; Kapelan, Z.; Vamvakeridou-Lyroudia, L.; Savic, D. Considering the mutual dependence of pulse duration and intensity in models for generating residential water demand. *J. Water Resour. Plan. Manag.* **2015**, *141*, 04015031.
19. Alvisi, S.; Franchini, M.; Marinelli, A. A stochastic model for representing drinking water demand at residential level. *Water Resour. Manag.* **2003**, *17*, 197–222.
20. Alcocer-Yamanaka, V.H.; Tzatchkov, V.; Buchberger, S.G. Instantaneous water demand parameter estimation from coarse meter readings. In Proceeding of the 8th Water Distribution Systems Analysis Symposium, Cincinnati, OH, USA, 27–30 August 2006; ASCE: Reston, VA, USA; pp. 1–14.
21. García, V.J.; García-Bartual, R.; Cabrera, E.; Arregui, F.; García-Serra, J. Stochastic model to evaluate residential water demands. *J. Water Resour. Plan. Manag.* **2004**, *130*, 386–394.
22. Creaco, E.; Kossieris, P.; Vamvakeridou-Lyroudia, L.; Makropoulos, C.; Kapelan, Z.; Savic, D. Parameterizing residential water demand pulse models through smart meter readings. *Environ. Model. Softw.* **2016**, *80*, 33–40.
23. Alvisi, S.; Franchini, M.; Marinelli, A. Generation of synthetic water demand time series at different temporal and spatial aggregation levels. *Urban Water J.* **2014**, *11*, 297–310.
24. Blokker, E.J.M.; Vreeburg, J.H.G.; van Dijk, J.C. Simulating residential water demand with a stochastic end-use model. *J. Water Resour. Plan. Manag.* **2010**, *136*, 19–26.
25. Blokker, E.J.M.; Pieterse-Quirijns, E.J.; Vreeburg, J.H.G.; van Dijk, J.C. Simulating Nonresidential Water Demand with a Stochastic End-Use Model. *J. Water Resour. Plan. Manag.* **2011**, *137*, 511–520.

26. Beal, C.; Stewart, R.A. Identifying Residential Water End-Uses Underpinning Peak Day and Peak Hour Demand. *J. Water Resour. Plan. Manag.* **2014**, *140*, 04014008.
27. Creaco, E.; Pezzinga, G.; Savic, G. On the choice of the demand and hydraulic modeling approach to WDN real-time simulation. *Water Resour. Res.* **2017**, *53*, 6159–6177.
28. Gargano, R.; Tricarico, C.; Granata, F.; Santopietro, S.; de Marinis, G. Probabilistic Models for the Peak Residential Water Demand. *Water* **2017**, *9*, 417.
29. Filion, Y.R.; Adams, B.; Karney, B. Cross correlation of demands in water distribution network design. *J. Water Resour. Plan. Manag.* **2007**, *133*, 137–144.
30. Magini, R.; Pallavicini, I.; Guercio, R. Spatial and temporal scaling properties of water demand. *J. Water Resour. Plan. Manag.* **2008**, *134*, 276–284.
31. Gato-Trinidad, S.; Gan, K. Characterizing maximum residential water demand. *Water* **2012**, *122*, 15–24.
32. Creaco, E.; Signori, P.; Papiri, S.; Ciaponi, C. Peak demand assessment and hydraulic analysis in WDN design. *J. Water Resour. Plan. Manag.* **2018**, *144*, 04018022.
33. Moughton, L.J.; Buchberger, S.G.; Boccelli, D.L.; Filion, Y.R.; Karney, B.W. Effect of time step and data aggregation on cross correlation of residential demands. In Proceeding of the 8th Annual Water Distribution Systems Analysis Symposium, Cincinnati, OH, USA, 27–30 August 2006; ASCE: Reston, VA, USA; pp. 1–14.
34. Boyle, T.; Giurco, D.; Mukheibir, P.; Liu, A.; Moy, C.; White, S.; Stewart, R. Intelligent metering for urban water: A review. *Water* **2013**, *5*, 1052–1081.
35. Cominola, A.; Giuliani, M.; Piga, D.; Castelletti, A.; Rizzoli, A.E. Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review. *Environ. Model. Softw.* **2015**, *72*, 198–214.
36. Willis, R.; Stewart, R.A.; Panuwatwanich, K.; Capati, B.; Giurco, D. Gold Coast domestic water end use study. *Water J. Aust. Water Assoc.* **2009**, *36*, 79–85.
37. Britton, T.; Stewart, R.; O'Halloran, K. Smart metering: Enabler for rapid and effective post meter leakage identification and water loss management. *J. Clean. Prod.* **2013**, *54*, 166–176.
38. Gargano, R.; Tricarico, C.; Del Giudice, G.; Granata, F. A stochastic model for daily residential water demand. *Water Sci. Technol. Water Supply* **2016**, *16*, 1753–1767.
39. Padulano, R.; Del Giudice, G. A mixed strategy based on Self-Organizing Map for water demand pattern profiling of large-size smart water grid data. *Water Resour. Manag.* **2018**, *32*, 3671–3685.
40. Padulano, R.; Del Giudice, G. Pattern Detection and Scaling Laws of Daily Water Demand by SOM: An Application to the WDN of Naples, Italy. *Water Resour. Manag.* **2019**, *33*, 739–755.
41. Creaco, E.; De Paola, F.; Fiorillo, D.; Giugni, M. Bottom-up generation of water demands to preserve basic statistics and rank cross-correlations of measured time series. *J. Water Resour. Plan. Manag.* **2020**, *146*, 06019011.
42. Padulano, R.; del Giudice, G. A nonparametric framework for water consumption data cleansing: An application to a smart water network in Naples (Italy). *J. Hydroinf.* Available online: https://iwaponline.com/jh/article/doi/10.2166/hydro.2020.133/72559/A-nonparametric-framework-for-water-consumption (accessed on 6 March 2020).
43. Frankel, M. *Handbook of Survey Research*; Academic Press: San Diego, CA, USA, 1983.
44. Kay, S. *Intuitive Probability and Random Processes Using MATLAB®*; Springer Science & Business Media: Berlin, Germany, 2006.
45. McCharty, P.J.; Snowden, C.B. *The Bootstrap and Finite Population Sampling*; Vital and Health Statistics Series 2, No. 95. DHHS Pub. No. (PHS) 85–1 369; Public Health Service, U.S. Government Printing Office: Washington, DC, USA, 1985.
46. Starsinic, M. Incorporating a Finite Population Correction Factor into American Community Survey Variance Estimates. In Proceedings of the Section on Survey Research Methods; American Statistical Association: Alexandria, VA, USA, 2011; pp. 3621–3631.
47. Payton, M.E.; Greenstone, M.H.; Schenker, N. Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? *J. Insect Sci.* **2003**, *3*, 1–6.
48. Kolmogorov, A.N. Confidence limits for an unknown distribution function. *Ann. Math. Stat.* **1941**, *12*, 461–463.

49. Kottegoda, N.T.; Rosso, R. *Applied Statistics for Civil and Environmental Engineers*, 2nd ed.; Blackwell: Malden, MA, USA, 2008.
50. Quintiliani, C.; Marquez-Calvo, O.; Alfonso, L.; Di Cristo, C.; Leopardi, A.; Solomatine, D.P.; de Marinis, G. Multiobjective Valve Management Optimization Formulations for Water Quality Enhancement in Water Distribution Networks. *J. Water Resour. Plan. Manag.* **2019**, *145*, 04019061.