Water Research 280 (2025) 123442

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Flow time series decomposition to identify non-revenue water components in drinking water distribution systems: A data-driven approach

Maria Almeida Silva [a,b,c,*], Conceição Amado [c], Dália Loureiro [d]

[a] Lusófona University, Portugal
[b] Center of Technology and Systems (UNINOVA-CTS) and Associated Lab of Intelligent Systems (LASI), Portugal
[c] CEMAT, Instituto Superior Técnico, University of Lisbon, Portugal
[d] National Laboratory for Civil Engineering, Portugal

ARTICLE INFO

Keywords:
Drinking water distribution systems
Flow time series decomposition
Hidden components
Kernel regression smoother
Singular Spectrum Analysis (SSA)

ABSTRACT

Water utilities face challenges in managing non-revenue water, which encompasses unbilled authorised consumption, leaks, bursts, authorised consumption errors, and unauthorised consumption. Several approaches have been developed to address these issues. Most existing methods focus on estimating individual components of non-revenue water, rather than considering all aspects comprehensively. The installation of smart water meters has significantly reduced unmetered billed consumption, addressing issues related to the absence of water meters in some customer locations or difficulties in systematic meter reading. Water utilities can obtain a comprehensive view of non-revenue water over time by combining the billed metered consumption time series obtained with smart meters with the network flow time series. Partitioning the non-revenue water time series into several components, each representing a different pattern in the data, can help one better grasp the underlying patterns. In this paper, time series decomposition techniques reveal hidden non-revenue water components, allowing the water utilities to create a network strategy to reduce water losses. Several decomposition methods were applied, and the best reliable results were achieved with Singular Spectrum Analysis.

1. Introduction

Efficient management of freshwater resources is a significant topic in the 2030 Agenda for Sustainable Development (Casale and Cordeiro Ortigara, 2019), especially in a water scarcity situation.

The water balance, published by the International Water Association (IWA) (Alegre et al., 2006) is a technique widely used by water utilities to quantify the system's input volume, authorised consumption, and water losses (Table 1). The water balance is usually computed with a periodicity of one year, due to the difficulty in computing and estimating all the components with an higher frequency (*e.g.*, monthly, daily). Water utilities use this approach for quantifying non-revenue water in their distribution systems. The non-revenue water represents a significant percentage of the system input volume in drinking water systems and substantially impacts the economic sustainability of water utilities. In the Portuguese water utilities, the non-revenue water reached 27.1% of the system input volume in 2022 (ERSAR, 2023). Nevertheless, for some water utilities, the annual water balance is the only way to estimate non-revenue water and its components (unbilled authorised consumption, apparent losses, and real losses).

The unbilled authorised consumption is divided into metered and unmetered and is usually due to water uses for network operation and maintenance, street cleaning, and garden watering. This component of non-revenue water is relatively easy to control by the water utilities since it is possible to estimate or measure the water use points. Apparent losses encompass both unauthorised consumption (e.g., due to illegal water uses or illegal connections to the public network) and authorised consumption errors (related to the systematic error in water meters' measurements due to their age). Real losses include all the leaks and pipe bursts occurring in the network up to the customers' measurement point. In general, real losses constitute the largest portion of non-revenue water and can have a significant impact on several dimensions, including service interruptions caused by pipe bursts and energy efficiency. Moreover, real losses are the most challenging component to estimate alongside the unauthorised consumption (Liemberger and Wyatt, 2019).

More and better tools to analyse non-revenue water and estimate its components are always welcomed by water utilities. Several literature methods aim to detect and account for the water volume lost due to

* Corresponding author at: Lusófona University, Portugal.
E-mail address: maria.almeida.silva@ulusofona.pt (M.A. Silva).

https://doi.org/10.1016/j.watres.2025.123442
Received 25 November 2024; Received in revised form 15 February 2025; Accepted 4 March 2025
Available online 15 March 2025

Table 1
Water balance defined according to IWA (Alegre et al., 2006).

System input volume	Authorised consumption	Billed authorised consumption	Billed metered consumption	Revenue water
			Billed unmetered consumption	
		Unbilled authorised consumption	Unbilled metered consumption	Non-revenue water
			Unbilled unmetered consumption	
	Water losses	Apparent losses	Unauthorised consumption	
			Authorised consumption errors	
		Real losses	Leakage and overflows at transmission and/or distribution storage tanks	
			Leakage on service connections up to the measurement point	
			Real losses on raw water mains and at the treatment works	
			Leakage on transmission and/or distribution mains	

leaks and bursts and authorised consumption errors (Alkasseh et al., 2013; Amoatey et al., 2014; Fu et al., 2022; McKenzie and Seago, 2005; Mutikanga et al., 2011; Palau et al., 2004, 2012; AL-Washali et al., 2016). However, to our knowledge, the literature does not commonly provide methods for estimating unauthorised consumption. Most methods focus on estimating only one component of non-revenue water, leaving the other components unaddressed. As a result, these components are typically estimated separately.

A review of the current methods for managing any type of water loss in a water distribution system is presented in Mutikanga et al. (2013). This paper reviews and presents tools and methodologies for managing leakages, apparent losses, and real losses.

In AL-Washali et al. (2016), three methods to assess real losses in a water supply system are compared: Minimum Night Flow Analysis (MNFA), Burst And Background Estimates (BABE), and the top-down water balance. The advantages and disadvantages of each one are presented. MNFA allows for evaluating the measurement accuracy and estimating the leakages. However, it requires intensive fieldwork and pressure information, and the legitimate nighttime flow should be estimated. MNFA was also used in Alkasseh et al. (2013) to understand which factors contribute most to the minimum night flow. On the other hand, the BABE methodology divides the losses into background losses (also known as leakages) and bursts based on the flow rate and duration. Due to the high number of assumptions, this method should only be applied when another option is not possible to use. McKenzie and Seago (2005) reviewed this methodology and its evolution over the years, aiming to assess real losses. In Amoatey et al. (2014), water losses due to leakages and bursts were investigated using the MNFA and BABE and applied to Ghana's urban water supply network with losses of around 40%. The top-down water balance does not depend either on the pressure or on extensive fieldwork. However, it requires an unauthorised consumption estimate, which is still a limitation for water utilities. Moreover, this method does not allow the estimation of leaks and bursts separately.

The method proposed in Almandoz et al. (2005) uses simulation to compute the water losses due to leaks and the volume of water consumed but not measured. However, this methodology is analytical, requiring multiple measurements and a mathematical model of the network.

Palau et al. (2004) used multivariate principal component analysis to detect moderate and severe outliers, *i.e.*, bursts. The flow time series is split into days, creating a multivariate problem, and the nighttime period between 0 h and 6 h is analysed. In Palau et al. (2012), the previous approach was extended for three day periods: from 0 h to 7 h, from 7 h to 16 h, and from 16 h to 24 h, separating working days from weekends.

A structured methodology (Mutikanga et al., 2011) was presented to estimate every component of apparent losses, using a sample of calibrated meters, auditors of the meter readers, data handling and billing errors, and inspections of suspected locals. The paper affirms that all the above methods can be used individually and occasionally; however, applying the methodology proposed is usually difficult for water utilities.

