

Article

Burst Detection in Water Distribution Systems: The Issue of Dataset Collection

Andrea Menapace ¹, Ariele Zanfei ¹, Manuel Felicetti ², Diego Avesani ¹,
Maurizio Righetti ^{1,*} and Rudy Gargano ³

¹ Faculty of Science and Technology, Free University of Bozen-Bolzano, Universitätsplatz 5, 39100 Bolzano, Italy; andrea.menapace@unibz.it (A.M.); ariele.zanfei@natec.unibz.it (A.Z.); diego.avesani@unibz.it (D.A.)

² Department of Civil, Environmental and Mechanical Engineering, University of Trento, via Mesiano 77, 38123 Trento, Italy; manuel.felicetti@unitn.it

³ Department of Civil and Mechanical Engineering, University of Cassino and Southern Lazio, via G. Di Biasio 43, 03043 Cassino, Italy; gargano@unicas.it

* Correspondence: Maurizio.Righetti@unibz.it

Received: 28 September 2020; Accepted: 17 November 2020; Published: 20 November 2020



Abstract: Developing data-driven models for bursts detection is currently a demanding challenge for efficient and sustainable management of water supply systems. The main limit in the progress of these models lies in the large amount of accurate data required. The aim is to present a methodology for the generation of reliable data, which are fundamental to train anomaly detection models and set alarms. Thus, the results of the proposed methodology is to provide suitable water consumption data. The presented procedure consists of stochastic modelling of water request and hydraulic pipes bursts simulation to yield suitable synthetic time series of flow rates, for instance, inlet flows of district metered areas and small water supply systems. The water request is obtained through the superimposition of different components, such as the daily, the weekly, and the yearly trends jointly with a random normal distributed component based on the consumption mean and variance, and the number of users aggregation. The resulting request is implemented into the hydraulic model of the distribution system, also embedding background leaks and bursts using a pressure-driven approach with both concentrated and distributed demand schemes. This work seeks to close the gap in the field of synthetic generation of drinking water consumption data, by establishing a proper dedicated methodology that aims to support future water smart grids.

Keywords: anomaly detection methods; hydraulic leakages modeling; stochastic water request analysis; synthetic dataset generation; water distribution systems

1. Introduction

The water distribution infrastructures, along with the energy grids, have undertaken an important renewal process in the last few years that aims to transform the current distribution networks in smart grids [1,2]. The reasons for this changing lie in the necessity of water distribution systems (WDS, hereafter) to tackle water scarcity because of climate change and to deal with water requests increasing due to population growth in urban areas [3]. Therefore, water-smart networks comprise smart sensors (e.g., flow, pressure and quality meters, and noise loggers) and communication and data storage devices together with data management and analysis routines [4]. Despite the potential benefits of smart systems [2], above all the enhancement of WDS management, proper handling of the data stream through the different components represents a real challenge nowadays [5].

Smart WDS enables improving the control of water leakages which are a demanding task well known in the literature. Water leaks comprise the background losses and pipes bursts [6,7]. The former

are characterised by a high number of small losses spread along the distribution network, which are mainly caused by the time deterioration of the pipes, valves, fittings, fire hydrants, pumps, and service connections, while the latter are defined as abrupt cracks with a considerable amount of water outflow and a limited duration over time. Thus, quick identification and localization of water pipe bursts is crucial for decreasing the amount of water losses and limiting the service interruptions [8]. At this aim, the data stream provided by the smart water grids plays a crucial role in providing the necessary inputs to leaks detection systems. Consequently, the efficiency of leaks detection depends on two main factors: on one hand, the quality and frequency of the data collected by the smart system, and, on the other, the accuracy and the reliability of the bursts detection algorithms.

In recent years, the growth of available data has given a significant boost to the development of suitable leaks detection algorithms. Several data-driven methods, which aim to identify abnormal water consumption, have been proposed based on inlet flow rates and pressure measurements. In fact, water consumption follows a typical pattern in an ideal scenario. However, sudden events, such as bursts and hydrants use, can generate sharp changes. Therefore, these unusual spikes are identified as outliers in the water consumption time series and can be detected through an anomaly detection approach [9].

According to We and Liu [10], the data-driven methods adopted in bursts detection can be divided into classification methods, prediction methods, and statistical methods. The first group of models is based on identifying the data with bursts among the normal data (normal data means data not affected by unusual behaviors). At this scope, for instance, static and time delay artificial neural networks (ANN, hereafter) are used by Mounce and Machell [11], while Aksela et al. [12] introduced an alarm system that relies on self-organizing map ANN. The second methods rely on a prediction phase of normal data followed by a classification: when a burst occurs, the prediction greatly differs from the measured data. A remarkable number of models has been developed belonging in this group, such as those ones based on Kalman filter [13,14], support vector regression [15,16], and ANN [11,17,18]. Instead, the third group includes the models based on statistic theories. In this group, the method mainly used is the statistical process control (SPC, hereafter) that allows for identifying variations in the data stream by means of control charts and graphs given control thresholds [19,20].

Despite the different classes of methods, suitable data are fundamental for the proper development of such detection models [21]. Specifically, normal data and data affected by bursts (called also abnormal data) are necessary for several purposes. The classification methods need lacking labels data, which means time series consisting by normal and abnormal data with information on the bursts occurrence, for the algorithm training. On the contrary, only normal data without outliers are used by prediction methods. In addition, statistic methods, which are set up for identifying outliers, require both normal and abnormal data. In addition, lacking labels are compulsory for the evaluation of the performance of the all classes of methods and the uncertainty analysis of the alarms.

In order to provide the massive amount of data required, three options are currently available. Firstly, burst events can be collected by opening fire hydrants or drain valves [22,23]. This process allows for having reliable data with a full knowledge of leakages in terms of location, duration, and magnitude. However, it has the drawbacks that this procedure is time-consuming, economically expensive, and above all causes disservice. Secondly, real hydraulic data with historical bursts report can be used, as [19,24]. Nevertheless, two important limitations affect real dataset: the knowledge of bursts time, amount of water losses, and the data uncertainty due to the sensors performance and the presence of missing data. Thirdly, another option for providing the data are the synthetic data generation as proposed by [25,26]. In this case, the main issue is that the methods trained with such data are not reliable for real applications due to scarce variation in the provided dataset.

Given that the large amount of data required is the main limitation of all categories of data-driven models [10], the authors deem that the only feasible way to provide the needed dataset is the synthetic data generation due to the difficulties in collecting accurate real data or in generating bursts by hydrants. However, no proper data generator methodology has already been proposed that is able to

provide suitable data for the development of anomaly detection methods because of scarce variation of water demand representation and poor accuracy of bursts modeling. Thus, this work aims to lay the foundations in the synthetic generation of water consumption data in order to support the development of bursts detection in WDS. Therefore, a methodology for the generation of water consumption time series is presented, which comprises all its components, i.e., the water request, the background losses and the bursts, in terms of mean and variation.

For the sake of clarity, the water request is defined in this paper as the water demand required by the aqueduct users. The term demand is used instead to indicate the water demand actually supplied by the water distribution system to the users. In addition, water consumption means the total amount of water delivered by the water supply system including both water demand and water losses along the distribution network.

Summarizing, the contribution of this study is a reliable methodology that provides suitable water consumption data. The final scope is to support the development of data-driven techniques for the bursts detection, which need a large amount of accurate data. Specifically, the proposed methodology, hereafter refers as stochastic-hydraulic time series generator (SHtsG, hereafter), relies on two principal phases: the water request modeling and the hydraulic WDS simulation. The first stage is based on a superimposition approach that takes into account the daily, weekly, and seasonal deterministic patterns together with the random component of the variation. Its output is the water request time series of the different group of WDS districts. On the other hand, the second stage involves the production of hydraulic data on the base of a distributed fully pressure driven model (DFPDM, hereafter) able to properly simulate the water demand, the background leakages, and the bursts. This model simulates each component of the water request with a proper pressure–demand relationship and suitable demand scheme: distributed along pipes for the water demand and background leakages and concentrated at nodes for bursts. In addition, DFPDM enables randomly generating different types of pipe bursts by varying the position along the networks, the amount of water losses, and the frequency and duration according to a defined occurrence probability. Therefore, DFPDM that allows for obtaining extended period simulations includes both the hydraulic and mechanical reliability of WDS through a proper simulation of the satisfied water request and the pipe breaks [27,28]. The results highlight the generation of a suitable hydraulic dataset of a lacking labels time series, which are closed to real data in both mean and variance. Eventually, the SHtsG is able to provide a massive amount of data that is crucial in the development of machine learning algorithms for anomaly detection.

The rest of the paper is organized as follows: Section 2 presents the framework of the proposed methodology; Section 3 depicts two test cases aimed at validating the methodology as well as enlightening its advantages and strengths; and Section 4 draws conclusions and final remarks.

2. Methodology

The proposed stochastic-hydraulic methodology, called SHtsG, consists of two phases: the first refers to the stochastic generation of a water request time series and the second regards the WDS simulation. Specifically, the stochastic part superimposes daily, weekly, and yearly water request patterns jointly with a random component in order to produce suitable water request time series. These are subsequently employed in the second part, where the hydraulic solver computes a WDN water consumption time series through an extended period simulation.

For the sake of clarity, Figure 1 illustrates the SHtsG. Specifically, the first part provides the water consumption time series for each district/group of users of the WDS (district is defined as a group of neighboring users with uniform characteristics) by means of a superimposition of deterministic and random trends. This SHtsG stage contemplates multiple time series according to different district typologies as well as various user aggregation. Given the water request distribution along the network, the WDS hydraulic model is built in the second part of the SHtsG. The hydraulic model exploits a fully pressure driven approach, called DFPDM, where water demands, background leakages, and water bursts are properly embedded in order to ensure extended period simulations accuracy and fidelity.

The final output is the consumption time series of the WDS inlet flow rate (or district metered areas flow rates) considering both hydraulic and mechanical reliability.

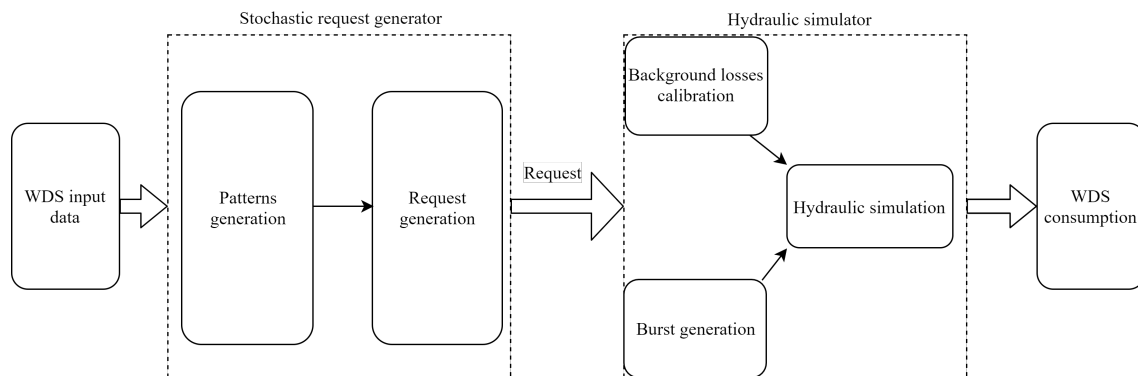


Figure 1. Flow chart of the proposed SHtsG for developing bursts detection models.

2.1. Water Request Stochastic Modeling

The first phase of the SHtsG deals with the modeling of the water request through a superimposition approach of different trends, as follows:

$$D_{ut} = \bar{D}_{uy}\alpha_{d,ut}\alpha_{w,ut}\alpha_{y,ut} + \epsilon_{ut} \quad \text{with } D_{ut} \in \mathbb{R}^+, \tag{1}$$

where the subscript u and t , respectively, represent the users' group/district and the time coordinate of the water request, while D_{ut} denotes the water request at the time t of the group of users u , and \bar{D}_{uy} is the annual average water request of group u . The daily water request pattern is $\alpha_{d,ut} = \bar{D}_{uh}/\bar{D}_{ud}$, which represents the dimensionless daily behavior of the water request with an hourly time interval expressed by hourly average request (\bar{D}_{uh}) over daily average request (\bar{D}_{ud}). Besides, the weekly and yearly dimensionless pattern are the ratio between daily average request and monthly average request ($\alpha_{w,ut} = \bar{D}_{ud}/\bar{D}_{um}$), and between monthly average request and yearly average request ($\alpha_{y,ut} = \bar{D}_{um}/\bar{D}_{uy}$), respectively. On the other hand, ϵ_{ut} represents the random component of the water request of the group of users u at the time t . Although ϵ_{ut} could be a negative number, the water request D_{ut} in Equation (1) (D_{ut} is the sum of one real non-negative term and another real term) has to belong to the real positive numbers due to the definition itself of water request that can not results negative.

The water request in Equation (1) consists of two parts: the first addend regards the deterministic terms and the second one the random term. This formula allows for properly modeling the water request of different users aggregation due to the different deterministic temporal trends and to the random component.

On one hand, the deterministic component uniquely defines the water request by merging the different dimensionless trends, e.g., $\alpha_{d,ut}$, $\alpha_{w,ut}$ and $\alpha_{y,ut}$ together with the yearly average request for each analyzed group of users. On the other one, the random component adds a variability effect at the deterministic request making the resulting time series more realistic. This component is calculated on the basis of the number of users aggregated and the first and second order moments of the water request itself. Given that the number of users aggregated and the water request mean are given (deterministic part of Equation (1)), the second order moment is defined through the Gargano et al. [29,30] approach, as follows:

$$CV_{ut} = 0.1 + \frac{6}{(0.25\bar{D}_{ut} N_u)^{3/4}}, \tag{2}$$

where CV is the coefficient of variation, which is defined as the ratio of the standard deviation to the mean (σ/μ), and N_u is the number of users aggregated. Equation (2) is valid for N_u ranges between 200 and 1250 users. Since the water request variation is defined, the random component of each time step for each group of users (ϵ_{ut}) is determined by a random generation of the normal distribution

$\mathcal{N}(\mu, \sigma^2)$. The μ and σ^2 represent the time step average request \bar{D}_{ut} and its variance, respectively. The variance is derived by Equation (2) as follows: $\sigma^2 = (CV_{ut} \bar{D}_{ut})^2$.

The water requests of the different district/groups of users u are implemented in the hydraulic model by distributing them uniformly along the pipes [31] belonging to the corresponding district area. The uniformly distributed requests are defined as the water request per unit length:

$$d_{ut} = \frac{D_{ut}}{\sum_{ij \in u} L_{ij}}, \tag{3}$$

where L_{ij} denotes the length of the ij -th pipe belonging to group u . The water request of each pipe of the u -th group thus result in being uniform.

2.2. Water Consumption Hydraulic Simulation

The second phase of the SHtsG aims to provide a synthetic consumption time series close to real ones due to both a suitable calculation of the satisfied water request, called demand, and a proper modeling of the water losses with the DFPDM. The water consumption for each group of users reads as:

$$Wt_{ut} = Wd_{ut} + Wl_{ut} + Wb_{ut}, \tag{4}$$

where the Wt_{ut} defines the water consumption for the group u ; Wd_{ut} , Wl_{ut} , Wb_{ut} are instead the water demand, the background losses, and the water bursts, respectively. The water consumption is represented by the distributed along pipe consumption scheme and the concentrated at the nodes demand scheme (for details, see [32]). The first is used for implementing water demand and background leakages, as follows:

$$p_{ij} = pd_{ij} + pl_{ij}, \tag{5}$$

where ij is the pipe index on which the withdrawals are spread, while p_{ij} denotes the part of the consumption assumed distributed along the ij -th pipe, pd_{ij} , and pl_{ij} are instead the distributed water demand and background leakages of the ij -th pipe. These variables are expressed as flow rate per unit length and are characteristic related to the pipes. On the contrary, the second component is the focus of lumped withdrawals at the i -th node, and reads as:

$$q_i = qb_i, \tag{6}$$

where q_i is the component of the water consumption concentrated at nodes which consists only of the water bursts qb_i . They are measured as flow rate drawn at the nodes of the network.

The mathematical equations on the base of the DFPDM are the mass balance at the node \hat{i} :

$$\sum_i \left(Q_{\hat{i}\hat{i}} - \int_0^{L_{\hat{i}\hat{i}}} p_{\hat{i}\hat{i}}(h(x)) dx \right) - \sum_j Q_{\hat{i}j} - q_{\hat{i}}(h_{\hat{i}}) = 0, \tag{7}$$

where i and j are the ending pipe nodes, x is the pipe longitudinal coordinate, \hat{i} is the considered node for the mass balance, h is the hydraulic head, and Q_{ij} is the flow rate at the node i . p_{ij} and q_i , which are defined above, denote the water consumption assumed as distributed and concentrated, respectively. The head loss equation, which is expressed according to the Darcy–Weisbach formula, for the ij -th pipe reads as follows:

$$h_i - h_j = r_{ij} \int_0^{L_{ij}} \left(Q_{ij} - \int_0^x p_{ij}(h(x)) d\xi \right) \left| Q_{ij} - \int_0^x p_{ij}(h(x)) d\xi \right| dx, \tag{8}$$

where h_i and h_j is the hydraulic head at node i and j , and r_{ij} denotes the hydraulic resistance per unit length of the ij -th pipe. Equation (8) is the generic formulation of the head loss drop of the pipes

that embraces all the possible cases. For example, Equation (8) in the case of null distributed water consumption along pipes assumes the following form:

$$h_i - h_j = r_{ij}L_{ij} |Q_{ij}| Q_{ij} \quad \text{if } p_{ij} = 0, \tag{9}$$

while the case of pipe water consumption different from zero in demand driven hydraulic conditions results as follows:

$$h_i - h_j = \frac{r_{ij}}{3p_{ij}} \left(|Q_{ij}|^3 - |Q_{ij} - p_{ij}L_{ij}|^3 \right) \quad \text{if } p_{ij} = c \{c \in \mathbb{R}^+\}. \tag{10}$$

In this case, the final water demand is equal to the water request ($p_{ij} = d_{ij}$) given the required pressure for regular working. The last possible case regards pipes working in a pressure driven condition with distributed water consumption. The distributed request p_{ij} is therefore function of the hydraulic head along the ij -th pipe and can be assumed as a second order polynomial function:

$$p_{ij} = w_{ij,1}x^2 + w_{ij,2}x + w_{ij,3}, \tag{11}$$

where $w_{ij,1}$, $w_{ij,2}$, and $w_{ij,3}$ are the three coefficients defined by means of the punctual pressure–consumption relationship evaluated at the top, middle, and end points of pipe. In this work, both the three coefficients of $p_{ij}(h(x))$ and the nodal $q_{ij}(h(x))$ consumption are evaluated by Siew and Tanyimboh [33]. The solver used for performing the hydraulic simulations is based on a Menapace and Avesani numerical scheme [34]. This hydraulic solver enables simulating both distributed along a pipe and lumped at nodes water withdrawals with a pressure-driven approach using the GGA method [35].

Since the mathematical equations on the basis of the DFPDM have been presented, the modeling of the different components of the water losses is herewith specified. Firstly, the water demand is modeled with the pressure–demand relationship [33] and the distributed along pipe demand scheme [34]. The accuracy of this approach, which has been proved in [36,37], is given by the uniformly widespread withdrawals along each pipe of the network depending on the pressure along it, as follows:

$$pd_{ij} = d_{ij}(h_{ij}). \tag{12}$$

In the case of normal operating pressure conditions, the water demand pd_{ij} of the ij -th pipe results in being equal to the water request d_{ij} , which is therefore fully satisfied. On the other hand, Equation (12) shows that the water request is only partially satisfied in case of scarce pressure conditions, and it depends on the pressure drop along the ij -th pipe.

Secondly, the background leakages also adopt the distributed scheme that is well suited to small and diffuse water losses. This type of losses are represented as

$$pl_{ij} = \beta_1(\overline{h_{ij}})^{\beta_2} \tag{13}$$

where β_1 and β_2 are the formula parameters representing the magnitude and the power, respectively, while $\overline{h_{ij}}$ is the average pressure of the ij -th pipe that is calculated as the difference between the hydraulic head h_{ij} and the pipe level z_{ij} . The background losses usually are settled by means of a calibration procedure involving an optimization algorithm on some parameters, e.g., pipe roughness, water demand, and losses of the hydraulic model.

Thirdly, the water bursts are implemented in the model using lumped withdrawals at the nodes with different formulations depending on the type of pipe breaks. Three types of pipe breaks are

considered: longitudinal, circumferential, or spiral cracks. According to [38–41], the universal formula of the three bursts for the i -th node reads as

$$qb_i = C_d \sqrt{2g} \left(A_{0,i} \hat{h}_i^{0.5} + m_i \hat{h}_i^{1.5} \right)^{0.5} \tag{14}$$

where C_d denotes the coefficient of discharge, g is the gravity acceleration, and \hat{h} is the pressure at the i -th node. Instead, $A_{0,i}$ and m_i define the initial leak area and the head-area slope of the pipe where the break is located (node i), respectively. The head-area slopes m_i of the three type of cracks are defined according to [41] through the following equations:

$$m_{long} = \frac{2.93157d^{0.3379} L_c^{4.80} 10^{0.5997(\log L_c)^2} \rho g}{Et^{1.746}}, \tag{15}$$

$$m_{spir} = \frac{3.7714d^{0.178569} L_c^{6.051} \sigma_l^{0.0928} 10^{1.05(\log L_c)^2} \rho g}{Et^{1.6795}}, \tag{16}$$

$$m_{circ} = \frac{1.6480210^{-5} L_c^{4.87992662} \sigma_l^{1.09182555} 10^{0.82763163(\log L_c)^2} \rho g}{Et^{0.33824224} d^{0.186376316}}, \tag{17}$$

for the longitudinal, circumferential, or spiral cracks, respectively. d is the pipe diameter, L_c is the crack length, σ_l is the longitudinal stress, ρ is the water density, E is modulus of elasticity, and t pipe thickness. The water bursts are therefore functions of the type of crack, the pipe characteristics (e.g., material), and the hydraulic pressure.

This thorough bursts implementation allows for generating realistic punctual water losses with a wide variety in type and magnitude. Indeed, a burst generator is coupled with the hydraulic DFPDM for supporting the random generation of pipes break due to the possibility of generating three types of bursts with their own characteristics: probability of failure, location, outflow, and duration. First of all, the probability of failure of an individual pipes at each time step is defined using the Poisson probability distribution [42], as follows:

$$P_{ij} = 1 - e^{-br_{ij}L_{ij}} \tag{18}$$

where L_{ij} and br_{ij} represent the length (as mentioned above) and the break rate of the ij -th pipe. The latter is defined adopting the Walski and Pelliccia formulation [43]:

$$br_{ij} = c_1 c_2 a e^{b(y-k)} \tag{19}$$

where c_1 and c_2 are two correction factors respectively depending on the pipe material and its state (presence of previous breaks), and on the pipe diameter. y is the current year and k is the installation year of the pipe, while a and b represent the regression coefficients [43] depending on the pipe material. When a burst is generated in a pipe according to Equations (18) and (19), the punctual water losses are lumped at one of the ending nodes of the selected pipe. Thereafter, the amount of water losses is chosen by randomly selecting a pair of geometric parameters L_c and A_0 able to generate an outflow ranges between a minimum and maximum flow rate thresholds. Eventually, the burst duration is fixed according to the magnitude of the flow rate loss. The realistic bursts modeling and the randomness introduced by the bursts generators allows for producing suitable time series for the training of anomaly detection algorithms.

3. Application

The scope of Section 3 is to test the performance of the SHtsG. To guide the reader into the proposed methodology, a summary of all the steps involved is hereafter described. The first step regards the water request generation. Hence, a suitable request time series is generated according to the formulation proposed in Section 2.1. This involves the modelling of both the deterministic and the random component through Equation (1). Therefore, the second step concerns the water consumption generation including the water leakages. This means that the previously generated water requests are used as input in the hydraulic simulations based on DFPDM. Specifically, the background losses in Equation (13) are defined through a calibration procedure given pressure and flow rate measurements in some points of the network. The bursts are indeed generated randomly according to the Poisson distribution in Equation (18) with a random hole shape selected among the three formulations proposed in Section 2.2. In the following two applications, the entire methodology is deeply described by practical applications.

3.1. Apulian

The first application used to test the performance of the SHtsG is the Apulian WDN, a well-known test case in literature. This WDS has been selected due to its scarce hydraulic pressure behavior, which ensures pressure driven conditions, and, due to the simple layout, which potentially emphasizes the weakness and strengths of the SHtsG. Figure 2 shows the network layout which consists of 34 pipes and 23 nodes. More details about the Apulian network can be found in [31]. In addition, the Apulian network is divided into five districts for highlighting the SHtsG capability in generating variable water requests for different parts of the network.

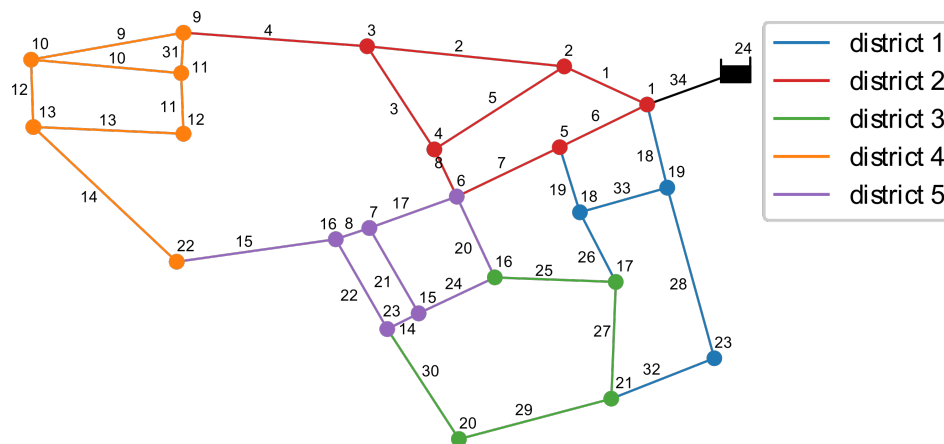


Figure 2. Layout of the Apulian network and its five districts.

In order to properly model the WDS, the total amount of water request has been reduced as in [44]. This means that the yearly average total users request results in 70 L/s. Therefore, this water request has been distributed along the pipes of the network. Then, the Apulian WDS has been simulated with a fully pressure driven approach according to Section 2.2. Nonetheless, the modeling of the stochastic behavior of the water requests in time is crucial for performing a realistic extended period simulation. Hence, the water request has been generated according to Section 2.1 by modelling the deterministic and the random components for each pipe of the network over a period of 4 years. These components for a single month are reported in Figure 3.

Starting from the top, the first plot in Figure 3 represents the dimensionless daily behaviour. To specifically model the water behaviour of each district, 5 different daily pattern time series have been generated and distributed in the corresponding district of the network (see the districts in Figure 2). Each daily district time series has been generated by randomly selecting different pattern combinations according to the main characteristics of the district water request, e.g., number of peaks (two or three),

main peak (morning, midday or evening) and occurrence time. The use of different district daily patterns helps to increase the variance of the final WDS water request, making it more realistic [45]. The second plot in Figure 3 represents the weekly patterns, and the third plot the monthly one, which is constant in the plot since only one month is displayed. The yearly pattern adopted corresponds to that of a typical seaside tourist region, characterized by a high water request during the summer (more details [44]). The fourth plot instead represents the dimensional random component of the water request (L/s), which is specific of each user aggregation (i.e., the water request is aggregated for each pipe of the WDS). Having the Apulian 34 pipe, the random components have been calculated for the distributed water request of each pipe. In this way, the random nature of the request is guaranteed together with a temporal and spatial variability of the request time series. Finally, the fifth plot shows the total water requests of the entire network in L/s.

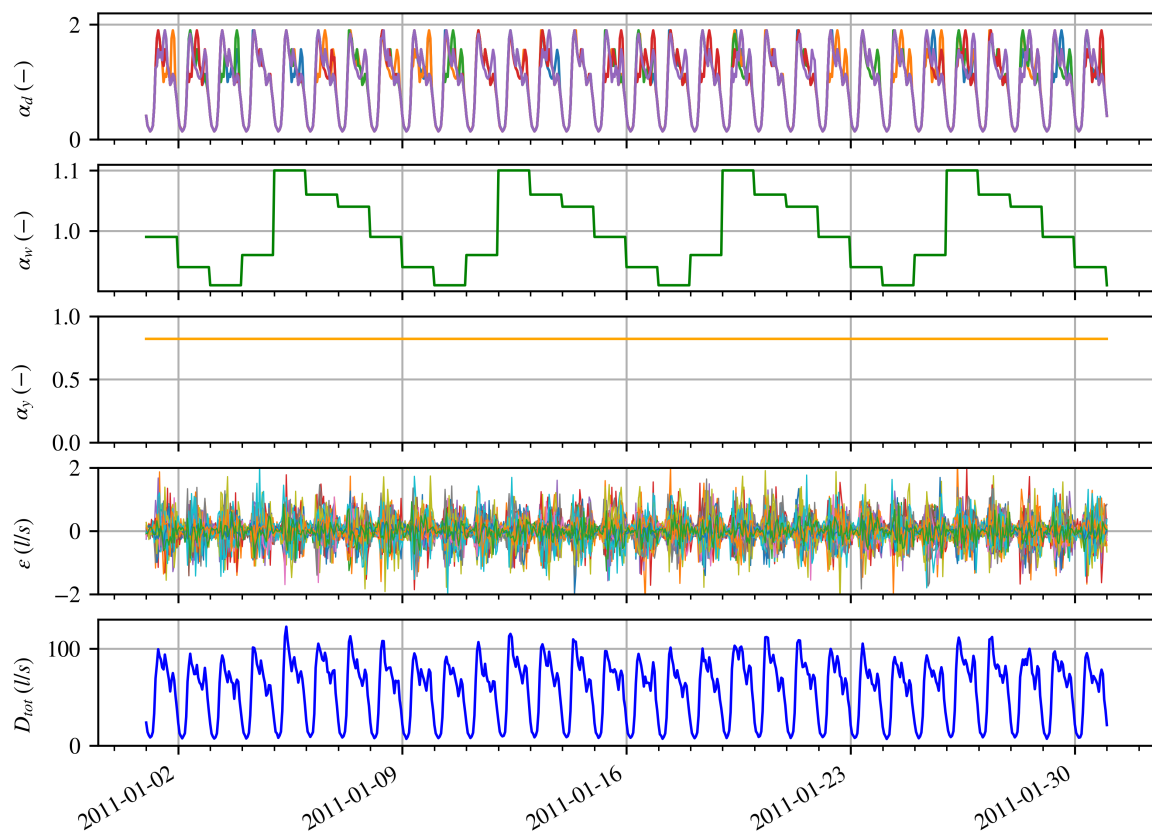


Figure 3. Water request components of the Apulian test case. Starting from the top, the first plot represents the daily pattern, the second the weekly pattern, and the third the yearly pattern. The fourth plot represents the random component and the fifth the resulting water request.

Calculated the water request in the first phase of the SHtsG, hereafter, the second phase is described. Therefore, the water request is the input of the DFPDM which, together with the bursts and the background losses, allows for producing a realistic time series of the water consumption. The total WDS consumption divided for its components is reported in Figure 4.

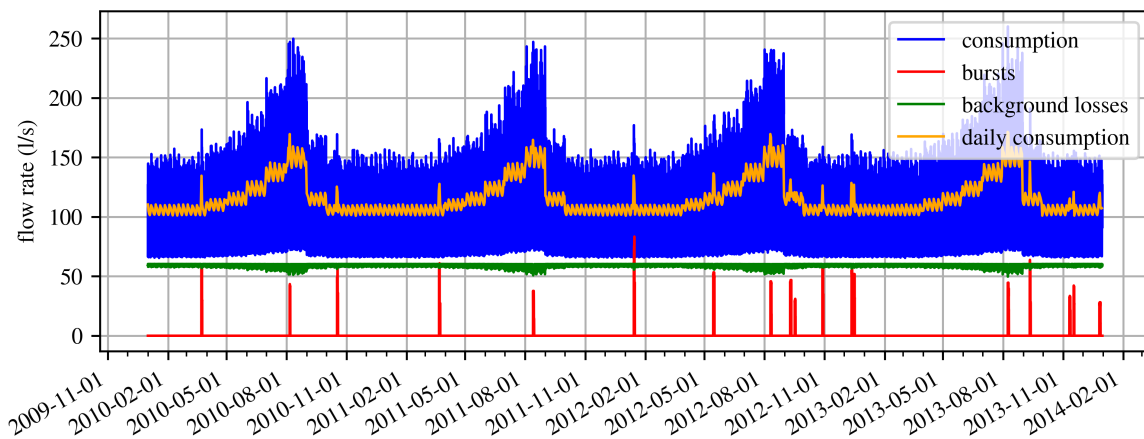
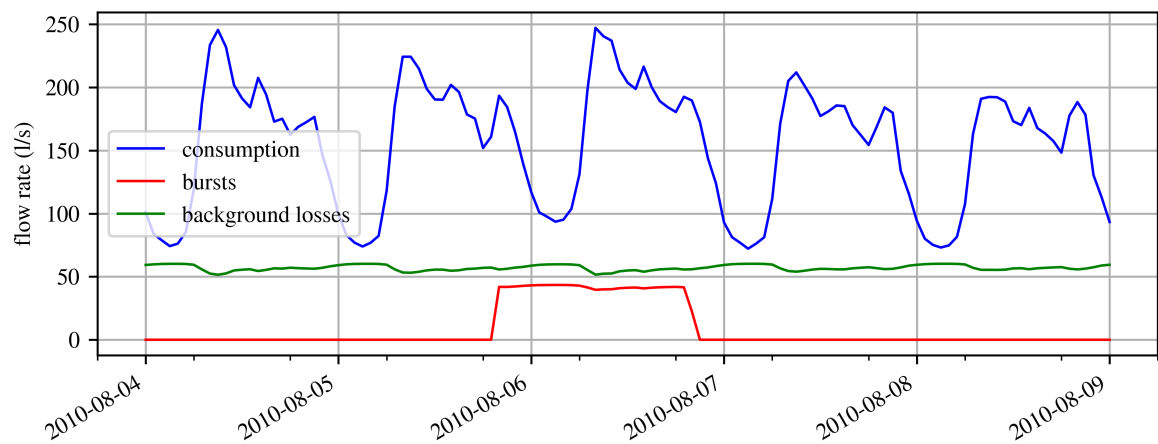


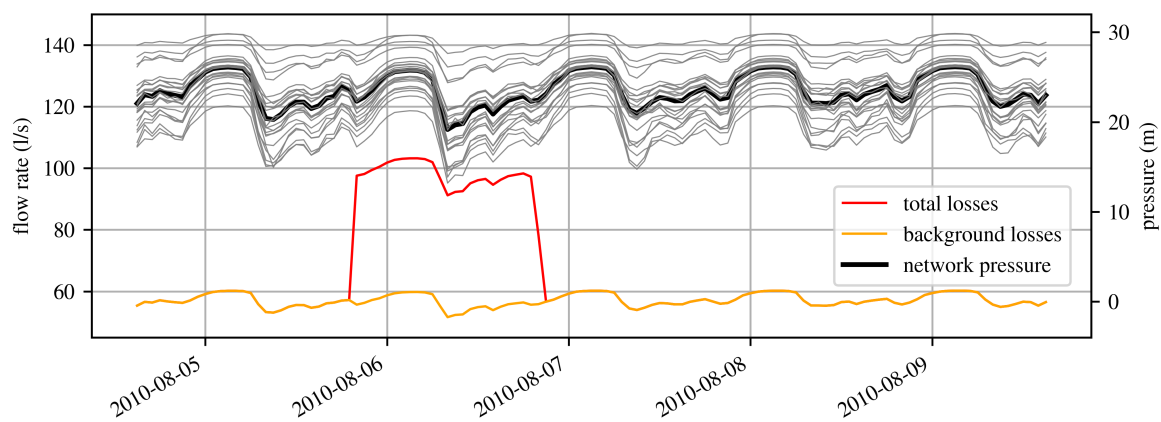
Figure 4. Total consumption of Apulian network generated for a period of four years. The plot also displays the bursts, the background losses, and the average daily consumption.

The first step of this second phase is the calibration of the background losses. In this work, it was assumed that the average background leakage corresponds to the 40% of the yearly average request. These type of losses have been uniformly spread along the network by distributing the same amount of water losses per unit length along each pipe. Due to the presence of background leakages along the entire network, it is interesting to note that such water losses follow the average pressure of the network. Once the background losses was modeled, the bursts generator has randomly introduced local pipe breaks along the networks and then the WDS has been simulated by DFPDM producing the time series of the consumption displayed in Figure 4. It is worth noting that the type of bursts generated, their characteristics, and frequency have been set according to the WSD characteristics, following Section 2.2.

To emphasize the robustness of the SHTsG by the generation of a high number of pipe failures, the average age of the pipes has been assumed to be 60 years, and also the state of all the pipes of the WDS has been classified as “pipes with one or more previous breaks” [43]. Hence, the c_1 coefficient of Equation (19) has been fixed to 7.364. The other coefficient has been fixed by assuming the value of pit cast iron pipes, which are 0.02577 and 0.0207 for a and b , respectively. Regarding the magnitude of the bursts, the geometric parameters of the crack in Equations (15)–(17) have been randomly generated to grant the range of lost flow rate between 10 L/s and 70 L/s. In addition, the repair time has also been random between 12 and 48 h. The detail of a random generated burst is given in Figure 5.



(a)



(b)

Figure 5. Detail of simulated bursts in the Apulian network. (a) shows a burst event with its consumption behavior, while (b) depicts the water losses (burst and background leaks) and the pressure of each network nodes with also the average network pressure.

Figure 5a shows the behavior of the network consumption when a burst is generated. The outlet flow rate of a burst depends on two main reasons: firstly, the characteristic of the burst itself (for details, see Section 2.2) and, secondly, the pressure where the bursts happen. This punctual pressure derives from the network capacity in terms of both pressure and flow rate. This dependency is well displayed in Figure 5b, where the burst is displayed with the available pressure at each network nodes during the time. In particular, a slight pressure drop at the nodes close to the pipe burst are present during the burst occurrence. Because the outlet flow rate fully depends on the local pressure, the behavior of the outlet flow rate follows the local pressure. These considerations are fundamental to properly model the consumption of the time series for anomaly detection purposes. In fact, the pressure dependency of a burst is mandatory to be modeled to catch the physics of this phenomenon, even more, when the network has a lack of pressure. These notes are also highlighted in Figure 6.

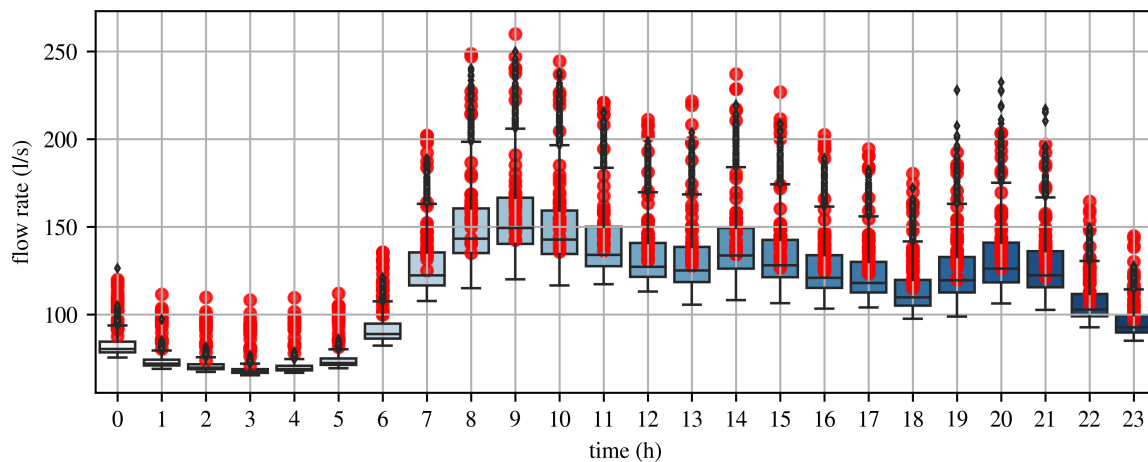


Figure 6. Box & Whiskers plot of the four-year Apulian consumption for each hour of the day. The red points represent hours where at least a burst is active, and black boxes and points stand for normal data.

Figure 6 shows the distribution of the generated consumption for each daily hour in the Apulian network. In particular, the black boxes represent the quantiles of the hourly water consumption of normal data, i.e., not affected by bursts, and the black points represent their outliers. The abnormal data are instead labeled with the red color. It is noteworthy that a lot of abnormal data lie inside the hourly boxes of the normal data during the daytime, and, vice versa, several outliers belong to normal data. This behavior has to be imputed to the scarce operating pressure of the WDS. This means that the scarce pressure condition in the network does not allow for sufficing to both burst and consumption. In conclusion, both the satisfied water demand and the outlet flow rate of a burst are dependent on the pressure; specifically, both of them decrease with a pressure drop. Thereby, a burst occurrence causes a decrease in the network pressure and consequently a reduction into the satisfied water request, resulting in a moderate changing of the final WDS flow rate distribution.

3.2. Eгна

The WDS of Eгна has been considered as the second test case to test the reliability of the proposed data generation methodology. Figure 7 reports the Eгна aqueduct, which consists of around 28 km of pipes with diameter between 50 and 150 mm and two tanks, which supply water to the WDS. In particular, the network model consists of 169 pipes and 149 junctions. More information about the network can be found in [46]. Differently from the previous test case, this WDN does not exhibit a pressure deficit, providing consequently further scenarios for evaluating the proposed methodology. In addition, this test results in being more interesting due to the availability of a small dataset of measured hourly water consumption during March and April 2019. It is worthwhile pointing out that the water demand always matches the water request due to the high operating pressure in the Eгна WDS.

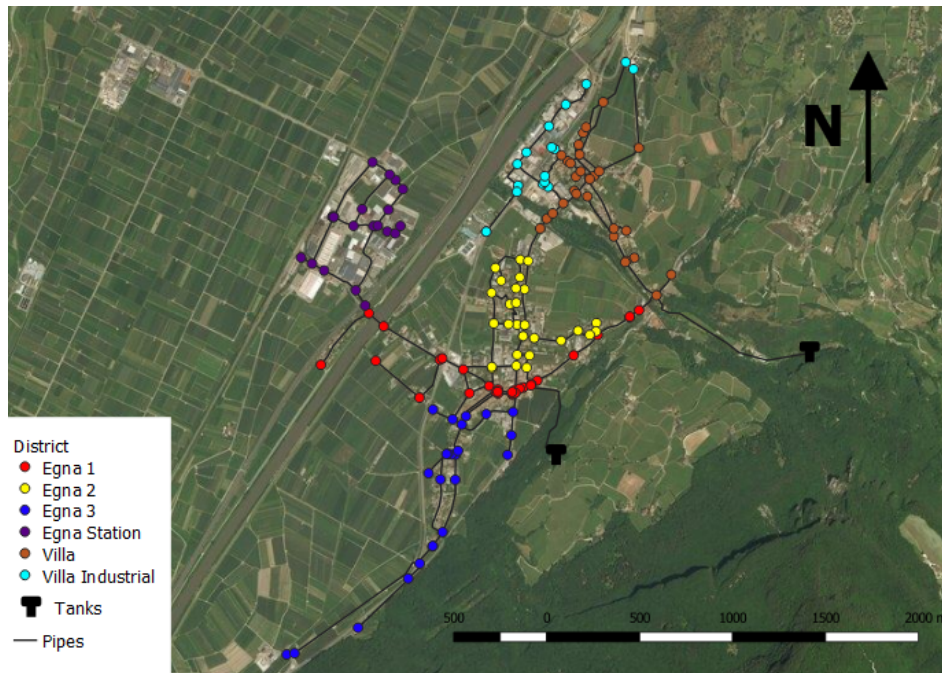


Figure 7. The water distribution system of Egna and its six districts.

As a preliminary step, to generate a proper water consumption time series, which are close to the real Egna WDS, the SHtsG has been set up with the data of March and then tested with the April data. Firstly, a data imputation has been performed adopting the Kalman filter jointly with Arima model according to [47] to fill 13 *na* values into the dataset. Once the dataset was fulfilled, the consumption time series of March has been decomposed into the different components as follows:

$$\widetilde{Wt} = \widetilde{Wd} + \widetilde{Wl} + \widetilde{Wb}, \tag{20}$$

where \widetilde{Wt} represents the monthly average water consumption, \widetilde{Wd} is the monthly average water demand, \widetilde{Wl} represents the monthly average background losses and \widetilde{Wb} the monthly average bursts.

By analysing the March consumption, it emerges that the minimum night flow (which ranges from 2:00 a.m. to 4:00 a.m.) is constant during March. This leads to the assumption that no bursts are present in this period. This follows that the \widetilde{Wb} term in Equation (20) is zero. In addition, the water demand \widetilde{Wd} of this month has been calculated given the water request of 123 $\frac{l}{\text{habitant}\cdot\text{day}}$ and the number of habitants equal to 5290. Thus, the average background losses \widetilde{Wl} has also been directly calculated through Equation (20). To decompose the signal hour per hour, the behavior of the background losses during the days has been evaluated. The relationship for undetected leaks [48] has been adopted as follows:

$$\frac{Wl_t}{\overline{Wl}} = \left(\frac{P_t}{\overline{P}} \right)^{1.5}, \tag{21}$$

where Wl_t is the background losses term at the *t*-th hour, the \overline{Wl} is the average background losses of March, the P_t is the average network pressure at the *t*-th hour, and the \overline{P} term is the average network pressure in the month of March. The pressure of the network is known because of a measurement campaign made in the same month. Hence, the behavior of the background losses was estimated through Equation (21) and the time series has been decomposed.

Differently from the March period, the time series of the consumption of April results in being not stationary during the night. This means that the term \widetilde{Wb} is different from zero during some days in April. To decompose the signal for April, the background losses have been assumed constant

in time assuming that the degradation of the network in a short period of a month is unchanged. Thus, the same values of $Q_{l,i}$ of March were also used for April. Moreover, the water demand has been estimated considering a water request per habitant for the April month of $131 \frac{l}{\text{habitant} \cdot \text{day}}$. Hence, the time series has been finally decomposed and the result reported in Figure 8.

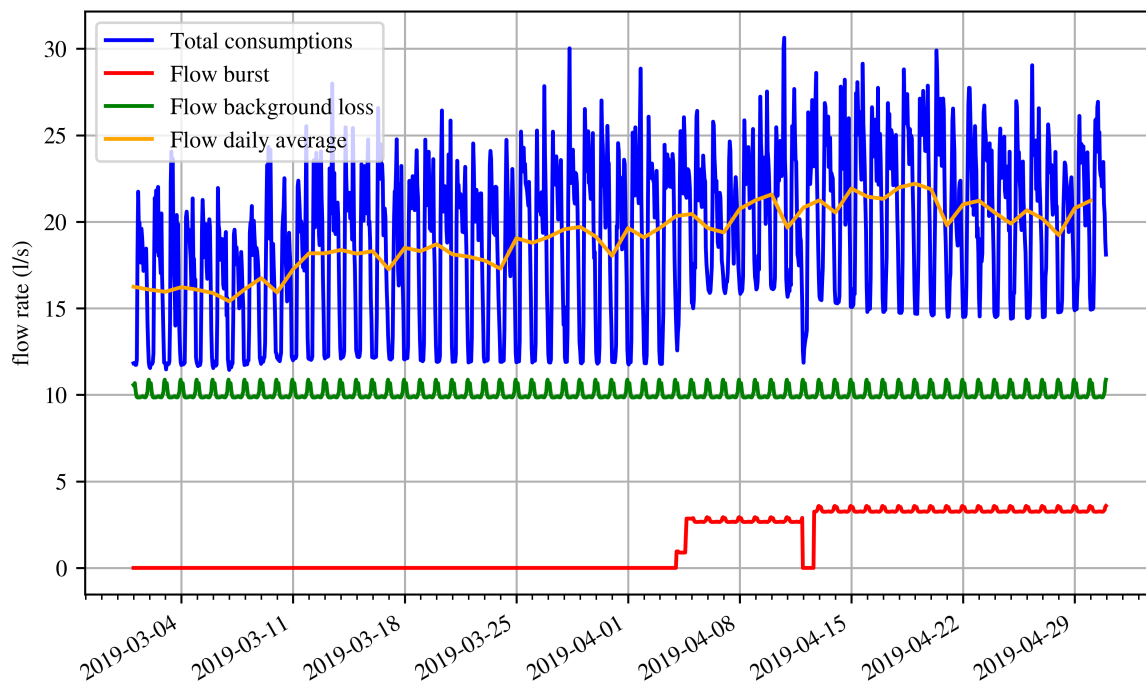


Figure 8. The decomposition of the metered flow rate of March and April for the Egna WDS.

Starting from the decomposed time series, the DFPDM has been modeled to properly reproduce the consumption time series of March. The first phase consists of calibrating the water request terms of Equation (1). As previously declared, the calibration of the model has been performed over the March data. An analysis of the March daily pattern has been carried out. Firstly, the structure of the daily pattern has been studied. In particular, the working days, which ranges from Monday to Friday, always show a 3 peaks behavior, while the Saturdays have three peaks in 60% of the cases and two peaks in the others. Differently from the previous, the Sundays always show two peaks. Secondly, an analysis of the peaks level and their occurrence has been performed. It was found that the main peak happens in the morning in 80% of the cases for the working days, and the remaining 20% happens in the evening. In particular, when the main peak is in the morning, it happens in 75% of the cases at 7.00 a.m. and in 25% of the cases at 8.00 a.m. Concerning the midday peaks occurrence, in 65% of the cases, the peak happens at 12.00 a.m., and for the remaining 35% at 1.00 p.m. Lastly, the evening peak occurs at 5.00 p.m. in 20% of the cases, at 6.00 p.m. in 30% of the cases and in the remaining 50% at 7.00 p.m. Regarding the Saturdays with a three-peaks structure, the morning, midday and evening peaks occurs at 9.00 a.m., 12.00 a.m., and 6.00 p.m., respectively. Instead, for the Sundays and the Saturdays with a two-peak structure, the first peak happen always at 9.00 a.m., and the second peak is at 5.00 p.m. or 6.00 p.m. in 50% of the cases.

Therefore, the aforementioned probabilities have been used to generate the daily pattern coherent with the real data analyzed. This generation has been performed for each district of the network shown in Figure 7. Moreover, the generation of the water requests has been made as described in Section 2.1, omitting the weekly and yearly components. Regarding the random component, it has been calculated for each district of Egna with its different user aggregation. Figure 9 depicts the resulting water request.

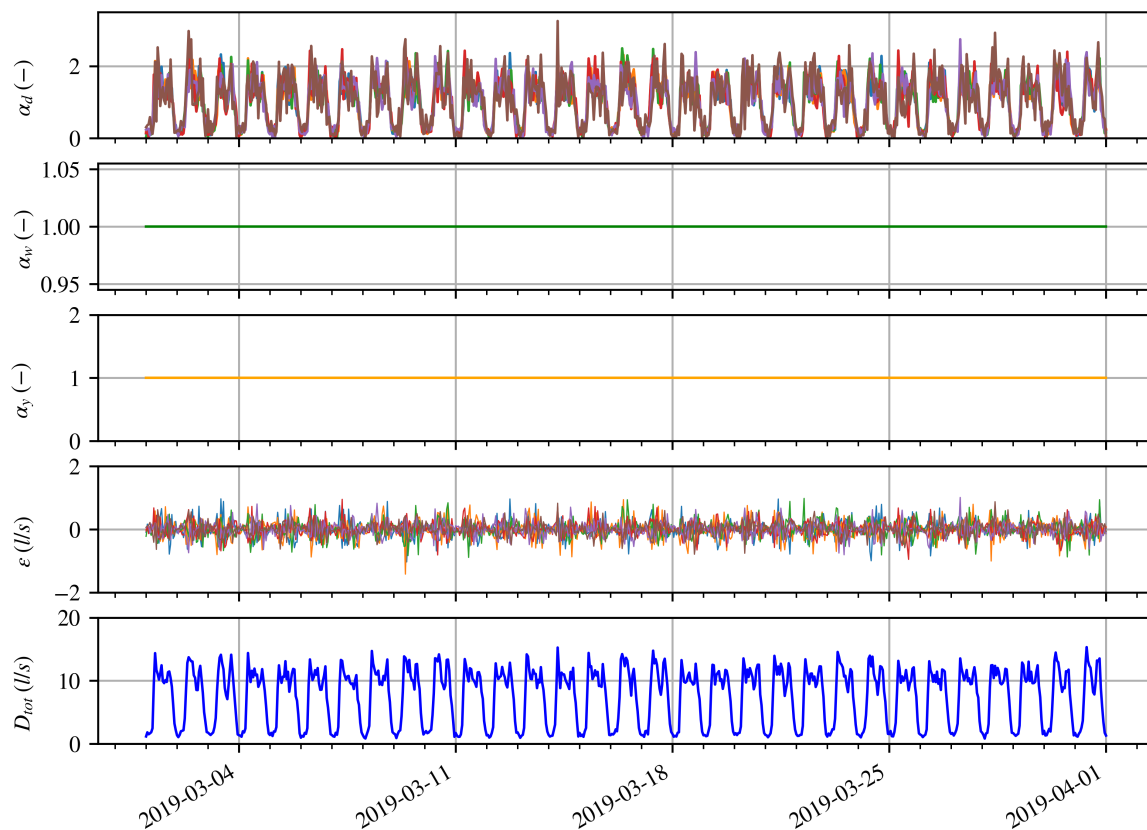


Figure 9. Water request components for the Egna test case. Starting from the top, the first represents the daily pattern, the second the weekly pattern, and the third the yearly one. The fourth plot represents the random components and the fifth the resulting water request.

Then, the request previously generated has been used as input into the DFPDM with both the bursts and background losses. Figure 10 shows the generated consumption for March month with no bursts and background losses coherent with the real data.

The model has been then used to generate the consumption also for the April month. In particular, 10 different generations of the consumption of April have been made to show the reliability of the methodology. Regarding the burst parameters, the geometric parameters A_0 and L_c of the cracks in Equations (15)–(17) have been bounded to generate values of outflow between 1.5 L/s and 10 L/s. This decision has been made due to characteristics of the Egna WDS. Regarding the duration, different intervals have been selected depending on the burst magnitude. In addition, the repair time ranges between 3 and 21 days in case of bursts outflow from 1.5 L/s to 4 L/s, while the bigger than 4 L/s leaks last from 12 to 72 h. These long repair times are due to the poor operational management of the small municipality analyzed. Moreover, the age of the pipes is known and the c_1 coefficient has been fixed to 7.364. The a and b coefficients are set to 0.02577 and 0.0207, respectively.

Figure 11 displays the 10 April generations under the Box & Whiskers plot for each hour of the day.

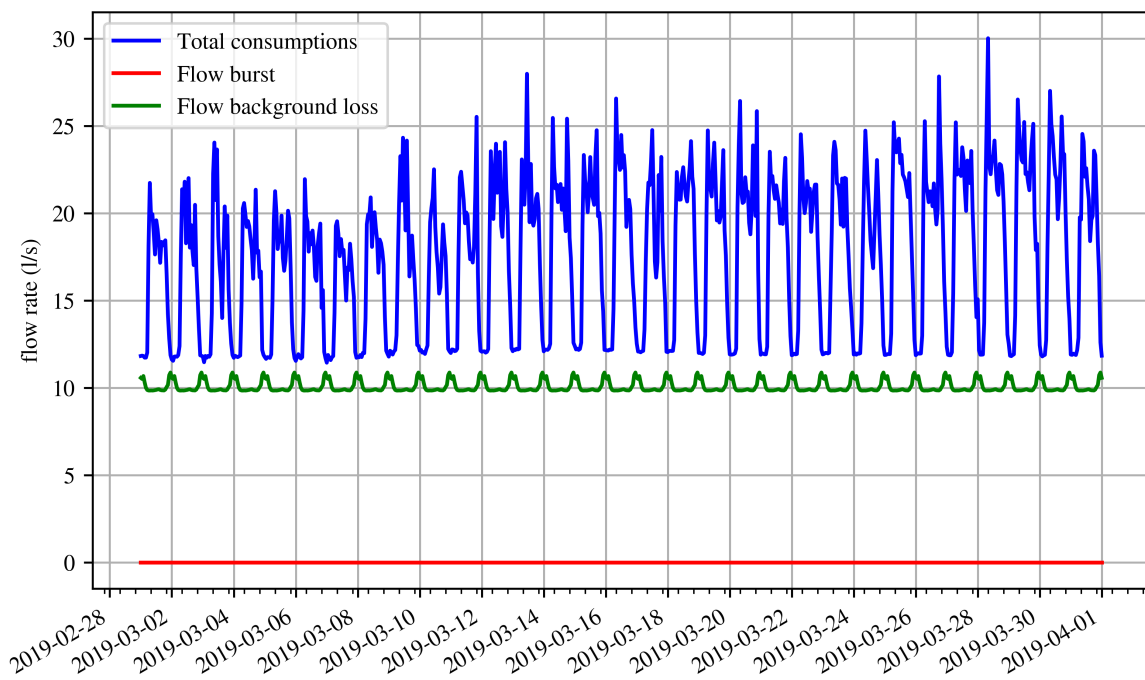


Figure 10. Total generated consumption of the Egna network for the March month, including the bursts and the calibrated background losses.

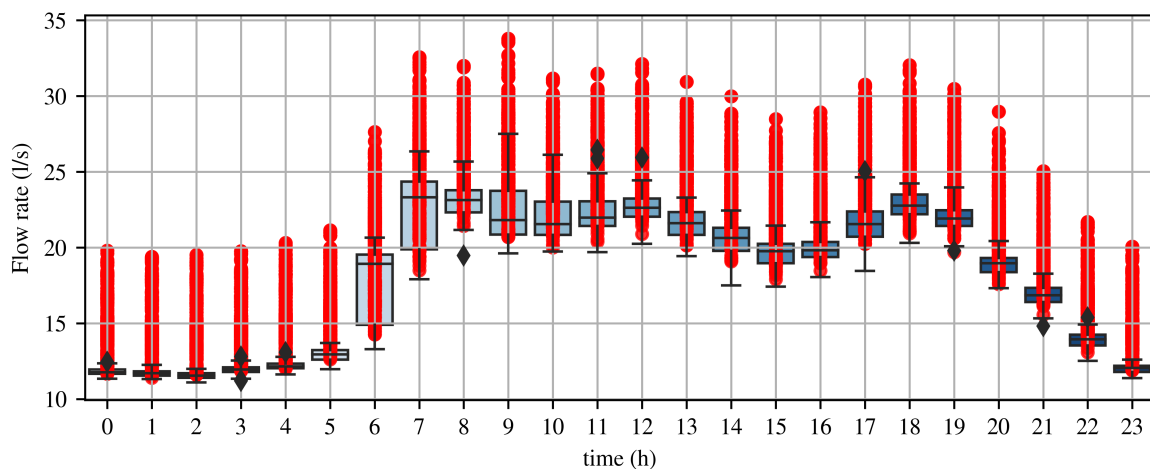


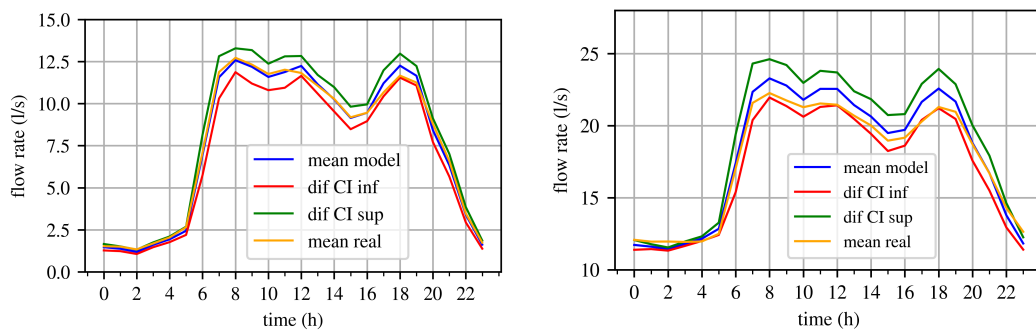
Figure 11. Box & Whiskers plot for each hour of the day of the 10 generated samples of April consumption of the Egna WDS.

It shows a few outliers regarding the normal data, while the abnormal data, indicated with the red color, appears to be more diffused in the outlier zone. This behavior is different from the Apulian test case (see Figure 6). In fact, the Egna network does not suffer from the pressure deficit condition. This means that the available pressure is always enough to suffice at both users' water request and bursts. These observations underline even more the crucial role played by the proper hydraulic solver adopted in the simulations. In fact, the generation of a time series of consumption for burst detection can not be made regardless of the pressure behavior of the WDS.

To show the robustness of the proposed methodology, the original time series and the data generated by SHtsG have been tested by means of the *t*-student test. Figure 12 and Table 1 display the corresponding results. Thus, the reliability of the consumption time series generated by SHtsG has been tested for both March and April, which are the months used for the model training and

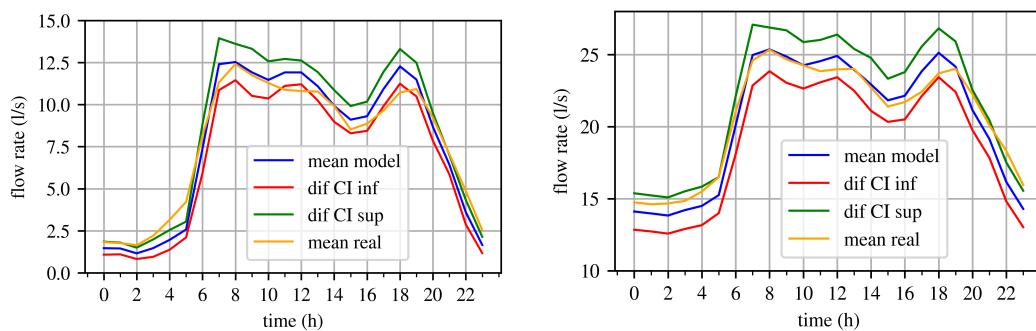
validation, respectively. The *t*-student test enables to define if there is a consistent difference between the original and the generated data at each hour of the day. In particular, the null hypothesis states that the mean of the real data are equal to the mean of the generated by SHtsG data. The results of the test are displayed through the bounds of Figure 12, which are defined according to a significance level of 1%, allow to easily identify the rejected hours, i.e., real values out of the limit lines.

Firstly, Figure 12a,b report the results of March data adopted in the training of the SHtsG. In this month, the null hypothesis is never rejected for the water request, while only 5 times the null hypothesis is rejected for the consumption. This means that the generated time series are able to well resemble the real data with 1% significance level. Secondly, Figure 12c,d show the results of the *t*-student test of April data. It is worth noting that the statistical test has been made between 1 month of real data and the 10 sample of the generated April data. The performance of 10 samples representing April data has been adopted in order to have a more robust evaluation of the SHtsG. In case of a single sample, the random part of the bursts modelling excessively affects the variability of the results even if the procedure has been properly set up. The test gives positive results even in the severe case of the validation dataset of April. Only in a few hours, the two dataset are significantly different from each other. It mainly happens during the night, where the difference between March and April is more significant. Table 1 lists the *p*-values of the *t*-student test applied on the 4 cases described above.



(a) T-student test applied to the generated and real request data of Egna in March.

(b) T-student test applied to the generated and real consumption data of Egna in March.



(c) T-student test applied to the generated and real request data of Egna in April.

(d) T-student test applied to the generated and real consumption data of Egna in April.

Figure 12. T-student test between generated and real data of March and April for both water request and consumption.

Table 1. *P*-values obtained by the *t*-student test of the null hypothesis such that the mean of the real data are equal to the mean of the generated by SHtsG data for each hour of the day. Specifically, real data and synthetic data are compared both for the water request and consumption in March (training) and April (validation), separately.

Hour	March Request	March Consumption	April Request	April Consumption
0:00	0.538	0.007	0.016	0.177
1:00	0.093	3.42×10^{-6}	0.019	0.172
2:00	0.038	4.20×10^{-14}	4.9×10^{-4}	0.078
3:00	0.269	0.033	0.001	0.188
4:00	0.197	0.002	6.0×10^{-6}	0.050
5:00	0.099	0.032	3.1×10^{-10}	0.010
6:00	0.813	0.634	0.156	0.208
7:00	0.626	0.290	0.063	0.635
8:00	0.688	0.043	0.763	0.959
9:00	0.793	0.055	0.740	0.766
10:00	0.665	0.238	0.646	0.995
11:00	0.771	0.033	0.001	0.201
12:00	0.175	0.012	1.9×10^{-4}	0.096
13:00	0.757	0.044	0.311	0.906
14:00	0.940	0.165	0.970	0.755
15:00	0.909	0.254	0.059	0.433
16:00	0.990	0.181	0.172	0.466
17:00	0.181	0.005	0.002	0.031
18:00	0.113	0.015	3.0×10^{-4}	0.028
19:00	0.181	0.116	0.134	0.821
20:00	0.309	0.846	0.076	0.041
21:00	0.477	0.972	0.005	0.052
22:00	0.759	0.111	5.3×10^{-5}	1.52×10^{-4}
23:00	0.353	1.11×10^{-5}	2.9×10^{-5}	0.001

4. Conclusions

This study proposes a methodology to generate water consumption time series affected by bursts for supporting the development of anomaly detection models in WDS. The SHtsG deals separately with water request and water consumption generation. The first consists of modeling the water requests through a superimposition approach, which considers both the different deterministic trends (e.g., seasonal, weekly and daily pattern) and the random components of the variance. The second phase consists of generating consumption with accurate WDS modeling in which background losses and bursts are properly introduced. The two phases methodology allows for providing a synthetic hydraulic dataset close to the real data in terms of both mean and variance. Moreover, the generated time series are not affected by uncertainty and missing data due to measurements and transmission system. These synthetic data concern normal and abnormal data with precise leaking labels including complete information about the bursts.

Two different applications are presented to validate and highlight the advantages of the SHtsG. The first consists of a WDS widely used in literature that is called the Apulian network. This test case aims to provide a detailed step by step explanation of the methodology applied to a WDS operating in scarce pressure conditions. The reliability of the SHtsG is underlined by the resulting time series of the WDS inlet flow rate which represents four years of the total consumption. Moreover, the variance of the consumption is shown by means of the random introduction of the different types of bursts properly labelled. It is noteworthy to mention how complex it is to distinguish normal data from abnormal data, i.e., affected by bursts, due to the pressure deficit condition of the analyzed WDS. Therefore, the use of proper pressure-driven hydraulic solver, which is able to properly simulate the different water request components (e.g., demand, background leakages, and bursts), results in being crucial for providing a final suitable dataset.

The second presents the Egna network which consists of a small mountain WDS with elevated operating pressures. This test case helps to understand how the SHtsG can be used to enlarge a real dataset maintaining the characteristics of the original time series. The accuracy of the methodology in reproducing real data are proven by a *t*-student test. Moreover, the elevated network pressures enhanced the different behavior of the burst, which can be more easily distinguished. The SHtsG depicts that it is possible to properly reproduce the variability of the real data and to generate a time series coherent with the real WDS behavior. The importance of the accurate hydraulic solver adopted is also highlighted.

To conclude, the presented methodology has shown promising results to generate suitable time series extrapolated from extended period simulations that include both hydraulic and mechanical WDS reliability. The ability of SHtsG to produce massive hydraulic data simply by selecting a flow rate time series of crucial WDS points is noteworthy, e.g., tanks outflow, district inlet, or districts representative pressure. Future efforts will involve the direct application of this methodology in the development of anomaly detection algorithms.

Author Contributions: Conceptualization, A.M., A.Z., M.F., M.R. and R.G.; Data curation, A.M., A.Z. and M.F.; Formal analysis, A.M., A.Z., M.F., M.R. and R.G.; Funding acquisition, M.R.; Investigation, A.M., A.Z. and M.F.; Methodology, A.M., A.Z., M.F. and R.G.; Project administration, M.R. and R.G.; Software, A.M., A.Z. and M.F.; Supervision, M.R. and R.G.; Validation, A.M., A.Z. and M.F.; Visualization, A.M., A.Z. and M.F.; Writing—original draft, A.M., A.Z. and M.F.; Writing—review & editing, A.M., A.Z., M.F., D.A., M.R. and R.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. Lee, S.W.; Sarp, S.; Jeon, D.J.; Kim, J.H. Smart water grid: The future water management platform. *Desalin. Water Treat.* **2015**, *55*, 339–346. [\[CrossRef\]](#)
2. Ramos, H.M.; McNabola, A.; López-Jiménez, P.A.; Pérez-Sánchez, M. Smart water management towards future water sustainable networks. *Water* **2020**, *12*, 58. [\[CrossRef\]](#)
3. Vörösmarty, C.J.; Green, P.; Salisbury, J.; Lammers, R.B. Global water resources: vulnerability from climate change and population growth. *Science* **2000**, *289*, 284–288. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Mutchek, M.; Williams, E. Moving towards sustainable and resilient smart water grids. *Challenges* **2014**, *5*, 123–137. [\[CrossRef\]](#)
5. Ramos, H.M.; Carravetta, A.; Nabola, A.M. New Challenges in Water Systems. *Water* **2020**, *12*, 2340. [\[CrossRef\]](#)
6. Zaman, D.; Tiwari, M.; Gupta, A.; Sen, D. A review of leakage detection strategies for pressurised pipeline in steady-state. *Eng. Fail. Anal.* **2020**, *109*. [\[CrossRef\]](#)
7. Puust, R.; Kapelan, Z.; Savic, D.A.; Koppel, T. A review of methods for leakage management in pipe networks. *Urban Water J.* **2010**, *7*, 25–45. [\[CrossRef\]](#)
8. Misiunas, D.; Lambert, M.; Simpson, A.; Olsson, G. Burst detection and location in water distribution networks. *Water Sci. Technol. Water Supply* **2005**, *5*, 71–80. [\[CrossRef\]](#)
9. Tan, P.N.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*, 1st ed.; Pearson Education India: Delhi, India, 2005.
10. Wu, Y.; Liu, S. A review of data-driven approaches for burst detection in water distribution systems. *Urban Water J.* **2017**, *14*, 972–983. [\[CrossRef\]](#)
11. Mounce, S.R.; Machell, J. Burst detection using hydraulic data from water distribution systems with artificial neural networks. *Urban Water J.* **2006**, *3*, 21–31. [\[CrossRef\]](#)
12. Aksela, K.; Aksela, M.; Vahala, R. Leakage detection in a real distribution network using a SOM. *Urban Water J.* **2009**, *6*, 279–289. [\[CrossRef\]](#)
13. Ye, G.; Fenner, R.A. Kalman filtering of hydraulic measurements for burst detection in water distribution systems. *J. Pipeline Syst. Eng. Pract.* **2011**, *2*, 14–22. [\[CrossRef\]](#)

14. Akkaya, A.; Talu, M. Extended kalman filter based IMU sensor fusion application for leakage position detection in water pipelines. *J. Fac. Eng. Archit. Gazi Univ.* **2017**, *32*, 1393–1404.
15. Mounce, S.R.; Mounce, R.B.; Boxall, J.B. Novelty detection for time series data analysis in water distribution systems using support vector machines. *J. Hydroinform.* **2011**, *13*, 672–686. [[CrossRef](#)]
16. Zhang, Q.; Wu, Z.; Zhao, M.; Qi, J.; Huang, Y.; Zhao, H. Leakage zone identification in large-scale water distribution systems using multiclass support vector machines. *J. Water Resour. Plan. Manag.* **2016**, *142*. [[CrossRef](#)]
17. Arsene, C.; Gabrys, B.; Al-Dabass, D. Decision support system for water distribution systems based on neural networks and graphs theory for leakage detection. *Expert Syst. Appl.* **2012**, *39*, 13214–13224. [[CrossRef](#)]
18. Fang, Q.; Zhang, J.; Xie, C.; Yang, Y. Detection of multiple leakage points in water distribution networks based on convolutional neural networks. *Water Sci. Technol. Water Supply* **2019**, *19*, 2231–2239. [[CrossRef](#)]
19. Loureiro, D.; Amado, C.; Martins, A.; Vitorino, D.; Mamade, A.; Coelho, S.T. Water distribution systems flow monitoring and anomalous event detection: A practical approach. *Urban Water J.* **2016**, *13*, 242–252. [[CrossRef](#)]
20. Jung, D.; Kang, D.; Liu, J.; Lansley, K. Improving the rapidity of responses to pipe burst in water distribution systems: A comparison of statistical process control methods. *J. Hydroinform.* **2015**, *17*, 307–328. [[CrossRef](#)]
21. Chan, T.K.; Chin, C.S.; Zhong, X. Review of Current Technologies and Proposed Intelligent Methodologies for Water Distributed Network Leakage Detection. *IEEE Access* **2018**, *6*, 78846–78867. [[CrossRef](#)]
22. Romano, M.; Kapelan, Z.; Savić, D.A. Automated Detection of Pipe Bursts and Other Events in Water Distribution Systems. *J. Water Resour. Plan. Manag.* **2014**, *140*, 457–467. [[CrossRef](#)]
23. Kang, J.; Park, Y.J.; Lee, J.; Wang, S.H.; Eom, D.S. Novel Leakage Detection by Ensemble CNN-SVM and Graph-Based Localization in Water Distribution Systems. *IEEE Trans. Ind. Electron.* **2018**, *65*, 4279–4289. [[CrossRef](#)]
24. Bakker, M.; Trietsch, E.A.; Vreeburg, J.H.G.; Rietveld, L.C. Analysis of historic bursts and burst detection in water supply areas of different size. *Water Supply* **2014**, *14*, 1035–1044. [[CrossRef](#)]
25. Eliades, D.G.; Polycarpou, M.M. Leakage fault detection in district metered areas of water distribution systems. *J. Hydroinform.* **2012**, *14*, 992–1005. [[CrossRef](#)]
26. Jung, D.; Lansley, K. Water Distribution System Burst Detection Using a Nonlinear Kalman Filter. *J. Water Resour. Plan. Manag.* **2015**, *141*, 04014070. [[CrossRef](#)]
27. Gargano, R.; Pianese, D. Influence of hydraulic and mechanical reliability on the overall reliability of water networks. In Proceedings of the 26th Convegno di Idraulica e Costruzioni Idrauliche, Catania, Italy, 9–12 September 1998.
28. Paez, D.; Fillion, Y. Water Distribution Systems Reliability under Extended-Period Simulations. *J. Water Resour. Plan. Manag.* **2020**, *146*, 04020062. [[CrossRef](#)]
29. Gargano, R.; Tricarico, C.; del Giudice, G.; Granata, F. A stochastic model for daily residential water demand. *Water Supply* **2016**, *16*, 1753–1767. [[CrossRef](#)]
30. Gargano, R.; Tricarico, C.; Granata, F.; Santopietro, S.; de Marinis, G. Probabilistic models for the peak residential water demand. *Water* **2017**, *9*, 417. [[CrossRef](#)]
31. Giustolisi, O.; Todini, E. On the approximation of distributed demands as nodal demands in WDN analysis. In Proceedings of the XXXI National Hydraulics and Hydraulic Construction Conference, Perugia, Italy, 9–12 September 2008; pp. 9–12.
32. Menapace, A.; Avesani, D.; Righetti, M.; Bellin, A.; Pisaturo, G. Uniformly Distributed Demand EPANET Extension. *Water Resour. Manag.* **2018**, *32*, 2165–2180. [[CrossRef](#)]
33. Siew, C.; Tanyimboh, T.T. Pressure-Dependent EPANET Extension. *Water Resour. Manag.* **2012**, *26*, 1477–1498. [[CrossRef](#)]
34. Menapace, A.; Avesani, D. Global Gradient Algorithm extension to distributed pressure driven pipe demand model. *Water Resour. Manag.* **2019**, *33*, 1717–1736. [[CrossRef](#)]
35. Todini, E.; Pilati, S. A gradient method for the solution of looped pipe networks. *Comput. Appl. Water Supply* **1988**, *1*, 1–20.
36. Menapace, A.; Righetti, M.; Avesani, D. Application of distributed pressure driven modeling in water supply system. In Proceedings of the 1st International WDSA/CCWI Joint Conference, Kingston, ON, Canada, 23–25 July 2018; pp. 1–8.

37. Zanfei, A.; Menapace, A.; Santopietro, S.; Righetti, M. Calibration Procedure for Water Distribution Systems: Comparison among Hydraulic Models. *Water* **2020**, *12*, 1421. [[CrossRef](#)]
38. Cassa, A.M.; van Zyl, J.E.; Laubscher, R.F. A numerical investigation into the effect of pressure on holes and cracks in water supply pipes. *Urban Water J.* **2010**, *7*, 109–120. [[CrossRef](#)]
39. Cassa, A.M.; van Zyl, J.E. Predicting the head-leakage slope of cracks in pipes subject to elastic deformations. *J. Water Supply Res. Technol. Aqua* **2013**, *62*, 214–223. [[CrossRef](#)]
40. Cassa, A.; van Zyl, J. Predicting the Leakage Exponents of Elastically Deforming Cracks in Pipes. *Procedia Eng.* **2014**, *70*, 302–310. [[CrossRef](#)]
41. van Zyl, J.E.; Cassa, A.M. Modeling Elastically Deforming Leaks in Water Distribution Pipes. *J. Hydraul. Eng.* **2014**, *140*, 182–189. [[CrossRef](#)]
42. Su, Y.C.; Mays, L.W.; Duan, N.; Lansey, K.E. Reliability-based optimization model for water distribution systems. *J. Hydraul. Eng.* **1987**, *113*, 1539–1556. [[CrossRef](#)]
43. Walski, T.M.; Pelliccia, A. Economic analysis of water main breaks. *J. Am. Water Work. Assoc.* **1982**, *74*, 140–147. [[CrossRef](#)]
44. Mazzolani, G.; Berardi, L.; Laucelli, D.; Simone, A.; Martino, R.; Giustolisi, O. Estimating leakages in water distribution networks based only on inlet flow data. *J. Water Resour. Plan. Manag.* **2017**, *143*, 04017014. [[CrossRef](#)]
45. Di Nardo, A.; Di Natale, M.; Gargano, R.; Giudicianni, C.; Greco, R.; Santonastaso, G.F. Performance of partitioned water distribution networks under spatial-temporal variability of water demand. *Environ. Model. Softw.* **2018**, *101*, 128–136. [[CrossRef](#)]
46. Zanfei, A.; Menapace, A.; Pisaturo, G.R.; Righetti, M. Calibration of Water Leakages and Valve Setting in a Real Water Supply System. In *Environmental Sciences Proceedings*; Multidisciplinary Digital Publishing Institute: Basel, Switzerland, 2020; Volume 2, p. 41.
47. Moritz, S.; Bartz-Beielstein, T. imputeTS: Time Series Missing Value Imputation in R. *R J.* **2017**, *9*, 207. [[CrossRef](#)]
48. Lambert, A. What do we know about pressure-leakage relationships in distribution systems. In *Proceedings of the IWA Systems Approach to Leakage Control and Water Distribution System Management*, Brno, Czech Republic, 16–18 May 2001.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).