

Article

# Exploring the Statistical and Distributional Properties of Residential Water Demand at Fine Time Scales

Panagiotis Kossieris \*  and Christos Makropoulos

Department of Water Resources, School of Civil Engineering, National Technical University of Athens, Heron Polytechniou 5, GR-157 80 Zographou, Greece; cmakro@mail.ntua.gr

\* Correspondence: pkossier@central.ntua.gr; Tel.: +30-210-772-2886

Received: 30 August 2018; Accepted: 17 October 2018; Published: 19 October 2018



**Abstract:** Residential water demand consists one of the most uncertain factors posing extra difficulties in the efficient planning and management of urban water systems. Currently, high resolution data from smart meters provide the means for a better understanding and modelling of this variable at a household level and fine temporal scales. Having this in mind, this paper examines the statistical and distributional properties of residential water demand at a 15-minute and hourly scale, which are the temporal scales of interest for the majority of urban water modeling applications. Towards this, we investigate large residential water demand records of different characteristics. The analysis indicates that the studied characteristics of the marginal distribution of water demand vary among households as well as on the basis of different time intervals. Both month-to-month and hour-to-hour analysis reveal that the mean value and the probability of no demand exhibit high variability while the changes in the shape characteristics of the marginal distributions of the nonzero values are significantly less. The investigation of performance of 10 probabilistic models reveals that Gamma and Weibull distributions can be used to adequately describe the nonzero water demand records of different characteristics at both time scales.

**Keywords:** residential water demand; smart meters; seasonal analysis; statistical analysis; urban water modeling; stochastic simulation; marginal distribution; uncertainty; hourly demand; 15-minute demand

## 1. Introduction

Residential water demand is characterized by high temporal and spatial variability consisting of the most influential source of uncertainty that poses extra difficulties in the efficient planning and management of urban water systems [1–3]. In the conventional modeling of such systems, this variability is typically neglected and the attention is mainly concentrated on the appropriate representation of other, also uncertain, parameters of the system such as the exact roughness, diameter, and length of the pipes [1,4,5]. In this case, a constant multiplier pattern that provides a coarse representation of the variation of demand within the day (typically with hourly time steps) is used to determine the water demand for every type of user in each part of the studied system (node). This deterministic approach is known as “top-down” allocation and has been widely applied in urban water applications [6] even though it fails to adequately reproduce the high variability of residential water demand [7–10].

Recently, the rising deployment of smart metering technologies, making available large amount of new water demand data at very fine temporal (down to 1 s) and spatial (household or even water appliance) scales, provides new perspectives towards a better analysis, understanding, and modeling of the water demand process (for a comprehensive review of smart urban systems see References [11] and references herein). This dataflow enables the deployment of stochastic simulation techniques,

which can be used to generate a large number of synthetic, yet statistically consistent, realizations of water demand events. Then, via a “bottom-up” approach, the demand series at higher temporal and spatial levels can be obtained by aggregating the fine synthetic data from individual users [6].

In this context, various models have been developed for the reproduction of the statistical properties and the stochastic structure of water demand at fine time scales. These models rely on Poisson pulse processes [5,12–18] as well as on Poisson-cluster processes such as the Neyman-Scott [19–21] and Bartlett-Lewis [22] mechanisms to reproduce the overall water demand at the level of a single-household or a group of households. Simulation schemes that generate pulses starting from individual water appliances and uses a level have also been proposed [23,24]. These models are typically parameterized on the basis of super-fine water demand data (typically from 1 s up to 1 min) while some of them require data at individual pulse levels [5,13,16]. A comparison and review of these models can be found in Creaco et al. [25].

Although, a lot of work has been done so far on the modeling of water demand at super-fine time scales to the best of authors’ knowledge the statistical and stochastic characteristics of residential water demand processing at higher, yet fine, time scales (e.g., from 15-minute up to hourly scale), which are the typical temporal resolutions in many urban water modeling applications, have never been examined before. This may be related with the absence of a methodology regarding the simulation of water demand at these time scales.

The present paper attempts to fulfil the former gap by conducting an extensive analysis of the statistical characteristics of residential water demand at 15-minute and hourly time scales. This statistical analysis, which precedes the modeling phase, provides new insights and paves the ground for the development of concrete tools toward the probabilistic representation and simulation of water demand across different time scales. Towards this, the present study takes also into account recent advances in stochastic simulation field [26–30], which alleviate the barriers [31] in the deployment of typical linear stochastic models in cases of non-Gaussian and intermittent processes such as residential water demand at fine time scales, allowing the reproduction of the whole marginal distribution of the process.

In this respect, 15-minute and hourly water demand should be treated and examined as an intermittent variable with a mixed-type marginal distribution, which is composed of a probability mass concentrated at zero (i.e., discrete part) and a continuous part that describes probabilistically the nonzero demand magnitudes. The discrete part is mathematically expressed by the probability of no demand while any continuous probabilistic model can be used to describe the continuous part, i.e., the nonzero values. This modeling approach deviates from that of the already developed schemes since the latter model the individual components of demand events (i.e., occurrence, duration, and intensity [5,12,13,15–18]) or reproduce the main summary statistics of the process [14,19–22] without posing any prerequisite to model and capture the whole marginal distribution of the process.

With respect to the above described modeling approach, this study focuses on the statistical properties and peculiarities of the positive magnitudes of the records (i.e., continuous part), but also on the discrete part of the process as it is expressed by the probability of no demand. Especially regarding the continuous part, it is of high importance to examine which probabilistic models are considered more suitable to describe the nonzero data. However, prior to that, the seasonal variation of the marginal distribution of the nonzero 15-minute and hourly demand should be investigated.

More specifically, the present study provides insights on the following questions based on a dataset of 15-minute and hourly residential water demand records hereinafter referred to as the “*iWIDGET dataset*” collected in the framework of the *iWIDGET* project [32].

1. How the main statistical characteristics of the nonzero 15-minute and hourly water demand vary from household-to-household?
2. Which statistical characteristics vary the most at different temporal levels (i.e., monthly, daily, hourly basis) and which can be regarded as essentially invariant?

3. Which are the most suitable probabilistic models to describe the 15-minute and hourly water demand of each household at each temporal level?
4. Whether there is a single model able to capture the process of all households, all hours and all months, both at a 15-minute and an hourly scale?
5. What are the typical values of parameters for the studied models?

To provide answers to the above questions, we present a general framework to preliminary screen different distributions as well as to fit and evaluate the most suitable of them. This framework is easily applied and extendable to analyze other large water demand datasets such as that studied in Reference [33] to detect water demand patterns on the basis of behaviors and life habits of residences.

Although this analysis seems to have local character, the size and the nature of the records allows for general conclusions and insights regarding the appropriate probabilistic representation of the residential water demand at the two understudy time scales. Especially the evaluation of the performance of several distributions and the discrimination of the most suitable of them consist of valuable prior knowledge regarding which of them should be considered as the first choice in the modeling of residential water demand. Additionally, the derived set of model parameters can be regarded and used either as default values in cases where no recorded data is available for a household or as prior estimations if the sample size is small and, subsequently, the sufficient fitting of a model is highly uncertain. Lastly, the insights on the seasonal variation of the main statistical characteristics at different temporal levels directly inform and determine the structure (e.g., stationary or cyclo-stationary) of the simulation models that should be employed to capture the variability at these time scales.

The paper is structured as follows: Section 2 provides an overview of the “*iWIDGET dataset*” and describes the modeling framework and presents a coarse analysis of the main statistical characteristics of the 15-minute and hourly water demand records (answering 1). Section 3 examines the seasonal variation of the marginal distribution of water demand at different temporal levels (answering 2). Section 4 includes the preliminary screening of the candidate models as well as the final identification, fitting, and evaluation of the most suitable to describe the marginal distribution of the understudy variables (answering 3, 4, and 5). Lastly, Section 5 concludes the present analysis summarizing the key findings.

## 2. General Properties of the “*iWIDGET Dataset*”

### 2.1. Preparation and Overview of the Dataset

The present study is based on a set of residential water consumption data collected within the framework of the *iWIDGET* project [32]. The monitoring program concerned the measurement of total water demand at 20 households with different socio-demographic characteristics and consumption profile in the city of Athens, Greece. The installation of smart meters was implemented in September–October 2014 and the first phase of the measuring campaign was completed at the end of October 2015, which is coincident with the completion of the *iWIDGET* project. As a post-project activity, 10 of the households were selected for further monitoring until October 2016.

The installed equipment logged the instantaneous 15-minute water demand, i.e., it records the total amount of water consumed within a 15-minute time interval. The collected data were transmitted and stored into a central database on a daily basis, which is also accessible by the householders via a web-application [34,35].

Considering the peculiarities of the conducted analysis, we focus on the water consumption records of 11 households that have appropriate quality standards, i.e., (a) total record length larger than 1 year, (b) missing values of 15-minute time series less than 20% for each hour of the day. Unfortunately, the remaining nine time series of the “*iWIDGET dataset*” do not meet the above criteria either due to a high number of missing values or other special conditions that distort the real consumption profile of the household (i.e., physical absence of the residents for a long time that had as a result

no recorded consumption for extended periods). In order to ensure the quality of the selected data set, the 15-minute time series were initially examined for negative or unrealistically large values due to smart meter malfunctions. All negative values and all values greater than 300 L/15 min were flagged as “missing” and were excluded from the analysis. The time series at an hourly time scale were obtained by summing up the 15-minute data of each hour, flagging as “missing” and excluding from the analysis the hours with at least one missing 15-minute value.

Further to the above criteria, our target was the development of a concrete and representative dataset that is composed by households with different socio-demographic characteristics and, hence, consumption profiles. This enables the analysis of time series properties on the basis of the key water consumption determinants (factors) that have direct and strong influence on the per capita water consumption of a household. According to the literature (e.g., see indicatively [36,37]), these factors are: the type (i.e., detached house, apartment) and size of the property, the number of people living in the household (occupancy), the household makeup, and the existence of kids in the family as well as the existence of water activity related to garden watering. The main characteristics of the understudy households as well as the length of hourly records are given in Table 1. The dataset is mainly composed by small families with three or less people while there are also four large families including two of them with kids. The age of adult residents varies from 20 to 70 years old. All households belong to the middle-income to high-income category while the educational level of adult residents is high (i.e., high school or higher level studies). Additionally, all of them are owners of their property working full-time or part-time while retired persons are living in five households (with House ID #2, 3, 7, 10, 11). Lastly, it should be mentioned that, in all households, the same typical water-related activities and appliances are experienced, i.e., washing machine, dishwasher, toilet (WC), bath, shower, and kitchen sink while seven of them use water for garden irrigation. No leakages or bursts were reported by the householders during the period of the collection of data while this was further verified by investigating the recorded series for extensive periods without zero values (indication for a persistent leakage) or suspicious high values for successive time intervals (indication for a pipe burst).

The present analysis was conducted on the basis of per capita consumption rather than the total consumption as it is recorded from the smart meters. Subsequently, the hourly and 15-minute time series, which are analyzed here, is derived by dividing the total water consumption records of the household by its corresponding occupancy. This simple transformation enables the direct comparison of consumption characteristics among different households without affecting the shape characteristics of marginal probability distributions and the dependence structure of the series (i.e., autocorrelation coefficients). Further to that, the analysis and parameterization of stochastic models for the simulation of residential water demand at fine time scales is typically conducted on the basis of per capita values rather on total volumes (e.g., [5,12,14,22]).

**Table 1.** Main characteristics of the households used in the analysis along with the size of records.

House ID	Type	Size (m <sup>2</sup> )	Occupancy	Kids	Garden	Sample Size (hours)
1	Detached house	350	6	Yes	Yes	9167
2	Detached house	230	5	No	Yes	9167
3	Apartment	105	3	No	Yes	8903
4	Detached house	100	3	No	No	19186
5	Apartment	145	3	Yes	Yes	9167
6	Detached house	85	3	Yes	No	11024
7	Apartment	80	2	No	Yes	10727
8	Apartment	110	3	Yes	No	18239
9	Apartment	100	2	No	No	9167
10	Detached house	200	3	No	Yes	18239
11	Apartment	105	4	No	Yes	18191

## 2.2. The Modeling Strategy

Residential water demand at fine temporal scales (e.g., at daily and mainly sub-daily scales) occurs as an intermittent process with the presence of zero values (indicating no-demand) in the observed series becoming more and more extensive as the temporal resolution of metering is getting lower. Such a process is described probabilistically via a mixed-type (also known as “discrete-continuous” or “zero-inflated”) marginal distribution, composed by a probability mass concentrated at zero (i.e., discrete part), and a continuous part for the positive nonzero values. Given that  $p_{ND} := P(\underline{x} = 0)$  is the probability of no demand (probability zero), then the probability density function (pdf) and the cumulative distribution function (cdf) of  $\underline{x}$  are given by the equations below.

$$F_{\underline{x}}(x) = \begin{cases} p_{ND}, & x = 0 \\ p_{ND} + (1 - p_{ND})G_{\underline{x}}(x), & x > 0 \end{cases} \quad (1)$$

$$f_{\underline{x}}(x) = \begin{cases} p_{ND}, & x = 0 \\ (1 - p_{ND})g_{\underline{x}}(x), & x > 0 \end{cases} \quad (2)$$

where  $G_{\underline{x}}$  and  $g_{\underline{x}}$  are the cdf and pdf, respectively, of values greater than zero, i.e.,  $G_{\underline{x}} := F_{\underline{x}|\underline{x}>0} = P(\underline{x}|\underline{x} > 0)$ . Subsequently, the formulation of  $F_{\underline{x}}(x)$  or  $f_{\underline{x}}(x)$  for an intermittent process such as water demand is based on two distinct parts: (a) the estimation of the discrete part,  $p_{ND}$ , calculated directly from observed data as the ratio of the number of zero values over the total number of values and (b) the selection and fit of a suitable continuous distribution  $G_{\underline{x}}$  to describe the probabilistic nature of nonzero values (e.g., [26,29]).

This modeling approach has been followed in statistical analysis and stochastic simulation [38] of hydrometeorological processes with intermittent nature (e.g., wind speed and rainfall intensities at fine time scales, discharge of intermittent flows) while a mixed-type model to simulate the total daily water demand of the group of households was formulated by Gargano et al. [39]. On the contrary, such an approach had never been adopted in the simulation of residential water demand at a single-household level since the already developed schemes either model the individual components of demand events (i.e., occurrence, duration, and intensity [5,12,13,15–18]) or reproduce the main summary statistics of the process [14,19–22] without posing any prerequisite to model and capture the whole marginal distribution of the process.

In the present study, taking into account the above along with recent advances in the stochastic simulation field [26–30], which alleviate the barriers [31] in the deployment of typical linear stochastic models under non-Gaussian assumptions, we treat and examine residential water demand as a random variable described by a mixed-type distribution. Therefore, our focus is on the study of the statistical properties and peculiarities of the positive magnitudes of the records (i.e., continuous part), but also on the discrete part of the process as it is expressed by probability of no demand.

The main statistics that are examined in the present analysis are: (a) the mean value, (b) the standard deviation, (c) the skewness, (d) the L-variation, (e) the L-skewness, and (f) the probability of no demand in a specific time interval. The latter statistic indicates whether there is a probability mass concentrated at zero and is calculated as the ratio of the number of zero values over the total number of values. The mean value is a typical measure of central tendency of the random variable and the standard deviation is used to quantify the amount of dispersion of the data values about their mean. Skewness is a typical measure of the asymmetry of the random variable about its mean. Statistics of higher order (i.e., kurtosis) were not included in the analysis since their estimation is highly uncertain.

During the last decades, L-moments [40] have gained great popularity in the analysis and statistical characterization of environmental data especially in cases of small sample sizes [41,42]. They are linear combinations of order statistics analogous to conventional product moments and they can be used to estimate quantities analogous to standard deviation, skewness, and kurtosis, which are termed the L-scale, L-skewness, and L-kurtosis, respectively. The L-mean is identical to the

conventional mean. The sampling properties for L-moments statistics unlike product moments are nearly unbiased even in small samples and are near normally distributed [43]. In the present study, we use the L-variation,  $\tau_2 = \lambda_2/\lambda_1$ , and L-skewness,  $\tau_3 = \lambda_3/\lambda_2$ , defined as ratios of L-moments  $\lambda_i$ . L-variation is analogous to the conventional coefficient of variation (defined as the ratio of standard deviation to the mean value) and expresses the variability of a sample. For positive random variables such as water demand L-variation takes values in the range [0, 1]. L-skewness is a dimensionless measure of asymmetry analogous to a skewness coefficient and can take positive or negative values in the range [−1, 1]. L-skewness equals to zero implies a symmetric distribution while the distribution is right or left skewed for positive and negative values, respectively. Both  $\tau_2$  and  $\tau_3$  can be regarded as measures of distributional shape since they depend only on the shape parameter of a distribution.

### 2.3. Statistical Characteristics of the Records

To obtain a general idea on the characteristics and peculiarities of the understudy variables, we initially examine the entire per capita 15-minute and 1-h records. Further to the analysis on the basis of the individual household, we also studied the properties of a unified series composed by merging the records of all 11 households. This series, indicated with “All data” in the next figures and tables, can be regarded as a “parent household” that comprises a large number of different realizations of the same random variable irrespective of the household and, hence, the determinants and consumption generation mechanisms.

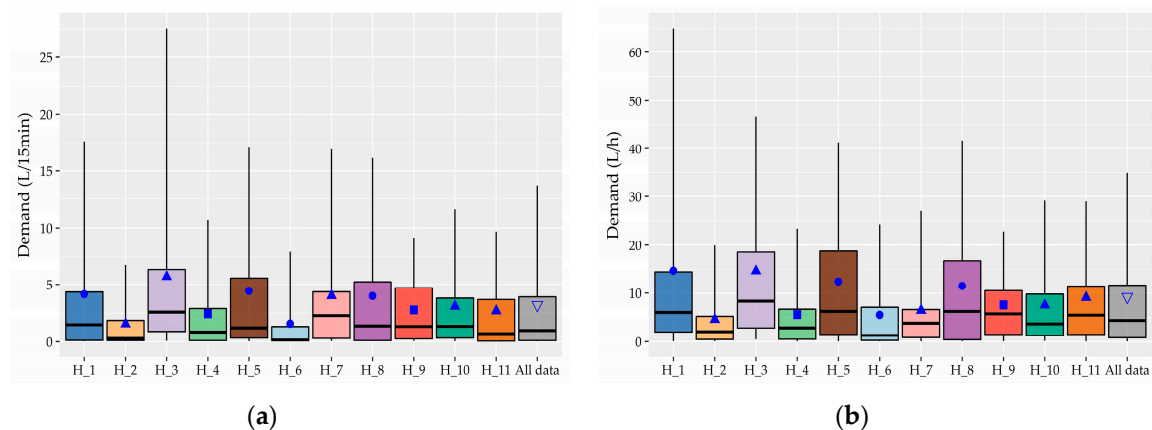
The representative statistics (mean, standard deviation, skewness, L-variation, L-skewness, probability of no demand) of the nonzero 15-minute and hourly water demand records of each household are presented in Table 2 while a graphical representation is provided via the box plots of Figure 1. In all box plots presented in this paper, the top and bottom lines of the boxes represent the 25% and 75% empirical quantile points, respectively, which defines the empirical interquartile range (IQR) that contains 50% of the central values. The middle horizontal line of the box indicates the median of the sample. Typically, a box plot also displays outliers, i.e., data values out of the range of whiskers (in the standard boxplot graph, the upper and lower fences of the whiskers extend  $1.5 \times$  IQR away from the top and bottom lines of the box, respectively). However, in the box plots of the present analysis, in order to provide a more clear and indicative representation of the sample, we choose to not present outliers and define the upper and lower fences of the whiskers as the 5% and 95% empirical quantile points (or the 90% empirical confidence interval (ECI)). Additionally, the mean values (indicated by the blue points) are also displayed in the box plots.

**Table 2.** Basic summary statistics of the nonzero 15-minute and hourly per capita demand records.

House #	1	2	3	4	5	6	7	8	9	10	11	All
<b>15-minute Records</b>												
<i>p</i> <sub>ND</sub>	0.16	0.50	0.58	0.68	0.46	0.15	0.81	0.36	0.55	0.65	0.25	0.48
Mean	4.14	1.56	5.70	2.45	4.43	1.57	4.01	3.99	2.78	3.10	2.70	3.21
St. Deviation	6.46	3.38	8.70	4.39	7.98	3.58	6.04	6.32	3.63	4.94	5.30	5.78
Skewness	2.32	4.38	2.99	3.83	5.58	4.62	3.09	3.04	2.83	3.47	5.63	4.25
L-variation, $\tau_2$	0.69	0.75	0.64	0.70	0.69	0.78	0.64	0.69	0.61	0.65	0.71	0.71
L-skewness, $\tau_3$	0.52	0.65	0.52	0.56	0.54	0.68	0.48	0.50	0.40	0.52	0.54	0.55
<b>Hourly Records</b>												
<i>p</i> <sub>ND</sub>	0.05	0.31	0.34	0.45	0.23	0.03	0.52	0.11	0.33	0.42	0.10	0.27
Mean	14.65	4.58	14.61	5.59	12.39	5.51	6.44	11.47	7.55	7.53	9.08	9.19
St. Deviation	21.93	6.88	20.05	7.91	15.67	9.17	8.83	14.41	8.07	10.18	14.18	13.92
Skewness	2.29	3.07	3.98	2.50	2.52	3.05	2.78	2.05	2.42	2.52	6.05	3.82
L-variation, $\tau_2$	0.66	0.65	0.58	0.64	0.60	0.71	0.60	0.62	0.52	0.61	0.59	0.64
L-skewness, $\tau_3$	0.52	0.50	0.45	0.47	0.39	0.54	0.47	0.39	0.31	0.46	0.44	0.48

As Table 2 and Figure 1 reveal, mean nonzero water demand varies among households while the households composed of a family with kids (House #1, 5, 8) as well as those with water needs for garden irrigation (House #1, 3, 5, 11) have clearly greater values especially at an hourly scale. These households

are also characterized by lower percentages of zero values,  $p_{ND}$ , values, which implies, in general, not only more intense but also more frequent use of water. Generally, in the great majority of records, the probability of no demand is especially high and varies around 50% at a 15-minute scale and around 30% at an hourly scale. As Figure 1 shows, the households with a greater mean value (blue points in the graph) have also wider IQR as well as 90% ECI, which indicates a greater dispersion of the nonzero data. However, as Table 2 presents, all nonzero water demand records, irrespective of their mean value, are characterized by a considerable high variability, which is expressed by the large L-variation values that vary among households in a narrow range, i.e., from 0.61 to 0.78 and from 0.52 to 0.71 at a 15-minute scale and hourly, respectively.



**Figure 1.** Box plots depicting the nonzero 15-minute (a) and hourly (b) per capita water demand of the 11 households and the unified series (indicated as “All data”). The blue points represent the mean value of the samples where dot is used for a “family with kids”, triangle is used for a “family of adults”, a square is used for “adults”, and an inverse triangle is used for the mean of the unified series.

A small variation among households is also noted in the median of the nonzero records with the estimated values being significantly smaller than the corresponding mean. Since it is evident in the above box plots, in most records, about 50% of the recorded 15-minute and hourly nonzero values are concentrated in the narrow intervals (0, 1.5) and (0, 3.5), respectively, which indicates the prevalence of small water consumption uses in the sample.

As we can further note, in all box plots, the median is much closer to the bottom line of the box than the top line, which indicates the much higher concentration of data between the 25% and 50% quantile. Additionally, in all cases, the upper whisker is much longer than the lower one, which is not even distinguishable in some records. These indicate the high positive skewness of nonzero water demand at both understudy time scales. Furthermore, especially the box plots of the records with a higher mean value are characterized by long upper whiskers, which far exceeds the length of the box (IQR) and indicates a heavy-tailed distribution. The level of asymmetry of the nonzero water demand records, in terms of L-skewness, is given in Table 2. As we see, L-skewness varies irregularly among households from 0.40 to 0.67 and from 0.31 to 0.54 at 15-minute and hourly scales, respectively.

As is already known, residential water demand exhibits significant seasonal variation at different temporal scales (i.e., daily, weekly, monthly), which is often attributed to various coexisting factors such as, among others, household makeup, the characteristics (frequency, duration, and volume) of individual uses, and climatological conditions [6]. This observation typically concerns the mean per capita water demand (including zero values) with the relevant studies examining how the latter varies from month-to-month and/or throughout the day (diurnal pattern of demand) against certain consumption habits.

However, whether and how the other characteristics (i.e., higher order moments) including the mean value of nonzero magnitudes of the marginal distribution vary have never been examined before. This is of high importance since such an analysis reveals which statistics vary the most and which can

be regarded as invariant at different time scales, which determine directly the type of statistical and stochastic models that should be employed to model properly residential water demand.

Having said that, the aim of the present section is twofold. On the one hand, to provide insights on how the understudy statistical characteristics of 15-minute and hourly water demand records (i.e., the probability of no demand as well as the mean value, the L-variation, and the L-skewness of nonzero data) vary among households from month-to-month, hour-to-hour, and day-to-day. On the other hand, to support the identification and selection of suitable distributions for the modeling of understudy variables, Section 4 is presented. The present section directly answers the questions 1 and 2, as posed in Section 1.

### 3. Analysis of Seasonal Variation

#### 3.1. Variation from Month-to-Month

To examine the variation of nonzero water demand from month-to-month for each household including the “parent household”, we formed 12 individual records (one record per month). The main statistical characteristics of these records are displayed in Figures 2 and 3 for a 15-minute and hourly scale, respectively. Indicatively, the variation of the nonzero demand records of House #4 on a monthly basis is presented in the form of boxplots in Figure A1 in Appendix A.

As observed, the mean value of the nonzero water demand in most households at both time scales (Figures 2b and 3b) exhibits a small variation among months while a significant increase during the summer months is observed in two of the households (House #1, 3), which is attributed to the intense use of water for garden irrigation. Deviation from this pattern is noted in House #5 where the continuous supply of grass with small amounts of water has as a result the lower mean values, i.e., small values prevail in the sample and significantly lower proportion of zero values during the summer months.

As Figures 2a and 3a depict, the households which are characterized by a small monthly change in the mean value (House #4, 7, 10) exhibit, in general, small variation in the  $p_{ND}$  while the slightly higher values observed in the summer months and especially in August is due to the absence of householders for vacations. This also stands for the variation pattern of  $p_{ND}$  of the “parent household”. Despite the fact that, in all cases, the major changes in  $p_{ND}$  values are noted in the summer months, which is a remarkable differentiation in the patterns of this variation studied along the mean value, and is observed. There are cases where the pattern of variation of the two statistics are consistent (House #5, 8, 11). The cases where the patterns show a contrasting behavior to each other (House #1) as well as cases where  $p_{ND}$  varies without any variation is noted in the corresponding mean values (House #2, 3, 6). This highlights the difficulty to associate the mechanism of variation of these statistics with specific water consumption habits even in households with similar characteristics and prevailing uses such as garden irrigation since this is, obviously, the result of different aspects of the water demand process, which are interrelated at these time scales.

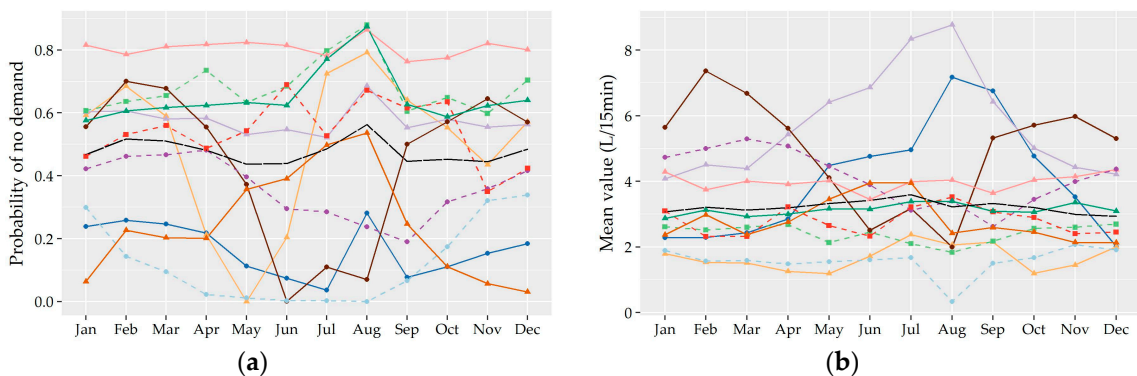


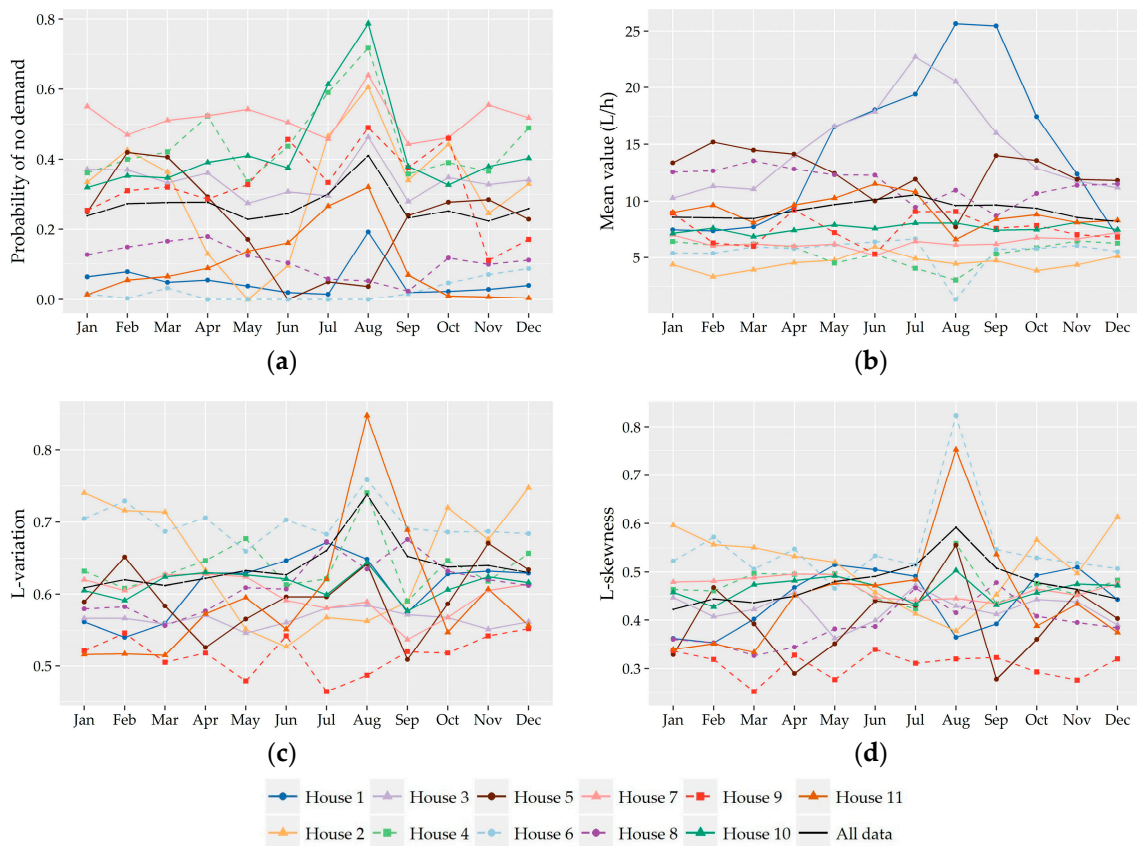
Figure 2. Cont.





**Figure 2. 15-minute scale:** Variation per month of the probability of no demand (a), mean value (b), L-variation (c), and L-skewness (d) of the 15-minute nonzero demand records analyzed.

Regarding  $\tau_2$  and  $\tau_3$ , as Figure 2c,d and Figure 3c,d present, a small fluctuation is noted from month-to-month in most of the households (note that  $y$ -axis of these plots spans the range covered by the estimated values to provide a more clear picture). It is worth noting that the two households with the significant higher mean value in the summer months (House #1, 3) do not exhibit any significant changes in terms of the two shape statistics. By comparing the variation of the two statistics, it is interesting to note that, at both time scales, they follow an identical pattern in all households with peaks appearing in the same months.



**Figure 3. Hourly Scale:** Variation per month of the probability of no demand (a), mean value (b), L-variation (c), and L-skewness (d) of the hourly nonzero demand records analyzed.

Summarizing the above findings, we can say that, at both time scales, the statistic with a higher monthly variation is  $p_{ND}$  followed by the mean value. The changes in these statistics as well as in the two shape characteristics are observed mainly during the summer months.

Furthermore, we examined if there are any significant changes of the above discussed statistical characteristics from day-to-day since there is evidence in the literature that residential water consumption exhibits changes especially during week-end days [6,44]. Towards this, seven individual records (one record per day) were formed for each household as well as the “parent household”. Figures A2 and A3, provided in Appendix A, show that the understudy statistics of all households remain almost invariant among days at both time scales. This is also illustrated in Figure A1 that presents the variation of the nonzero 15-minute and hourly water demand records, respectively, of House #4.

### 3.2. Variation from Hour-to-Hour (Diurnal Pattern)

As is widely known, the total residential water demand exhibits significant variation throughout the day as the result of the different consumption habits experienced during the day. Typically, the diurnal consumption pattern is composed by low flows during night hours, a sharp morning peak and a moderate flow during the day until an evening and/or early night peak [6,44]. However, this fluctuation concerns the mean water demand (including zero values) while the diurnal pattern of other statistical characteristics including the mean of nonzero values have never been examined.

The analysis of previous section showed that the understudy statistics especially the mean value and the  $p_{ND}$  of some households exhibit a non-negligible variation especially during the summer months. This observation lead us to initially examine whether the diurnal pattern and the statistics also varies month-to-month. Towards this, for each month, we formed 24 individual records (one record per hour) by assuming that, within a specific hour, the main statistical characteristics of a household remain invariant (i.e.,  $12 \times 24$  records for each household). The individual record of each hour without monthly discrimination was also formed (i.e., 24 records for each household assuming no variation of the diurnal pattern among months) and the statistics of these records were estimated and plotted to reveal any changes of the diurnal pattern of the studied statistics from month-to-month and per household. An indicative example of such plots is provided in Figure A4 of Appendix A, which present the diurnal pattern of variation of the four studied statistics for the above 15-minute records of House #1.

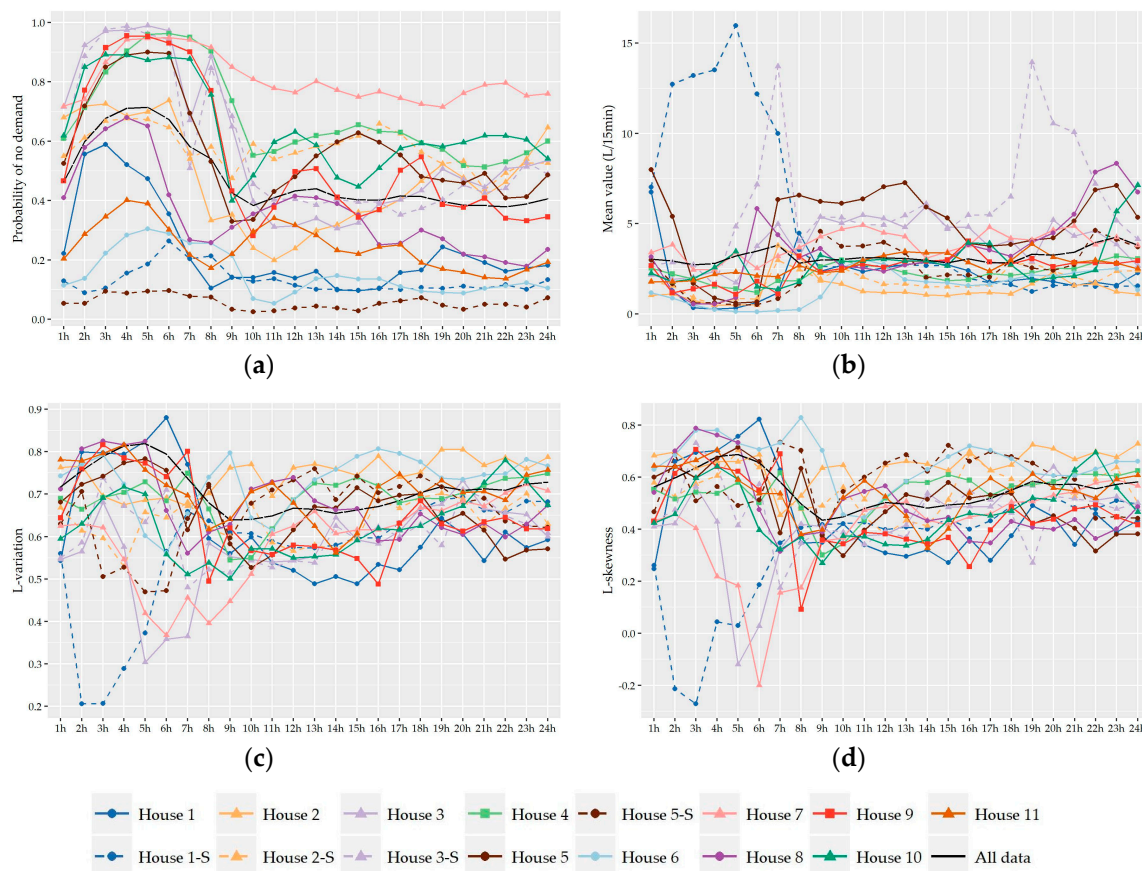
The visual analysis of these plots shows that only the four households (House #1, 2, 3, 5) whose statistics exhibit a significant differentiation among winter and summer months (see Section 2.1) exhibit a differentiation in the diurnal pattern of the statistics between the low and high consumption periods. To take into account this differentiation, the hour-to-hour analysis for these households was conducted by assuming two different periods: a winter and a summer (i.e., with irrigation use) diurnal pattern of variation of the statistics in which the statistical characteristics of each specific hour correspond to the same process. The “summer period” for House #1 extends from April to November and for House #2 from June to August., from May to October for House #3 and from June to August for House #5. Regarding the remaining households (House #4, 6, 7, 8, 9, 10, 11), the analysis will be conducted assuming that the marginal distribution of each specific hour remains invariant among the months since no significant deviation of the statistics of the individual month records from those without monthly separation is noted.

According to the above, for the seven households without monthly fluctuation, we formed 24 individual records (one record per hour) using the data of all months, assuming that within a specific hour the main statistical characteristics remain invariant. Regarding the remaining four households, the 24 individual records were formed separately for the winter and summer periods. The understudy statistics of these records at 15-minute and hourly scales are graphically presented in Figures 4 and 5, respectively, where the dotted lines in the plots correspond to the data of the summer months. Figures 4b and 5b present the diurnal pattern of variation of the mean value of the nonzero records at 15-minute and hourly records, respectively. It is interesting to note that,

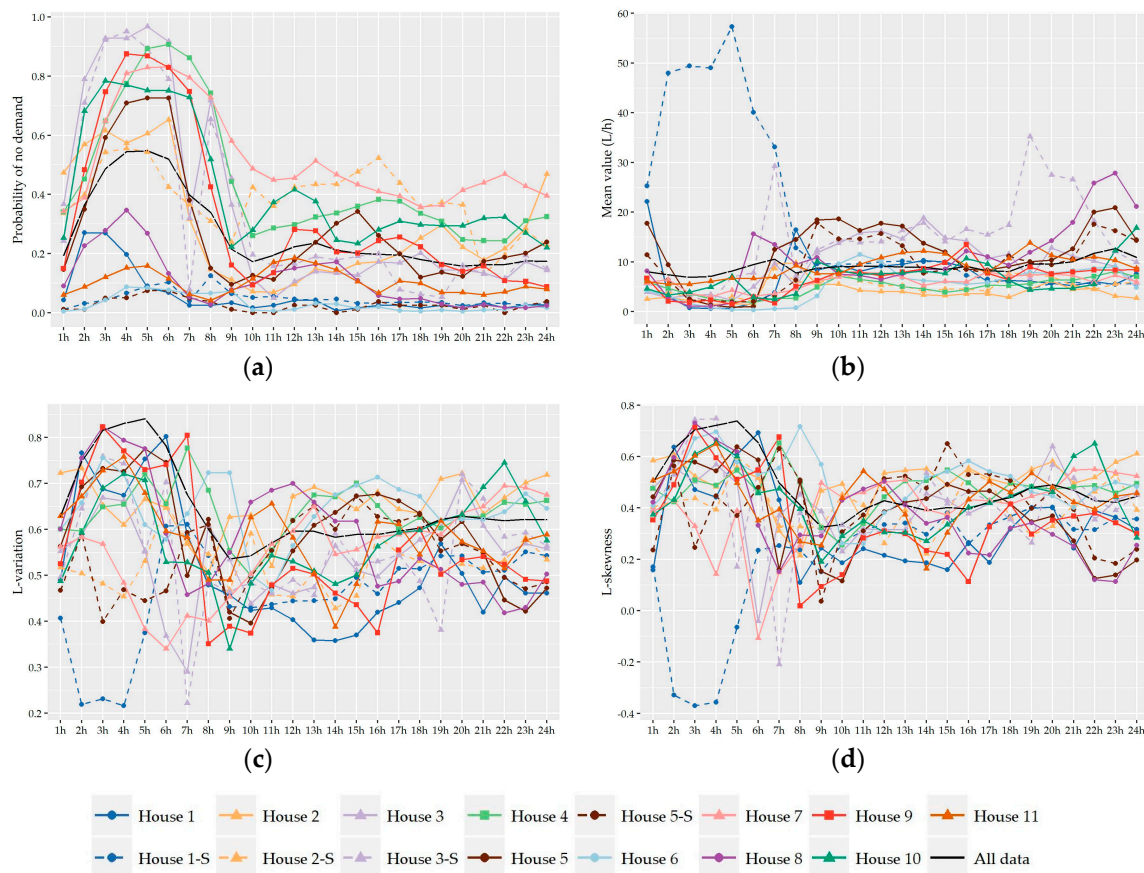
in almost all households, this fluctuation deviates from the typical diurnal pattern of the mean demand (i.e., series including zero values, which is described in the beginning of Section 4). This concerns the small difference in the mean values of nonzero data between daily and night hours and, mainly, the fact that, within these periods, this statistic does not exhibit significant shifts. On the contrary,  $p_{ND}$  (see Figures 4a and 5a) is in closer agreement, although contrasting with the known diurnal consumption pattern and exhibits higher hour-to-hour variability with sharper shifts during the daily hours. The combined study of these two statistics reveals that, in most understudy households, the hour-to-hour variation of total demand is the result of the fluctuation of the frequency of water use, as expressed via  $p_{ND}$ , rather of the intensity of use, as expressed via the mean of nonzero values.

The extremely large mean values observed during the night hours of the summer period for House #1 is attributed to the extensive grass irrigation while the extensive presence of large nonzero values in this sample has as a result low L-variation and L-skewness values with the latter becoming negative (i.e., left-skewed data).

Studying the diurnal pattern of the shape statistics, it is evident that the main differentiation is observed between the daily and night hours. As we see in Figures 4c and 5c, most values of L-variation are concentrated in a narrow interval, which range from 0.6 to 0.8 and from 0.5 to 0.7 for 15-minute and hourly records, respectively, while the variation from hour-to-hour is not very large especially during the daily hours (note that the y-axis of these plots spans the range covered by the estimated values to provide a more clear picture). Regarding L-skewness (see Figures 4d and 5d), most of the estimated values are gathered in a wider range that spans from 0.4 to 0.7 and from 0.2 to 0.6 for 15-minute and hourly records, respectively.



**Figure 4.** 15-minute scale: Variation per hour of the probability of no demand (a), mean value (b), L-variation (c), and L-skewness (d) of the 15-minute nonzero demand records analyzed.



**Figure 5. Hourly scale:** Variation per hour of the probability of no demand (a), mean value (b), L-variation (c), and L-skewness (d) of the hourly nonzero demand records analyzed.

Summarizing the above, we can say that, at both time scales, the statistic with higher hour-to-hour variation is  $p_{ND}$  while the mean values do not exhibit any significant changes during the hours of the day. Significant changes in these statistics as well as in the two shape characteristics are mainly observed between the daily and night hours.

#### 4. Probabilistic Models for Residential Water Demand

The analysis of previous sections revealed that the main statistical characteristics of the nonzero 15-minute and hourly water demand vary among households while significant variation was also noted on hourly and, in some cases, on a monthly basis. Further to that, as discussed in Section 1, the already developed simulation schemes for water demand incorporate different distributions to model either the individual components of demand events (i.e., occurrence, duration, and intensity) or the total demand magnitudes [5,12–18,22]. However, it is worth mentioning that these studies focus on the demonstration of the structure of the simulation model, typically fitted and validated on a short series from one household, rather on the provision of general insights regarding the suitability of specific distributions to model water demand and, hence, they have a local character in terms of statistical characterization of this process. Additionally, these models address the simulation problem at super-fine time scales (i.e., 1-s up to 1-min), which are not the temporal resolution of interest of the present study.

The above may lead one to assume that different probabilistic models should be employed to capture adequately the statistical behavior of water demand process at different households and time intervals (i.e., for different hours of the day or even months). Having said that, it is of high importance to provide rigorous insights regarding which distributions, from the vast amount available in the literature, can be considered more suitable to describe the understudy variables based on a large

dataset. This process, from a practical point-of-view, informs and facilitates future water demand modeling studies by restricting the search of candidate distributions to a small set and by highlighting which of them should be considered the first choice. Further to that, to provide a representative set of model parameters for the most suitable distributions, it is of high importance since these sets can be regarded and used either as default values in cases where no recorded data is available for a household or as prior estimations if the sample size is small and, subsequently, the sufficient fitting of a model is highly uncertain. The present section provides answers to the questions 3, 4, and 5, as posed in Section 1.

#### 4.1. Preliminary Screening of Models

A widely used approach for the preliminary discrimination of most consistent distributions to describe the understudy variable is the so-called “moment-ratio diagrams”. This type of analysis has been widely applied in the probabilistic study of hydrological variables especially in cases where a large number of records are examined since it provides a quick and direct evaluation of the suitability of several distributions from a statistical point-of-view by using just a single graphical instrument [38,41,42,45]. In the present work, we first transfer this framework in the analysis of large datasets of water demand records.

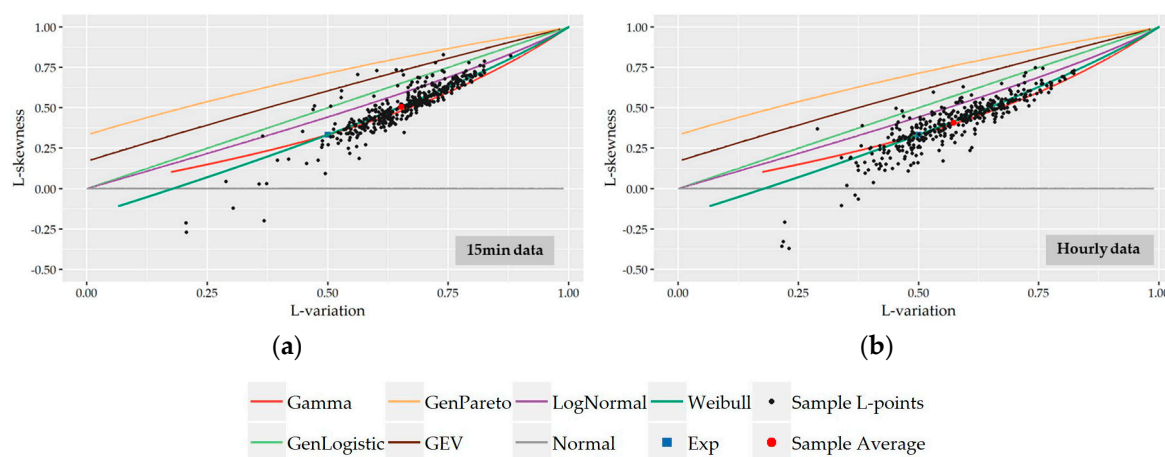
The moment-ratio diagrams provide a graphical comparison between the sample statistics, calculated from the records, and the theoretical ones as they are given by the formulas of the candidate parametric distributions. The proximity of the sample statistics to the theoretical locus (i.e., point, curve, or area) of a particular distribution provides an indication of the appropriateness of this distribution to describe the understudy records, which implies whether a specific set of a distribution parameter exists so as to reproduce the understudy moments. In such diagrams, a coefficient of variation vs. a coefficient of skewness or a coefficient of skewness vs. a coefficient of kurtosis are typically depicted while any pair of standardized moments could be also used. A comprehensive review of the conventional product moment-ratio diagrams summarizing 37 theoretical probabilistic models can be found in Vargo and Leemis [46].

During the last decades, the L-moments ratio diagrams, which were first introduced by Hosking [40], have gained popularity against conventional ones and have been widely used in many hydrological applications due to the better statistical properties of L-moments [42]. L-variation vs. L-skewness and L-skewness vs. L-kurtosis pairs are mostly used while the latter is more popular since L-variation is not well defined for distributions with a mean value equals zero or is negative. The theoretical locus of a distribution in L-ratio diagrams as well as in conventional moments diagrams can be a point, a curve, or an area depending on the number of parameters of the distribution. In L-skewness vs. L-kurtosis diagram, the type of the locus depends on the number of shape parameters, i.e., distributions without a shape parameter form a point (e.g., Normal distribution) with one shape parameter forming a curve (e.g., Gamma distribution) and with two shape parameters forming an area (e.g., Burr Type XII). On the contrary, the theoretical locus of distributions in an L-variation vs. an L-skewness diagram depends also on the location and scale parameters (if they are adjusted to take any value) and, subsequently, the number of distributions whose locus is a point and is lower. The present analysis was based on the L-variation vs. L-skewness diagram for positive random variables such as nonzero water demand is well defined and more robust than L-kurtosis [38].

The special characteristics of the 15-minute and an hourly nonzero demand process, as derived from the above analysis, determine the initial set of candidate probabilistic models that are going to be examined as more suitable to describe the understudy variables. First, the studied distributions should be defined in the real positive axis. Additionally, the high L-skewness and L-variation coefficients along with the great variability of those shape characteristics from household-to-household and hour-to-hour imply the use of distributions whose shape parameters support a range of values for those statistics. Further to that, we took into account the probabilistic models that are used by the simulation schemes to model water demand at super-fine time scales (e.g., the Exponential [1,19–22],

Weibull [5], Lognormal [15,17,47], and Normal [14]) as well as in the simulation of hydrological variables with similar characteristics.

Having said that, as candidate models to describe the understudy variables, we selected the Gamma (G) distribution, the Generalized Logistic (GL) bounded at zero from below, the Generalized Pareto (GP) bounded at zero from below, the Generalized Extreme Value (GEV) distribution bounded at zero from below, the Lognormal (LN) distribution and the Weibull (W) distribution. The Normal (N) and Exponential (EXP) distribution were also included in the analysis since they are very popular and widely used models. As we see in the L-variation vs. L-skewness diagrams of Figure 6, the locus of all understudy distributions is a curve except from that of the Exponential distribution, which is controlled via just one scale parameter and subsequently it has constant L-variation ( $\tau_3 = 0.50$ ) and L-skewness ( $\tau_3 = 1/3$ ). Normal distribution is symmetric about the mean, i.e.,  $\tau_3 = 0$ , and their theoretical locus is a horizontal line since L-variation can be adjusted to take any value by controlling the location and scale parameters.



**Figure 6.** L-variation vs. L-skewness ratio diagrams that compare the observed L-Points of the nonzero 15-minute (a) and hourly (b) water demand records of all households per hour against the theoretical L-points of the distributions under investigation.

Following the analysis of the variation of the two shape characteristics of the nonzero water demand at different temporal levels of analysis (i.e., monthly, daily, hourly), the preliminary identification of most suitable probabilistic models as well as their fitting and further performance evaluation was conducted on the records used in the hour-to-hour analysis, i.e., 24 individual records for the seven households without significant monthly variation and a different set of individual records for the winter and summer period for the remaining four households (i.e.,  $24 \times 7 + 24 \times 2 \times 4 = 360$  records). Since L-variation and L-skewness do not exhibit significant variation from hour-to-hour especially among successive hours, the following analysis could be conducted for forming larger individual records composed of 15-minute and hourly data of more than one hour intervals (e.g., assuming a variation between night and daily hours). However, we chose to keep the seasonal discretization of the previous section in order to study a larger cluster of L-points.

Subsequently, the evaluation of the suitability of various distributions was not conducted on the basis of 11 series of households but on the basis of 360 individual records.

The observed L-points ( $\tau_2, \tau_3$ ) of the understudy nonzero 15-minute and hourly water demand records along with the theoretical locus of the examined distributions are presented in Figure 6. Visually, it is apparent that Gamma and Weibull distribution provide the best approximation to the great majority of observed L-points at both time scales while a smaller cluster of points is concentrated closer to the Lognormal distribution especially at an hourly scale. As we see, the average position of the sample 15-minute L-points (0.63,0.50) lies upon the theoretical curve of Weibull distribution while, at the hourly scale, the average point (0.58,0.40) is among the curves of Gamma and Weibull

distributions close to the point of their interception. It is worth noting that the sample L-points, which are out of the main body of the cluster (i.e.,  $\tau_2 \leq 0.50$  and  $\tau_3 \leq 0.30$ ) and away to the curves of the above distributions, correspond to the records of night hours, i.e., 12:00 to 8:00 am, where the probability of zero demand is high and the sample size of nonzero records is smaller. Since the variability of statistics is closely associated with sample size, it is expected that these points would be closer to the main scatter of L-points and closer to the distributions if a greater number of records was available.

The preliminary selection of the three above discussed distributions (i.e., Gamma, Weibull, and Lognormal) as more suitable to describe nonzero demand is further supported by the complementary L-variation vs. L-skewness diagrams provided in the Appendix B. Figure A5 present the sample L-points ( $\tau_2, \tau_3$ ) of the individual records of each hour assuming also monthly separation (i.e.,  $12 \times 24 = 288$  points for each household), at a 15-minute and hourly scale, respectively. As shown, the sample L-plots are mostly concentrated around the three above mentioned distributions while the average L-point lies upon the curve of Weibull distribution. Additionally, Figure A5 present the L-points ( $\tau_2, \tau_3$ ) of the entire nonzero record of each household (12 points) without assuming seasonal discrimination at 15-minute and hourly time scales, respectively, while Figure A5 contain the L-points of the records of each hour of the day including the records of all households at 15-minute and hourly scale, respectively (24 points). In all cases, the sample L-points are very close to the curve of either Gamma or Weibull distribution.

#### 4.2. Evaluation of the Models

The preliminary evaluation from the previous section showed that Gamma and Weibull distribution seem to be the most suitable probabilistic models to describe at least in terms of the first three L-moments. The great majority of the nonzero 15-minute and hourly water demand records of all households and all hours of the day. At the same time, it seems that a smaller percentage of records is better captured by the Lognormal distribution. The performance of the three distributions is further evaluated and compared after their actual fitting to the above described records. In this section, we present the evaluation of the three models on the basis of fitting error measures while the fitting procedure along with the analysis of the values of distribution parameters are presented in the next section.

The three probabilistic models, which are examined, have two parameters and more specifically one scale and one shape parameter with the latter controlling the asymptotic behavior of the distribution tail. The two-parameter Gamma distribution is one of the most commonly used probabilistic models in hydrological applications due to its special characteristics, i.e., it is defined only for non-negative values of the variable and is positively skewed. Its pdf is given by the equation below.

$$f_G(x) = \frac{1}{\beta\Gamma(\gamma)} \left(\frac{x}{\beta}\right)^{\gamma-1} \exp\left(-\frac{x}{\beta}\right), x > 0 \quad (3)$$

where  $\beta > 0$  is the scale parameter and  $\gamma > 0$  is the shape parameter. For  $\gamma < 1$ , the pdf is J-shaped with an infinite ordinate at  $x = 0$  and its right tail decreases slightly faster than the exponential tail. For  $\gamma = 1$  the distribution is identical with exponential whereas, for  $\gamma > 1$ , the pdf is bell-shaped and it has a tail that decreases slightly slower than the exponential one. For large values of shape parameters ( $\gamma > 15$ ) Gamma distribution approaches the normal.

The Weibull distribution is another popular probabilistic model and its pdf is given by the equation below.

$$f_W(x) = \frac{\gamma}{\beta} \left(\frac{x}{\beta}\right)^{\gamma-1} \exp\left(-\left(\frac{x}{\beta}\right)^\gamma\right), x > 0 \quad (4)$$

where  $\beta > 0$  is the scale parameter and  $\gamma > 0$  is the shape parameter. Like Gamma, Weibull distribution is positively skewed and defined for non-negative values of the variable while the value of shape

parameter controls the form of the pdf. For  $\gamma < 1$ , the pdf is J-shaped with an infinite ordinate at  $x = 0$  while, for  $\gamma = 1$ , the pdf tends to  $1/\gamma$  as  $x$  approaches 0 and then decreases exponentially. For  $\gamma > 1$ , the pdf tends to 0 at  $x = 0$  from above, increases until the mode, and decreases after the mode. Regarding the asymptotic behavior of the tail, for  $\gamma < 1$ , the Weibull distribution has a heavier tail (i.e., tending to zero less rapidly) than the exponential one while, for  $\gamma > 1$ , the tail decreases faster than the exponential.

Lastly, the pdf of Lognormal distribution is given by the equation below.

$$f_{LN}(x) = \frac{1}{\sqrt{\pi}\gamma x} \exp\left(-\ln^2\left(\frac{x}{\beta}\right)^{1/\gamma}\right), x > 0 \quad (5)$$

where  $\beta > 0$  is the scale parameter and  $\gamma > 0$  is the shape parameter. A random variable  $x$  is lognormally distributed if random variable  $y = \ln(x)$  is normally distributed. Due to this transformation, Lognormal distribution is defined for non-negative values of the variable, is positively skewed, and has a bell-shaped pdf. It is worth noting that the Lognormal distribution is commonly parameterized by using the mean and standard deviation of the distribution on the log scale. In the present work, we use the shape and scale parameterization concept to facilitate the direct comparison with the other two understudy distributions. Regarding extremes, Lognormal distribution is considered a heavy-tailed distribution.

The three probabilistic models are characterized by different tail behaviors, which have been extensively assessed and compared in the reproduction of extremes of different hydrometeorological variables such as rainfall and runoff (e.g., see Reference [48] and the literature herein). Summarizing the above, further to the Lognormal distribution that is considered heavy-tailed distribution, Weibull can be both heavy-tailed and light-tailed depending on the shape parameter while Gamma is considered essentially as an exponential tail distribution. The ordering of these distributions from a heavier to a lighter tail is Lognormal, Weibull with  $\gamma < 1$ , Gamma and Weibull distribution with  $\gamma > 1$ .

It is worth noting that the three above described distributions as well as exponential distributions are special cases of the Generalized Gamma distribution [49]. This distribution is very flexible but more complicated. It is comprised of one scale and two shape parameters that govern the form of the pdf and the behavior of the left and right tail. Although this model can be clearly regarded as a “universal model” for the 15-minute and hourly nonzero water demand, we prefer in this scenario to focus our study on two parameter models for the sake of parsimony and simplicity. Furthermore, the estimation of three parameters instead of two is a more uncertain, sensitive, and complicated procedure, requiring very long records of data. The appropriate reproduction of extremes is also of high importance in the sufficient modeling of residential water demand since the correct estimation of peak flow demand plays a crucial role in the efficient design, planning, and management of urban water systems [6,50]. Subsequently, a naïve selection of the probabilistic model may result in the overestimation or underestimation of maximum water demand magnitudes. In this respect, further to the standardized mean square error (MSE) that was used as the main “goodness-of-fit” measure to evaluate the overall fitting performance of the above three distributions, two additional error measures that indicate the performance in terms of extreme values were employed. A similar set of measures were employed by Papalexiou and Koutsoyiannis [38] to evaluate the distributions’ performance on daily precipitation.

The MSE considers the relative error between the empirical and the theoretical quantile values.

$$ER_1 = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_u}{x_{(i)}} - 1 \right)^2 \quad (6)$$

where  $x_u$  is the value predicted by the theoretical quantile function of the distribution where  $u$  equals the empirical probability  $p_i$  of  $x_{(i)}$  (the  $i$  indicates the position of the  $i$ th element in the sample of length



$n$ , ranked in ascending order), according to the Weibull plotting position, i.e.,  $p_i = i/(n + 1)$ . In this case, taking into account the observations on the form of empirical pdf of nonzero water demand records, we preferred to use a standardized measure instead of the classical MSE in order to avoid the dominance of large values in the estimation of the error. The fitted distribution with the smallest  $ER_1$  value is considered as better.

Complementary to the MSE, the first measure, focused on the extremes, considers the relative error between the  $m$  largest sample values and the corresponding theoretical values.

$$ER_2 = \frac{1}{m} \sum_{i=n-m+1}^n |x_u - x_{(i)}| \quad (7)$$

In the present work,  $ER_2$  was estimated on the 10 largest values for each of the understudy record and distribution while the better fitted distribution is that with the smallest  $ER_2$  value. The second measure assesses the percentage difference between the maximum sample value and the predicted maximum value.

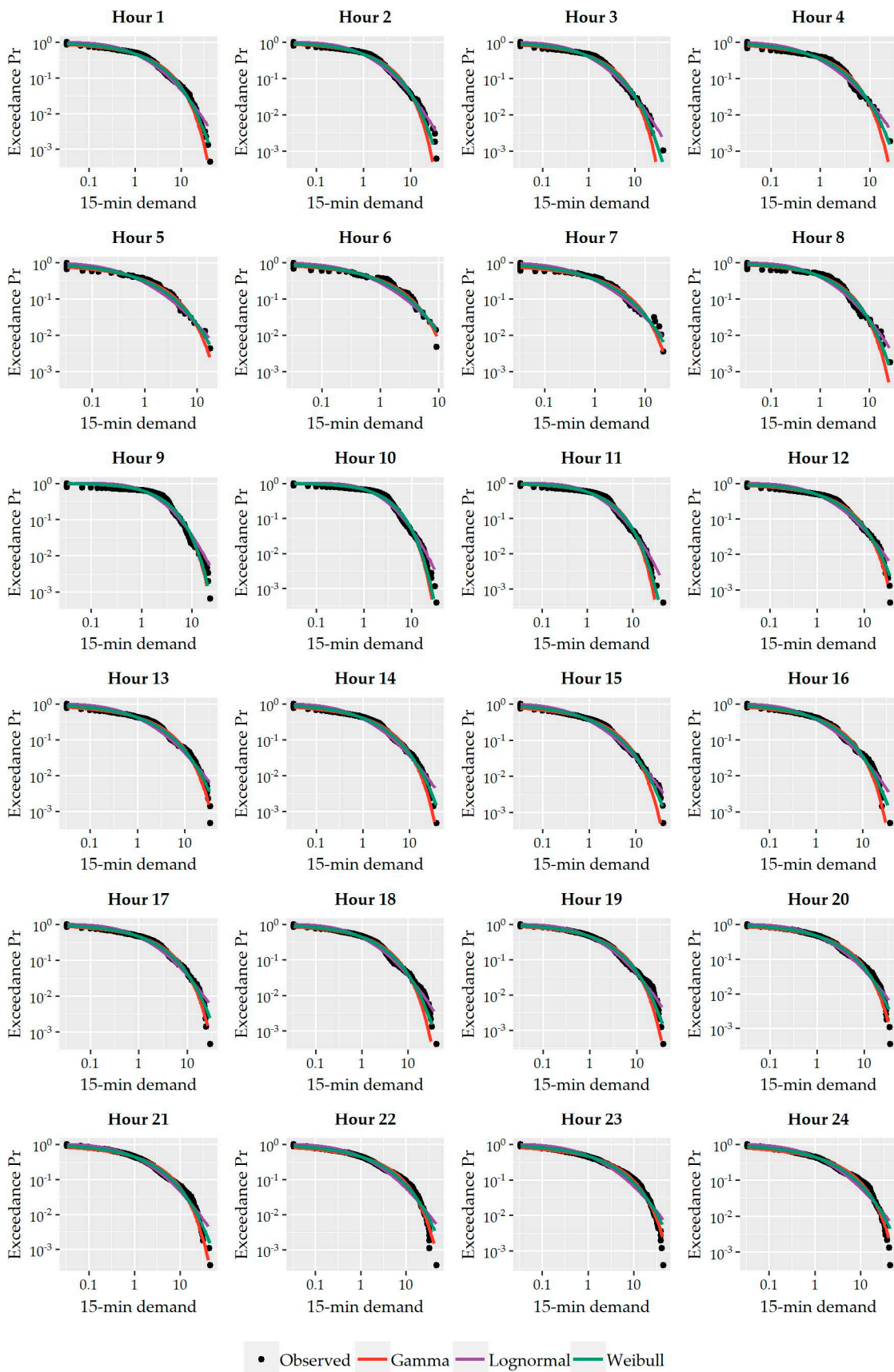
$$ER_3 = \left( \frac{x_u}{x_{(n)}} - 1 \right) 100 \quad (8)$$

The negative or positive percentage difference in  $ER_3$  implies, respectively, underestimation or overestimation of the maximum value by the fitted distribution. This error measure along with  $ER_2$  provides a clear indication on which type of tail behavior of the three understudy probabilistic is in better agreement with the observed water demand extremes.

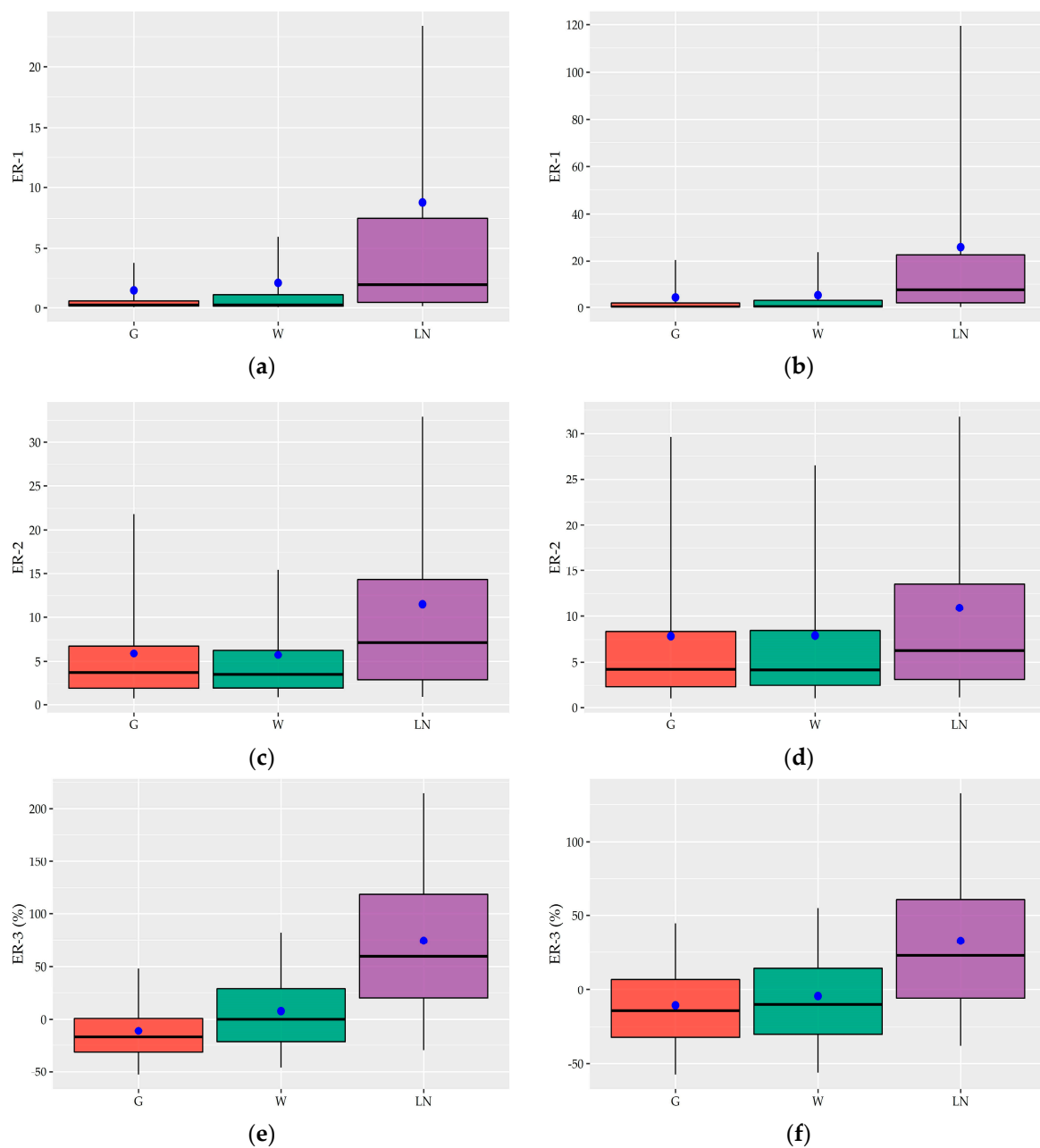
The three distributions were fitted on the individual records described in the previous section as well as on the entire record of each household (i.e., 11 series) and each hour of the day (i.e., data from all households for each hour of the day, 24 series) for 15-minute and hourly time scale. Since all three distributions that are studied here are positively skewed, it was not possible to obtain a set of parameters for the individual records with negative or zero L-skewness. As discussed in Section 3.2, this is the case for seven 15-minute records and 10-h records, which correspond mainly to night hours (02:00–07:00 am) during the summer period. These records were excluded from the below evaluation of fitting, which is flagged as “NA” (i.e., Not Available) in Tables A1 and A2 in Appendix B that present the best fitted distribution on the understudy 15-minute and hourly nonzero water demand records, respectively, according to  $ER_1$  measures.

Figure 7 presents, indicatively, the empirical probability plots per hour of the three fitted distributions to the nonzero 15-minute water demand values of House #4. All three distributions fit well the sample data while the different right tail behavior of each distribution is clearly illustrated with a Lognormal distribution having a heavier tail than the other two. Similar behavior was observed on the plots of the remaining households at both time scales.

The fitting performance of G, W, and LN distribution on the individual records are graphically illustrated in terms of box plots (Figure 8) while Table A3 in Appendix B summarizes the main statistical characteristics of the three error measures along with the percentage of records in which each distribution exhibits a better performance. Figure 8 and Table A3 show that G and W distributions perform much better than the LN. According to the  $ER_1$  measure that assesses the overall fitting, the former two distributions achieve a very good fit to the great majority of individual 15-minute and hourly records (i.e., 85% of 15-minute records and 95% of hourly records) while the analysis of individual values revealed that they fitted equally well with the LN distribution in the records in which the latter is appears to be superior (i.e., mainly in the records of House #5 and #2). As we see in Figure 8a,b, the LN distribution is characterized by very high mean  $ER_1$  values, i.e., 8.78 and 25.81 in 15-minute and hourly records, respectively, along with much wider 90% ECI. In Tables A1 and A2, it becomes evident that the three study distributions do not dominate in specific time intervals or households in terms of the  $ER_1$  measure since G and W distributions are the best fitted models in the records of all households.



**Figure 7.** Empirical probability plots per hour of the three fitted distributions to the nonzero 15-minute water demand of House #4.



**Figure 8.** Box plots depicting the error measures that evaluate the fitting performance of the Gamma (G), Weibull (W), and Lognormal (LN) distribution to the nonzero 15-minute (**left column, a,c,e**) and hourly (**right column, b,d,f**) of 384 individual records. The blue points represent the mean value of the sample.

Focusing on the fitting measures regarding extremes, G and W distributions perform equally well with mean  $ER_2$  equal to 5.92 and 5.77, respectively, in 15-minute records, and 7.85 and 7.93, respectively, in hourly records. On the contrary, the LN distribution has higher mean values such as 11.53 and 10.6 at 15-minute and hourly scale, respectively, and wider 90% ECI. However, its performance seems to be improved in the hourly records. In terms of the  $ER_3$  measure that evaluates the difference between the predicted and observed maximum value, LN distribution, which is considered a heavy-tailed model, tends to overestimate the maximum on average 74.31 and 32.76 at 15-minute and hourly scale, respectively, which have very wide 90% ECI, ranging from  $-29.73\%$  to  $214.49\%$  and from  $-38.26\%$  to  $132.62$ , respectively. In contrast, both G and W distributions are able to produce maximum values that are in closer agreement with the observed one with mean  $ER_3$  equals to  $-11.27$  and  $7.72$ ,

respectively, in 15-minute records and  $-10.65$  and  $-4.45$ , respectively, in hourly records while the 90% ECI range of variation is much narrower than the corresponding LN distribution. Figure 8e shows that G distribution tends to underestimate the predicted maximum value of 15-minute records while the  $ER_3$  values of W distribution have a uniform dispersion around the median that is very close to zero. Regarding hourly records, as Figure 8 presents, the two distributions show similar performance, underestimating the maximum. Taking into account the strictness of the last assessment criterion that examines values with exceedance probability very close to 0 along with the results of the  $ER_2$  measure, the general performance of G and W distribution in terms of extremes can be characterized as exceptionally good at both time scales.

Further to individual records, the three distributions were also fitted and evaluated on the entire record of each household as well as on the data of the “parent household” (i.e., 12 series). Table A4 in the Appendix B reveals that G and W distributions outperform LN in terms of all error measures in all cases and achieve an equally good fitting to the observed data. Focusing especially on  $ER_2$  and  $ER_3$  measures, it is evident that the fitting of LN distribution to the sample tail is poor, overestimating the maximum value. Since the entire records of each household are much larger than the individual one allowing a more accurate estimation of parameters, this can be regarded as an additional important indication of the non-suitability of this heavy-tailed distribution to model the nonzero 15-minute and hourly demand values.

The above analysis provided insights in the two of the questions posed in the beginning of this sections, i.e., which are the most suitable models to simulate the understudy variables and whether there is one or a set of models that prevail the others. As shown G, W and LN distributions are able to describe the great majority of nonzero water demand records, while the more thorough evaluation of the three models revealed the superiority of the two former models compared to the LN distribution. At the same time, G and W distributions show an almost equally good performance, at least on the basis of three error measures.

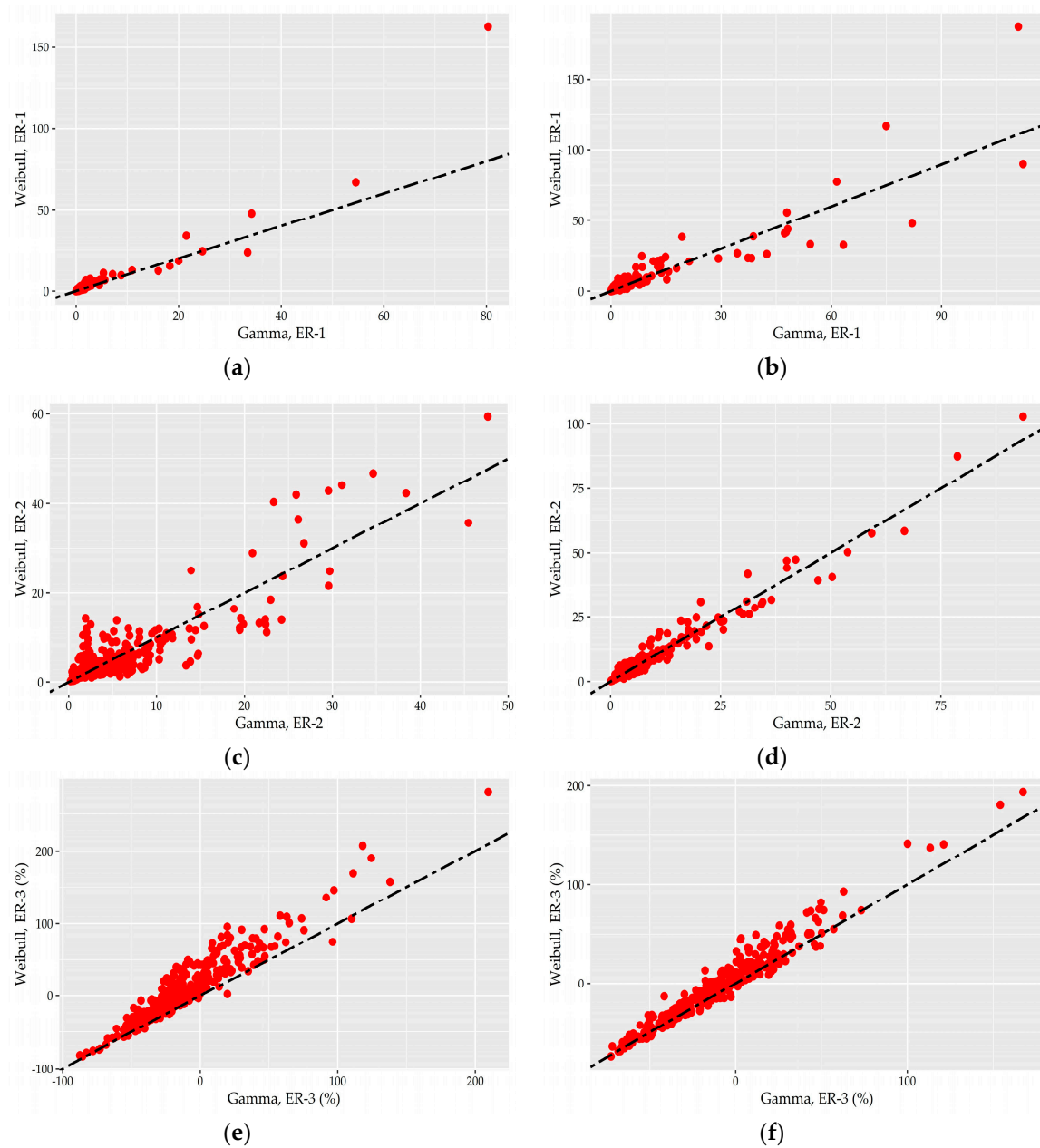
However, it is of high importance to investigate whether G and W distributions are complementary in the sense that one model achieves a good fitting on the records in which the other model has poor performance and vice versa or the two models have similar performance in the same records. This directly answers the third question posed in the beginning of Section 4, i.e., whether we can use just one probabilistic model to reproduce the characteristics of nonzero water demand records of all households and all hours of the day, indicating if there is just one distribution that should be the first choice in the modeling of nonzero water demand at these time scales.

A direct comparison of the G and W distribution is provided in the scatter plots of Figure 9 that depict the error measures of the fitting of the two models on individual records. In the plots corresponding to  $ER_1$  and  $ER_2$  measures, a point above the diagonal black dashed line implies better performance for G distribution for this record while the points below the line corresponds to records in which W distribution achieved a better fit. In the plots depicting  $ER_3$  values, the points with both coordinates greater than zero imply an overestimation of the maximum from the two models and can be interpreted as the plots of the other two error measures while the opposite stands for the points with both coordinates lower than zero.

As is apparent in all box plots, the performance of G and W distributions follows a similar pattern in the sense that they exhibit a poor fit in the same few records, which correspond to the night hours of summer and winter period of House #1 while an almost identical good fitting was achieved for the remaining records. Similar scatter plots presenting the points of the LN distribution show that the fitting error for the “poor-fitted” points is even higher. Furthermore, we can see that especially the  $ER_2$  and  $ER_3$  points at the hourly scale are dispersed closely around the diagonal line, which implies an almost identical fitting performance.

The above observations provide strong evidence that the G and W distributions cannot be regarded as complementary models but as essentially equally good models. However, if there was a necessity for the discrimination of just one model, then Weibull distribution should be the first choice

to represent the nonzero water demand data both at 15-minute and hourly time scales due to its ability to better reproduce the observed extremes.

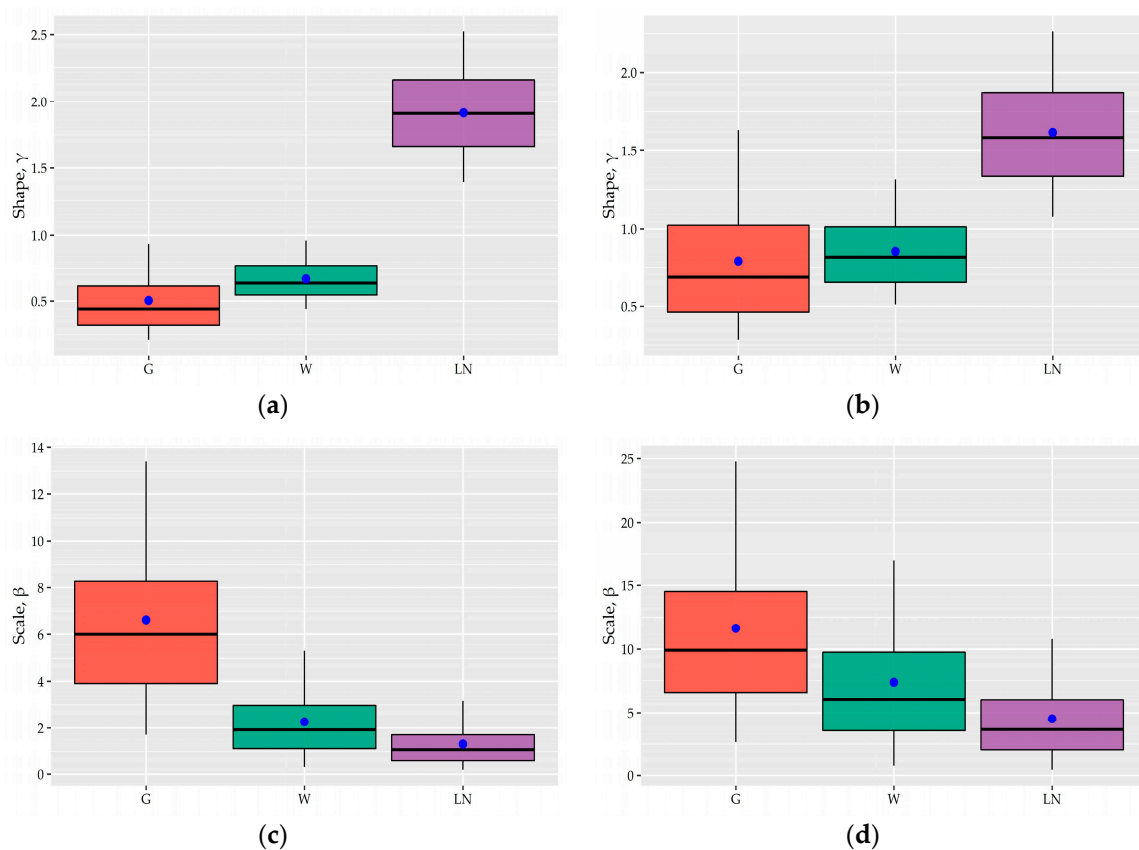


**Figure 9.** Scatter plots that compare the fitting performance of the Gamma and Weibull distributions for the individual records at 15-minute (left column, a,c,e) and hourly (right column, b,d,f) scale. Red points represent the values of the goodness-of-fit measures of individual series.

### 4.3. Fitting and Parameters of the Models

In the present section, we describe the fitting procedure of the G, W, and LN distributions to the nonzero records water demand records, providing the actual values of their scale and shape parameters. From a practical point-of-view, the derive sets of parameters consist of valuable prior information, e.g., for instance, using the mean values of the parameter as the initial estimation especially in cases where no recorded data is available for a household or the sample size is small and, subsequently, the sufficient fitting of a model is highly uncertain.

To fit the three models, the method of L-moments was followed by using the analytical formulas for the first three L-moments for each distribution. This method is analogous to the usual method of moments, according to which the parameter estimates are obtained by equating the first  $p$  sample L-moments to the corresponding distribution moments. Then, by solving the system of equations, we obtain the shape and scale parameters of the distribution. The first three L-moments for the three understudy distribution (e.g., [40]) as a function of scale parameter  $\beta$  and shape parameter  $\gamma$  are provided in Table A5 in Appendix B. The method of L-moments were implemented in the R programming environment [51] with the use of the “lmom” package [52]. Scale and shape parameters were obtained for the nonzero 15-minute and hourly time scales and the individual records described in the previous section as well as on the entire record of each household (i.e., 11 series) including each hour of the day (i.e., data from all households for each hour of the day, 24 series). As explained in the previous section, the records with negative or zero skewness on which it was not possible to fit the three distributions were excluded from the analysis. The scale and shape parameters for 15-minute and hourly individual records are presented in the form of box plots in Figure 10 while their summary statistics are given in Table A6 in Appendix B. With regard to G and W distributions, we also included in the present analysis the shape and scale parameters for the LN distribution.



**Figure 10.** Box plots depicting the error measures that evaluate the fitting performance of the Gamma (G), Weibull (W), and Lognormal (LN) distribution to the nonzero 15-minute (left column, a,c) and hourly (right column, b,d) 384 individual records. The blue points represent the mean value of the sample.

As evident in Figure 10, the shape parameters of the two former models and especially those of W distributions vary in a very narrow range at both time scales (i.e., consider that the theoretical range of this parameter is  $(0, \infty)$ ). The values of shape parameters for both W and G distribution are lower than 1 for the great majority of 15-minute records, i.e., the 90% ECI is  $(0.44, 0.96)$  and  $(0.21, 0.93)$ , which indicates J-shaped density functions. Regarding the hourly records, a greater number of records

results in bell-shaped densities with  $\gamma > 1$  since the 90% ECI of  $\gamma$  for W and G distribution is (0.51,1.32) and (0.28,1.63), respectively. As is also apparent in Table A7, the variability of the shape parameter of the fitted distributions to the entire records of each household is even lower with all values being significantly less than 1.

As expected, the scale parameters vary in a wider range since this type of parameter depends on the unit of measurement and the values of a variable. The effect of the scale parameter is to stretch out the pdf of the distribution. As we can see in general, the scale parameters of hourly records are greater than the corresponding 15-minute records while, for both time scales, the empirical 90% ECI of the  $\beta$  parameter of Weibull distribution, i.e., (0.34,5.31) and (0.81,16.99) for 15-minute and hourly records, respectively, is much narrower than that of Gamma distribution, i.e., (1.71,13.41) and (2.67,24.77) for 15-minute and hourly records, respectively.

## 5. Discussion and Conclusions

The present study examines the statistical characteristics and peculiarities of 15-minute and hourly residential water demand based on the “*iWIDGET dataset*” consisting of 11 long records (i.e., from one up to two years data) from single-households in Greece. The dataset is composed by representative households with different socio-demographic characteristics allowing the statistical characterization of water demand under different consumption patterns and profiles. Furthermore, in order to provide more rigorous and concrete insights on the distributional properties of water demand, the analysis was conducted on the basis of 360 individual records, which are derived after the study of seasonal variation of the main statistical characteristics of 15-minute and hourly water demand data.

In the analysis, water demand was treated and examined as an intermittent process with a mixed-type marginal distribution, which is composed by a discrete part and mathematically expressed by the probability of no demand and a continuous part that describes probabilistically the nonzero demand magnitudes. Following this modeling strategy, we examined, (a) which statistical characteristics of the marginal distribution of water demand (i.e., probability of no demand as well as the mean value, L-variation, and L-skewness of the nonzero values) vary the most at different temporal levels (i.e., monthly, daily, hourly basis) and which can be regarded as essentially invariant and (b) which probabilistic models are considered more suitable to describe the continuous part (i.e., nonzero values) of the distribution.

Both questions are of high importance within the context of probabilistic modeling and stochastic simulation of residential water demand. From a practical point-of-view, the first question provides insights regarding the appropriate assumptions on the structure (e.g., stationary or cyclo-stationary) of the simulation scheme that should be employed to model the variability at these time scales while the second one determines the type of models from the vast amount available in the literature that should be considered as the first choice in the probabilistic description of residential water demand.

A first coarse comparison of the above described statistical characteristics on the basis of households indicated that the statistics with higher variation include the mean value and the probability of no demand, following the typical patterns observed in the literature, i.e., households with kids or water needs for garden irrigation are characterized by higher mean nonzero water demand and lower probabilities of no demand. Regarding the shape characteristics, the records of all households are characterized by especially high variability (in terms of L-variation) and positive skewness (in terms of L-skewness) irrespective of the corresponding mean value. As the analysis revealed, these two statistics vary in a very narrow range despite the differences in the socio-demographic characteristics of the understudy households.

Furthermore, we studied the seasonal variation of the above characteristics for all the records of all households on a monthly, a daily, and an hourly basis. The analysis on a monthly basis showed that significant changes in the statistics are observed between the summer and winter months especially in the households with irrigation needs while, within the two periods, no significant variation is observed.

At both time scales, the statistic with a higher monthly variation is the probability of no demand,  $p_{ND}$ , followed by the mean value while the changes of L-variation and L-skewness is much less.

The studied statistics of the demand records of all households remain almost invariant from day-to-day at both time scales, which support the assumption of stationarity at these temporal levels.

Lastly, we examined the diurnal pattern (hour-to-hour) of variation of the above properties. As on the monthly basis,  $p_{ND}$  exhibits the highest fluctuation with sharper shifts during the daily hours. This observation along with the fact that, in most cases, the mean value of nonzero demand remains almost invariant during the day indicates that the widely known diurnal pattern of water demand (including zero values) is the result of the fluctuation of the frequency of water use, as expressed via  $p_{ND}$ , rather than of the intensity of use, as expressed via the mean of nonzero values. On the contrary, the two shape characteristics shift between the daily and night hours.

Next, we examined the suitability of different probabilistic models to describe the nonzero water demand of all households at a 15-minute and an hourly scale. Towards this, we use the L-moments ratio diagrams to preliminary screen 10 distributions in terms of reproducing the shape characteristics of the understudy records. This analysis highlighted the superiority of Gamma and Weibull distributions in terms of the first three L-moments to model the great majority of the nonzero 15-minute and hourly water demand records of all households and all hours of the day. At the same time, a smaller percentage of records is better described by the Lognormal distribution.

The performance of the three distributions were further evaluated on the basis of three error measures. The comparison of models' performance shows that Weibull and Gamma distribution achieve a very good fit to the great majority of 15-minute and hourly records even in the cases where the Lognormal is appeared as superior. Focusing on the fitting measures regarding extremes, Gamma and Weibull distributions perform equally well while Lognormal distribution, which is considered a heavy-tailed model, tends to overestimate the observed extremes.

A more detailed comparison of the performances of the two distributions revealed that the Gamma and Weibull distributions cannot be regarded as complementary models but as equally good models. If there is a necessity for a clear indication of just one model, the better performance of Weibull distribution in terms of the maximum value consists of the latter model as more suitable to describe the 15-minute and hourly nonzero water demand records of all households and time intervals.

As a rule of thumb, Weibull distribution can be the first choice in the probabilistic modeling of water demand records at these time scales.

The actual fit of the above distributions to the understudy records provided ranges for the parameters of the models. In the great majority of 15-minute records, the shape parameter  $\gamma$  of Weibull and Gamma distribution is much lower than 1 indicating J-shaped density functions while, at an hourly scale, a greater number of records have bell-shaped densities with  $\gamma > 1$ . It is worth noting that the range of variation of Weibull parameters is much narrower than the corresponding of Gamma distribution, which adds an extra argument to favor the former model in the statistical modeling of water demand. The derived set of model parameters with a relative small variation of shape statistics among households and time intervals can be regarded and used either as prior estimations in cases where no recorded data is available for a household or the sample size is small and, subsequently, the sufficient fitting of a model is highly uncertain. In such cases, the mean parameter values can be used, i.e., for a 15-minute time scale,  $\gamma = 0.67$  and  $\gamma = 0.5$  for the Weibull and the Gamma distributions, respectively, and for the hourly time scale,  $\gamma = 0.79$  and  $\gamma = 0.85$  for the two models, respectively.

**Author Contributions:** This manuscript is the result of the research of P.K. under the supervision of C.M.

**Acknowledgments:** The dataset used in the present analysis was collected in the framework of iWIDGET Project (2012–2015), received funding from the European Union Seventh Framework Programme (Grant Agreement No. 318272). Part of the present work was carried out in the framework of “EFFICIENT, EFFECTIVE, ECONOMIC WATER DEMAND MANAGEMENT IN THE GROWING OSLO CITY” (E<sup>3</sup>WDM) project (2015–2018), funded by the Norwegian Regional Research Fund and Oslo Water and Sewerage Works (Contract N<sup>o</sup> 245652). All analyses were implemented in the R programming environment [51] while all graphs were produced with the use of ggplot2

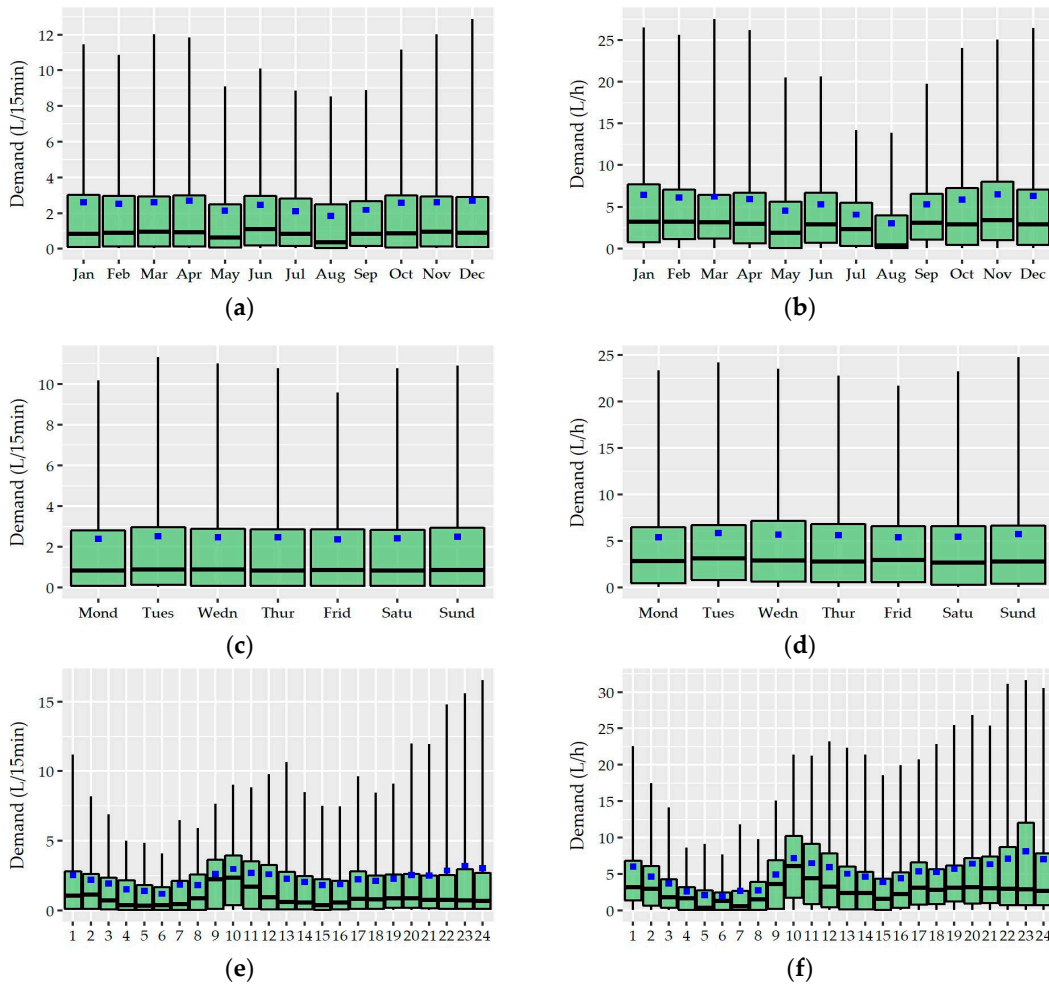


package [53]. Lastly, we would like to sincerely thank I. Tsoukalas for his valuable comments and suggestions, which substantially improved the present paper.

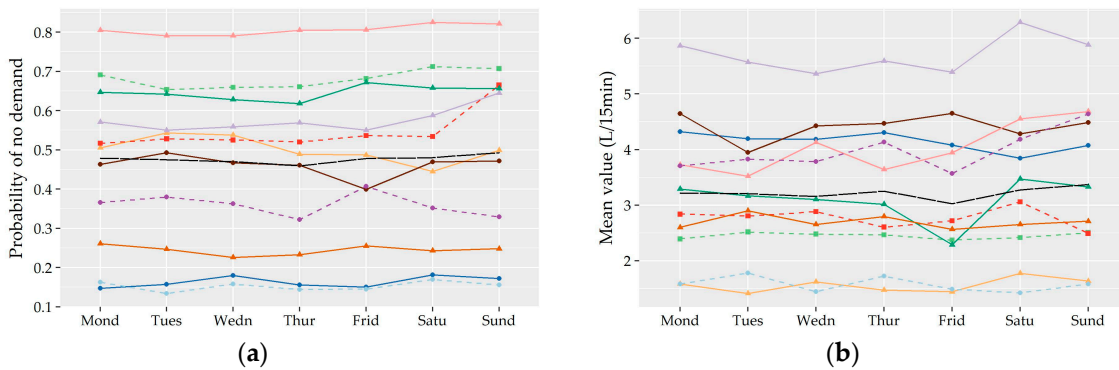
**Conflicts of Interest:** The authors declare no conflict of interest.

**Appendix A**

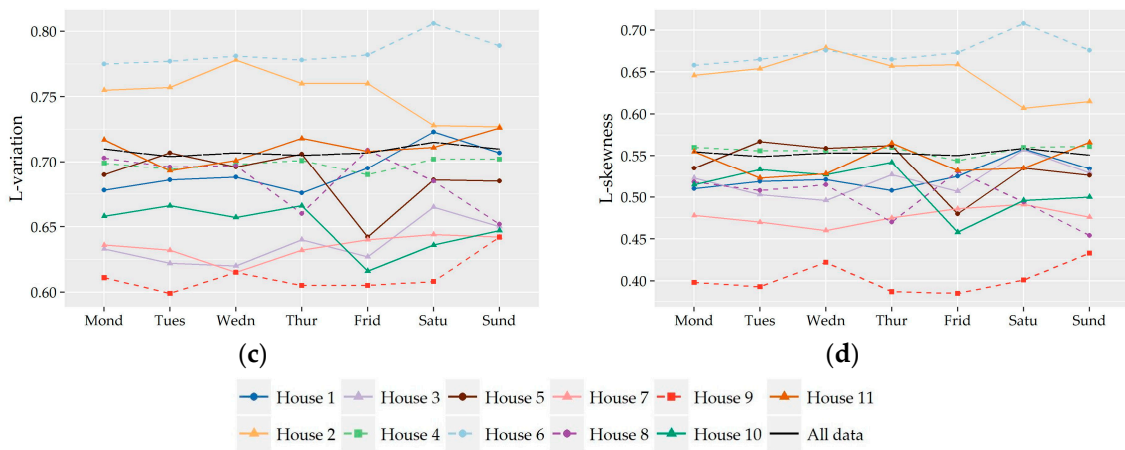
Appendix A presents complementary graphs associated with the analysis of seasonal variation of nonzero water demand, which is presented in Section 3.



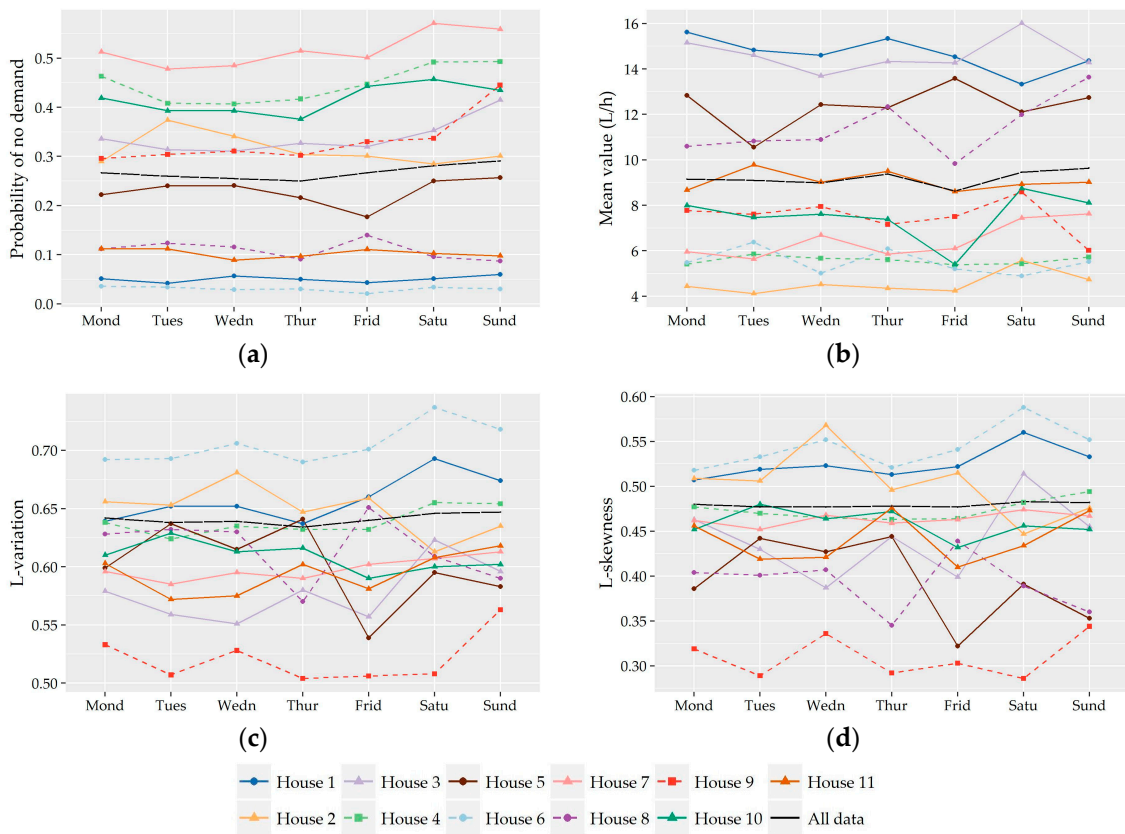
**Figure A1.** Box plots depicting the nonzero 15-minute (left column) and hourly (right column) water demand records of House #4, on monthly (a,b); daily (c,d) and hourly basis (e,f).



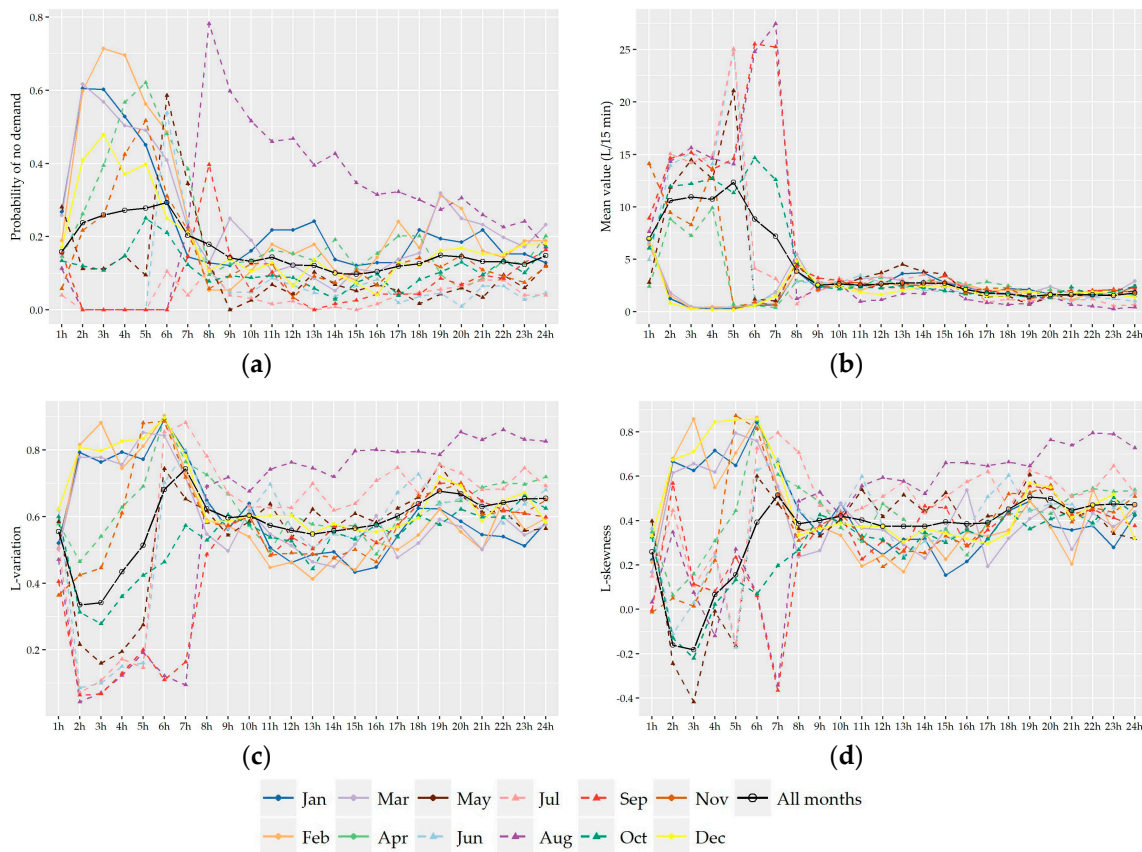
**Figure A2.** Cont.



**Figure A2.** 15 min scale: Variation per day of the probability of no demand (a), mean value (b), L-variation (c), and L-skewness (d) of the 15-minute nonzero demand records.



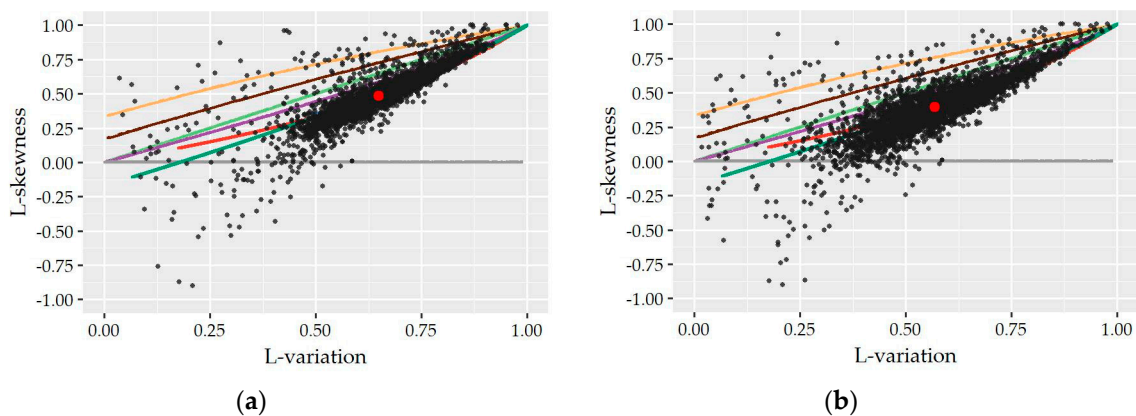
**Figure A3.** Hourly scale: Variation per day of the probability of no demand (a), mean value (b), L-variation (c), and L-skewness (d) of the hourly nonzero demand records.



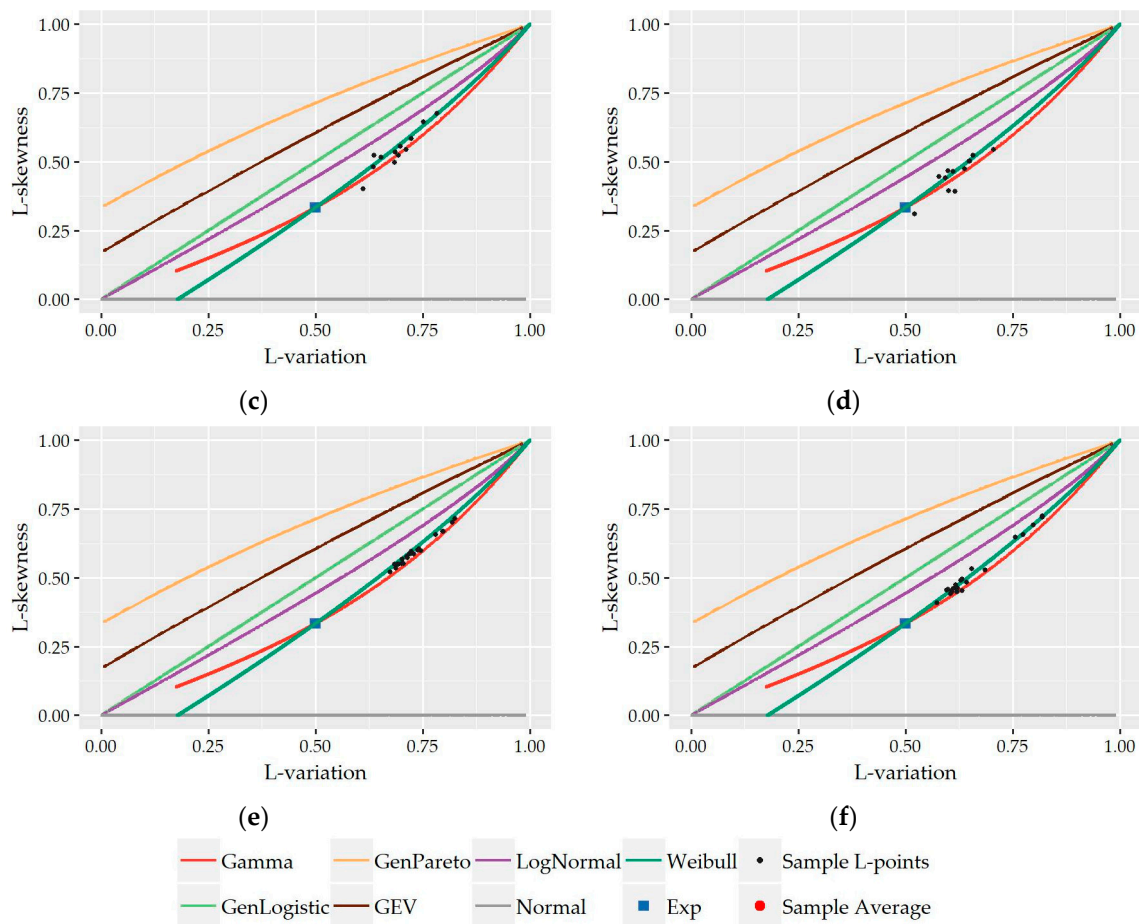
**Figure A4.** 15-minute scale: Daily variation of the probability of no demand (a), mean value (b), L-variation (c), and L-skewness (d) of the 15-minute nonzero demand records of House #1 for each month.

**Appendix B**

Appendix B includes complementary material associated with the screening, evaluation, and fitting of probabilistic models to nonzero water demand presented in Section 4.



**Figure A5.** Cont.



**Figure A5.** Sample L-points of the individual records of each hour assuming also monthly separation (i.e.,  $12 \times 24 = 288$  points for each household) at 15-minute (a) and hourly (b) scale. Sample L-points of the entire nonzero record of each household (12 points) at 15-minute (c) and hourly (d) scale. Sample L-points of the records of each hour of the day (24 points) including the data of all households at 15-minute (e) and hourly (f) scale along with the theoretical lines of the examined distributions.

**Table A1.** Best fitted distributions at the nonzero 15-minute demand records per household and per hour according to  $ER_1$  measure. “W” stands for Weibull distribution, “G” for Gamma distribution, and “LN” Lognormal distribution while “NA” used in the records where it was not possible to fit the models.

	1-W	1-S	2-W	2-S	3-W	3-S	4	5-W	5-S	6	7	8	9	10	11	All
1 h	G	G	LN	LN	G	G	G	W	W	W	G	G	G	G	G	G
2 h	G	NA	LN	LN	G	W	G	W	LN	LN	G	W	G	G	G	G
3 h	G	NA	LN	LN	W	W	G	G	LN	LN	G	LN	G	W	G	G
4 h	G	NA	LN	W	G	G	G	W	LN	LN	G	LN	G	W	G	G
5 h	G	NA	LN	LN	NA	W	G	W	LN	LN	W	W	G	W	G	G
6 h	G	NA	LN	W	W	LN	G	W	LN	LN	NA	G	G	G	G	G
7 h	G	G	G	W	W	W	G	G	LN	LN	W	G	G	G	G	G
8 h	G	G	G	W	G	G	G	W	LN	LN	W	G	W	G	G	G
9 h	G	G	W	W	G	W	G	G	G	LN	W	G	G	G	G	G
10 h	G	G	W	W	G	G	G	G	LN	G	G	G	G	G	G	G
11 h	G	G	G	W	G	G	G	G	LN	G	G	G	G	G	G	G
12 h	G	G	W	G	G	G	G	W	LN	G	W	G	G	G	G	G
13 h	W	G	W	W	W	G	G	W	LN	G	G	G	G	G	G	W
14 h	G	G	W	W	W	W	G	G	LN	W	W	G	G	G	G	G
15 h	W	G	W	LN	W	W	G	G	LN	W	G	G	G	W	G	G
16 h	G	G	LN	LN	W	W	G	G	LN	W	G	G	W	W	G	G
17 h	G	G	W	LN	W	W	G	G	LN	W	G	G	G	G	G	G
18 h	G	G	W	W	W	W	W	G	LN	W	G	G	G	W	G	W

Table A1. Cont.

	1-W	1-S	2-W	2-S	3-W	3-S	4	5-W	5-S	6	7	8	9	10	11	All
19 h	G	G	LN	W	W	G	W	G	LN	W	G	G	G	W	G	W
20 h	G	G	LN	W	W	W	W	G	LN	W	G	G	G	W	G	W
21 h	G	G	W	LN	W	W	W	G	LN	G	G	G	W	W	G	W
22 h	G	G	LN	LN	W	W	W	G	W	W	W	G	W	LN	G	W
23 h	G	G	LN	LN	W	W	W	G	W	G	W	G	G	W	G	W
24 h	G	G	LN	W	G	W	W	G	W	W	W	G	G	G	G	W
Whole Record	G	G	W	W	W	W	G	W	W	W	G	G	G	W	G	G

Table A2. Best fitted distributions at the nonzero hourly demand records per household and per hour according to the ER<sub>1</sub> measure. “W” stands for Weibull distribution, “G” for Gamma distribution, and “LN” Lognormal distribution while “NA” was used in records where it was not possible to fit the models.

	1-W	1-S	2-W	2-S	3-W	3-S	4	5-W	5-S	6	7	8	9	10	11	All
1 h	W	W	G	G	W	W	G	G	W	W	G	G	G	W	G	W
2 h	G	NA	W	G	G	W	G	G	W	W	G	G	G	G	G	W
3 h	G	NA	W	W	G	W	G	G	W	LN	G	G	G	W	G	G
4 h	G	NA	W	W	W	W	G	G	LN	LN	W	G	G	W	G	W
5 h	G	NA	W	G	G	G	G	G	W	LN	W	G	G	W	G	W
6 h	G	G	G	G	NA	LN	G	G	G	LN	NA	G	G	G	G	W
7 h	G	G	G	G	W	NA	G	NA	LN	LN	W	NA	G	G	G	G
8 h	W	G	G	G	G	W	G	G	LN	LN	W	W	W	W	W	G
9 h	W	W	G	W	G	G	G	W	W	W	W	G	W	W	W	G
10 h	W	W	G	G	W	W	W	W	W	G	G	G	W	W	G	G
11 h	W	W	G	W	W	W	G	W	G	W	G	G	W	G	G	G
12 h	W	W	W	W	W	W	G	G	LN	G	G	G	G	G	G	G
13 h	W	W	W	W	W	W	G	G	LN	G	G	G	G	G	G	G
14 h	W	W	W	W	W	W	G	G	W	W	G	G	W	W	W	G
15 h	W	W	G	W	G	W	G	G	LN	W	G	G	W	G	W	G
16 h	W	W	W	G	W	W	G	G	LN	W	G	W	W	W	G	G
17 h	W	G	W	G	G	G	G	G	LN	W	G	W	G	G	G	G
18 h	W	G	G	G	W	W	G	G	LN	W	G	G	G	W	G	G
19 h	G	G	W	G	W	W	G	G	W	G	G	G	G	W	G	G
20 h	G	G	W	G	W	W	G	G	LN	G	G	W	G	W	G	G
21 h	W	G	W	G	W	W	G	G	W	G	W	G	W	W	G	G
22 h	W	G	W	G	G	W	G	NA	W	W	W	W	G	LN	G	G
23 h	W	G	W	W	G	G	G	W	W	G	W	W	W	G	G	G
24 h	W	G	W	G	W	W	G	W	W	W	G	G	W	G	G	W
Whole Record	G	G	G	G	W	W	G	G	G	W	G	G	G	W	G	G

Table A3. Summary statistics of the three error measures that evaluate the fitting performance of the distributions along with the percentage of the records in which each distribution is better fitted. Q<sub>5</sub> and Q<sub>95</sub> is the 5% and 95% empirical quantile points, respectively.

Time Scale	Statistics, ER—1	Gamma	Weibull	Lognormal
15-minute	Perc. of better fitted (%)	55%	30%	15%
	Mean	1.48	2.1	8.78
	St. Deviation	6.22	10.1	39.34
	Min	0.01	0.01	0.02
	Q <sub>5</sub>	0.05	0.04	0.11
	Q <sub>95</sub>	3.73	5.94	23.39
	Max	80.33	162.41	605.68

Table A3. Cont.

Time Scale	Statistics, ER—1	Gamma	Weibull	Lognormal
Hourly	Perc. of better fitted (%)	53%	42%	5%
	Mean	4.5	5.41	25.81
	St. Deviation	13.47	17.5	54.49
	Min	0.01	0.01	0.02
	Q <sub>5</sub>	0.04	0.03	0.11
	Q <sub>95</sub>	20.53	23.8	119.5
	Max	112.24	187.11	488.49
Statistics, ER—2				
15-minute	Perc. of better fitted (%)	33%	40%	27%
	Mean	5.92	5.77	11.53
	St. Deviation	6.82	7.56	14.76
	Min	0.22	0.23	0.42
	Q <sub>5</sub>	0.72	0.85	0.93
	Q <sub>95</sub>	21.85	15.42	32.92
	Max	47.68	59.37	105.06
Hourly	Perc. of better fitted (%)	34%	33%	33%
	Mean	7.85	7.93	10.9
	St. Deviation	11.02	11.18	14.67
	Min	0.13	0.2	0.31
	Q <sub>5</sub>	1	1.02	1.11
	Q <sub>95</sub>	29.66	26.51	31.8
	Max	93.69	102.73	132.83
Statistics, ER—3				
15-minute	Perc. of better fitted (%)	39%	37%	24%
	Mean	−11.27	7.72	74.31
	St. Deviation	35.07	44.73	82.64
	Min	−87.53	−83.38	−70.23
	Q <sub>5</sub>	−53.43	−46.02	−29.73
	Q <sub>95</sub>	48.21	82	214.49
	Max	209.52	280.78	552.91
Hourly	Perc. of better fitted (%)	34%	25%	41%
	Mean	−10.65	−4.45	32.76
	St. Deviation	32.91	37.17	58.25
	Min	−72.6	−74.04	−68.18
	Q <sub>5</sub>	−57.97	−56.1	−38.26
	Q <sub>95</sub>	44.72	54.83	132.62
	Max	167.48	193.41	311.92

Table A4. Fitting performance of the three distributions on the entire nonzero records of each household at 15-minute (upper part) and hourly (lower part) scale. “W” stands for Weibull distribution, “G” for Gamma distribution, and “LN” for Lognormal distribution.

House	ER—1			ER—2			ER—3 (%)		
	G	W	LN	G	W	LN	G	W	LN
1	0.21	1.49	18.66	35.92	70.90	239.76	115.60	242.20	1013.92
2	0.33	0.17	0.27	1.54	19.72	89.95	−6.41	72.14	410.84
3	0.11	0.04	0.82	13.91	33.66	152.70	29.48	72.03	388.67
4	0.11	0.19	2.30	2.67	17.91	106.26	8.37	71.57	429.72
5	0.15	0.08	0.91	46.58	16.45	122.63	−34.77	−0.55	198.81
6	0.32	0.28	0.51	2.26	39.26	164.49	18.05	154.93	748.37
7	0.11	0.25	2.86	1.78	12.29	72.30	5.72	37.01	248.89
8	0.71	2.07	11.61	6.30	40.68	230.51	−1.35	56.17	447.53

Table A4. Cont.

House	ER—1			ER—2			ER—3 (%)		
	G	W	LN	G	W	LN	G	W	LN
9	0.20	0.43	3.14	3.59	3.51	54.65	2.81	28.71	263.83
10	0.10	0.04	0.66	4.94	20.12	108.39	29.71	82.40	462.47
11	0.56	2.05	10.69	10.97	20.92	173.96	−8.71	56.56	454.54
All	0.14	0.18	2.63	26.96	49.10	527.12	−20.81	56.70	673.84
Hourly Data									
1	0.10	0.21	11.03	75.53	129.16	381.78	76.62	141.23	522.41
2	0.06	0.06	4.25	7.94	8.28	69.20	−20.49	4.88	153.43
3	0.03	0.01	2.82	59.35	44.08	93.65	−33.62	−24.11	79.61
4	0.77	2.55	21.72	12.68	29.24	120.82	28.03	67.52	338.12
5	1.64	4.49	58.72	2.09	18.96	159.43	−1.05	18.59	189.44
6	0.17	0.15	1.23	7.87	43.88	174.81	15.99	79.28	374.18
7	0.22	0.55	7.50	4.42	5.59	66.78	−19.84	−4.77	121.37
8	9.14	19.66	122.86	25.44	49.34	262.70	−7.09	17.44	231.07
9	4.50	5.10	30.24	9.35	7.71	44.78	−10.20	−7.03	112.24
10	0.06	0.02	1.39	17.11	33.19	142.74	42.83	76.90	361.08
11	1.20	2.54	29.17	106.48	90.21	47.17	−49.88	−39.85	62.82
All	0.06	0.44	12.55	53.65	12.13	496.02	−21.98	14.24	342.29

Table A5. L-moments of Gamma, Weibull, and Lognormal distribution.

Distribution	L-Moments
Gamma	$\lambda_1 = \beta\gamma, \lambda_2 = \pi^{-\frac{1}{2}}\beta\Gamma\left(\gamma + \frac{1}{2}\right)/\Gamma(\gamma),$ $\tau_3 = 6_{1/3}(\gamma, 2\gamma) - 3$ where $I_x(p, q)$ is the incomplete beta function
Weibull	$\lambda_1 = \beta\Gamma\left(1 + \frac{1}{\gamma}\right), \lambda_2 = \beta\left(1 - 2^{-\frac{1}{\gamma}}\right)\Gamma\left(1 + \frac{1}{\gamma}\right)$ $\tau_3 = 3 - 2\left(1 - 3^{-\frac{1}{\gamma}}\right)/\left(1 - 2^{-\frac{1}{\gamma}}\right)$
Lognormal	$\lambda_1 = \exp\left(\log(\beta) + \frac{\gamma^2}{4}\right),$ $\lambda_2 = \exp\left(\log(\beta) + \frac{\gamma^2}{4}\right)\text{erf}\left(\frac{\gamma}{2\sqrt{2}}\right)$ $\tau_3 = 6\pi^{-\frac{1}{2}} \int_0^{\frac{\gamma}{\sqrt{3}}} \text{erf}\left(\frac{x}{\sqrt{3}}\right) \exp(-x^2) dx / \text{erf}\left(\frac{\gamma}{2\sqrt{2}}\right)$ where $\text{erf}(x)$ is the error function

Table A6. Summary statistics of the shape and scale parameters of the three understudy distributions fitted on the individual records.

Time Scale	Statistic	Gamma		Weibull		Lognormal	
		$\gamma$	$\beta$	$\gamma$	$\beta$	$\gamma$	$\beta$
15-minute	Mean	0.5	6.62	0.67	2.26	1.92	1.31
	St. Dev.	0.27	4.15	0.18	1.72	0.35	1.05
	Min	0.1	0.16	0.33	0.09	0.93	0.05
	Q <sub>5</sub>	0.21	1.71	0.44	0.34	1.39	0.18
	Q <sub>95</sub>	0.93	13.41	0.96	5.31	2.52	3.17
	Max	2.22	31.35	1.56	14.03	3.12	9.07
Hourly	Mean	0.79	11.61	0.85	7.38	1.62	4.51
	St. Dev.	0.44	8.8	0.25	5.58	0.37	3.56
	Min	0.17	0.54	0.4	0.32	0.74	0.18
	Q <sub>5</sub>	0.28	2.67	0.51	0.81	1.08	0.43
	Q <sub>95</sub>	1.63	24.77	1.32	16.99	2.26	10.78
	Max	3.52	85.53	2.02	38.84	2.71	27.52

**Table A7.** Shape and scale parameters of the three understudy distributions fitted on the entire records of each household.

House	15-minute Scale						Hourly Scale					
	Gamma		Weibull		Lognormal		Gamma		Weibull		Lognormal	
	$\gamma$	$\beta$	$\gamma$	$\beta$	$\gamma$	$\beta$	$\gamma$	$\beta$	$\gamma$	$\beta$	$\gamma$	$\beta$
1	0.37	11.14	0.59	2.67	2.04	1.46	0.45	32.74	0.65	10.65	1.90	5.94
2	0.27	5.86	0.50	0.77	2.31	0.41	0.47	9.77	0.66	3.41	1.87	1.92
3	0.50	11.41	0.68	4.40	1.82	2.49	0.67	21.74	0.80	12.93	1.61	7.67
4	0.36	6.75	0.58	1.55	2.06	0.85	0.50	11.23	0.68	4.31	1.82	2.44
5	0.39	11.49	0.60	2.92	2.02	1.60	0.60	20.69	0.75	10.46	1.69	6.08
6	0.22	7.19	0.45	0.64	2.47	0.34	0.35	15.88	0.57	3.38	2.10	1.83
7	0.50	7.99	0.69	3.11	1.81	1.76	0.60	10.72	0.76	5.45	1.68	3.17
8	0.39	10.29	0.60	2.65	2.01	1.45	0.56	20.62	0.72	9.36	1.74	5.38
9	0.57	4.89	0.73	2.29	1.72	1.32	0.89	8.45	0.94	7.34	1.42	4.57
10	0.46	6.75	0.65	2.29	1.88	1.28	0.57	13.30	0.73	6.19	1.73	3.57
11	0.34	7.98	0.56	1.62	2.12	0.88	0.62	14.65	0.77	7.77	1.66	4.55
All	0.31	10.90	0.54	1.95	2.18	1.05	0.47	20.56	0.66	7.11	1.87	3.99

## References

- Magini, R.; Pallavicini, I.; Guercio, R. Spatial and temporal scaling properties of water demand. *J. Water Resour. Plan. Manag.* **2008**, *134*, 276–284. [[CrossRef](#)]
- Magini, R.; Capannolo, F.; Ridolfi, E.; Guercio, R. Demand uncertainty in modelling WDS: Scaling laws and scenario generation. *WIT Trans. Ecol. Environ.* **2016**, *210*, 735–746. [[CrossRef](#)]
- Bao, Y.; Mays, L.W. Model for water distribution system reliability. *J. Hydraul. Eng.* **1990**, *116*, 1119–1137. [[CrossRef](#)]
- Alvisi, S.; Ansaloni, N.; Franchini, M. Generation of synthetic water demand time series at different temporal and spatial aggregation levels. *Urban Water J.* **2014**, *11*, 297–310. [[CrossRef](#)]
- Garcia, V.J.; Garcia-Bartual, R.; Cabrera, E.; Arregui, F.; Garcia-Serra, J. Stochastic model to evaluate residential water demands. *J. Water Resour. Plann. Manag.* **2004**, *130*, 386–394. [[CrossRef](#)]
- Walski, T.; Chase, D.; Savic, D.; Grayman, W.; Beckwith, S.; Koelle, E. *Advanced Water Distribution Modeling and Management*; University of Dayton: Dayton, OH, USA, 2003; ISBN 0971414122.
- Blokker, E.J.M.; Vreeburg, J.H.G.; Beverloo, H.; Klein Arfman, M.; van Dijk, J.C. A bottom-up approach of stochastic demand allocation in water quality modelling. *Drink. Water Eng. Sci.* **2010**, *3*, 43–51. [[CrossRef](#)]
- Filion, Y.R.; Karney, B.W.; Adams, B.J. Stochasticity of Demand and Probabilistic Performance of Water Networks. In *Impacts of Global Climate Change*; American Society of Civil Engineers: Reston, VA, USA, 2005; Volume 40792, pp. 1–12.
- Filion, Y.; Adams, B.; Karney, B. Cross correlation of demands in water distribution network design. *J. Water Resour. Plann. Manag.* **2007**, *133*, 137–144. [[CrossRef](#)]
- Alvisi, S.; Ansaloni, N.; Franchini, M. Comparison of parametric and nonparametric disaggregation models for the top-down generation of water demand time series. *Civ. Eng. Environ. Syst.* **2016**, *33*, 3–21. [[CrossRef](#)]
- 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. [[CrossRef](#)]
- Buchberger, S.G.; Wells, G.J. Intensity, duration, and frequency of residential water demands. *J. Water Resour. Plann. Manag.* **1996**, *122*, 11–19. [[CrossRef](#)]
- Buchberger, S.G.; Wu, L. Model for instantaneous residential water demands. *J. Hydraul. Eng.* **1995**, *121*, 232–246. [[CrossRef](#)]
- Guercio, R.; Magini, R.; Pallavicini, I. Instantaneous residential water demand as stochastic point process. *Water Resour. Manag.* **2001**, *48*. [[CrossRef](#)]
- Creaco, E.; Farmani, R.; Kapelan, Z.; Vamvakieridou-Lyroudia, L.; Savic, D. Considering the mutual dependence of pulse duration and intensity in models for generating residential water demand. *J. Water Resour. Plann. Manag.* **2015**, *141*, 04015031. [[CrossRef](#)]



16. Creaco, E.; Alvisi, S.; Farmani, R.; Vamvakeridou-Lyroudia, L.; Franchini, M.; Kapelan, Z.; Savic, D.A. Preserving duration-intensity correlation on synthetically generated water demand pulses. *Procedia Eng.* **2015**, *119*, 1463–1472. [[CrossRef](#)]
17. 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. [[CrossRef](#)]
18. Gargano, R.; Di Palma, F.; de Marinis, G.; Granata, F.; Greco, R. A stochastic approach for the water demand of residential end users. *Urban Water J.* **2016**, *13*, 569–582. [[CrossRef](#)]
19. Alvisi, S. A Stochastic model for representing drinking water demand at residential level. *Water Resour. Manag.* **2003**, *17*, 197–222. [[CrossRef](#)]
20. 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. [[CrossRef](#)]
21. Alcocer-Yamanaka, V.H.; Tzatchkov, V.; Buchberger, S. Instantaneous water demand parameter estimation from coarse meter readings. In *Water Distribution Systems Analysis Symposium 2006*; American Society of Civil Engineers: Reston, VA, USA, 2008.
22. Kossieris, P.; Makropoulos, C.; Creaco, E.; Vamvakeridou-Lyroudia, L.; Savic, D.A. Assessing the applicability of the bartlett-lewis model in simulating residential water demands. *Procedia Eng.* **2016**, *154*, 123–131. [[CrossRef](#)]
23. 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. Plann. Manag.* **2010**, *136*, 19–26. [[CrossRef](#)]
24. Kofinas, D.T.; Spyropoulou, A.; Laspidou, C.S. A methodology for synthetic household water consumption data generation. *Environ. Model. Softw.* **2018**, *100*, 48–66. [[CrossRef](#)]
25. Creaco, E.; Blokker, M.; Buchberger, S. Models for generating household water demand pulses: Literature review and comparison. *J. Water Resour. Plann. Manag.* **2017**, *143*, 04017013. [[CrossRef](#)]
26. Tsoukalas, I.; Efstratiadis, A.; Makropoulos, C. Stochastic periodic autoregressive to anything (SPARTA): Modeling and simulation of cyclostationary processes with arbitrary marginal distributions. *Water Resour. Res.* **2018**, *54*, 161–185. [[CrossRef](#)]
27. Papalexioiu, S.-M.; Koutsoyiannis, D.; Montanari, A. Can a simple stochastic model generate rich patterns of rainfall events? *J. Hydrol.* **2011**, *411*, 279–289. [[CrossRef](#)]
28. Kossieris, P.; Makropoulos, C.; Onof, C.; Koutsoyiannis, D. A rainfall disaggregation scheme for sub-hourly time scales: Coupling a Bartlett-Lewis based model with adjusting procedures. *J. Hydrol.* **2016**, *556*, 980–992. [[CrossRef](#)]
29. Papalexioiu, S.M. Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency. *Adv. Water Resour.* **2018**, *115*, 234–252. [[CrossRef](#)]
30. Papalexioiu, S.M.; Markonis, Y.; Lombardo, F.; AghaKouchak, A.; Foufoula-Georgiou, E. Precise temporal disaggregation preserving marginals and correlations (DiPMaC) for stationary and non-stationary processes. *Water Resour. Res.* **2018**. [[CrossRef](#)]
31. Tsoukalas, I.; Papalexioiu, S.; Efstratiadis, A.; Makropoulos, C. A cautionary note on the reproduction of dependencies through linear stochastic models with non-gaussian white noise. *Water* **2018**, *10*, 771. [[CrossRef](#)]
32. Savić, D.; Vamvakeridou-Lyroudia, L.; Kapelan, Z. Smart meters, smart water, smart societies: The iWIDGET project. *Procedia Eng.* **2014**, *89*, 1105–1112. [[CrossRef](#)]
33. 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. [[CrossRef](#)]
34. Kossieris, P.; Kozanis, S.; Hashmi, A.; Katsiri, E.; Vamvakeridou-Lyroudia, L.S.; Farmani, R.; Makropoulos, C.; Savic, D. A web-based platform for water efficient households. *Procedia Eng.* **2014**, *89*, 1128–1135. [[CrossRef](#)]
35. Kossieris, P.; Panayiotakis, A.; Tzouka, K.; Gerakopoulou, P.; Rozos, E.; Makropoulos, C. An eLearning approach for improving household water efficiency. *Procedia Eng.* **2014**, *89*, 1113–1119. [[CrossRef](#)]
36. Willis, R.M.; Stewart, R.A.; Giurco, D.P.; Talebpour, M.R.; Mousavinejad, A. End use water consumption in households: Impact of socio-demographic factors and efficient devices. *J. Clean. Prod.* **2013**, *60*, 107–115. [[CrossRef](#)]
37. Beal, C.; Stewart, R.A.; Spinks, A.; Fielding, K. Using smart meters to identify social and technological impacts on residential water consumption. *Water Sci. Technol. Water Supply* **2011**, *11*, 527–533. [[CrossRef](#)]

38. Papalexiou, S.M.; Koutsoyiannis, D. A global survey on the seasonal variation of the marginal distribution of daily precipitation. *Adv. Water Resour.* **2016**, *94*, 131–145. [[CrossRef](#)]
39. 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. [[CrossRef](#)]
40. Hosking, J.R.M. L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *J. R. Stat. Soc. B* **1990**, *52*, 105–124.
41. Peel, M.C.; Wang, Q.; Vogel, R.M.; McMAHON, T.A. The utility of L-moment ratio diagrams for selecting a regional probability distribution. *Hydrol. Sci. J.* **2001**, *46*, 147–155. [[CrossRef](#)]
42. Vogel, R.M.; Fennessey, N.M. L moment diagrams should replace product moment diagrams. *Water Resour. Res.* **1993**, *29*, 1745–1752. [[CrossRef](#)]
43. Hosking, J.R.M. Moments or L-Moments? An example comparing two measures of distributional shape. *Am. Stat.* **1992**, *46*, 186–189. [[CrossRef](#)]
44. Momen, F.A.; Butler, D. Water consumption trends and demand forecasting techniques. In *Water Demand Management*; IWA Publishing: London UK, 2006.
45. Papalexiou, S.M.; Koutsoyiannis, D. Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resour. Res.* **2013**, *49*, 187–201. [[CrossRef](#)]
46. Vargo, E.; Leemis, L. Moment-ratio diagrams for univariate distributions. *J. Qual. Technol.* **2010**, *42*, 276–286. [[CrossRef](#)]
47. Buchberger, S.G.; Carter, J.T.; Lee, Y.H.; Schade, T.G. *Random Demands, Travel Times, and Water Quality in Dead Ends*; AWWA Research Foundation: Denver, CO, USA, 2003.
48. Papalexiou, S.M.; Koutsoyiannis, D.; Makropoulos, C. How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 851–862. [[CrossRef](#)]
49. Stacy, E. A generalization of the gamma distribution. *Ann. Math. Stat.* **1962**, *33*, 1187–1192. [[CrossRef](#)]
50. Tricarico, C.; de Marinis, G.; Gargano, R.; Leopardi, A. Peak residential water demand. *Proc. Inst. Civ. Eng. Water Manag.* **2007**, *160*, 115–121. [[CrossRef](#)]
51. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2017.
52. Hosking, J.R.M. *L-Moments*; R Package Vignette: Madison, WI, USA, 2017.
53. Wickham, H. *GGPLOT2: Elegant Graphics for Data Analysis*; Statistics and Computing; Springer: New York, NY, USA, 2016.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).