Bragalli et al. (2019) proposed a methodology to assess the water losses in real-time using incomplete readings of smart meters. Random sampling and the evolution of the corresponding error are used to study the impact of the lack of reading from an increasing number of smart meters on water losses.

Covas et al. (2008) presented a methodology for assessing real and apparent losses in a DMA, using MNFA combined with pressure data and hydraulic simulations of the system through EPANET.[1] Although this paper includes an approach for assessing different components of the non-revenue water at the same time, it is strongly dependent on fieldwork.

Therefore, a gap persists in data-driven approaches that facilitate the simultaneous assessment of all non-revenue water components. Recent technology, such as smart water meters, could greatly assist in developing a methodology to achieve this goal.

The installation of smart water meters greatly reduces unmetered billed consumption (mainly when there are difficulties in accessing water meters for systematic reading). The availability of billed metered consumption time series, when combined with network flow time series, enables the extraction of non-revenue water time series data. Therefore, it is crucial to have methods for separating non-revenue water into its constituent parts: unbilled authorised consumption, apparent losses, and real losses. These parts could then be used to fill in the water balance, reducing uncertainty about estimates.

[1] EPANET is widely used software for modelling water distribution systems worldwide.

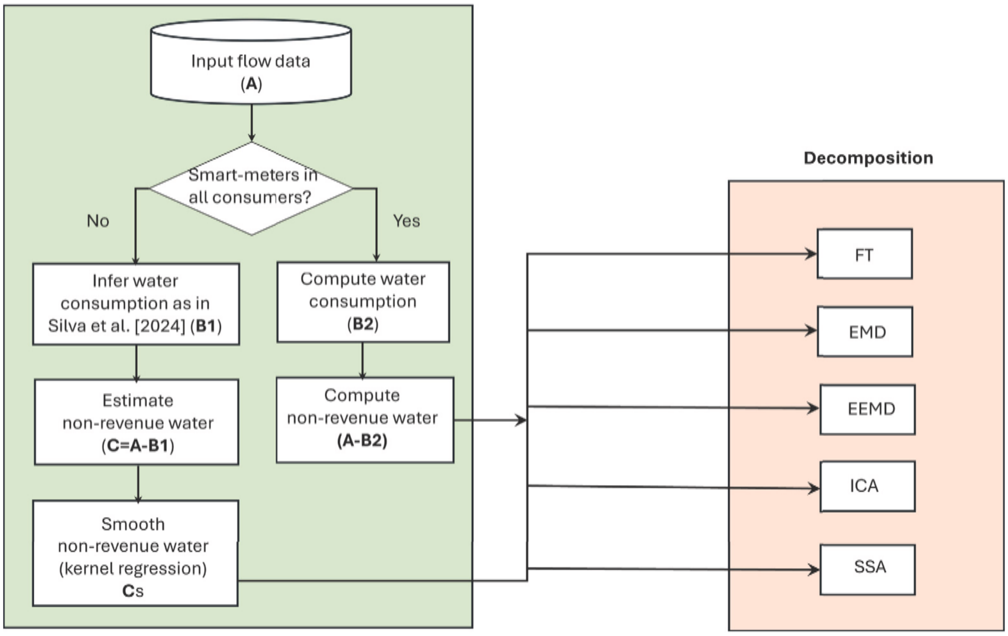


Fig. 1. Diagram for decomposition of non-revenue water time series.

In time series analysis, the data can show some patterns; therefore, splitting a time series into several components, each indicating a different underlying pattern, can be useful to understand the dynamical process. This paper aims to propose a data-driven approach for non-revenue water time series decomposition based on a combination of smart metering consumption, network flow data, and statistical methods.

2. Time series decomposition

The most usual and basic decomposition of time series is through their natural components — trend, seasonal, cyclical and random (Fuller, 2009), but it is unable to decompose the non-revenue water time series into the desired components of the water balance. In order to achieve the desired decomposition, it is necessary to employ other decomposition methods. One option is the time–frequency representation, which maps a one-dimensional signal of time into a two-dimensional signal of time and frequency. Two examples of this type of time series decomposition are the Fourier Transform (FT) and the Wavelet Transform (WT).

FT is a classical technique for decomposing data into linear, stationary, and harmonic components, but it has limitations due to its linearity and lack of time-domain information. WT is an alternative method that uses mathematical functions to decompose time series at different scales and resolutions. It is helpful for non-stationary processes, such as historical price data (Khandelwal et al., 2015).

Besides the time–frequency decomposition methods, empirical methods can also be used to decompose time series into unusual components, such as the Empirical Mode Decomposition (EMD) or Independent Component Analysis (ICA). EMD is a key part of the Hilbert–Huang Transform (HHT) (Huang et al., 1998) which decomposes any data set into Intrinsic Mode Functions (IMFs). However, there is a lack of mathematical understanding of EMD algorithms, and mode mixing becomes a problem when the same IMF exhibits significantly different amplitudes over time, or when different IMFs exhibit similar oscillations in amplitudes. A new method called Ensemble Empirical Mode Decomposition (EEMD) was developed by Wu and Huang (2009) to solve the mode mixing problem.

Several variations of EMD and EEMD (Chu et al., 2012; Jianwei et al., 2017; Moore et al., 2018) have emerged, including single-channel signal data analysis, multivariate extensions, and higher-order statistical techniques like Independent Component Analysis (ICA).

Singular Spectrum Analysis (SSA) is another common technique for decomposing time series. The arbitrary number of additive components that are obtained can be grouped and interpreted as the slowly varying trend, oscillatory and noise components. A single parameter, the window length, is involved in the algorithm, but the decomposition results depend on the choice of this parameter. Kume and Nose-Togawa (2015) proposed a decomposition of the power spectrum with filters useful for monitoring how the SSA algorithm works for the time series decomposition and the proper choice of the window length. This method was applied in Kume and Nose-Togawa (2015) to daily currency exchange rate EUR/USD time series.

3. Methodology

This section presents the suggested approach, along with a concise description of all implemented procedures. The proposed methodology aims to decompose a non-revenue time series into its latent components; the steps are summarised in the following:

1. Collect the system input flow from the DMA, for some time (*e.g.*, one month);
2. Compute the non-revenue water: Does the water utility have smart meters installed on all customers?

   (Yes) Calculate the non-revenue water by subtracting the metered water consumption from the system input flow using synchronous series, normalised with the same time step;

   (No)
   (a) Infer the water consumption over time for the same period, using a representative sample of customers with smart water meters installed (Silva et al., 2024);
   (b) Compute the estimated non-revenue water through the subtraction of the estimated metered water consumption from the system input flow;
   (c) Smooth the estimated non-revenue water time series using the kernel regression smoother;

3. Decompose (smoothed estimated) non-revenue water time series.

Fig. 1 summarises the proposed methodology.

Although the statistical decomposition method is applied to the non-revenue water time series, in fact the whole methodology proposed allows to decompose the system input flow time series, since the authorised metered consumption is estimated using the smart metering data.

If the water utility has smart water meters in all billed customers, the decomposition method can be applied to the non-revenue water time series to obtain its components. In this scenario, neither the authorised billed metered consumption estimation nor the smoothing of the estimated non-revenue water are necessary, as it is possible to obtain the real non-revenue time series.

A more detailed description of the methods used in the methodology steps 2a, 2c and 3 is presented below.

3.1. Inference of the water consumption over time

The following procedure is only required if the water utility does not have smart water meters for all of its customers. The paper by Silva et al. (2024) introduced a comprehensive approach to infer the water consumption over a period of time in that case. The method can be summarised in the following steps:

1. Acquisition of water billing data obtained by the water utility' responsible person at the local from a predominantly domestic network area;
2. Cleaning of water billing data, *e.g.*, through the elimination of the locals without active customers in the analysis period;
3. Clustering of water billing time series data;
4. Obtain a stratified sample of customers by the minimisation of the variance of the error estimator of the total billed consumption. The strata corresponds to clusters obtained in the previous step.
5. Inference of the total metered water consumption for each time instant t: using the sample of n customers with smart metering ($\{\mathbf{y}_{11},\ldots,\mathbf{y}_{1n_1},\ldots,\mathbf{y}_{k1},\ldots,\mathbf{y}_{kn_k}\}$ where $\mathbf{y}_{ij} = \{y_{ij1}, y_{ij2},\ldots,y_{ijM}\}$ represents the smart metering time series of customer j from cluster i and M is the length of the time series), the total consumption is given by:

$$y_{Te,t} = \sum_{i=1}^{k} N_i \overline{y}_{i,t}, \tag{1}$$

where $\overline{y}_{i,t} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ijt}$, n_i is the number of selected customers from stratum i, N_i is the size of stratum i and $t \in \{1,\ldots,M\}$.

3.2. Kernel regression smoother for smoothing the non-revenue water time series

The water balance presented in Table 1 shows that the flow time series has different components: metered authorised water consumption (billed or unbilled), unmetered authorised water consumption (billed or unbilled), unauthorised water consumption, water losses due to leaks and bursts, and measuring errors regarding the water meters.

The sampling approach allowed inferring the metered authorised water consumption over time for all customers supplied by a DMA. Thus, the total metered authorised water consumption can be subtracted from the flow time series, and the remaining flow (the non-revenue water) could be decomposed.

Nevertheless, the sampling approach was designed to infer the total water consumption in a given period. The estimated non-revenue water time series has greater variability than the real non-revenue water time series, which is explained by the sampling approach used to estimate the water consumption. Therefore, in order to achieve a more effective decomposition, it is suggested that smoothing is used as an essential step. The kernel-regression smoother was selected over moving averages due to its superior results.

Considering a time series $\{x_1, x_2,\ldots,x_T\}$ of dimension T, the purpose is to compute a smoother time series $\{\tilde{x}_1,\tilde{x}_2,\ldots,\tilde{x}_T\}$. Kernel regression is a non-parametric technique to estimate the conditional expectation of a variable Y relative to a variable X:

$$E(Y|X) = m(X), \tag{2}$$

where m is a function.

Nadaraya (1964) and Watson (1964) proposes that m is defined as:

$$m_h(x) = \frac{\sum_{i=1}^{n} K_h(x-x_i)y_i}{\sum_{i=1}^{n} K_h(x-x_i)}, \tag{3}$$

where K_h is a kernel with a bandwidth h, *i.e.*, $K_h(x) = K(\frac{x}{h})$. A Gaussian kernel could be considered. When applied to a time series, the variable X is considered the time and Y is the value of the time series. The bandwidth, h, needs to be chosen and validated in practice.

3.3. Decomposition of non-revenue water into hidden components

The time series decomposition techniques mentioned in Section 2 can be used to carry out the process of splitting the non-revenue time series into its hidden components. This section briefly discusses the most common decomposition strategies that can be useful in this application domain.

3.3.1. Fourier Transform (FT)

Considering a sequence of N complex numbers $\{\mathbf{x}_n\} = \{x_0, x_1,\ldots,x_{N-1}\}$, the Discrete Fourier Transform (DFT) transforms this sequence into another sequence of complex numbers $\{\mathbf{X}_k\} = \{X_0, X_1,\ldots,X_{N-1}\}$ given by:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{i2\pi}{N}kn} \tag{4}$$

$$= \sum_{n=0}^{N-1} x_n \left[cos\left(\frac{2\pi}{N}kn\right) - i\ sin\left(\frac{2\pi}{N}kn\right)\right]. \tag{5}$$

The DFT transforms a sequence of values in the time domain into a sequence of values in the frequency domain. After identifying the predominant frequencies and removing the undesired ones, an inverse transformation is necessary to return to the time domain. The Inverse Fourier Transform (IFT) is such that:

$$x_n = \frac{1}{N}\sum_{k=0}^{N-1} X_k e^{\frac{i2\pi}{N}kn}. \tag{6}$$

For the application of FT, the system must be linear and the data must be strictly periodic or stationary.

3.3.2. Empirical Mode Decomposition (EMD)

The Empirical Mode Decomposition (EMD) method decomposes any data set into Intrinsic Mode Functions (IMFs) (Huang et al., 1998). An IMF is a function that satisfies two conditions: (1) in the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one; (2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

The method proposed for the decomposition into IMF components is intuitive, direct, *a posteriori* and adaptive, only based on the data. The systematic process for the extraction of IMFs is designated as the sifting process. In the end, the data can be decomposed into n-empirical modes and a residue, either the mean or a constant.

EMD has been tested and validated, but only empirically. This method greatly benefits from the true physical meanings found in many of the examined data sets. Furthermore, it is a data-driven algorithm and works in the temporal domain directly rather than in the corresponding frequency domain (Mijovic et al., 2010).

3.3.3. Ensemble Empirical Mode Decomposition (EEMD)

Despite the EMD method has been widely adopted to decompose time series (Hong, 2011; Qiu et al., 2016; Yaslan and Bican, 2017; Nava et al., 2018), there is a lack of mathematical understanding of the EMD algorithms, *e.g.*, the dependence of IMFs on the number of sifting, the convergence property and the stability to noise perturbation (Singh et al., 2017; Moore et al., 2018). Moreover, a problem was pointed out when the same IMF has very different amplitudes along time or different IMFs have similar amplitude oscillations, called mode mixing. In addition, applying EMD often results in more IMFs than the number of characteristic time scales actually present.

As previously mentioned, Wu and Huang (2009) introduced the Ensemble Empirical Mode Decomposition (EEMD) method, which addresses the issue of mode mixing in EMD by enhancing the high-frequency component through the addition of random white Gaussian noises to the original signal.

Adding white noise improves the accuracy of the decomposed signal and preserves its original characteristics. EEMD depends on the amplitude of the added noise and the ensemble times. When the amplitude of the added white noise is too low, the mode mixing problem cannot be suppressed, while if the amplitude is too high, more pseudo components will appear (Yang et al., 2017). Relative to the ensemble times, if the noise is added in the EEMD more times, then the noise of the average result is smaller, and the result is closer to the real value (Zhang et al., 2018).

3.3.4. Independent Component Analysis (ICA)

The Independent Component Analysis (ICA) is a technique aiming to find a linear representation of non-Gaussian data, through a set of statistically uncorrelated components (Hyvärinen and Oja, 2000; Hyvärinen et al., 2001).

When the data are time series, the model is defined as:

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t, \tag{7}$$

where $\mathbf{A}$ is a square matrix and the Independent Components (ICs) s_{it} (the hidden components) are uncorrelated. The assumption is not that the ICs are non-Gaussian, but rather that they have different auto-covariances, particularly non-null ones, or their variance is not stationary. The method based on auto-covariances is applicable for temporal dependent data, but only works when auto-covariances are different. An alternative method uses non-stationary data, which works better for data with changing variance over time (Hyvärinen et al., 2001).

ICA requires multivariate data. However, if the data is only one time series, a typical approach in the literature is to use EMD or EEMD before applying ICA to the set or a subset of IMFs (Mijovic et al., 2010; Xian et al., 2016; Yu et al., 2018).

3.3.5. Singular Spectrum Analysis (SSA)

Singular Spectrum Analysis (SSA) is one technique whose primary objective is to separate the original time series into an interpretable set of time series. The arbitrary number of additive components that are obtained can be grouped and interpreted as the slowly varying trend, oscillatory, and noise components. SSA is a nonparametric, data-adaptive spectral technique that is essentially an application of principal component analysis in the time domain. A single parameter, the window length, is involved in the algorithm, and the decomposition results depend on its choice.

Considering a time series $\{x_1, x_2, \ldots, x_T\}$ of dimension T, the algorithm of SSA can be summarised as follows (Hassani, 2007):

1. **Embedding:** Built the trajectory matrix $\mathbf{X}$ of the time series;
2. **Singular Value Decomposition (SVD):** Perform the SVD of the trajectory matrix $\mathbf{X}$;
3. **Grouping:** Considering m disjoint subsets $I_1, \ldots, I_m$ of the indices $\{1, 2, \ldots, rank(\mathbf{X})\}$, and $I_k = \{i_{k_1}, \ldots, i_{k_p}\}$, compute the matrix $\mathbf{X}_{I_k}$ corresponding to the group I_k;
4. **Diagonal averaging:** Each matrix $\mathbf{X}_{I_k}$ is transformed in a Hankel matrix and, then, into a new series of length T through the correspondence between Hankel matrices and time series. Diagonal averaging applied to the resultant matrix $\mathbf{X}_{I_k}$ produces a reconstructed time series $\{y_1^k, \ldots, y_T^k\}$. Therefore, the original time series can be decomposed into the sum of m reconstructed subseries:

$$x_n = \sum_{k=1}^{m} y_n^k, \; n = 1, \ldots, T. \tag{8}$$

4. From theory to practice: Methodology application

4.1. Case study description

A predominantly domestic water distribution network area was used in this work to illustrate the proposed methodology. The water distribution network is organised into network areas where the water utility has invested in smart metering to monitor water losses. According to the latest available data (ERSAR, 2023), unbilled water for this water utility in 2022 was 19.2%. This figure is relatively low compared to the last reported average by ERSAR for the entire country, which was 27.1%. Furthermore, real losses averaged 62 liters per service connection per day, significantly below the IWA's upper limit for good performance, which is 100 liters per service connection per day. Results from this study indicate that there are few pipe bursts in the network, aligned with the water utility's overall performance of 13 failures per 100 km per year — this figure is within the acceptable limits for good performance.

Locals without customers were removed from the analysis because it makes no sense to include them. Thus, from the total 582 locals of the DMA, only 496 with active customers were considered.

System input flow data were collected between March 2017 and February 2019, while water billing data was collected from April 2017 to March 2019.[2] During this period, no smart meters were installed in the customers and the water utility manually collected the monthly consumption of each customer for invoicing. This case study was used in Silva et al. (2024) to infer water consumption over time using a representative sample of customers with smart meters.

The approach presented in Section 3.1, which only uses billing data obtained without smart metering for the representative sample selection of customers and uses smart data for the subsequent revenue water estimation, was used to estimate the revenue water. Afterwards, the non-revenue water was computed as the subtraction between the system input flow and the total water metered consumption for decomposition. The non-revenue water volume accounts 19% of the total system input volume (which can be considered a low value (Liemberger and Wyatt, 2019), and is aligned with the value reported by the water utility for the entire system) in the period between the 26th of October, 2019 and the 4th of January, 2020. This was the period considered for the revenue water inference since there are smart metering data from all customers for this period, as well as the system input time series. Thus, the estimated non-revenue water (based on the estimated revenue water) can be compared with the real non-revenue water (based on the smart metering data from all customers).

In this case study, all the smart water meters were new, and it was assumed that the measurement errors could be neglected. Therefore, the negative values in the non-revenue water (without a physical meaning) come from to the fact that the flowmeter, that controls the flow into and out of the distribution network, has systematic errors (due to under-registration, for example) that are not known (Ribeiro et al., 2018). To prevent this from happening, it would be necessary to calibrate this equipment either in situ or in the laboratory, a process that was not possible in this study. The next sections present the smoothing and decomposition results obtained.

[2] Billing data has a delay of one month since the invoice corresponds to the last month's consumption.

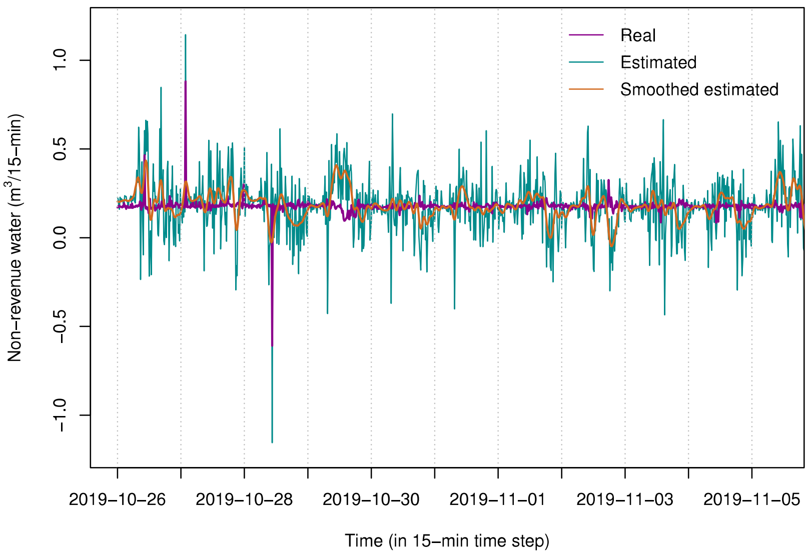


Fig. 2. Real, estimated and smoothed estimated non-revenue water using the kernel regression smoother with a bandwidth equal to eight.

4.2. Smoothing of the estimated non-revenue water

The sampling approach used to infer the total water consumption in a given period also allowed inferring the metered authorised water consumption over time for all customers supplied. However, this approach generates greater variability in the estimated non-revenue water time series, as can be observed in Fig. 2, where the real and estimated non-revenue water are represented. Therefore, the kernel-regression smoother was suggested as an essential step.

A bandwidth (h) must be defined to apply kernel regression smoother (Eqs. (2) and (3)). To choose the best option for the bandwidth, a Euclidean distance between the real and the smoothed estimated non-revenue water was computed for several values of the parameter:

$$d(\mathbf{X}, \tilde{\mathbf{X}}) = \sqrt{\sum_{i=1}^{T} (x_i - \tilde{x}_i)^2}, \tag{9}$$

where $\mathbf{X} = \{x_1, x_2, \dots, x_T\}$ is the real non-revenue water time series and $\tilde{\mathbf{X}} = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_T\}$ is the smoothed estimated non-revenue water time series.

The Euclidean distance between the real and smoothed estimates of non-revenue water was calculated for bandwidths ranging from 2 to 20. Typically, when selecting a bandwidth, one should consider the curve's elbow. However, the elbow is not evident in this situation, and the distance ranges from 12.5 to 10.5. Therefore, the first bandwidth with a distance below 11 was considered, *i.e.*, a bandwidth of 8. Fig. 2 illustrates the difference between the real, estimated, and smoothed estimated non-revenue water over a period of one and a half weeks. The discrepancies between the real non-revenue water and the smoothed estimates are minimal, with deviations primarily arising from the estimation process of the non-revenue water itself.

4.3. Decomposition of the smoothed estimated non-revenue water

The next goal is to decompose the smoothed estimated non-revenue water using the methods discussed in Section 3.3. In this section, the decomposition results of the application of each method or a combination of methods to the smoothed estimated non-revenue water are presented and discussed. Since the real components of the non-revenue water are not known, the analysis is based on the components' shape, common sense and the opinion of knowledgeable people. To validate these approaches, it is essential to understand the history of bursts, the unmetered uses, and the illegal uses, based on data available at the water utility (e.g., work orders, identified illegal water uses, identification and characterisation of authorised unmetered consumption, namely for garden watering, fire fighting).

4.3.1. Fourier Transform

To decompose the estimated smoothed non-revenue water using the FT, the first step was to compute the existing frequencies and respective amplitudes of the time series in the study (`spec.fft` function from *spectral* R package).

This series presented 6789 different frequencies with a different amplitude associated, which corresponds to its length. Fig. 3 shows the spectrum of the time series, *i.e.*, the frequencies and the respective amplitudes. Although the null frequency has a prominent amplitude, there are also several frequencies around zero with important values of amplitude. For the decomposition using the FT, the predominant frequencies should be considered, and the undesired ones should be removed. However, observing the spectrum of the smoothed estimated non-revenue water, there are a huge number of predominant frequencies. Thus, the decomposition procedure will result in many components, one per frequency. Furthermore, each of these components will exhibit harmonic characteristics. Grouping frequencies is one possible solution. Nevertheless, the number of possible combinations makes this procedure impractical.

4.3.2. Empirical Mode Decomposition

EMD is a technique that decomposes data into IMFs based solely on the data, without the need for defined parameters. Applying EMD to the smoothed estimated non-revenue water (`emd` function from *EMD* R package) resulted in 10 IMFs and the residue, presented in Fig. 4.

The residue serves as a representation of the data trend and may indicate the occurrence of small leaks spread through the network. Furthermore, the first four components may indicate the occurrence of bursts, authorised unmetered consumption and unauthorised consumption. Nevertheless, it is believed that the shape of the remaining IMFs does not correspond to the shape of any desired component of the smoothed estimated non-revenue water decomposition. Some combinations of IMFs could be more meaningful. However, the high number of IMFs, the lack of information about possible combinations, and the difficulty in understanding what the remaining components represent complicate this process.

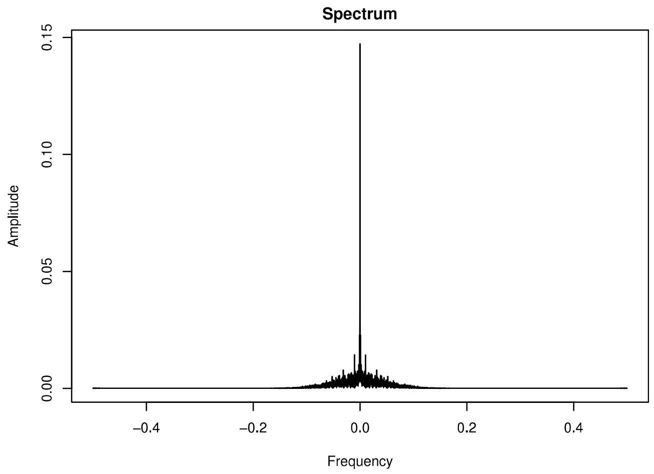


Fig. 3. Spectrum of the smoothed estimated non-revenue water (frequency vs. amplitude).

4.3.3. Ensemble Empirical Mode Decomposition

For the EEMD application, two parameters are necessary to be defined: the amplitude of the noise and the ensemble times. When the amplitude of the noise is too high, pseudo components will appear. When the amplitude is too low, the mode mixing problem cannot be suppressed. Some values for the amplitude were considered without significant changes in the results. Regarding the ensemble times, when the noise is added more times, the result is closer to the real value. In this case, the ensemble times was defined as 50. The results presented from now on were obtained with a noise equal to 0.0001. The application of EEMD (`EEMD` function from *hht* R package) with these parameters resulted in nine average IMFs and the residue, presented in Fig. 5.

EEMD allowed reducing one IMF relative to the application of EMD. Moreover, the last two average IMFs together with the residue could represent the leaks, while the first three, maybe four, average IMFs could be the bursts, authorised unmetered consumption and the unauthorised consumption. Still, it is difficult to understand the meaning of the remaining elements.

4.3.4. Independent Component Analysis

A multivariate time series is required for the application of ICA. Because the goal is to decompose the smoothed estimated non-revenue water, a standard combination of procedures was employed. Thus, EMD/EEMD was used first to produce the set of IMFs, and then ICA was applied to the IMFs. Three choices were explored based on the literature (Mijovic et al., 2010; Xian et al., 2016; Yu et al., 2018): using all IMFs, using the original time series and a set of IMFs less correlated with it, and using the set of IMFs more correlated with it. The three alternatives were repeated, with EMD and EEMD as the first step.

ICA was run using the FastICA algorithm that uses non-gaussianity measured by the negentropy (`fastICA` function from *fastICA* R package with default values).

Using the 10 IMFs obtained with EMD, ICA was applied to the set of all IMFs, obtaining 10 ICs represented in Fig. 6. There are some ICs (2nd, 3rd and 10th) that could represent a trend, *i.e.*, the leaks that occur in the system. The 1st, 5th, 6th and 9th ICs could be the bursts, authorised unmetered consumption and the unauthorised consumption. Nevertheless, it is still hard to understand the remaining components and their physical meaning.

Applying EMD to the original time series, either to the set of IMFs less correlated with the original time series or to the set of IMFs more correlated with the original time series, rendered interpreting the IMFs challenging. Although fewer ICs were obtained, there was no physical meaning for the components from the non-revenue water decomposition. In some cases, the IMFs are unable to represent the desired components of the decomposition of the smoothed estimated non-revenue water. Even though some components could be combined (although knowing which components should be combined could be difficult), this will be a manual operation without a practical guide to follow. Since the goal is to give practical tools for water utilities, combining components based on common sense may not be a beneficial option.

The previous three ICA applications were redone using the average IMFs obtained from EEMD. Similar difficulties were encountered in all applications, thus the results are not reported. In this case, certain combinations make no physical sense, or possible combinations are hard to define. Furthermore, it seems that the usual leaks are not represented in some of the decompositions.

4.3.5. Singular Spectrum Analysis

For the SSA application, besides the window length, the number of eigenvalues considered is also a parameter to be selected in SSA. Initially, the window length was defined as $L = \frac{T+1}{2}$, where T is the length of the time series, and the number of eigenvalues as the same default value of the R function, *i.e.*, 50.

The norm of each component linked to each eigenvalue was calculated (Fig. 7) in order to figure out how the components should be grouped. Observing this plot, the components with similar norms should be grouped. In this case, the 1st component stands alone, the 2nd and 3rd should be grouped as a 2nd final component, and the 4th to 50th component should also be grouped as a 3rd final component. This means the smoothed estimated non-revenue water is decomposed into three components plus the residuals.

Fig. 8 presents the three components and the residuals after the grouping and diagonal averaging. These components characterise non-revenue water (authorised unbilled unmetered consumption, apparent losses due to unauthorised consumption, and real losses due to leaks and pipe bursts). Component 1 stands for a trend almost constant, probably representing the set of small continuous leaks and background leakage in the system. The remaining components may be attributed to variations in consumption behaviour (unbilled unmetered consumption and unauthorised consumption) rather than pipe bursts. Based on the identified components, it is unlikely that pipe bursts have occurred. Such events are typically associated with a sudden and sustained increase in flow until the pipe is repaired. These bursts usually lead to prolonged disruptions and abnormal flow increases that alert utility managers to potential failures. Component 2 represents some oscillations, maybe due to some weekly water consumption variations. Component 3 and residuals could represent authorised unbilled unmetered or unauthorised consumption. It is pinpointed that the negative values in components 2, 3 and residuals could be due to flowmeter errors. A characterisation was conducted for the negative values observed in components 2 and 3. It was found that their magnitudes range from -10^{-7} to -10^{-2}, which is significantly lower compared to non-revenue water values. Therefore, these negative values can be considered negligible (Simões et al., 2024). Additionally, the regression line associated with the calibration of the flowmeters could be utilised to correct these values.

Tests with the window length varying between 25% and 75% of the time series length in steps of 250 were run. Some differences were verified in the cut-off point between the 2nd and 3rd components in terms of the 4th and 5th eigenvalues. However, observing the corresponding components makes it difficult to identify the differences. Thus, $L = \frac{T+1}{2}$ continued to be the choice.

In terms of eigenvalues, a range of 5 to 50 was investigated, with steps of 5. When the number of eigenvalues is reduced, the variability of the 3rd component reduces, reaching almost a trend when five eigenvalues are used. Since this makes no physical sense, it was decided that 50 eigenvalues would be the best option.

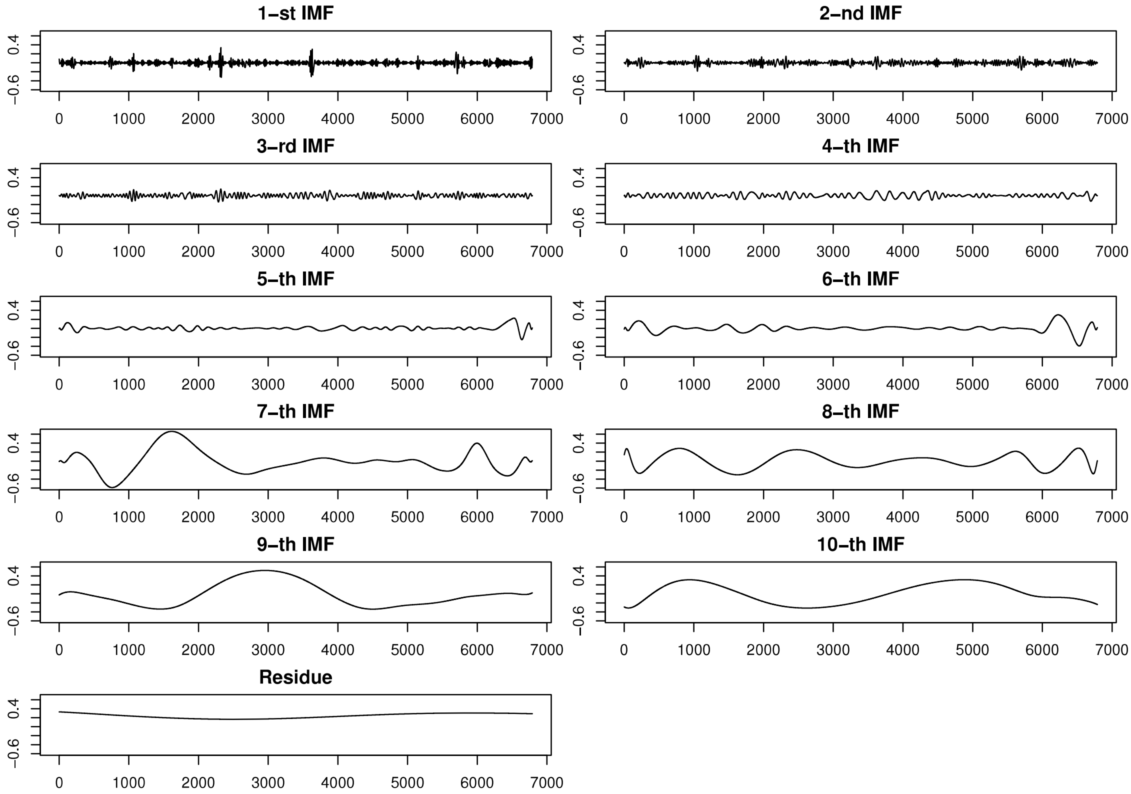


Fig. 4. EMD decomposition of the smoothed estimated non-revenue water into 10 IMFs and the residue. The *y*-axis is equal in all subplots.

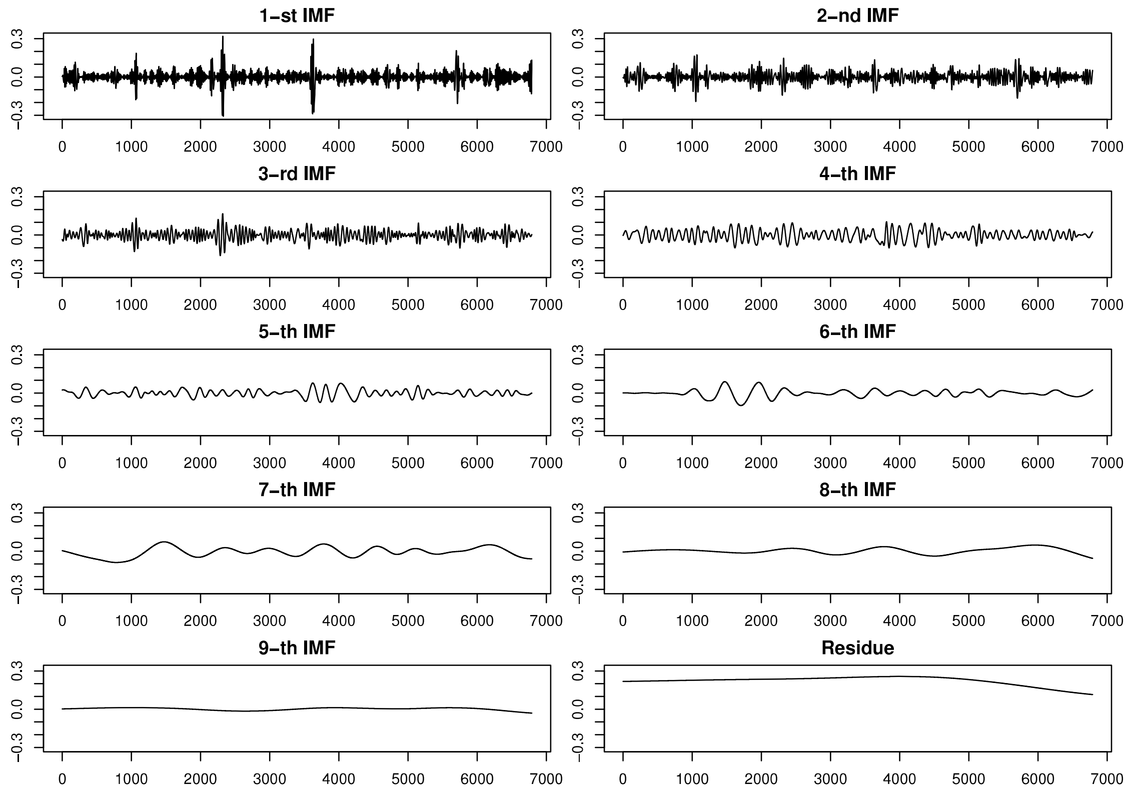


Fig. 5. EEMD decomposition of the smoothed estimated non-revenue water into nine average IMFs and the residue. The *y*-axis is equal in all subplots.

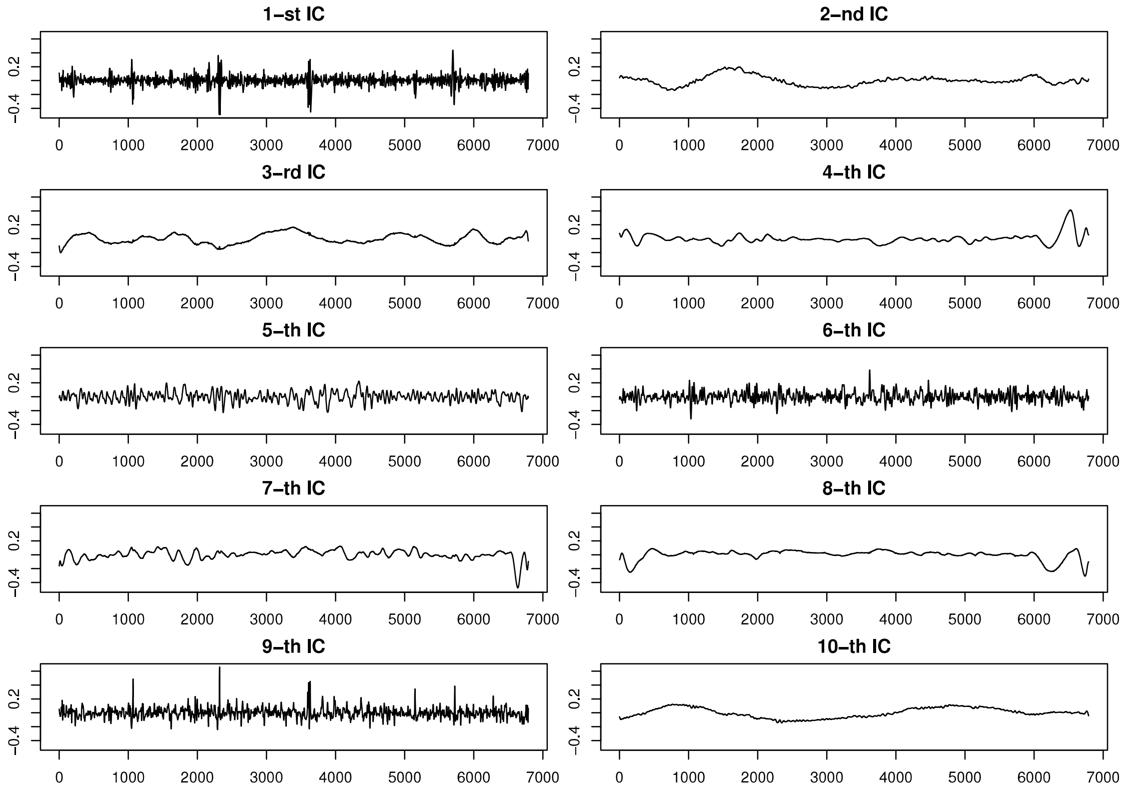


Fig. 6. ICA decomposition of the smoothed estimated non-revenue water into 10 ICs, using all IMFs obtained with EMD.

Thus, of all the decompositions obtained previously, the one presented in Fig. 8 has the most physical meaning. Fig. 9 presents the decomposition of the real non-revenue water. There are some differences, especially in component 2 and the residuals. These differences are essentially visible in the variability and peaks, the result of the smoothing that had to be applied previously and which were already visible when comparing the real non-revenue water time series with the smoothed estimated series. However, the general behaviour of the components seems to be in line with the expected behaviour of the non-revenue water components and the results obtained are therefore considered a satisfactory approximation. Table 2 presents the system input volume series decomposition obtained with SSA and according to the water balance. The estimated real losses due to small continuous leaks and background leakage result from the Component 1 and Component 2, while the estimated authorised unmetered and unauthorised consumption result from Component 3 and residuals. The identified base loss (small continuous leaks and background leakage) seems reasonable for a network with the characteristics of the water utility and could likely be managed through pressure regulation or network rehabilitation. However, the current rehabilitation rate is only 0.4% per year, which is below the recommended rate.

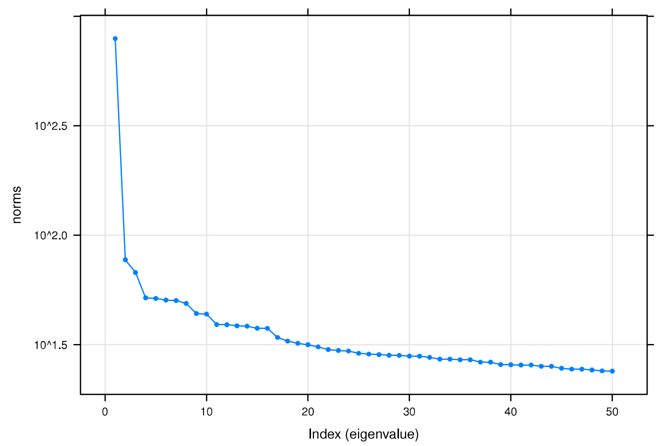


Fig. 7. Norm of each component associated with each eigenvalue in the SSA analysis.

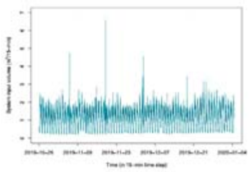
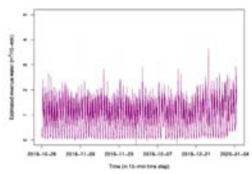
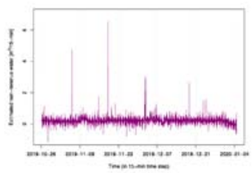
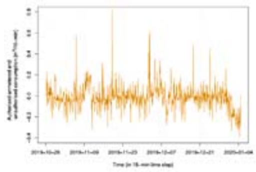
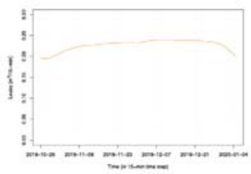
5. Conclusion

This work introduces and evaluates an approach for decomposing the flow time series, specifically those related to non-revenue water. When the water utility does not have smart water meters for all consumers, the non-revenue water needs to be estimated based on stratified sampling and time series smoothing.

A real-case study of a drinking water distribution system was considered. Smart water meters were already installed in this network area. For achieving the non-revenue water time series, the metered water consumption time series was estimated using smart metering data from a representative sample of customers. Because the stratified sampling approach uses global statistics, the inherent time variability is not taken into account. This variation is spread when the estimated metered consumption time series is subtracted from the system input flow to get the estimated unmetered consumption. Therefore, we needed to smooth out the estimated unmetered consumption before using decomposition methods, and we considered a kernel regression smoother for that purpose. Since the bandwidth definition is required, many values for this parameter were tried. The parameter value that produced the smallest Euclidean distance between the real and smoothed estimated non-revenue water was chosen. In this case, the kernel regression smoother with a bandwidth of eight was used to smooth the estimated non-revenue water.

Afterwards, different methods for time series decomposition into unusual components were applied to the smoothed estimated non-revenue water. The results with SSA were considered satisfactory according to

Table 2
System input volume series decomposition according to the water balance.

<table>
<tr><td rowspan="3">System input volume
</td><td colspan="2">Estimated revenue water
</td></tr>
<tr><td rowspan="2">Estimated non-revenue water
</td><td>Estimated authorised unmetered and unauthorised consumption
</td></tr>
<tr><td>Estimated real losses due to small continuous leaks and background leakage
</td></tr>
</table>

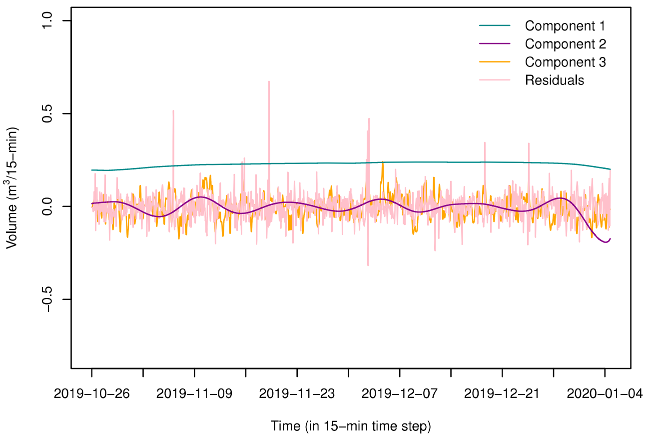


Fig. 8. SSA decomposition of the smoothed estimated non-revenue water into three components and the residuals.

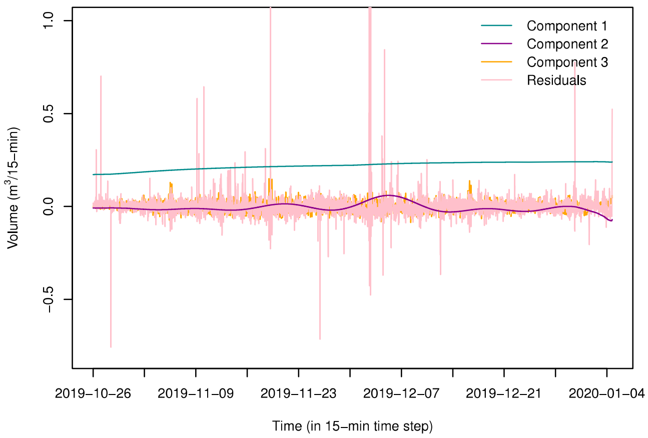


Fig. 9. SSA decomposition of the non-revenue water into three components and the residuals.

the water utility. Thus, the smoothed estimated non-revenue water was decomposed into three components plus the residuals. Component 1 could be considered the **usual small continuous leaks and background leakage** since it is a trend, while component 2 presents some oscillations, maybe due to **weekly water consumption variations**. Component 3 and residuals might be related to **authorised unbilled unmetered consumption** or **unauthorised consumption**. It is noted that the negative values in components 2, 3 and residuals could be due to flowmeter errors. Although SSA is applied to the non-revenue water, the entire methodology is able to decompose the system input flow into its components: authorised metered consumption, authorised unmetered consumption, unauthorised consumption, authorised consumption errors, and real losses due to leaks.

According to the water balance, the decomposition achieved is aligned with the goal of breaking down the non-revenue water into its individual parts. These findings are crucial for the water utility as they contribute to fill the water balance and effectively managing water losses. It is important to note that the water company only needs to use SSA (Singular Spectrum Analysis) on the real non-revenue water time series for the decomposition, provided that all customers already have smart water meters. For future development, it is advisable to implement the SSA decomposition in other network sectors that have higher levels of non-revenue water (both apparent and real losses) and various problematic components. In addition, calibrating the existing flowmeters in the network and correcting the respective flow series for systematic errors should be one of the first steps in the time series decomposition process.

CRediT authorship contribution statement

Maria Almeida Silva: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Conceição Amado:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization. **Dália Loureiro:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Maria Almeida Silva reports financial support was provided by Foundation for Science and Technology (FCT), Portugal. Conceicao Amado reports financial support was provided by Foundation for Science and Technology (FCT), Portugal. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The first author acknowledges the Fundação para a Ciência e Tecnologia (FCT), Portugal, for the PhD Grant SFRH/BD/131382/2017. The first two authors also acknowledge the financial support of the FCT through the projects UIDB/04621/2020 and UIDP/04621/2020 of CEMAT/IST-ID (Center for Computational and Stochastic Mathematics), Instituto Superior Técnico, Universidade de Lisboa.

Data availability

The data that has been used is confidential.

References

AL-Washali, T., Sharma, S., Kennedy, M., 2016. Methods of assessment of water losses in water supply systems: a review. Water Resour. Manag. 30 (14), 4985–5001. http://dx.doi.org/10.1007/s11269-016-1503-7.

Alegre, H., Baptista, J.M., Cabrera, Jr., E., Cubillo, F., Duarte, P., Hirner, W., Merkel, W., Parena, R., 2006. Performance Indicators for Water Supply Services, second ed. IWA Publishing.

Alkasseh, J.M.A., Adlan, M.N., Abustan, I., Aziz, H.A., Hanif, A.B.M., 2013. Applying minimum night flow to estimate water loss using statistical modeling: a case study in Kinta Valley, Malaysia. Water Resour. Manag. 27 (5), 1439–1455. http://dx.doi.org/10.1007/s11269-012-0247-2.

Almandoz, J., Cabrera, E., Arregui, F., Cabrera, Jr., E., Cobacho, R., 2005. Leakage assessment through water distribution network simulation. J. