water

MDPI

Article

Simulating Marginal and Dependence Behaviour of Water Demand Processes at Any Fine Time Scale

Panagiotis Kossieris [1,*], Ioannis Tsoukalas [1], Christos Makropoulos [1,2] and Dragan Savic [2,3]

[1] Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Heroon Polytechneiou 5, 15780 Zographou, Greece; itsoukal@mail.ntua.gr (I.T.); cmakro@mail.ntua.gr (C.M.)
[2] KWR, Water Cycle Research Institute, 3433 PE Nieuwegein, The Nederlands; Dragan.Savic@kwrwater.nl
[3] College of Engineering, Mathematics and Physical Sciences, University of Exeter, North Park Road, Exeter EX4 4QF, UK
* Correspondence: pkossier@central.ntua.gr; Tel.: +30-210-772-2886

Received: 8 April 2019; Accepted: 25 April 2019; Published: 27 April 2019

Abstract: Uncertainty-aware design and management of urban water systems lies on the generation of synthetic series that should precisely reproduce the distributional and dependence properties of residential water demand process (i.e., significant deviation from Gaussianity, intermittent behaviour, high spatial and temporal variability and a variety of dependence structures) at various temporal and spatial scales of operational interest. This is of high importance since these properties govern the dynamics of the overall system, while prominent simulation methods, such as pulse-based schemes, address partially this issue by preserving part of the marginal behaviour of the process (e.g., low-order statistics) or neglecting the significant aspect of temporal dependence. In this work, we present a single stochastic modelling strategy, applicable at any fine time scale to explicitly preserve both the distributional and dependence properties of the process. The strategy builds upon the Nataf's joint distribution model and particularly on the quantile mapping of an auxiliary Gaussian process, generated by a suitable linear stochastic model, to establish processes with the target marginal distribution and correlation structure. The three real-world case studies examined, reveal the efficiency (suitability) of the simulation strategy in terms of reproducing the variety of marginal and dependence properties encountered in water demand records from 1-min up to 1-h.

Keywords: residential water demand; stochastic simulation; non-Gaussian distributions; intermittency; correlation structure; linear stochastic models; Nataf's joint distribution model; copula; urban water management

1. Introduction

In the planning and management of urban water systems, the temporal and spatial variability of residential water demand is one of the most influential sources of uncertainty [1,2]. Conventional modelling approaches often neglect this aspect, by involving on the one hand, identical multiplier patterns to provide a coarse representation of the diurnal variation of water demand (typically from hour-to-hour) and on the other hand, simplified allocation techniques to express the variation of demand across different points (e.g., nodes) of the network. Despite its practical simplicity, this approach fails to capture adequately the high spatio-temporal variability and uncertainty of water demand (e.g., [3–7]), which becomes more and more intense as the scales of analysis become lower [8]. Furthermore, such a modelling concept does not allow the design, evaluation and management of water distribution systems (WDS) within an uncertainty-aware framework since it implies the use of deterministic water demand patterns as inputs in, typically, deterministic simulation models. A way to address some of

these shortcomings, is the study, analysis and modelling of the uncertain nature of water demand in a probabilistic framework by treating the latter as a stochastic process. In essence, this allows the development of stochastic water demand forecasting tools (e.g., see [9–12] and the references herein) and simulation methodologies. In this work, we focus on simulation that allows the generation of an arbitrarily large number of synthetic, yet statistical consistent, realizations of water demand events, which can then serve as non-deterministic inputs to a WDS, allowing for the probabilistic assessment of its performance under different input scenarios.

During the last decades, the rising deployment of smart metering systems provided new modelling perspectives by delivering large amounts of detailed water demand measurements (e.g., [13–15]). Taking advantage of this dataflow, the research community focused on the modelling of residential water demand at very fine temporal (down to 1 s) and spatial (i.e., household or even water appliance) scales. Then, when necessary, synthetic records at higher levels can be derived via a bottom-up approach that involves the aggregation of fine-resolution data [2].

As literature reveals, the notion of the rectangular pulse has prevailed in the stochastic modelling of residential water demand. This may be associated with an attempt to provide a physical realism in the representation of the real process. In effect, this mechanism, studied at a micro-scale level, can be regarded as a sequence of individual or clustered pulses (events) with constant duration and intensity. Stochastic simulation techniques, such as the Poisson rectangular pulse (PRP) processes [16–23] and Poisson-cluster processes, i.e., the Neyman–Scott [24–26] and Bartlett–Lewis [27] mechanisms, have been employed to generate synthetic water demand events, which occur at continuous time, at the level of a single-household or group of households. Moving from continuous to discrete-time processes, the aggregated demand series can be obtained by summing up all pulses which are active in each discrete time interval. Furthermore, simulation schemes, both stochastic and deterministic, that produce demand signals starting from the level of individual micro-components (i.e., water appliances and uses) have also been developed [28–31].

The above models, and especially PRP and appliance-based schemes, focus mainly on the generation of pulses, whose characteristics (frequency, intensity, duration) are statistically consistent with those of the observed demand events. This is also implied by the adopted procedures for the estimation of their parameters. More specifically, the structure of PRP, implying a one-to-one correspondence between a pulse and an event, allows the fitting of the underlying model's distributions of pulse characteristics directly upon the observed events, if instantaneous demand measurements (e.g., 1-sec time resolution) are available [16–18,20–22]. The schemes based on the simulation of micro-components provide a detailed and elegant description of the real water demand mechanism, but their proper parameterisation is a difficult task since it requires deep knowledge of the consumption habits of the users and the characteristics of water usage [23,32,33].

However, typical WDS modelling applications require demand records across a wide range of higher temporal and spatial levels where the characteristics of individual events are no longer identifiable (e.g., [32]). Thus, from a practical point-of-view, in real-world applications, apart from those that explicitly focus on the influence of individual uses at a micro-scale level (see e.g., [28,34,35]), the key task is not the modelling of individual demand events *per se*, but the adequate reproduction of the statistical and stochastic properties of the discrete-time demand process at different temporal (e.g., 1 min or 15 min scale etc.) and spatial (e.g., household, group of various number of users, nodes etc.) levels. The above mentioned models, though fitted on the basis of individual pulse characteristics, essentially embrace this necessity since their evaluation is also conducted upon statistical parameters of the aggregated demand series which are typically involved in the WDS applications (i.e., main summary statistics or cumulative distribution function of demand, water volume per day, maximum flow at different time scales and probability of no demand).

These models can be regarded as *implicit methods* in terms of reproducing the marginal and dependence characteristics (i.e., the stochastic structure) of the discrete-time water demand process, implying that this reproduction can be achieved via an adequate representation of the

characteristics of individual events. In fact, empirical evidence suggests that they provide a satisfactory, though not explicit, approximation of some statistical parameters of the observed aggregated series. The performance and effectiveness of these models is closely associated with the availability of super-fine observed data that allow the decomposition of demand measurements into single equivalent pulses with identifiable characteristics (see e.g., [16,36]). However, such data can be typically found only at a small number of pilot households, while the commercially available energy-efficient, cost-effective and with long lifetime smart metering devices provide water demand records with higher time step (i.e., 1 min or higher).

Another parameterisation approach, with direct reference to the discrete-time process, consists in obtaining the parameters of PRP and Poisson-cluster models so as to preserve specific properties of the aggregated demand series. Since this approach does not require data at the level of individual pulses, it can be also applied when data of larger time step are available. This method has been used to fit the original PRP model [19], and is the standard parameterisation procedure for the Poisson-cluster processes [24,27]. The structure of the latter (i.e., each event is formulated by a cluster of pulses) does not allow to obtain the parameters directly from the characteristics of observed events. This approach typically involves the theoretical equations of the models which express the main statistical properties (i.e., first- and second-order moments and the probability of no demand) of the discrete-time aggregated process as a function of model parameters, while numerical schemes in cases where the analytical expressions are not available (e.g., PRP model with correlated pulse intensity and duration) have also been studied [22]. Using this concept, pulse-based models are able to generate synthetic water demand data that explicitly reproduce specific properties of the observed aggregated series. However, these are limited to low-order summary statistics (i.e., mean, variance and probability of no demand) regarding the marginal properties of the process and lag-1 autocorrelation coefficient regarding its dependence structure. In this regard, the problem of stochastic simulation of residential water demand is partially addressed since neither the complete marginal distribution nor the whole stochastic structure of the process are explicitly preserved. This also stands for the models which are fitted upon pulse characteristics where the aggregated series are derived implicitly.

Further to the models based on the notion of pulse, a limited number of alternative stochastic procedures for the simulation of water demand process can be found in the literature. More specifically, Gargano et al. [32] proposed a method for the probabilistic representation of the daily trend of residential water demand for different number of users, using a mixed-type distribution to describe the whole process without, however, accounting for the underlying temporal dependences. Alvisi et al. [37,38], using a bottom-up approach, employed random polynomial processes along with reordering techniques to enable the generation of synthetic water demand data which are statistically consistent (in terms of mean, variance and spatio-temporal correlations) with observed the time series at lower and higher spatial and temporal scales. Following the opposite procedure (i.e., top-down allocation), stochastic disaggregation methods, typically applied in the field of hydrology, have also been used to allocate, probabilistically, total amounts of water of an area to its enclosed nodes [6]. Furthermore, and though not proposing stochastic simulation models, it is worth to mention the work of Kossieris and Makropoulos [39] who study the statistical and distributional properties of residential water demand at various scales, as well as the study of Gargano et al. [40] on the evaluation of probabilistic models to describe the peak residential water demand. Lastly, Magini et al. [8,41] and Vertommen et al. [42] studied the scaling properties of water demand at different spatial and temporal levels, while recently Magini et al. [43] employed those properties to generate large number of scenarios of spatially correlated demand series at the node of a network.

In WDS modelling, the adequate reproduction of these two aspects, i.e., the marginal distribution and stochastic structure, of the discrete-time water demand process at different temporal and spatial scales is of high importance. Regarding the first, the probabilistic behaviour of a process is fully described by its marginal distribution. On the contrary, low-order statistics provide only a part of the picture, without explicitly accounting for other important aspects of the marginal properties such as

tail behaviour and hence the reproduction of extremes. This is crucial especially in the design of WDS that requires a proper characterisation of peak flows at different temporal resolutions (e.g., [2,40,44]). This is further highlighted in the work of Kossieris and Makropoulos [39] who demonstrate that the use of some low-order moments (i.e., up to third moment) is not sufficient, since different distributions, having the same moments, result in totally different behaviour in terms of maximum water demands. Additionally, a complete description of the statistical behaviour (e.g., in terms of marginal distribution) of residential water demand is also important in WDS quality modelling applications where different flows, further to maximum values, are used in hydraulic models (e.g., see [3] and references herein).

The importance of appropriately reproducing the temporal and spatial dependencies, i.e., stochastic structure, encountered in water demand process has also been highlighted by many researchers [3–7, 24,37,38,45,46]. More specifically, autocorrelation structure expresses the dynamics of water demand process in time, affecting many parameters in the modelling of WDS, e.g., peak flows, stagnation time, velocities in the pipes, restoration time after pipe breaks etc (e.g., see [37] and references herein). On the other hand, cross-correlation expresses how water demands co-vary across different nodes of the network, and it is also associated with many aspects of the system, e.g., system's performance on the basis of the main statistics of the pressure heads and cost, as well as water quality (e.g., [5,7]). The influence of both auto- and cross-correlation structures have been also studied in the context of bottom-up approaches (see [2]). In this case, the main objective is for the aggregated data to exhibit the desired statistical properties at the temporal and spatial scales under study. As Alvisi et al. [24] pointed out, in order to reproduce the variance of the observed aggregated series, the auto- and cross-correlation structures of the lower-level data should be adequately captured.

In the present work we propose a stochastic modelling strategy for the simulation of residential water demand at any fine time scale, i.e., from 1-min up to 1-h. The proposed strategy combines the widely-used class of linear stochastic models (e.g., autoregressive models) with the concept of Nataf's joint distribution model [47] to enable the explicit reproduction of the marginal distribution and the dependence structure of the process. Nataf-based schemes have been applied in the field of operational research (e.g., [48–50]), while very recently this approach was transferred to the hydrological domain for the stochastic simulation of physical processes [51–56]. Based on these developments, here, we transfer this approach in the domain of residential water demand (a non-physical process) employing, for the first time, linear stochastic models, rather than pulse-based schemes, to explicitly account for the peculiarities of that process at different fine time scales (i.e., intermittency, skewed distributions, periodicity and strong temporal dependence). Furthermore, in the framework of this approach, various innovative modelling aspects are introduced and discussed, for the first time, in the field of water demand simulation, such as, the use of theoretical autocorrelation functions instead of empirical estimates as well as the use of mixed-type distributions with any arbitrary marginal distribution to describe the continuous part of the process. Especially regarding the latter, we also extend the above mentioned Nataf-based schemes by employing three-component mixed-type distributions to capture adequately both the discrete-continuous character (i.e., intermittent nature) of residential water demand as well as its tail behaviour.

A detailed description of the proposed modelling strategy can be found in Section 2 which is further subdivided into eight subsections presenting the components of the method. Specifically, Section 2.1 provides a literature review of the Nataf-based approach and a high-level description of its rational both in terms of random vectors and stochastic processes, while Sections 2.2 and 2.3 present the theoretical background of the methodology. Section 2.4 provides insights on the probability distributions that can be used to describe the marginal properties of residential water demand, while Section 2.5 illustrates an example of the method. Moving from the field of random variables to stochastic processes, Section 2.6 describes the auxiliary linear stochastic models. Next, the use of theoretical autocorrelation functions is presented in Section 2.7. Lastly, Section 2.8 summarises the overall approaches giving the steps of the generation procedure. The method was evaluated in terms of reproducing the marginal distribution and stochastic structure of residential water demand at three

different temporal levels, i.e., fine, medium and high (Section 3). Finally, Section 4 provides a synopsis and discussion of the presented modelling approach.

2. Methodology and Key Components

Nataf-based schemes enable the explicit preservation of the marginal and stochastic properties of a process or, in other words, its distribution functions and correlation structure. They are based upon the simple, yet pivotal, idea of Nataf [47] according to which the joint distribution of random variables with any target arbitrary marginal distributions can be obtained by mapping an auxiliary multivariate standard Gaussian distribution via the inverse cumulative distribution functions (ICDF). Following the same rational, by mapping an auxiliary (stationary or cyclostationary) Gaussian process with zero mean and unit variance (e.g., simulated by a linear stationary or cyclostationary autoregressive model), we can obtain a stochastic process with a target marginal distribution and correlation structure.

2.1. Modelling Rationale and Literature Review

In the core of the described methodology lies the problem of generating random variables with arbitrary marginal distributions and specific correlation matrix. According to Nataf's pivotal work, correlated random vectors can be obtained on the basis of multivariate standard normal variables. This general approach is known as the Nataf's joint distribution model (NDM) or Nataf transformation [57], while later the work of Cario et al. [49] established the term *NORmal To Anything* (NORTA) to describe a generalized procedure that builds upon and extends the NDM rationale to account also for random vectors comprised by continuous and discrete marginal distributions.

Conveniently, NDM can be regarded as a two-step procedure: first the multivariate normal variables are mapped to uniform domain, and then the multivariate uniform variables are mapped to the desire distributions using their inverse cumulative function. Since this procedure implies the establishment of the variables' joint distribution through the joint distribution of normal variables passing also from uniform domain, it is closely associated with copula theory. This connection was first noted by Cario et al. [49], as well as by Tsoukalas et al. [51], while an extensive discussion on the relation of NDM with Gaussian copula was provided later by Lebrun et al. [58].

The main challenge in NDM is the identification of the so-called equivalent correlations of normal variables that will result in the target correlations after applying the mapping procedure. The suitability of NDM to describe a wide range of correlations was investigated by Liu et al. [57] who also developed empirical formulas that relate the equivalent and target correlations for specific types of continuous marginal distributions. Further to this classical work, various, yet often cumbersome, methods have been proposed to establish a relationship between equivalent and target correlations (e.g., [59–62]).

Further to the modelling of random vectors, NDM has been also studied and applied in the field of stochastic processes for the generation of synthetic time series which requires, among others, to account for dependencies in time (i.e., stochastic structure). In this case, a Gaussian process is enrolled to generate serially correlated Gaussian random variables which are mapped according to NDM to provide the target process. In this concept, Cario et al. [48] combined an autoregressive linear model with the idea of NDM to simulate auto-correlated univariate stationary processes with arbitrary marginal distributions. This model, known as *AutoRegressive To Anything* (ARTA), was further extended for the generation of multivariate time series by Biller et al. [50] who developed the *Vector AutoRegressive To Anything* (VARTA) procedure.

Recently, NDM-based approaches have drawn the attention of hydrological community, since they provide the tools for the reproduction of the peculiarities of hydrometeorological variables, i.e., cyclostationarity, highly skewed character, intermittency, short or long-range auto-dependence structures as well as spatial dependency. In this concept, Tsoukalas et al. [51,54] developed the *Stochastic Periodic AutoRegressive To Anything* (SPARTA) scheme that is a generalization of ARTA and VARTA models for the simulation of univariate and multivariate cyclostationary (i.e., periodic) processes with arbitrary marginal distributions. Furthermore, the *Symmetric Moving Average To*

Anything (SMARTA) model [52] combines NDM with the symmetric moving average model [63] to simulate non-Gaussian processes that exhibit any-range dependence structure. Papalexiou [53], using autoregressive models, proposed a unified approach for the stochastic modelling of hydroclimatic processes characterised by different correlation structures and marginal distributions, with focus on the use of mixed-type distributions to model intermittency.

Further to the above models, as Tsoukalas et al. [51] pointed out, many other well-known simulation schemes for hydrological variables (e.g., [64–66]), with the most characteristic being the so-called Wilks' type weather generators [67], actually share a common rationale, and can be retrospectively viewed as NDM-based approaches.

2.2. The bivariate Nataf Distribution Model

As discussed above, NDM approach was initially developed for the generation of correlated random variables (RVs) with arbitrary marginal distributions, while next it was applied in the simulation of stochastic processes after certain extensions and modifications. Although this work focuses on the latter case, in order to keep things simple, we prefer to present NDM's theoretical background on the basis of a bivariate problem that studies the generation of two random variables with predefined marginal distributions and correlation. Essentially, this problem can be regarded as representative since it is also involved when we aim to apply NDM for the modelling of stochastic processes using linear models, due to the fact that the latter are also based on the Pearson's correlation coefficient which is a two-point dependence measure.

Given that our target is to generate correlated RVs X_1 and X_2 with predefined target marginal distributions $F_{X_1}(x_1)\mathrm{P}(X_1 \le x)$ and $F_{X_2}(x_2)\mathrm{P}(X_2 \le x)$, respectively, and target correlation $\rho_{X_1X_2}\mathrm{Corr}[X_1,\ X_2]$ which is the Pearson's correlation coefficient between the two variables, hereinafter abbreviated as $\rho_{1,2}$.

Let us initially to define two auxiliary correlated RVs Z_1 and Z_2 which both have the standard normal marginal distribution and specific Pearson correlation coefficient $\widetilde{\rho}_{Z_1Z_2}\mathrm{Corr}[Z_1, Z_2]$, herein after termed as equivalent correlation, for reasons revealed below, and abbreviated as $\widetilde{\rho}_{1,2}$. The joint distribution of the two auxiliary variables is the bivariate normal with zero mean, unit variance and correlation $\widetilde{\rho}_{1,2}$.

The RVs X_1 and X_2 can be obtained by mapping the auxiliary normal variables to the target distributions, according to the following operations:

$$X_1 = F_{X_1}^{-1}(\Phi(_1))\ ,\ \ X_2 = F_{X_2}^{-1}(\Phi(_2)), \tag{1}$$

where $F_{X_1}^{-1}(\cdot)$ and $F_{X_2}^{-1}(\cdot)$ denote the ICDF of the two target distributions and $\Phi(\cdot)$ stands for the standard Normal cumulative distribution function (CDF).

Recall that since $U = \Phi(\cdot)$ is uniformly distributed in the interval (0, 1), the use of the ICDF of the target distribution as a mapping function ensures that the final variables will have the desired marginal properties. This concept, as mathematically expressed in Equation (1), is based upon the lemma of probability integral transformation which allows the representation of any random variable as a transformation of a uniform random variable, i.e., if $U \sim U_{[0,1]}$, then the random variable $X = F^{-1}(U)$ has the distribution F_X (e.g., see, ([68], p. 139); and ([69], p. 36).

Having said the above, one may assume that the preservation of the target correlation matrix is also satisfied by setting $\widetilde{\rho}_{1,2} = \rho_{1,2}$. However, since ICDF typically imposes a nonlinear and monotonic transformation, it is not possible to preserve the linear correlation as expressed by the Pearson correlation coefficient and hence, $\rho_{1,2}$ will typically differ to $\widetilde{\rho}_{1,2}$ after the mapping of Equation (1). This transformation usually leads to underestimated correlation coefficients, while as the target distribution deviates from the Normal case, the larger will be the underestimation. This implies that in order to attain the desire target correlation coefficient $\rho_{1,2}$, we need to technically assign an inflated

value to $\widetilde{\rho}_{1,2}$. Therefore, the key challenge of the methodology is to determine the equivalent correlation $\widetilde{\rho}_{1,2}$ that, after applying the mapping procedure, will result to the desired target correlation $\rho_{1,2}$.

The establishment of the relationship between the target $\rho_{1,2}$ and the equivalent correlation $\widetilde{\rho}_{1,2}$ is based on the theoretical background of Nataf's joint distribution model. Elaborating, given Equation (1), the correlation of two RVs X_1 and X_2 can be written as:

$$\rho_{1,2}Corr[X_1,\ X_2] = Corr\Big[F_{X_1}^{-1}(\Phi(_1)),\ F_{X_2}^{-1}(\Phi(_2))\Big]. \tag{2}$$

According to the definition of Pearson correlation coefficient, the following equation also stands:

$$\rho_{1,2}Corr[X_1,\ X_2] = \frac{E[X_1\ X_2] - E[X_1]\ E[X_2]}{\sqrt{Var[X_1]\ Var[X_2]}}, \tag{3}$$

where $\mathrm{E}[X_1]$, $\mathrm{E}[X_2]$ and $\mathrm{Var}[X_1]$, $\mathrm{Var}[X_2]$ are the mean and variance of X_1 and X_2, respectively. Since the associated marginal distributions are already known (and have finite variance), these quantities can be directly estimated and hence our attention is restricted to adjusting $\mathrm{E}[X_1\ X_2]$. Taking the first cross-product moment of X_1 and X_2 and using Equation (1), we obtain:

$$\begin{aligned} E[X_1\ X_2] = \ & E\Big[F_{X_1}^{-1}(\Phi(_1))\ F_{X_2}^{-1}(\Phi(_2))\Big] \\ & = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_{X_1}^{-1}(\Phi(z_1))\ F_{X_2}^{-1}(\Phi(z_2))\varphi_2(z_1,z_2;\widetilde{\rho}_{1,2})dz_1dz_2, \end{aligned} \tag{4}$$

where $\varphi_2(_{1,1}\,;\widetilde{\rho}_{1,2})$ is the standard bivariate normal probability density function (PDF) with correlation $\widetilde{\rho}_{1,2}$. By substituting Equation (4) to Equation (3) we obtain:

$$\rho_{1,2} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_{X_1}^{-1}(\Phi(z_1))\ F_{X_2}^{-1}(\Phi(z_2))\varphi_2(z_1,z_2;\widetilde{\rho}_{1,\,2})dz_1dz_2 - E[X_1]\ E[X_2]}{\sqrt{Var[X_1]\ Var[X_2]}}. \tag{5}$$

In Equation (5) we see that the target correlation $\rho_{1,2}$ is a function of the equivalent correlation $\widetilde{\rho}_{1,2}$, given the target marginal distributions $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$. The latter equation can be compactly expressed as:

$$\rho_{1,2} = \mathcal{T}\Big(\widetilde{\rho}_{1,2}\Big|F_{X_1}(x_1), F_{X_2}(x_2)\Big) \tag{6}$$

where $\mathcal{T}(\cdot)$ is the abbreviation of the function defined in Equation (5). By inverting Equation (6), we express the problem as stated initially, i.e., what $\widetilde{\rho}_{1,2}$ will give the desired $\rho_{1,2}$ after applying the mapping procedure:

$$\widetilde{\rho}_{1,2} = \mathcal{T}^{-1}\Big(\rho_{1,2}\Big|F_{X_1}(x_1), F_{X_2}(x_2)\Big) \tag{7}$$

Equation (6) has analytical solutions only for a few cases where the variables have the same target marginal distributions, such as the Uniform [60], Exponential [49] and Log-Normal [70]. In the general case, numerical schemes are required to solve the integral in Equation (5) in order to establish a relationship among $\rho_{1,2}$ and $\widetilde{\rho}_{1,2}$ (see Section 2.3 for more details). In order to avoid tedious calculations and achieve a more efficient numerical search, the fundamental properties of Equation (6) are typically employed in the numerical procedures. The key properties, as provided by Liu and Der Kiureghian [57] and Cario and Nelson [48], are:

Lemma 1. $\rho_{1,2}$ is a strictly increasing function of $\widetilde{\rho}_{1,2}$.

Lemma 2. $\widetilde{\rho}_{1,2} = 0$ for $\rho_{1,2} = 0$ and $\widetilde{\rho}_{1,2} \geq (\leq) 0$ if $\rho_{1,2} \geq (\leq)\ 0$.

Lemma 3. $\big|\rho_{1,2}\big| \leq \big|\widetilde{\rho}_{1,2}\big|$, where the equality stands when $\rho_{1,2} = 0$ or when both marginal distributions are normal.

Lemma 4. The feasible minimum and maximum values for $\rho_{1,2}$, which can be obtained for a given set of target distributions, are given for $\widetilde{\rho}_{1,2} = -1$ and $\widetilde{\rho}_{1,2} = 1$, respectively.

2.3. Establishing the Target-Equivalent Correlation Relationship

In the core of the NDM approach lies the problem of establishing a relationship between the target and equivalent correlation, $\rho_{1,2}$ and $\widetilde{\rho}_{1,2}$, respectively, given the target marginal distributions of the RVs. In other words, to define $\mathcal{T}(\cdot)$, or alternatively $\mathcal{T}^{-1}(\cdot)$, that will give the value of $\widetilde{\rho}_{1,2}$ so as to attain the desired $\rho_{1,2}$ after mapping Gaussian variables to the real domain (i.e., Equation (1)). To solve the double integral appearing in Equation (5) and establish $\mathcal{T}(\cdot)$ numerical schemes are typically employed, since an analytic solution is feasible only for some few special cases. Among others, crude search procedures [48], root-finding techniques [60,62] as well as quadrature methods and Monte–Carlo procedures [59] have been used. These methods are often based on iterative methods while some of them are specifically designed for certain types of distributions (e.g. continuous).

Recently, Tsoukalas et al. [51] and Papalexiou [53] proposed generic hybrid schemes that share a common notion regarding the establishment of a suitable relationship between $\rho_{1,2}$ and $\widetilde{\rho}_{1,2}$. These schemes can be regarded as generic methodologies since on the one hand, they aim to capture the whole form of $\mathcal{T}(\cdot)$, and not to provide specific point estimates of $\widetilde{\rho}_{1,2}$, and on the other hand, they are easily-extendable and applicable irrespective of the type of marginal distributions of studied variables (i.e., continuous, discrete or mixed-type of distributions).

More specifically, in these methods, Equation (6) is solved for a specific set of $\widetilde{\rho}_{1,2}$ values, and the corresponding target $\rho_{1,2}$ values are obtained. To solve the double integral of Equation (5) typical integration techniques [53] or Monte–Carlo [51] methods are employed. In this step, the methods also take advantage of the properties of Equation (6), e.g., Lemma 2, to define the range of the $\widetilde{\rho}_{1,2}$ values that should be studied (e.g., if the target correlation is positive then $\widetilde{\rho}_{1,2}$ values will range from 0 up to 1). This procedure leads to a set of ($\rho_{1,2}$, $\widetilde{\rho}_{1,2}$) anchor points upon which a suitable function (e.g., polynomial or parametric) is fitted to establish an approximation of the true $\mathcal{T}(\cdot)$. This relationship can be used as an "interpolator" that provides the equivalent $\widetilde{\rho}_{1,2}$ given the target $\rho_{1,2}$. Papalexiou [53] proposes various parametric functions to establish directly $\mathcal{T}^{-1}(\cdot)$, depending on the type of RVs. On the other hand, Tsoukalas et al. [51] uses a polynomial function to approximate the $\mathcal{T}(\cdot)$, while the equivalent correlation $\widetilde{\rho}_{1,2}$, given a target correlation $\rho_{1,2}$, is obtained by inverting the fitted polynomial.

2.4. Modelling the Marginal Behaviour of Water Demand

As described above, the NDM approach requires the definition of probability distributions to describe the marginal properties of correlated random variables or stochastic processes (Section 2.6). Essentially, the method aims to reproduce the target distributions which have assigned *a priori* to the variables under study. Tsoukalas et al. [51,71] highlight the importance and benefits of this approach against the classical stochastic modelling of hydrometeorological processes, which typically focuses on the resemble of a series of specific statistical characteristics. As discussed in Section 1, this is also crucial in residential water demand modelling given that WDS applications require a proper reproduction of various characteristics (e.g., maximum demands) which cannot derive explicitly by the preservation of some low-order statistics.

Another key advantage of the NDM approach is its flexibility regarding the selection and fitting of the target distributions, given that the only prerequisite of the method regarding the latter is to have finite variance. In this respect, several candidate distributions, fitted with alternative methods (e.g., classical moments, L-moments, maximum likelihood estimation, weighted moments), can be evaluated and the most suitable for the variables under study can be identified on the basis of certain criteria.

Residential water demand series with a fine time step (i.e., at daily and mainly sub-daily scales) are characterised by the presence of individual or grouped zero values, whose proportion becomes more and more higher as the metering resolution is getting lower, implying more extensive time

intervals with no demand. To describe probabilistically such an intermittent process, a mixed-type (also known as discrete-continuous or zero-inflated) distribution can be employed, composed by a probability mass concentrated at zero (i.e., discrete part) and a continuous part that describes the nonzero positive demand magnitudes. This modelling approach, i.e., use of a unique distribution to model the whole water demand process, differs from that of the pulse-based schemes (see Section 1) which incorporate distributions for the individual components of demand events (i.e., occurrence, duration and intensity) without any reference to the marginal properties of the whole process.

Mixed-type distributions have been widely applied to model hydrometeorological processes (e.g., [51,53,72–75]) with intermittent behaviour (e.g., rainfall at fine time scales, discharge of intermittent, flows, wind speed), while recently they employed for the probabilistic representation of residential water demand [32,39].

More specifically, the cdf of a mixed-type distribution is given by:

$$F_X(x) = \begin{cases} p_{ND}, & x = 0 \\ p_{ND} + (1 - p_{ND})G_X(x), & x > 0 \end{cases} \quad (8)$$

where $p_{ND} := P(X = 0)$ is the probability of no demand (probability zero), while G_X and g_X are the cdf and pdf, respectively, of values greater than zero, i.e., $G_X F_{X|X\rangle 0} = P(X \leq x|X\rangle 0)$. Therefore, the establishment of $F_X(x)$ or $f_X(x)$ requires: (i) the estimation of p_{ND} which can be directly estimated from the observed data as the ratio of the number of zero values over the total number of observations, and (ii) the selection and fit of a continuous distribution G_X that describes adequately the nonzero values.

As mentioned above, any probability distribution, defined on the real positive axis, could be a candidate to model the continuous part of the process. At the same time, residential water demand at fine time scales is characterised by high variability and skewness and hence, distributions that will be able to capture these characteristics should be employed. In this respect, Kossieris and Makropoulos [39] recently examined the suitability of 10 distributions to describe the nonzero 15-min and hourly water demand, showing that Weibull, Gamma and Lognormal distributions have a good performance in terms of reproducing the statistical behaviour of the under study records, while the latter two perform also well in terms of maxima. Further to that, the probability distributions employed by the pulse-based schemes to model pulse intensities could be potential candidates to represent G_X, i.e., the Exponential [8,24,26,27], Weibull [18], Normal [19] and Lognormal [36] distribution.

In several cases, the above classical distributions appear as suitable models to describe the entire probabilistic behaviour of the under study continuous variable. On the other hand, as indicated by many hydrological studies, the probabilistic structure of observed data (e.g., rainfall, flows) in several cases is more complex (e.g., multimodal behaviour) and, hence, cannot be entirely captured with the use of a single distribution. An alternative solution to this consists the use of nonparametric (e.g., [76]) or mixture distributions. The former, being a data-driven approach, has extrapolation limitations and hence is inappropriate to model tail behaviour. Furthermore, mixture type distributions provide enhanced flexibility, at a cost of few parameters, to account simultaneously for low, moderate and high amounts of the under study variable. This approach has found fertile ground to adequately model hydrological extremes which is of high importance in many engineering applications (e.g., [77–79]). Furthermore, as discussed in Section 1, the whole range of water demand flows, further to the peak values, are often involved in WDS applications, and thus the reproduction of the entire probabilistic behaviour of residential water demand consists a *conditio sine qua non*.

To this end, in the present study we also employ a two-component mixture distribution for the description of nonzero water demand values. This enables to capture the main body of observed data as well as a more suitable representation of extreme behaviour, in cases where classical distribution models are found inadequate. Particularly, the employed mixture model for nonzero water amounts reads [80]:

$$G_X(x) = \begin{cases} (1-u_c)\frac{G_1(x|\boldsymbol{\theta}_1)}{G_1(c|\boldsymbol{\theta}_1)}, & 0 < x \le c \\ (1-u_c) + u_c G_2(x|\boldsymbol{\theta}_2), & x > c \end{cases} \quad (9)$$

where $G_1(x)$ denotes a cdf, with parameter vector $\boldsymbol{\theta}_1$, which describes the body of the data up to threshold c with corresponding quantile u_c. Furthermore, $G_2(x)$ stands for the cdf of the Generalised Pareto distribution (GPD), with vector parameter $\boldsymbol{\theta}_2$, which consists the model of the upper tail. The above mixture model has been implemented with the use of various continuous parametric models, such as those mentioned above, to represent $G_1(x)$, e.g., Weibull, Gamma and Lognormal distributions [81–83], while nonparametric methods have also been applied [80]. Furthermore, the use of GPD consists a well-established modelling approach for extremes in many disciplines, e.g., in hydrological domain this model has been used to probabilistically describe rainfall maxima. This model has the flexibility to represent different types of upper tail behaviour depending on the value of shape parameter γ (the cdf of GPD is given in Appendix A, Table A1). More specifically, for $\gamma > 0$ the GPD has a heavy (sub-exponential) upper tail, for $\gamma = 0$, it convergences to the exponential tail, while for $\gamma < 0$ the distribution has a finite upper bound at $u_c - \beta/\gamma$, where β and u_c denote the scale and location (threshold) parameter, respectively.

In the framework of the present study, the evaluation of different probabilistic models in terms of adequately representing the entire marginal distribution of the finely resolved observed water demand records showed that, in several cases, the latter was not possible to be achieved with the use of a single distribution. More specifically, the investigated single distributions, though resembling the main statistical characteristics upon which they are fitted (e.g., via the method of moments or L-moments), diverge significantly from the empirical marginals both in the body as well as the right tail of the distribution. On the contrary, when those models are combined with a GPD to formulate a mixture type distribution, i.e., Equation (9), a much better fitting is achieved.

This is further illustrated graphically in Figure 1 that depicts the empirical distribution of the nonzero 1-min water demand observations, recorded in time interval 20:00–21:00 (see Section 3.1 for more information on the dataset used in the present case studies), along with the corresponding theoretical distributions of three single models (i.e., Gamma, Weibull, Lognormal) and one mixture model composed by Gamma and GPD. The graph presents in double logarithmic plot the probability of exceedance, i.e., $\overline{G}_X(x) = P(X > x) = 1 - G_X(x)$. The single distribution models (their cdf are given in Appendix A, Table A1) were fitted with the method of L-moments [72,84], while the parameters of the mixture model were obtained via the maximum likelihood method (see [80,83] for further details). As Table 1 shows, single distribution models perfectly resemble the observed statistical characteristics upon which they are fitted (i.e., L-mean and L-variation), providing also a good approximation to observed L-skewness and L-kurtosis which are not involved in the fitting procedure. Despite this, Figure 1 reveals that these models show a significant divergence from the empirical distribution of the observed data for values greater than the quantile $x = 7$ L/min. On the contrary, the Gamma-GPD mixture model provides a much better approximation both to the body and the tail of the observed distribution, resembling also well the under study statistics (see Table 1).

Table 1. Comparison between observed and theoretical summary statistics of the five distributions under study for the indicative example.

	Observed	**G**	**W**	**LN**	**G-GPD**
Mean, μ	1.751	1.751	1.751	1.751	1.742
L-variation, τ_2	0.720	0.720	0.720	0.720	0.695
L-skewness, τ_3	0.592	0.559	0.592	0.658	0.577
L-kurtosis, τ_4	0.317	0.295	0.358	0.474	0.342

In the present study, we exploit the flexibility of Nataf-based models, allowing the use of any distribution model, and we employ this mixture-type distribution, in cases where simpler models

cannot provide a good fit. Further to this, to describe the entire marginal behaviour of water demand process at fine time scales via a single cdf model, we embed Equation (9) within Equation (8) to account simultaneously for intermittency (discrete part) as well as the body and tail behaviour of the continuous part (see Section 3).

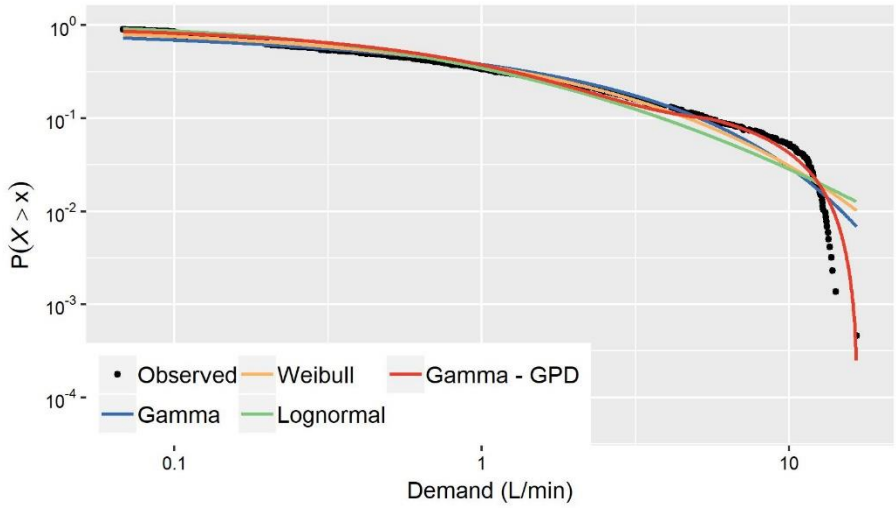

Figure 1. Comparison between observed and theoretical cumulative distribution functions (Weibull plotting positions) of the five distributions under study for the indicative example.

2.5. Demonstrating NDM Approach

Prior to studying the application of NDM for the simulation of water demand stochastic processes, which is the main subject of the present work, here we provide a comprehensive demonstration of the entire approach, as described in Sections 2.2 and 2.3, via a numerical example. More specifically, we examine the problem of generating RVs X_1 and X_2 which both follow the Weibull distribution and have target correlation $\rho_{1,2} = 0.75$. We assume that the parameter values for the distribution of both variables are $\beta = 1.01$ and $\gamma = 0.54$ (see Table A1 in Appendix A for more details on Weibull distribution).

Given the target marginal distribution and correlation, initially, the equivalent correlation $\widetilde{\rho}_{1,2}$ of the auxiliary standard normal variables Z_1 and Z_2 is obtained. This is achieved by first establishing $\mathcal{T}(\cdot)$ that provides the relationship between the target and equivalent correlations for the specified distributions. In the present work, we employed the numerical scheme proposed by Tsoukalas et al. [51] according to which the true $\mathcal{T}(\cdot)$ is approximated by a polynomial function (see Section 2.3 for more information). In the present example, the following quadratic function is derived:

$$\rho_{1,2} = \mathcal{T}\left(\widetilde{\rho}_{1,2} \middle| F_{X_1}(x_1), F_{X_2}(x_2)\right) = 0.5465\widetilde{\rho}_{1,2}^2 + 0.4457\widetilde{\rho}_{1,2} \quad (10)$$

The established relationship between target and equivalent correlations is further illustrated in Figure 2a. The equivalent correlation $\widetilde{\rho}_{1,2}$ of the Gaussian variables can be then obtained by solving the simple Equation (10) for the given target correlation. For this case, for $\rho_{1,2} = 0.75$, we get $\widetilde{\rho}_{1,2} = 0.834$. Next, we generated 100,000 values from the bivariate standard normal distribution having the specified equivalent correlation. The scatter plot as well as the marginal distributions of these auxiliary variables are presented in Figure 2b. These variables are first mapped to the uniform domain via the cdf of the standard normal distribution $\Phi(\cdot)$, providing correlated values that have uniform marginal distributions (see Figure 2c). The correlated uniformly distributed variables are then mapped to the actual domain (see Figure 2d), according to Equation (1), to provide variables with the target marginal distributions and correlation. As explained above, the use of the ICDF of the target distribution (the Weibull in this example) as a mapping function ensures that the final variables will have the desired

marginal properties, while the assignment of a technically inflated value to $\widetilde{\rho}_{1,2}$ allows to attain the target correlation $\rho_{1,2}$ after the mapping operation.

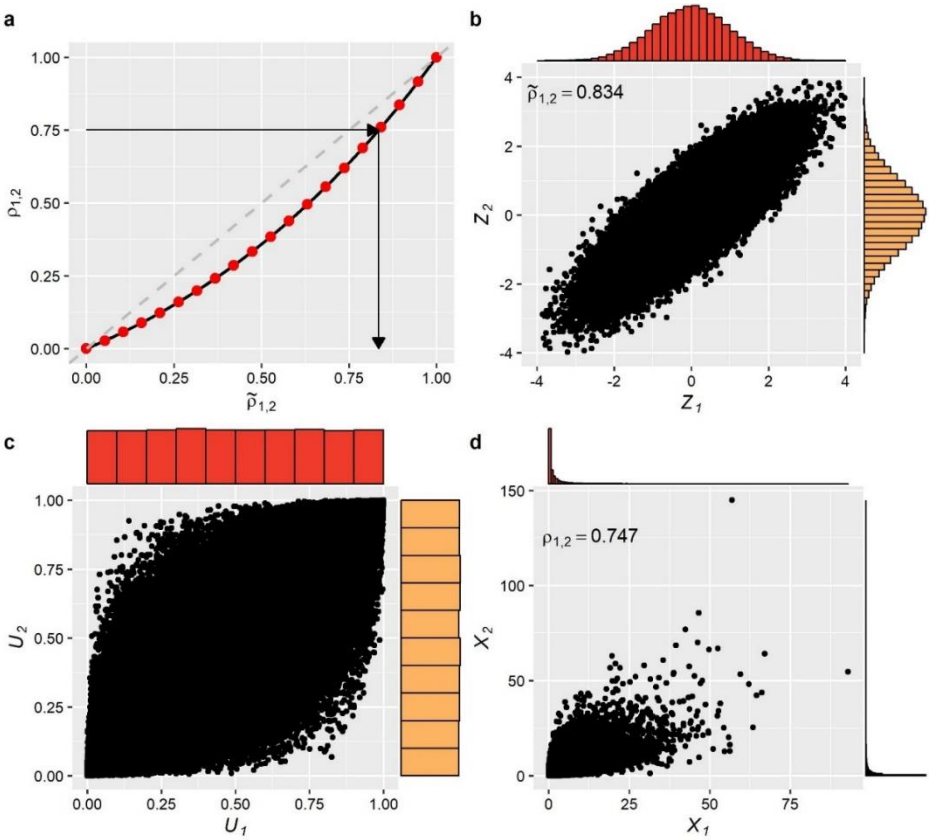

Figure 2. Numerical example demonstrating the generation of correlated random variables according to NDM approach. (**a**) Established relationship between target and equivalent correlations as well as scatter plots and marginal distributions of correlated variables at normal (**b**), uniform (**c**) and real (**d**) domain.

2.6. Moving from Random Variables to Stochastic Processes

As already mentioned above, NDM can be also applied to generate synthetic time series with the desired marginal distributions and stochastic structure (in terms of autocorrelation and cross-correlation coefficients). In this case, a Gaussian process is employed to generate auxiliary time series which are then mapped via the ICDF of the target distributions to obtain the target process. Although Gaussian process can be regarded as an intermediate auxiliary step in the whole procedure, its role is crucial since its structure determines that of the target process, e.g., to simulate a stationary auto-correlated process then a stationary Gaussian process should be employed, while the generation of cyclostationary time series requires the use of a cyclostationary Gaussian model as auxiliary. Summarising, the choice of the auxiliary process that is going to be implemented in the NDM concept is a matter of modelling requirements and convenience.

As mentioned in Section 2.1, different auxiliary models have been used in the existing Nataf-based schemes. In fact, as discussed in Tsoukalas et al. [52], any linear stochastic model could be used to play this role. For instance, Papalexiou [53] used the sum of independent autoregressive processes (e.g., see [85–87]) to generate Gaussian time series.

In the present work, taking into account the peculiarities of residential water demand process at the timescales under study (i.e., hourly and sub-hourly scales) and without loss of generality, we employ stationary and cyclostationary autoregressive models to simulate the auxiliary Gaussian

process. Prior describing these models, let us introduce some notations that allow the transition from random variables to processes.

Beginning from the stationary case, let X_t be a target process, where $t = 1, 2, \ldots$, is the time index, with target marginal distribution F_X and autocorrelation structure $\rho_\tau \text{Corr}[X_t, X_{t+\tau}]$, where τ is the time lag. Furthermore, let Z_t be the stationary Gaussian auxiliary process with zero mean and unit variance and autocorrelation structure $\widetilde{\rho}_\tau \text{Corr}[Z_t, Z_{t+\tau}]$. In order to align the bivariate NDM, as presented in Section 2.2 for the case of random variables, with stationary stochastic processes, we have to set throughout Equations (1)–(7), X_1X_t and $X_2X_{t+\tau}$ for the target process, as well as Z_1Z_t and $Z_2Z_{t+\tau}$ for the auxiliary, accordingly.

The most prominent linear stochastic model, used in a variety of applications and scientific fields, is the autoregressive model of order p (i.e., AR(p)). This can be attributed to its parsimony and simplicity as well as to its intuitive structure since the model implies that the present value of the process is obtained as the weighted sum of p past values and a random component. According to AR(p), a standard Gaussian process $Z_t \sim N(0,1)$, with specific autocorrelation structure $\widetilde{\rho}_\tau$, is obtained by the following recursive formula:

$$Z_t = \widetilde{a}_1 Z_{t-1} + \widetilde{a}_2 Z_{t-2} + \ldots + \widetilde{a}_p Z_{t-p} + \widetilde{\beta}\varepsilon_t, \tag{11}$$

where ε_t are i.i.d variables (white noise) that follow the standard Normal distribution, i.e., $\varepsilon_t \sim N(0,1)$, while $\widetilde{a}_i$, with $i = 1, \ldots, p$, and $\widetilde{\beta}$ are model parameters. The parameters $\widetilde{a}_i$ are derived analytically on the basis of the Yule–Walker system, given the values of $\widetilde{\rho}_\tau$ up to lag p. This reads as:

$$\widetilde{\boldsymbol{a}} = \widetilde{\boldsymbol{P}}^{-1}\widetilde{\boldsymbol{\rho}} \tag{12}$$

where:

$$\widetilde{\boldsymbol{a}} = \begin{bmatrix} \widetilde{a}_1 \\ \widetilde{a}_2 \\ \vdots \\ \widetilde{a}_p \end{bmatrix}, \qquad \widetilde{\boldsymbol{P}} = \begin{bmatrix} 1 & \widetilde{\rho}_1 & \widetilde{\rho}_2 & \cdots & \widetilde{\rho}_{p-1} \\ \widetilde{\rho}_1 & 1 & \widetilde{\rho}_1 & \cdots & \widetilde{\rho}_{p-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \widetilde{\rho}_{p-1} & \widetilde{\rho}_{p-2} & \widetilde{\rho}_{p-3} & \cdots & 1 \end{bmatrix}, \qquad \widetilde{\boldsymbol{\rho}} = \begin{bmatrix} \widetilde{\rho}_1 \\ \widetilde{\rho}_2 \\ \vdots \\ \widetilde{\rho}_p \end{bmatrix}. \tag{13}$$

In order for an AR(p) process to be stationary, certain conditions for the parameters $\widetilde{a}_i$ should be fulfilled (e.g., see [88]).

Parameter $\widetilde{\beta}$ essentially expresses the variance of the random component in Equation (11) and it can be estimated by:

$$\widetilde{\beta} = \sqrt{1 - \widetilde{a}_1\widetilde{\rho}_1 + \widetilde{a}_2\widetilde{\rho}_2 + \ldots + \widetilde{a}_p\widetilde{\rho}_p} \tag{14}$$

In the AR(p) model, since ε_t are generated from the Normal distribution, the random process Z_t will be also Gaussian. Furthermore, this model ensures the preservation of a given autocorrelation structure up to lag p, whereas for higher lags the model gives a gradually decay autocorrelation structure, i.e., at any lag $\tau \geq p+1$ the modelled autocorrelation is given by $\widetilde{\rho}_\tau^{AR} = \sum_{i=1}^{p} \widetilde{a}_i \widetilde{\rho}_{\tau-i}$.

The stationary AR(p) model can be easily adjusted to account for cyclostationarity, i.e., in cases where the distribution and correlation structure of the process vary periodically from season-to-season. In this case, the model is termed as periodic autoregressive model of order p (i.e., PAR(p)), and further to the time index t, it is convenient to employ another index $s = 1, \ldots, S$ that denotes the season (e.g., $S = 12$ for processes exhibiting month-to-month variation or $S = 24$ for hour-to-hour variation). Subsequently, the cyclostationary process reads as $Z_{s,t}$. Accordingly, $\widetilde{\rho}_{s,s-\tau}$ now represents the correlation between season s and $s-\tau$. Reasonably, model parameters will also depend on season s. Particularly, in the case of PAR(1) model, the model generating equation reads (the index t is omitted for simplicity):

$$Z_s = \widetilde{a}_s Z_{s-1} + \widetilde{\beta}_s \varepsilon_s \tag{15}$$

where $\widetilde{a}_s = \widetilde{\rho}_{s,s-1}$, $\widetilde{\beta}_s = \sqrt{1-\widetilde{a}_s^2}$ and $\varepsilon_s \sim N(0,1)$.

As presented in Section 3, in the present work, AR(p) was used as an auxiliary model in the simulation of residential water demand processes at sub-hourly time scales (see Sections 3.1 and 3.2), while to capture the hour-to-hour variation of hourly water demand (see Section 3.3) we employ the PAR(1) model.

2.7. Modelling the Dependence Behaviour of Water Demand

Further to the marginal distributions, Nataf-based models aim to capture and reproduce the dependence behaviour of a process as expressed by its stochastic structure. As explained above, the use of Pearson's correlation coefficient, as linear measure of dependence, arises naturally since linear stochastic models are employed to establish dependence in time. However, as it is known the estimation of empirical correlations, as well as of the other second-order statistics, subjects to important bias (e.g., [89,90]), especially in cases of small samples, large lags and intense (persistent) autocorrelation structures, i.e., large autocorrelation values that extend for large lags. This issue has been studied in the domain of hydrological variables (e.g., [63,91]), while Dimitriadis and Koutsoyiannis [92] compare the suitability of other tools to identify the auto-dependence structure of a process. These sources of bias are also encountered in residential water demand processes since the available observed records are typically short, while their autocorrelation structure exhibits a more persistent behaviour as the time scale becomes finer. Due to this, a theoretical autocorrelation structure (ACS) is typically preferred over the direct use of the empirical sample estimates. Further to the above, the use of a suitable ACS is favored for a series of other reasons. Firstly, it ensures the stability of the linear stochastic models given that a properly defined ACS provides positive definite autocorrelation structures, which is a prerequisite so as the latter to be valid (see [68] for more details). Furthermore, it can be regarded as a parsimonious procedure since the whole structure is described via a small set of parameters, enabling the extrapolation of correlation coefficients for lags as high as desired (1/3 of the sample size is the typically suggested as the maximum lag up to which empirical correlations can be properly estimated). Lastly, it enables the transferability of ACS parameters in cases where the sample size does not allow for a proper identification of the correlation structure and stable estimation of coefficients. The latter is of high importance in water demand modelling since observed records at fine time scales are typically limited both in terms of the number of meters and the length of records. In the present work, it is the first time that an ACS is employed to characterise the auto-dependence properties of residential water demand processes.

Various theoretical models to represent the correlation structure of stationary processes can be found in the literature (e.g., [53,86,89,90]). Here, we employ the Cauchy-type autocorrelation structure (CAS) proposed by Koutsoyiannis [63] as a simple and parsimonious model, able to capture a wide range of short- and long-range dependence structures. The power-type CAS, for positive ACF, is given by:

$$\rho_\tau^{CAS}(\kappa,\beta) = (1+\kappa\beta\tau)^{-1/\beta}, \qquad \tau \geq 0 \tag{16}$$

where $\beta \geq 0$ and $\kappa > 0$ are model parameters that control the form of the function and hence the degree of dependence. For $\beta = 0$, CAS provides a short-range autocorrelation structure, i.e., an exponential structure that decays rapidly and diminishes after few time lags. On the contrary, for $\beta > 1$, CAS resembles long-range autocorrelation structures, i.e., slowly decreasing structure that extends for large lags. For other values of parameters, the function provides the flexibility to represent autocorrelation structures between, or even outside of, the two cases explained above (see [63] for further details).

The estimation of CAS parameters can be obtained by minimizing the error between the sample and theoretical correlation coefficients, on the basis of the least squares method. In the present study, our target was to achieve a good overall fit to a specific number of autocorrelations but with an extra care regarding the preservation of empirical lag-1, i.e., by setting a larger weight in the objective function for this quantity.

The above modelling strategy has been recently employed within NDM-based approaches for hydrological process modelling. In the NDM concept, Tsoukalas et al. [52] used CAS model, while Papalexiou [53] proposed a series of parametric ACS models exploiting their structural similarity with the form of common complementary cumulative distribution functions.

2.8. Summary of the Modelling Approach Step-by-Step

In the present section, we summarise the stochastic modelling framework for residential water demand processes via the following steps:

1. Select and fit the suitable marginal distributions (i.e., continuous, mixed; see Section 2.4) and autocorrelation function (see Section 2.7) that better capture the distributional properties and autocorrelation structures, respectively, of the observed time series.
2. Given the target autocorrelation structure ρ_τ (derive either from a theoretical autocorrelation model or as it is estimated directly from the data), establish the correlation transformation function and estimate the equivalent correlation coefficient $\widetilde{\rho}_\tau$ up to the maximum specified lag τ (see Section 2.3).
3. Identify the suitable auxiliary Gaussian linear stochastic model (e.g., AR(p) and PAR(p); see Section 2.6) and fit it on the basis of $\widetilde{\rho}_\tau$.
4. Generate a realization z_t of the auxiliary process Z_t at the Gaussian domain.
5. Use the ICDF of the fitted distribution to map the synthetic time series z_t to the actual domain (i.e., Equation (1)) and obtain the final realization x_t of the target process X_t.

The above described steps, summarising the methods and tools of Section 2, have been implemented in the R programming language [93], in the form of an R package, named *anySim* (the package is available at: http://www.itia.ntua.gr/en/softinfo/33/; [94]).

3. Case Studies

The stochastic modelling strategy was applied and evaluated in the simulation of residential water demand at three different temporal scales which are typically involved in WDS applications, i.e., 1-min (Section 3.1), 15-min (Section 3.2) and 1-h (Section 3.3). Furthermore, these three case studies allow the assessment of the modelling approach on the basis of a variety of different marginal behaviours and correlation structures which are usually observed in the water demand records at these time scales. This further demonstrates that a single modelling strategy suffices for the reproduction of the peculiarities of residential water demand at any time scale, without requiring the use of ad-hoc techniques or scale-specific adjustments.

The observed data used in the present study concerns a total water demand record of 1-min resolution from a single household in Athens (Greece). The data was collected via a smart metering equipment installed in the framework of the iWIDGET project [95] and concern an observation period of 5 months, spanning from 1 October 2014 to 31 March 2015. This record was selected due to the quality of data (i.e., small percentage of missing values) and the regular consumption of water, while a shorter part of this dataset has been also used in past studies [22,27]. The employed dataset was initially examined for negative or unrealistically large values due to smart meter malfunctions. All negative values and the values greater than 50 L/min were excluded from the analysis. The 15-min and hourly records were obtained by aggregating temporally the 1-min data at the corresponding time intervals, flagging as missing and excluding from the analysis the values with at least one missing 1-min value.

The performance of the modelling strategy was evaluated by comparing several relevant properties and characteristics of the simulated series with the observed one. Since the stochastic methodology enables the explicit reproduction of the marginal distribution and the dependence structure of the process under study, we initially examine the capability of the model to resemble the empirical cumulative distribution functions (both probability of no demand and the distribution of nonzero values) and the Pearson's autocorrelation coefficients of the observed series. To further assess and verify

the model as a fully operational tool, we employed a series of different empirical characteristics which are not directly involved in the model's fitting, but their reproduction is regarded of high importance in WDS applications. These characteristics have been also used in the evaluation of pulse-based schemes [18,28,33] and concern: (1) the total water demand per day, V(L/d), which is the total volume of water consumed within a day; (2) the maximum flow per day which is the highest consumption of the day at the temporal scale of measurement, e.g., Q_{max} (L/min) or Q_{max} (L/15-min); and (3) the maximum hourly demand, Q_{max} (L/h), which is the highest hourly consumption of the day.

3.1. Simulation of 1-min Water Demand

The first application concerns the simulation of residential water demand at a fine temporal scale (1-min resolution). At this scale the process is characterised by especially high proportion of zero values p_{ND}, highly skewed distributions and strong autocorrelation structures. Furthermore, water demand characteristics exhibit a significant variation within the day, and, thus, the analysis and modelling is typically conducted by subdividing the day into shorter time intervals (e.g., typically with 1 or 2 h length) where the process can be assumed as stationary. In the present study, we treat 1-min water demand of each hour as a separate stationary stochastic process, and hence we apply the modelling strategy twenty-four times, varying the distribution function and autocorrelation structure from hour-to-hour. This implies that the series of each hour of the day has a different distribution and correlation function, while for the sake of simplicity and parsimony we do not assume different processes for working and weekend days.

Following the steps of the modelling strategy described in Section 2.8, initially, the most suitable probabilistic models to describe the marginal properties of 1-min water demand records at each time interval is identified. Due to the intermittent nature of the process, a mixed model (Section 2.4), given in Equation (8), was employed to describe simultaneously both the discrete and continuous part of the marginal distribution. The discrete part of the model, i.e., probability of no demand p_{ND}, was obtained directly by dividing the number of zero values with the total number of recorded data. Regarding the continuous part G_X of the mixed model, in the present study we used the two-parameter Gamma (G), Weibull (W) and Lognormal (LN) distributions as well as their combination with the Generalised Pareto distribution (GPD), i.e., the mixture model of Equation (9). Furthermore, the three-parameter Generalised Gamma distribution (GG) was also employed. The above models were fitted to the nonzero 1-min water demand data (either via the method of L-moments or maximum likelihood; see Section 2.4) and the most suitable to describe the record of each individual time interval was identified in terms of approximating adequately the entire empirical distribution of the observed data. These distributions were selected for the demonstration purposes of the present case study and they could be replaced by any other model as implied by the flexible structure of the modelling strategy. The estimated p_{ND} as well as the empirical probability distributions of the fitted models along with their parameters for the 1-min demand records of each hour of the day are given in Figure A1 of Appendix B.

Further to the marginal distributions, to model the autocorrelation structures we employed the power-type CAS given in Equation (16). The parameters of CAS were obtained for each time interval by minimizing the mean square error (MSE) between the sample and theoretical correlation coefficients. The empirical autocorrelation structures along with the fitted CAS and its parameters are graphically presented in Figure A2. Given the target autocorrelation structure ρ_τ and the marginal distributions, the $\mathcal{T}(\cdot)$, given in Equation (7), was established following the numerical scheme proposed by Tsoukalas et al. [51] and the equivalent correlation coefficients ρ_Z were obtained for each time interval. The visual inspection of autocorrelation structures of 1-min water demand records showed that ρ_τ is practically equal to zero for $\tau > 10$ in most hourly intervals and, thus, an AR(10) model was selected and fitted to $\widetilde{\rho}_\tau$, for $\tau = 1, 2, \ldots, 10$, to generate the auxiliary time series. The final time series was then obtained by mapping the Gaussian series to the selected distribution for each time interval. In the present case study, we generated synthetic data of 4 000 days length (4 000 × 24 × 60 values).

The application of the methodology is summarised in Figure 3 that presents the results of the simulation of 1-min water demand of three representative hourly intervals of the day, i.e., a night, a morning and an evening hour. More specifically, plots Figure 3a,b depict the observed and synthetic 1-min water demand data of a typical day. The efficiency of the method in terms of reproducing exactly the target distributions of the nonzero demand values as well as the probability of no demand is illustrated in the panels (c) and (e) of Figure 3 as well as in the probability plots of Figure A1. Furthermore, panels (f)–(h), along with plots in Figure A2, present graphically the capability of the model to resemble the target autocorrelation structures, while panels (i)–(k) depict the relationships between the target and equivalent correlation coefficients for the mixed-type model as well as for the distribution fitted to the nonzero demands. As we see, the presence of zero values has as a result the intense divergence of the relationship from linearity.

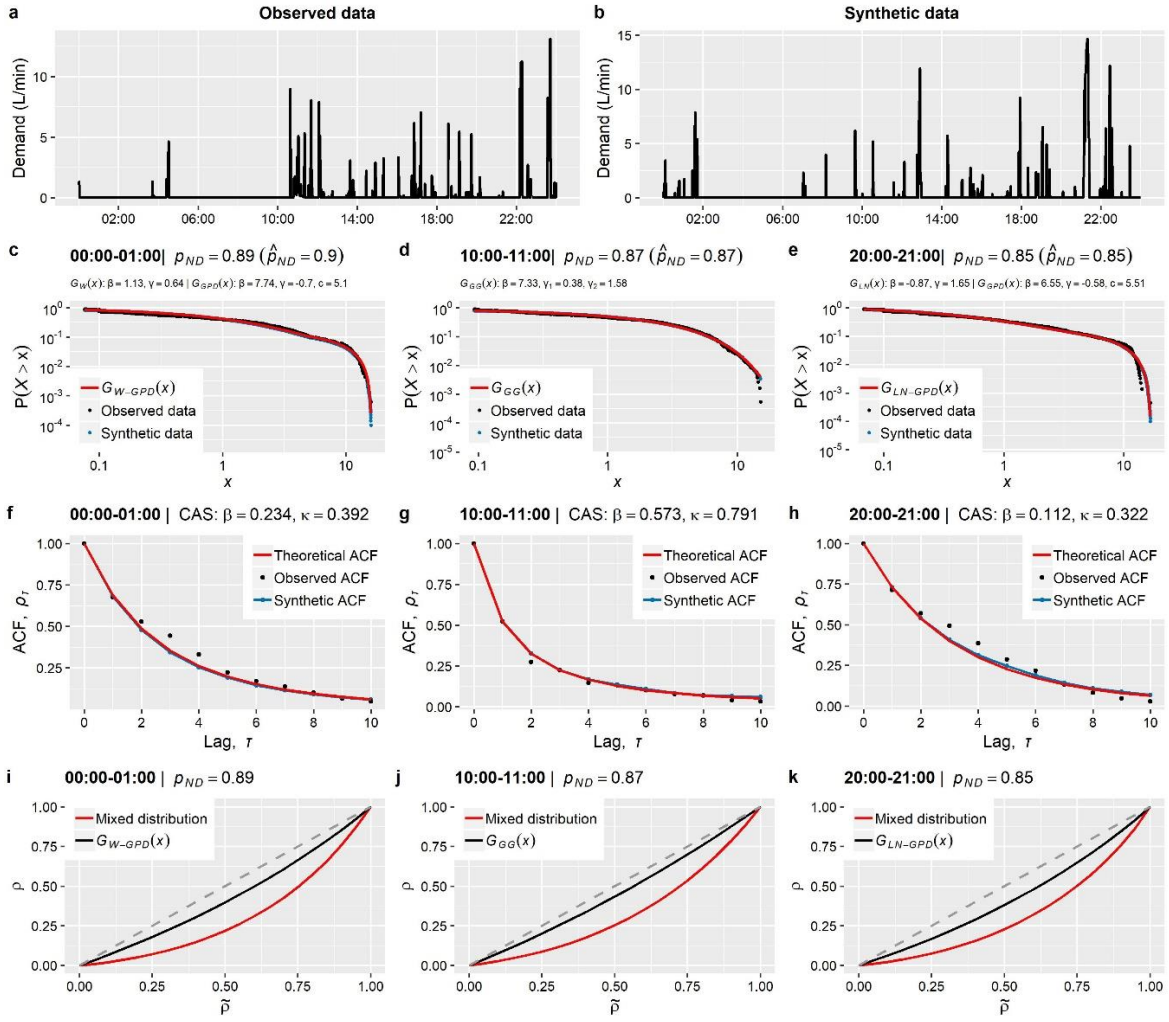

Figure 3. Simulation of 1-min residential water demand: (**a**) Observed 1-min data of a typical day. (**b**) Synthetic 1-min time series of a randomly selected day. (**c**)–(**e**) Observed, theoretical and simulated distribution functions (Weibull plotting positions and in logarithmic scale) of the nonzero water demand values for three representative hours of the day; each graph also displays the parameters of the fitted distribution as well as the observed p_{ND} and simulated $\hat{p}_{ND}$ values of probability of no demand. (**f**)–(**h**) Observed, theoretical and simulated autocorrelation functions of three representative hours of the day; each graph also displays the parameters of CAS. (**i**)–(**k**) the established relationship between the equivalent ρ_Z and target ρ_X correlation for the mixed-type distribution as well as the distribution fitted on nonzero demand values.

Furthermore, Figure 4 presents the performance of the stochastic methodology in a series of important characteristics which are not explicitly modelled by the former. More specifically, as we see in Figure 4a the cdf of the entire simulated 1-min water demand series resembles the observed one, while as Figure 4b shows that the model achieved a very good performance in terms of reproducing the observed daily volumes. Regarding the reproduction of extremes (Figure 4c,d), the cdfs of both nonzero maximum 1-min demand per day, Q_{max} (L/min), as well as maximum hourly demand per day, Q_{max} (L/h), produced by the model are in high agreement with the observed one, overestimating slightly the amounts with very low probability of exceedance.

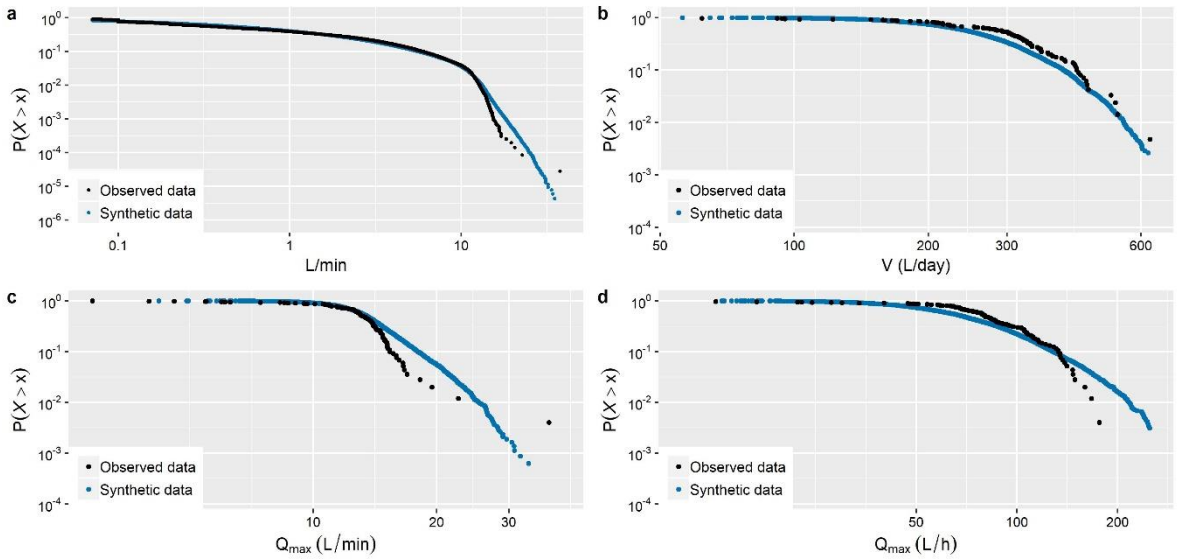

Figure 4. Observed and simulated cumulative distribution functions (Weibull plotting positions and in logarithmic scale) of: (**a**) the entire 1-min water demand series, (**b**) total water volume per day, (**c**) maximum nonzero 1-min demand per day, (**d**) maximum nonzero hourly demand per day.

3.2. Simulation of 15-min Water Demand

The second case study involves the simulation of residential water demand at a medium resolution scale, i.e., 15-min. This is the typical metering resolution of the available energy-efficient, cost-effective and with long lifetime smart metering devices, and it consists the typical temporal scale of many WDS applications. In the present case study, we followed the same approach as in the simulation of 1-min water demand (see Section 3.1) regarding the effect of seasonality, treating the process of each hour of the day as stationary and hence varying the distribution function and autocorrelation structure from hour-to-hour.

The mixed model of Equation (8) was employed to describe the marginal distribution of the intermittent process, while the distributions described in previous case study were also examined as candidates to model the nonzero 15-min demands of each time interval. The estimated p_{ND} as well as the empirical probability distributions of the fitted models along with their parameters for the 15-min demand records of each hour of the day are given in Figure A3 Appendix B. Again, the power-type CAS, given in Equation (16), was used to model the autocorrelation structure and $\mathcal{T}(\cdot)$, given in Equation (7), was established to provide the equivalent correlation coefficients $\widetilde{\rho}_\tau$ for each time interval. In the simulation of 15-min demand, an AR(3) model was selected to generate the Gaussian series, since the analysis of autocorrelation structures of the records showed ρ_τ is practically equal to zero for $\tau \geq 4$ in most hourly intervals.

For demonstration, we generated synthetic data of 4 000 days length (4 000 × 24 × 4 values), conducting a similar analysis with the above case study. The results are summarised in Figure 5 that depicts three representative hourly intervals of the day, i.e., a night, a morning and an evening hour. More specifically, Figure 5a,b present the observed and synthetic 15-min water demand data, respectively, of a typical day. In the plots (c)–(e) of Figure 5 along with probability plots of Figure A3, we can see that the method reproduces exactly the target distributions of the nonzero demand values as well as the probability of no demand. Additionally, the capability of the model to resemble the target autocorrelation structures is graphically illustrated in panels (f)–(h) and in the plots of Figure A4, while panels (i)–(k) depict the established relationships between the target and equivalent correlation coefficients for the mixed model as well as for the distribution fitted to the nonzero demands.

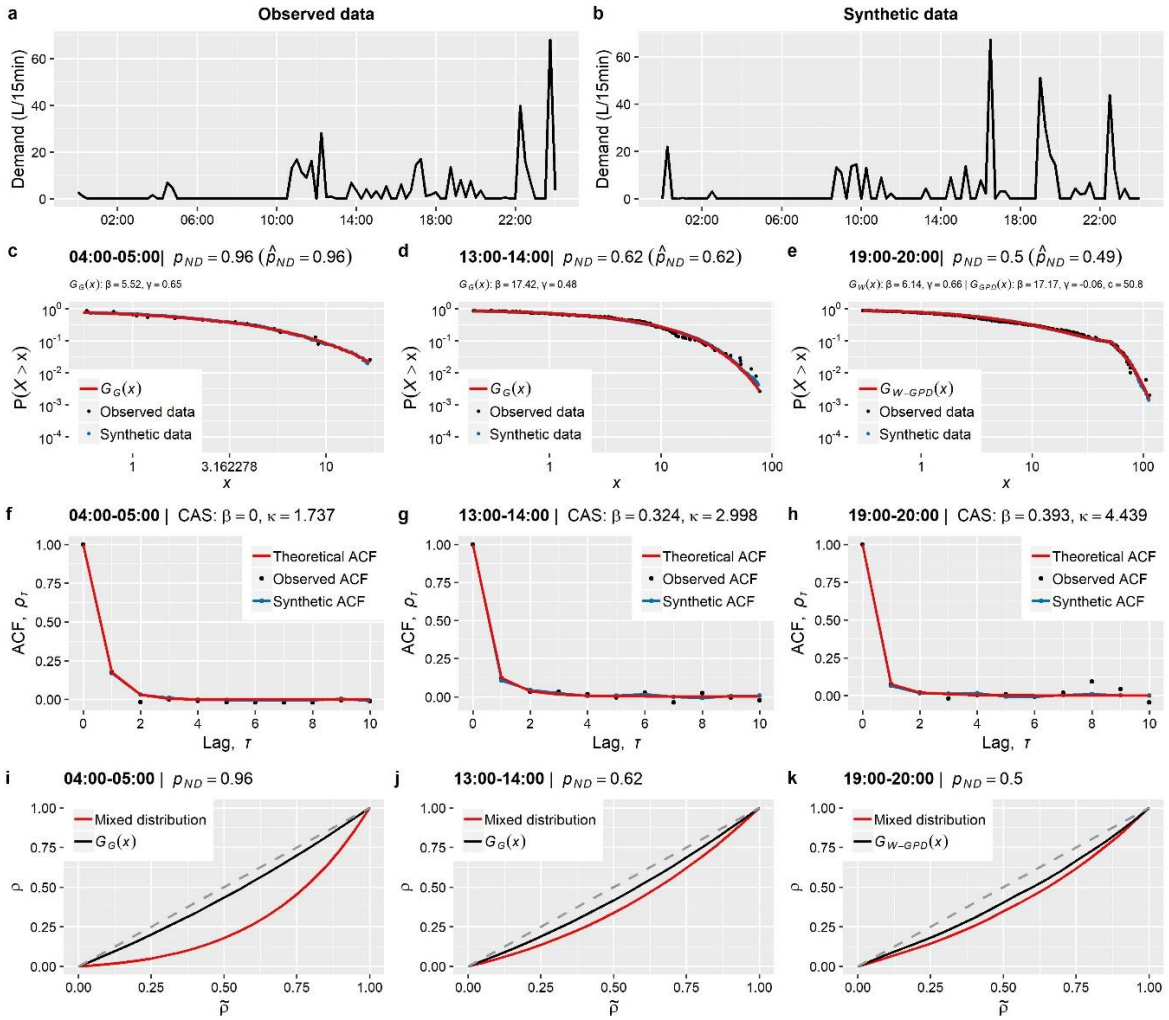

Figure 5. Simulation of 15-min residential water demand: (**a**) Observed 15-min data of a typical day. (**b**) Synthetic 15-min time series of a randomly selected day. (**c**)–(**e**) Observed, theoretical and simulated distribution functions (Weibull plotting positions and in logarithmic scale) of the nonzero water demand values for three representative hours of the day; each graph also displays the parameters of the fitted distribution as well as the observed p_{ND} and simulated $\hat{p}_{ND}$ values of probability of no demand. (**f**)–(**h**) Observed, theoretical and simulated autocorrelation functions of three representative hours of the day; each graph also displays the parameters of CAS. (**i**)–(**k**) the established relationship between the equivalent ρ_Z and target ρ_X correlation for the mixed-type distribution as well as the distribution fitted on nonzero demand values.

As it is further demonstrated in Figure 6a,b, the cdf of the entire simulated 15-min water demand series as well as the cdf of synthetic daily volumes resemble with precision the observed ones. Furthermore, the model has the capability to reproduce the behaviour of observed extremes as revealed by Figure 6c,d that present the cdf of the nonzero maximum 15-min demand per day, Q_{max} (L/15-min), and the cdf of maximum hourly demand per day, Q_{max} (L/h), respectively.

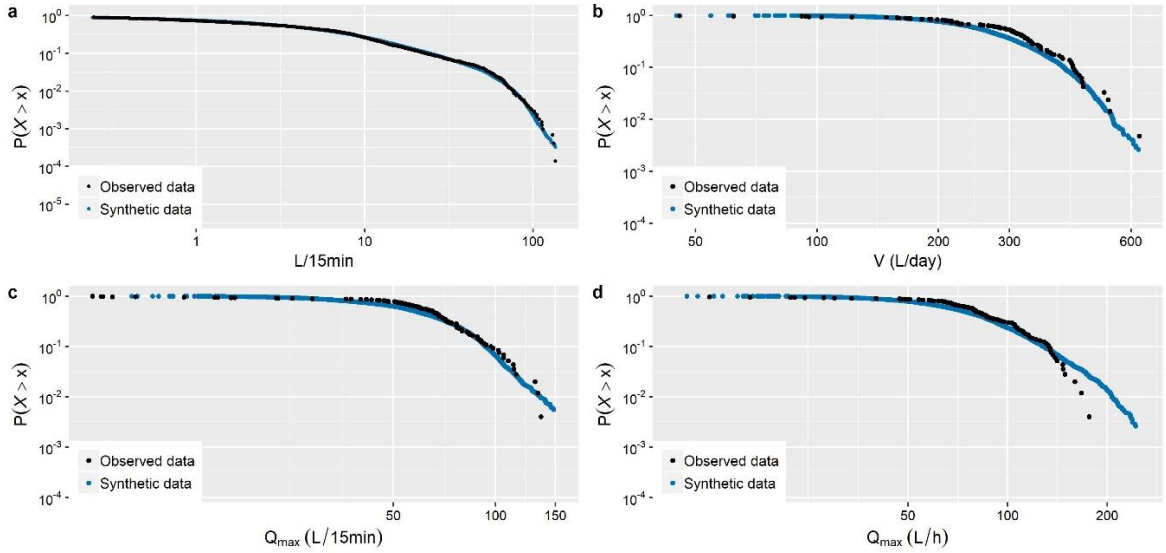

Figure 6. Observed and simulated cumulative distribution functions (Weibull plotting positions and in logarithmic scale) of: (**a**) the entire 15-min water demand series, (**b**) total water volume per day, (**c**) maximum nonzero 15-min demand per day, (**d**) maximum nonzero hourly demand per day.

3.3. Simulation of 1-h Water Demand

The final case study concerns the simulation of residential water demand process at hourly scale, which is the typical temporal scale of analysis of many real-world WDS applications and distribution network simulation models. At this time scale, the process is characterised by strong seasonality and hence it is treated as cyclostationary with marginal distributions and correlation structures that vary periodically from season-to-season. Having said this, we follow the same approach as in the previous two case studies treating hourly water demand at each hour of the day as a separate stochastic process with common marginal distribution, while in the present case the correlation coefficients express the dependence of hourly demand among successive hours. To capture this hour-to-hour variation of hourly water demand and establish this stochastic structure we built upon the SPARTA approach [51,54], and particularly, we employ a 24-season PAR(1) model (see Section 2.6) for the generation of auxiliary Gaussian series. To demonstrate the stochastic modelling strategy in this concept, we generated synthetic hourly water demand data of 4 000 days length (4 000 × 24 values), performing a similar analysis as in the previous two case studies.

Regarding the marginal properties of the hourly water demands, the process still have an intermittent nature, despite the fact that the observed p_{ND} values are much smaller compared to those of 1-min and 15-min scales, while the distributional behaviour of the nonzero values deviates significantly from Gaussianity. Due to this, the cyclostationary stochastic model is coupled with the mixed-type distribution of Equation (8) to describe the entire marginal behaviour of the process. More specifically, we fit and evaluate the two-parameter Gamma, Weibull and Lognormal distributions as well as the three-parameter Generalised Gamma distribution (GG) to model the positive demand quantities. Figure 7 displays the observed and simulated p_{ND} values as well as the observed, theoretical and simulated empirical probability plots of the nonzero hourly water demand records of each hour of the day. As we see, the mapping operation, i.e., Equation (1), employed by the method ensures

the perfect reproduction of the hourly-varying probabilities of no demand as well as the perfect resemblance of the assumed target theoretical distribution models by the simulated data.

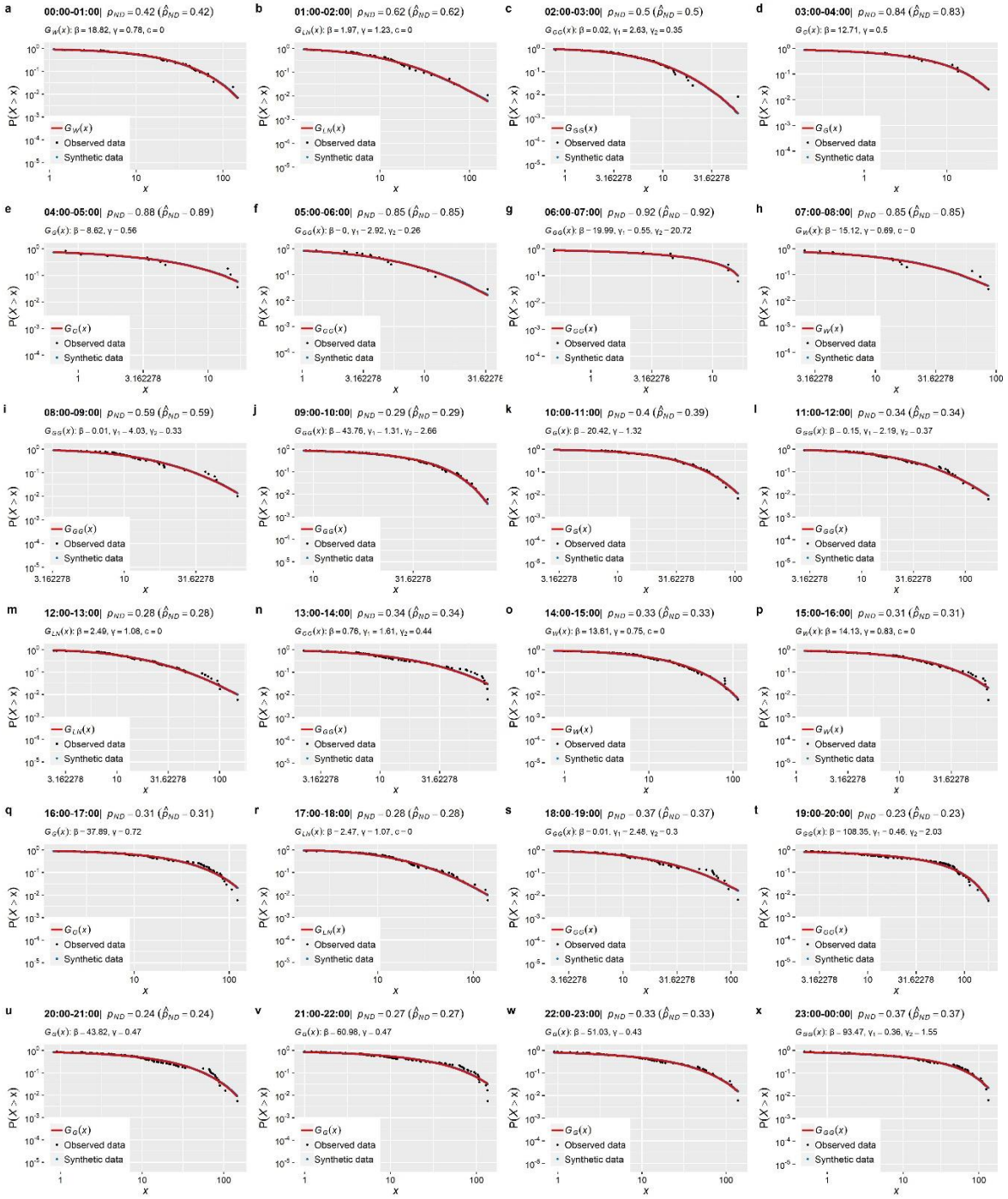

Figure 7. Observed, theoretical and simulated empirical probability plots (Weibull plotting positions and in logarithmic scale) of the nonzero hourly water demand for each hour of the day. Each plot displays the parameters of the best fitted distribution as well as the observed p_{ND} and simulated $\hat{p}_{ND}$ probability of no demand.

Furthermore, the performance of the stochastic model in terms of reproducing the target hour-to-hour correlation coefficients is graphically presented in Figure 8. As shown, the model, further to the marginal distributions, captures perfectly the great variety of lag-1 correlations of hourly water demand which is observed within a day.

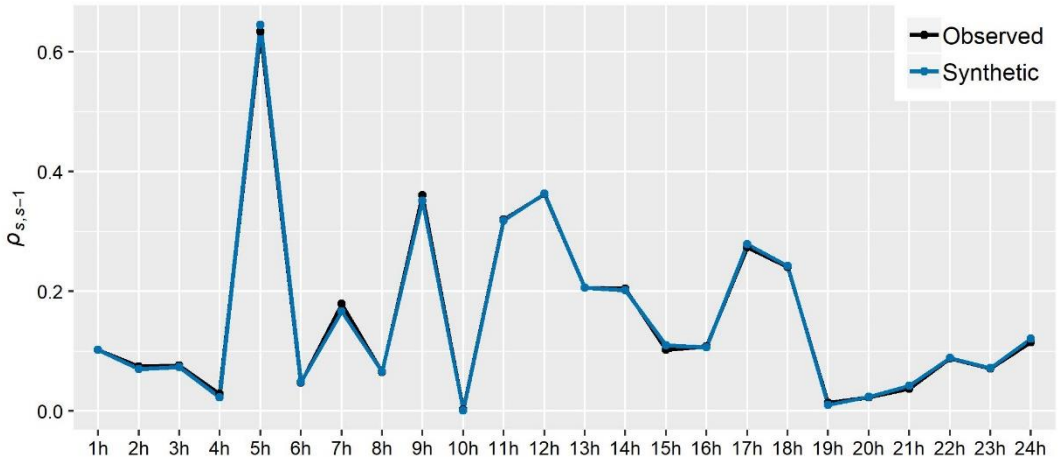

Figure 8. Comparison between observed and simulated lag-1 h-to-hour correlation coefficients.

Finally, for demonstration purposes, Figure 9a,b depict the time series of a sequence of 10 randomly selected consecutive days of observed and synthetic hourly demands, respectively. Additionally, the capability of the model to capture the cdf of the entire observed hourly water demand series and the series of the observed total daily volumes is graphically presented in Figure 9c,d, respectively. It is noted that the model adequately captures the empirical distribution of the whole series, although not directly modelled by it. A fact that can be attributed to the preservation of the cyclostationary behavior of the process, in terms of both marginal and dependence properties.

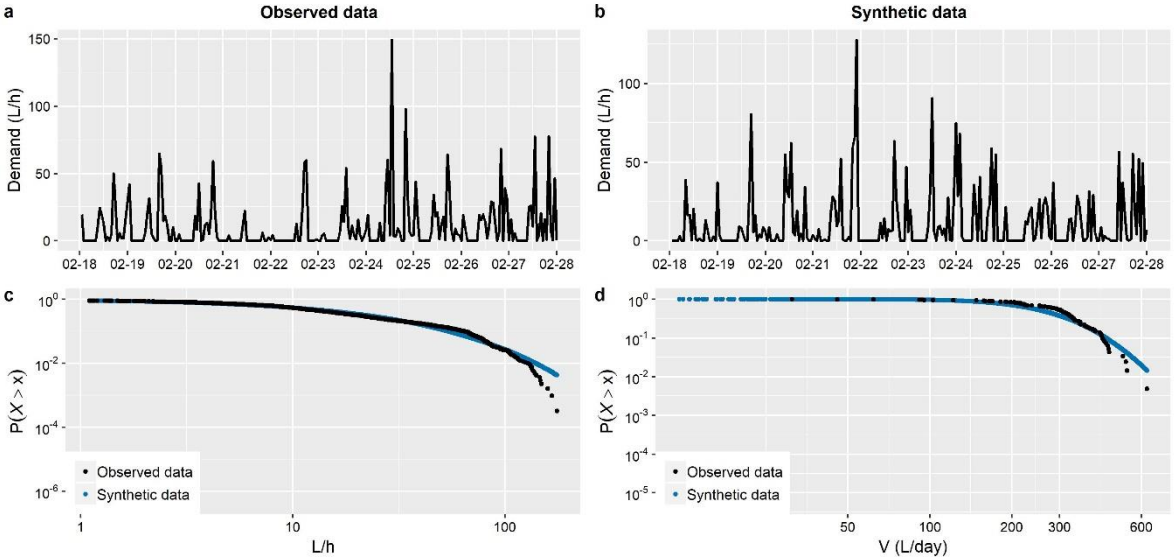

Figure 9. (**a**) Observed hourly water demand data of 10 days. (**b**) Synthetic hourly water demand data of 10 days. (**c**) Observed and simulated cumulative distribution functions (Weibull plotting positions and in logarithmic scale) the entire hourly water demand series. (**d**) Observed and simulated cumulative distribution functions (Weibull plotting positions and in logarithmic scale) of total water volume per day.

4. Summary and Conclusions

Due to the key role of water demand processes in the uncertainty-aware planning and management of urban water systems, the pursuit of adequately representing and simulating such processes has been a challenge that occupied many researchers over the years and resulted in the development of numerous simulation schemes. Depending on the time scale of analysis (which is typically a fine one), water demand processes exhibit a variety of marginal and dependence (expressed via auto-or cross-correlations) characteristics (e.g., non-Gaussianity, intermittency, auto-dependence) that should be reproduced by the stochastic model. However, already existing stochastic models provide a partial solution to the problem since they focus either on the marginal behavior of a process or on its dependence structure, and do not consider simultaneously both aspects.

Aiming to provide a remedy, and eventually a satisfactory modelling and simulation strategy for water demand processes, this work employs for the first time in the WDS modelling domain, the so-called Nataf-based stochastic models that entail the coupling of linear stochastic models with quantile mapping procedures. This type of models, recently employed to address challenging stochastic simulation problems for non-Gaussian hydrometeorological processes [51–53], build upon the notion of Nataf's joint distribution model [47] which provides a sound theoretical background, as well as the necessary flexibility to account for a wide range of processes with arbitrary marginal distributions and correlation structures.

The generality, as well as the flexibility provided by the Nataf-based modelling strategy (detailed in Section 2 and summarized in Section 2.8) is stress-tested through three distinct simulation studies that are of particular interest for urban water systems (Section 3). These concern the stochastic simulation of water demand processes at three characteristic time scales (i.e., 1-min, 15-min and 1-h), which arguably exhibit a variety of peculiarities, thus emphasizing different aspects of the method.

The main outcome from the simulation studies is that the employed modelling strategy is sufficient for the simulation of water demand process (stationary or cyclostationary) at any time scale (from 1 h down to 1 min), proving capable of reproducing simultaneously both the marginal and dependence properties of the processes without sacrificing neither precision nor parsimony. Particularly, it is highlighted that depending on the observed process characteristics, the method can be combined with a three-component mixed-distribution model (an additional contribution of this work) to provide an explicit representation of the entire target marginal distribution, accounting also for intermittency and tail behavior. Further to this, we move beyond the reproduction of empirical temporal correlations (typically involving few time lags) to a more complete and parsimonious description of the process's temporal dependence properties through the use of theoretical correlation structures.

We argue that this work provides an interesting alternative to the prominent pulse-based methods (that mainly provide a moments-based representation of a process) for simulating water demand processes, introducing, for the first time within water demand process simulation, classic and easy to implement linear stochastic models. In our view, this approach, exploiting the flexibility (in terms of admissible distributions and correlations structures) of Nataf's model, as well as knowledge derived from large scale studies (e.g., in the spirit of Kossieris and Makropoulos [39]) which allow for a better understanding of the probabilistic laws and correlations that describe water demand, can greatly contribute to the widespread use of alternative stochastically generated water demand realizations, for a more uncertainty-aware design and management of urban water systems. Potential domains of future research could focus on the simulation of spatio-temporal water demand processes at node or household levels, thus accounting also for spatial dynamics. Furthermore, the proposed methodology could be also extended and studied in a disaggregation framework to allow the capturing of probabilistic and stochastic behaviour of water demand across multiple time scales, as well as for short-term and long-term forecasting.

Author Contributions: This manuscript is the result of the research of P.K. and I.T., under the supervision of C.M. and D.S.

Funding: This research received no external funding.

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). All analyses were implemented in the R programming environment [93], while all graphs were produced with the use of *ggplot2* package [96]. The methodology presented in the present work is implemented in the form of an R package, named *anySim*, available at: http://www.itia.ntua.gr/en/softinfo/33/ [94].

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

Appendix A

This Appendix includes the probability density functions of the probabilistic models used in the present paper.

Table A1. Probability density functions of the distributions used in the present work.

Distribution	Probability Density Function
Gamma (G)	$f_G(x) = \frac{1}{\beta\Gamma(\gamma)}\left(\frac{x}{\beta}\right)^{\gamma-1} exp\left(-\frac{x}{\beta}\right),\ x > 0$ where $\beta > 0$ is the scale parameter and $\gamma > 0$ the shape parameter.
Weibull (W)	$f_W(x) = \frac{\gamma}{\beta}\left(\frac{x}{\beta}\right)^{\gamma-1} exp\left(-\left(\frac{x}{\beta}\right)^{\gamma}\right),\ x > 0$ where $\beta > 0$ is the scale parameter and $\gamma > 0$ the shape parameter.
Lognormal (LN)	$f_{LN}(x) = \frac{1}{\sqrt{\pi}\gamma x} exp\left(-ln^2\left(\frac{x}{\beta}\right)^{1/\gamma}\right),\ x > 0$ where $\beta > 0$ is the scale parameter and $\gamma > 0$ the shape parameter.
Generalised Gamma (GG)	$f_{GG}(x) = \frac{\gamma_2}{\beta\Gamma(\gamma_1/\gamma_2)}\left(\frac{x}{\beta}\right)^{\gamma_1-1} exp\left(-\left(\frac{x}{\beta}\right)^{\gamma_2}\right),\ x > 0$ where $\beta > 0$ is the scale parameter, while $\gamma_1 > 0$ and $\gamma_2 > 0$ are the shapes parameters that control the behaviour of the left and right tail, respectively.
Generalised Pareto (GPD)	$f_{GPD}(x) = \frac{1}{\beta}\left(1 + \frac{\gamma(x-c)}{\beta}\right)^{(-1/\gamma-1)},\ x > c$ where $\beta > 0$ is the scale parameter, γ the shape parameter and c the location parameter (threshold).

Appendix B

This Appendix comprises complementary graphs associated with the case studies presented in Sections 3.1 and 3.2.

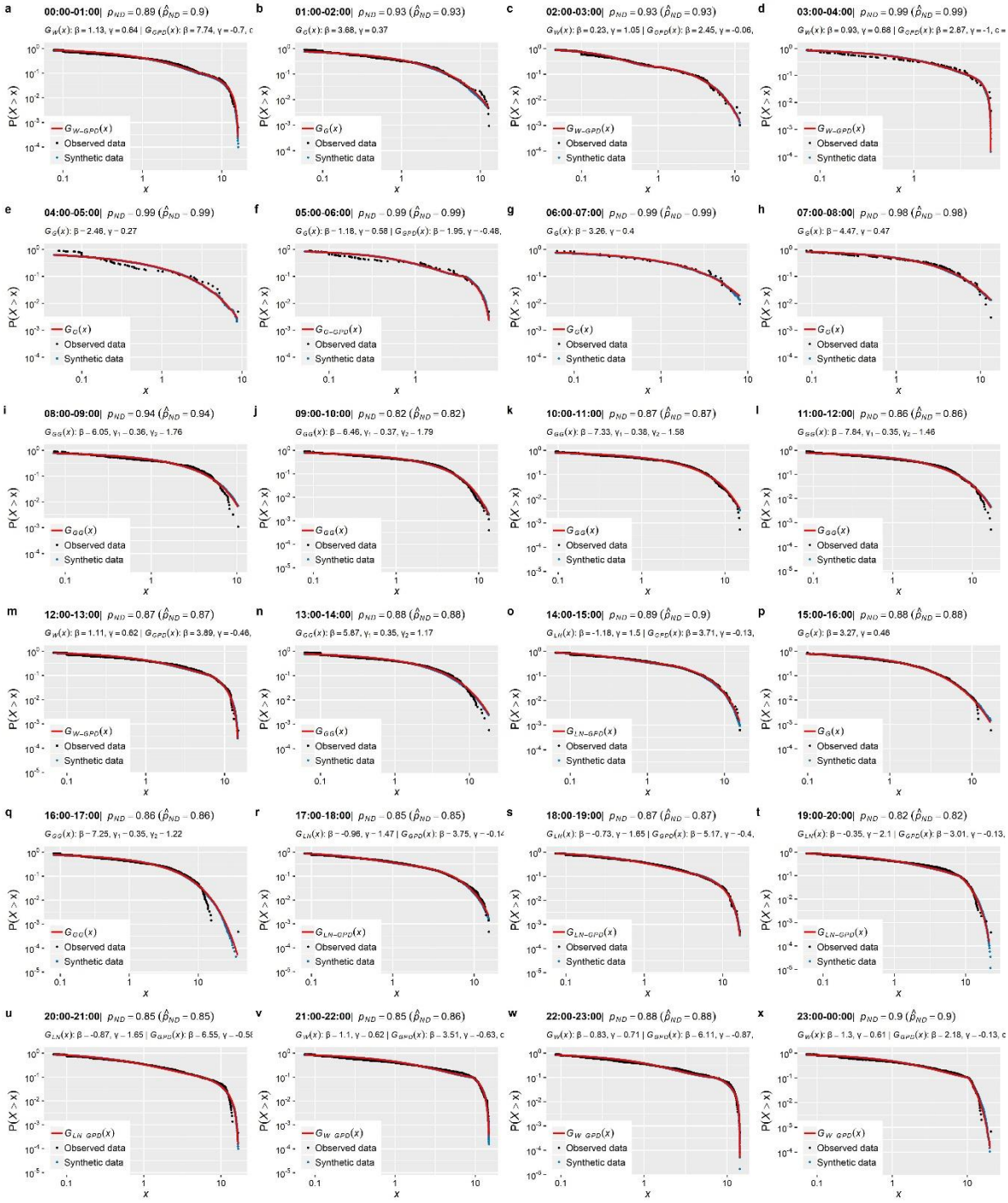

Figure A1. Observed, theoretical and simulated empirical probability plots (Weibull plotting positions and in logarithmic scale) of the nonzero 1-min water demand for each hour of the day. Each plot displays the parameters of the best fitted distribution as well as the observed p_{ND} and simulated $\hat{p}_{ND}$ probability of no demand.

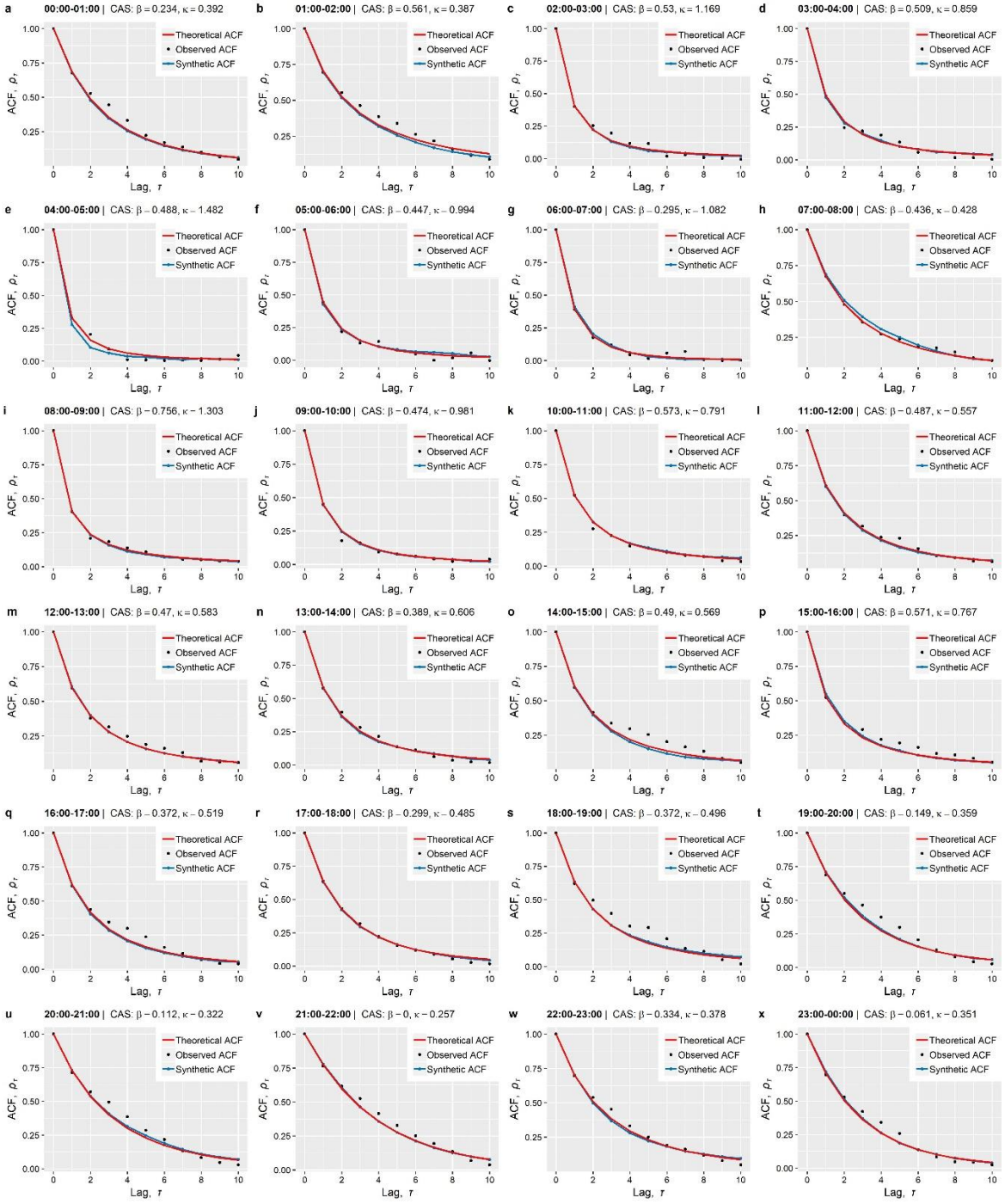

Figure A2. Observed, theoretical and simulated autocorrelation functions of 1-min water demand for each hour of the day. Each graph also displays the parameters of the fitted CAS.

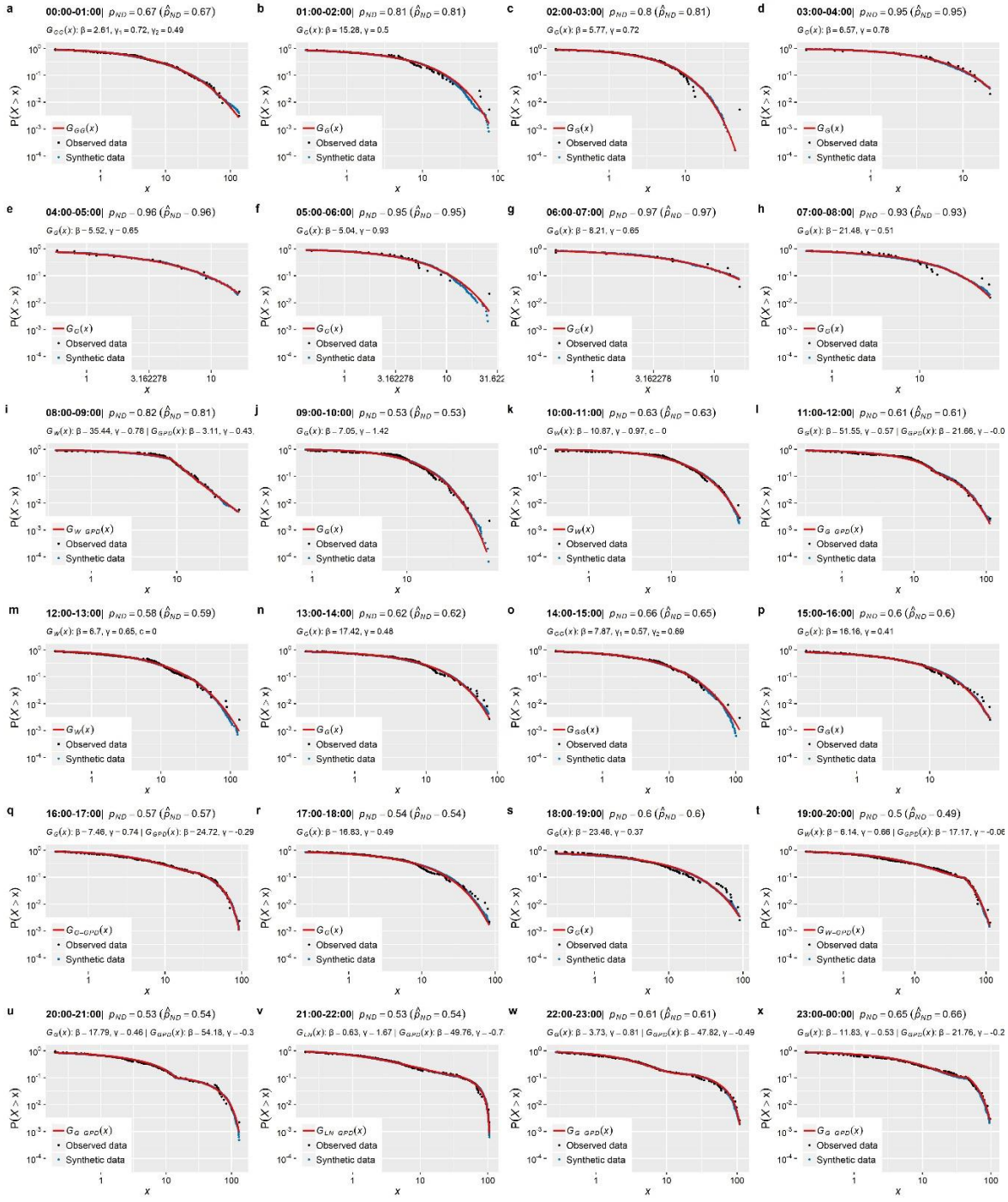

Figure A3. Observed, theoretical and simulated empirical probability plots (Weibull plotting positions and in logarithmic scale) of the nonzero 15-min water demand for each hour of the day. Each plot displays the parameters of the best fitted distribution as well as the observed p_{ND} and simulated $\hat{p}_{ND}$ probability of no demand.

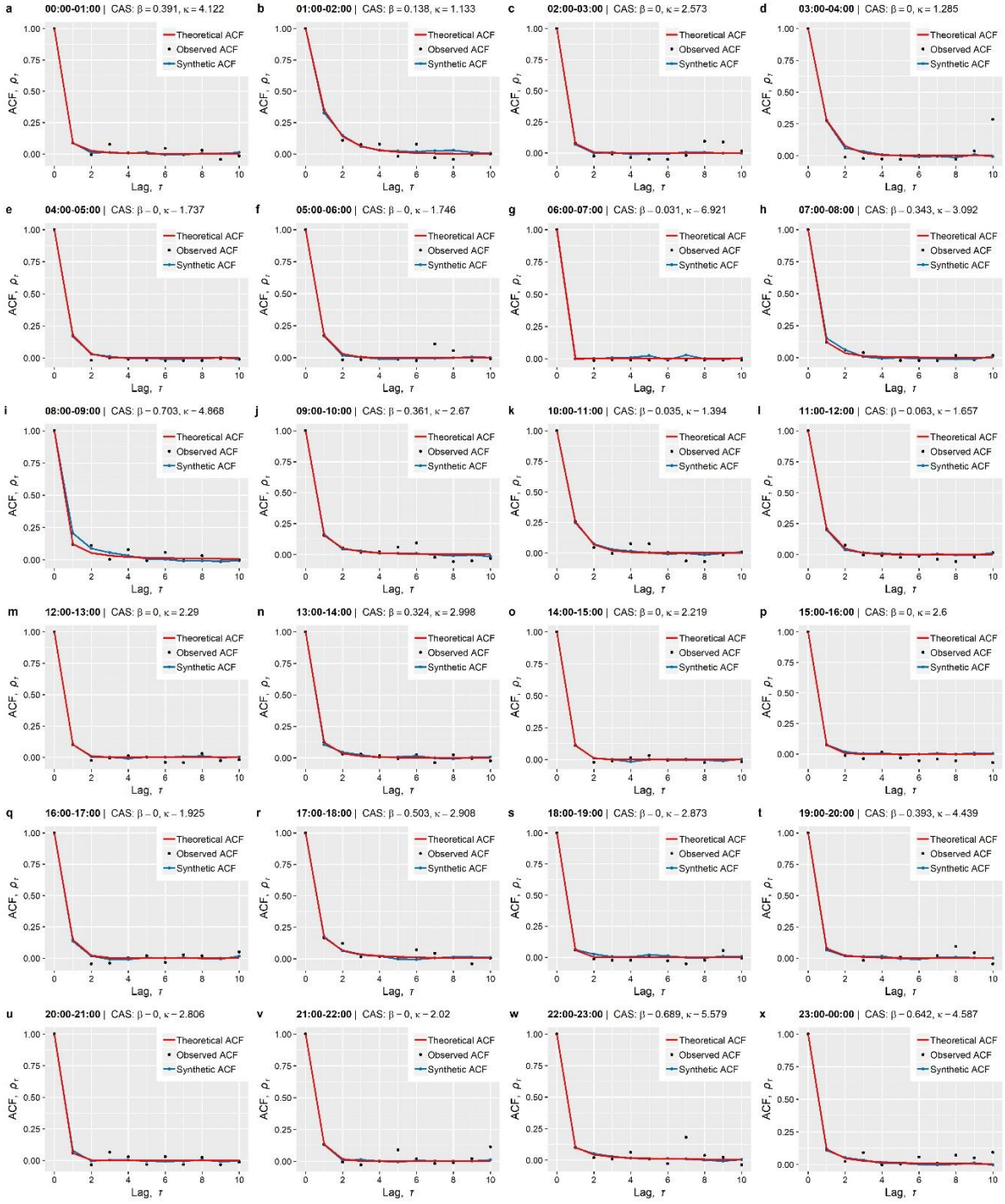

Figure A4. Observed, theoretical and simulated autocorrelation functions of 15 min water demand for each hour of the day. Each graph also displays the parameters of the fitted CAS.

References

1. Bao, Y.; Mays, L.W. Model for Water Distribution System Reliability. *J. Hydraul. Eng.* **1990**, *116*, 1119–1137. [CrossRef]
2. Walski, T.; Chase, D.; Savic, D.; Grayman, W.; Beckwith, S.; Koelle, E. *Advanced Water Distribution Modeling and Management*; Haestead Press: Waterbury, CT, USA, 2003; ISBN 0971414122.
3. Blokker, E.J.M.; Vreeburg, J.H.G.; Buchberger, S.G.; van Dijk, J.C. Importance of demand modelling in network water quality models: A review. *Drink. Water Eng. Sci. Discuss.* **2008**, *1*, 1–20. [CrossRef]

4. 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.
5. Filion, Y.; Adams, B.; Karney, B. Cross Correlation of Demands in Water Distribution Network Design. *J. Water Resour. Plan. Manag.* **2007**, *133*, 137–144. [CrossRef]
6. 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]
7. 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]
8. Magini, R.; Pallavicini, I.; Guercio, R. Spatial and Temporal Scaling Properties of Water Demand. *J. Water Resour. Plan. Manag.* **2008**, *134*, 276–284. [CrossRef]
9. Pacchin, E.; Alvisi, S.; Franchini, M. A Short-Term Water Demand Forecasting Model Using a Moving Window on Previously Observed Data. *Water* **2017**, *9*, 172. [CrossRef]
10. Gagliardi, F.; Alvisi, S.; Kapelan, Z.; Franchini, M. A Probabilistic Short-Term Water Demand Forecasting Model Based on the Markov Chain. *Water* **2017**, *9*, 507. [CrossRef]
11. Alvisi, S.; Franchini, M.; Marinelli, A. A short-term, pattern-based model for water-demand forecasting. *J. Hydroinformatics* **2006**, *9*, 39–50. [CrossRef]
12. Donkor, E.A.; Mazzuchi, T.A.; Soyer, R.; Alan Roberson, J. Urban Water Demand Forecasting: Review of Methods and Models. *J. Water Resour. Plan. Manag.* **2012**, *140*, 146–159. [CrossRef]
13. 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]
14. Boyle, T.; Giurco, D.; Mukheibir, P.; Liu, A.; Moy, C.; White, S.; Stewart, R. Intelligent metering for urban water: A review. *Water* **2013**, *5*, 1052–1081. [CrossRef]
15. Arregui, F.; Cabrera, E.; Cobacho, R. *Integrated Water Meter Management*; IWA Publishing: London, UK, 2006; ISBN 1843390345.
16. Buchberger, S.G.; Wells, G.J. Intensity, Duration, and Frequency of Residential Water Demands. *J. Water Resour. Plan. Manag.* **1996**, *122*, 11–19. [CrossRef]
17. Buchberger, S.G.; Wu, L. Model for instantaneous residential water demands. *J. Hydraul. Eng.* **1995**, *121*, 232–246. [CrossRef]
18. Garcia, V.J.; Garcia-Bartual, R.; Cabrera, E.; Arregui, F.; Garcia-Serra, J. Stochastic Model to Evaluate Residential Water Demands. *J. Water Resour. Plan. Manag.* **2004**, *130*, 386–394. [CrossRef]
19. Guercio, R.; Magini, R.; Pallavicini, I. Instantaneous residential water demand as stochastic point process. *Water Resour. Manag.* **2001**, *48*. [CrossRef]
20. Creaco, E.; Farmani, R.; Kapelan, Z.; Vamvakeridou-Lyroudia, L.; Savic, D. Considering the Mutual Dependence of Pulse Duration and Intensity in Models for Generating Residential Water Demand. *J. Water Resour. Plan. Manag.* **2015**, *141*, 04015031. [CrossRef]
21. 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]
22. Creaco, E.; Kossieris, P.; Vamvakeridou-Lyroudia, L.; Makropoulos, C.; Kapelan, Z.; Savic, D. Parameterizing residential water demand pulse models through smart meter readings. *Environ. Model. Softw.* **2016**, *80*, 33–40. [CrossRef]
23. 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]
24. Alvisi, S.; Franchini, M.; Marinelli, A. A stochastic model for representing drinking water demand at residential level. *Water Resour. Manag.* **2003**, *17*, 197–222. [CrossRef]
25. 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]
26. 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; pp. 1–14.

27. 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]
28. Blokker, E.J.M.; Vreeburg, J.H.G.; van Dijk, J.C. Simulating Residential Water Demand with a Stochastic End-Use Model. *J. Water Resour. Plan. Manag.* **2010**, *136*, 19–26. [CrossRef]
29. 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]
30. Murray, M.G.; Murray, P.M. WatSup model: A high resolution water supply modelling system. *J. Hydroinformatics* **2005**, *7*, 79–89. [CrossRef]
31. Makropoulos, C.K.; Natsis, K.; Liu, S.; Mittas, K.; Butler, D. Decision support for sustainable option selection in integrated urban water management. *Environ. Model. Softw.* **2008**, *23*, 1448–1460. [CrossRef]
32. Gargano, R.; Tricarico, C.; Del Giudice, G.; Granata, F. A stochastic model for daily residential water demand. *Water Sci. Technol. Water Supply* **2016**, *in press*, 1753–1767. [CrossRef]
33. Creaco, E.; Blokker, M.; Buchberger, S. Models for Generating Household Water Demand Pulses: Literature Review and Comparison. *J. Water Resour. Plan. Manag.* **2017**, *143*, 04017013. [CrossRef]
34. Rozos, E.; Makropoulos, C.; Butler, D. Design Robustness of Local Water-Recycling Schemes. *J. Water Resour. Plan. Manag.* **2010**, *136*, 531–538. [CrossRef]
35. Makropoulos, C.; Nikolopoulos, D.; Palmen, L.; Kools, S.; Segrave, A.; Vries, D.; Koop, S.; van Alphen, H.J.; Vonk, E.; et al. A resilience assessment method for urban water systems. *Urban Water J.* **2018**, *15*, 316–328. [CrossRef]
36. Buchberger, S.G.; Carter, J.T.; Lee, Y.H.; Schade, T.G. Random demands, travel times, and water quality in dead ends. *Awwarf* **2003**, 470.
37. 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]
38. Alvisi, S.; Ansaloni, N.; Franchini, M. A Procedure for Spatial Aggregation of Synthetic Water Demand Time Series. *Procedia Eng.* **2014**, *70*, 51–60. [CrossRef]
39. Kossieris, P.; Makropoulos, C. Exploring the Statistical and Distributional Properties of Residential Water Demand at Fine Time Scales. *Water* **2018**, *10*, 1481. [CrossRef]
40. Gargano, R.; Tricarico, C.; Granata, F.; Santopietro, S.; de Marinis, G. Probabilistic models for the peak residential water demand. *Water (Switzerland)* **2017**, *9*, 417. [CrossRef]
41. Magini, R.; Capannolo, F.; Ridolfi, E.; Guercio, R. Demand uncertainty in modelling WDS: scaling laws and scenario generation. *WIT Trans. Ecol. Environ.* **2016**. [CrossRef]
42. Vertommen, I.; Magini, R.; da Conceicao Cunha, M.; Guercio, R. Water Demand Uncertainty: The Scaling Laws Approach. In *Water Supply System Analysis-Selected Topics*; InTech: Lexington, KY, USA, 2012; ISBN 978-953-51-0889-4.
43. Magini, R.; Boniforti, M.; Guercio, R. Generating Scenarios of Cross-Correlated Demands for Modelling Water Distribution Networks. *Water* **2019**, *11*, 493. [CrossRef]
44. Tricarico, C.; de Marinis, G.; Gargano, R.; Leopardi, A. Peak residential water demand. *Proc. Inst. Civ. Eng. Water Manag.* **2007**, *160*, 115–121. [CrossRef]
45. Kapelan, Z.S.; Savic, D.A.; Walters, G.A. Multiobjective design of water distribution systems under uncertainty. *Water Resour. Res.* **2005**, *41*, 1–15. [CrossRef]
46. Moughton, L.J.; Buchberger, S.G.; Boccelli, D.L.; Filion, Y.R.; Karney, B.W. Effect of Time Step and Data Aggregation on Cross Correlation of Residential Demands. In *Water Distribution Systems Analysis Symposium 2006*; American Society of Civil Engineers: Reston, VA, USA, 2006; pp. 1–11.
47. Nataf, A. Statistique mathematique-determination des distributions de probabilites dont les marges sont donnees. *Comptes Rendus l'Academie des Sci.* **1962**, *255*, 42–43.
48. Cario, M.C.; Nelson, B.L. Autoregressive to anything: Time-series input processes for simulation. *Oper. Res. Lett.* **1996**, *19*, 51–58. [CrossRef]
49. Cario, M.C.; Nelson, B.L. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. *Ind. Eng.* **1997**, 1–19.
50. Biller, B.; Nelson, B.L. Modeling and generating multivariate time-series input processes using a vector autoregressive technique. *ACM Trans. Model. Comput. Simul.* **2003**, *13*, 211–237. [CrossRef]

51. 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]
52. Tsoukalas, I.; Makropoulos, C.; Koutsoyiannis, D. Simulation of stochastic processes exhibiting any-range dependence and arbitrary marginal distributions. *Water Resour. Res.* **2018**. [CrossRef]
53. Papalexiou, 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]
54. Tsoukalas, I.; Efstratiadis, A.; Makropoulos, C. Stochastic simulation of periodic processes with arbitrary marginal distributions. In Proceedings of the 15th International Conference on Environmental Science and Technology, Rhodes, Greece, 31 August–2 September 2017.
55. Papalexiou, 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]
56. Tsoukalas, I.; Efstratiadis, A.; Makropoulos, C. Building a puzzle to solve a riddle: A multi-scale disaggregation approach for multivariate stochastic processes with any marginal distribution and correlation structure. Unpublished manuscript.
57. Liu, P.-L.; Der Kiureghian, A. Multivariate distribution models with prescribed marginals and covariances. *Probabilistic Eng. Mech.* **1986**, *1*, 105–112. [CrossRef]
58. Lebrun, R.; Dutfoy, A. An innovating analysis of the Nataf transformation from the copula viewpoint. *Probabilistic Eng. Mech.* **2009**, *24*, 312–320. [CrossRef]
59. Xiao, Q. Evaluating correlation coefficient for Nataf transformation. *Probabilistic Eng. Mech.* **2014**, *37*, 1–6. [CrossRef]
60. Li, S.T.; Hammond, J.L. Generation of Pseudorandom Numbers with Specified Univariate Distributions and Correlation Coefficients. *IEEE Trans. Syst. Man. Cybern.* **1975**, *SMC-5*, 557–561. [CrossRef]
61. Li, H.; Lü, Z.; Yuan, X. Nataf transformation based point estimate method. *Sci. Bull.* **2008**, *53*, 2586–2592. [CrossRef]
62. Chen, H. Initialization for NORTA: Generation of Random Vectors with Specified Marginals and Correlations. *INFORMS J. Comput.* **2001**, *13*, 312–331. [CrossRef]
63. Koutsoyiannis, D. A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. *Water Resour. Res.* **2000**, *36*, 1519–1533. [CrossRef]
64. Kelly, K.S.; Krzysztofowicz, R. Precipitation uncertainty processor for probabilistic river stage forecasting. *Water Resour. Res.* **2000**, *36*, 2643–2653. [CrossRef]
65. Matalas, N.C. Mathematical assessment of synthetic hydrology. *Water Resour. Res.* **1967**, *3*, 937–945. [CrossRef]
66. Klemeš, V.; Borůvka, L. Simulation of Gamma-Distributed First-Order Markov Chain. *Water Resour. Res.* **1974**, *10*, 87–91. [CrossRef]
67. Wilks, D.S. Multisite generalization of a daily stochastic precipitation generation model. *J. Hydrol.* **1998**, *210*, 178–191. [CrossRef]
68. Papoulis, A. *Probability, Random Variables, and Stochastic Processes*; McGraw-Hill Series in Electrical Engineering: New York, NY, USA, 1991; ISBN 0-07-366011-6.
69. Robert, C.P.; Casella, G. *Monte Carlo Statistical Methods*; Springer Texts in Statistics; Springer: New York, NY, USA, 2004; Volume 42, ISBN 978-1-4419-1939-7.
70. Mostafa, M.D.; Mahmoud, M.W. On the problem of estimation for the bivariate lognormal distribution. *Biometrika* **1964**, *51*, 522–527. [CrossRef]
71. Tsoukalas, I.; Papalexiou, 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]
72. 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]
73. Bárdossy, A.; Pegram, G.G.S. Space-time conditional disaggregation of precipitation at high resolution via simulation. *Water Resour. Res.* **2016**, *52*, 920–937. [CrossRef]
74. Williams, P.M. Modelling seasonality and trends in daily rainfall data. *Sch. Cogn. Comput. Sci.* **1998**, *10*, 985–991.

75. Cannon, A.J. Probabilistic Multisite Precipitation Downscaling by an Expanded Bernoulli–Gamma Density Network. *J. Hydrometeorol.* **2008**, *9*, 1284–1300. [CrossRef]
76. Sharma, A.; Tarboton, D.G.; Lall, U. Streamflow simulation: A nonparametric approach. *Water Resour. Res.* **1997**, *33*, 291–308. [CrossRef]
77. Szulczewski, W.; Jakubowski, W. The Application of Mixture Distribution for the Estimation of Extreme Floods in Controlled Catchment Basins. *Water Resour. Manag.* **2018**, *32*, 3519–3534. [CrossRef]
78. Naveau, P.; Huser, R.; Ribereau, P.; Hannart, A. Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resour. Res.* **2016**, *52*, 2753–2769. [CrossRef]
79. Li, C.; Singh, V.P.; Mishra, A.K. Simulation of the entire range of daily precipitation using a hybrid probability distribution. *Water Resour. Res.* **2012**, *48*, 1–17. [CrossRef]
80. MacDonald, A.; Scarrott, C.J.; Lee, D.; Darlow, B.; Reale, M.; Russell, G. A flexible extreme value mixture model. *Comput. Stat. Data Anal.* **2011**, *55*, 2137–2157. [CrossRef]
81. Behrens, C.N.; Lopes, H.F.; Gamerman, D. Bayesian analysis of extreme events with threshold estimation. *Stat. Model. An Int. J.* **2004**, *4*, 227–244. [CrossRef]
82. Solari, S.; Losada, M.A. A unified statistical model for hydrological variables including the selection of threshold for the peak over threshold method. *Water Resour. Res.* **2012**, *48*. [CrossRef]
83. Hu, Y.; Scarrott, C. evmix: An R package for Extreme Value Mixture Modeling, Threshold Estimation and Boundary Corrected Kernel Density Estimation. *J. Stat. Softw.* **2018**, *84*. [CrossRef]
84. 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. [CrossRef]
85. Koutsoyiannis, D. The Hurst phenomenon and fractional Gaussian noise made easy. *Hydrol. Sci. J.* **2002**, *47*, 573–595. [CrossRef]
86. Papalexiou, S.-M.; Koutsoyiannis, D.; Montanari, A. Can a simple stochastic model generate rich patterns of rainfall events? *J. Hydrol.* **2011**, *411*, 279–289. [CrossRef]
87. Mandelbrot, B.B. A Fast Fractional Gaussian Noise Generator. *Water Resour. Res.* **1971**, *7*, 543–553. [CrossRef]
88. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. *Time Series Analysis*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2008; Volume 37, ISBN 9781118619193.
89. Gneiting, T.; Schlather, M. Stochastic Models That Separate Fractal Dimension and the Hurst Effect. *SIAM Rev.* **2004**, *46*, 269–282. [CrossRef]
90. Beran, J. *Statistics for Long-Memory Processes*; CRC Press Book: Boca Raton, FL, USA, 1994; ISBN 0412049015.
91. Koutsoyiannis, D. Climate change, the Hurst phenomenon, and hydrological statistics. *Hydrol. Sci. J.* **2003**, *48*, 3–24. [CrossRef]
92. Dimitriadis, P.; Koutsoyiannis, D. Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst–Kolmogorov processes. *Stoch. Environ. Res. Risk Assess.* **2015**, *29*, 1649–1669. [CrossRef]
93. R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing 2017. Available online: https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=2144573 (accessed on 20 February 2019).
94. Tsoukalas, I.; Kossieris, P. anySim: Stochastic simulation of processes with any marginal distribution and correlation structure. Available online: http://www.itia.ntua.gr/en/softinfo/33/ (accessed on 20 February 2019).
95. Savić, D.; Vamvakeridou-Lyroudia, L.; Kapelan, Z. Smart Meters, Smart Water, Smart Societies: The iWIDGET Project. *Procedia Eng.* **2014**, *89*, 1105–1112. [CrossRef]
96. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Statistics and Computing; Springer-Verlag: New York, NY, USA, 2016.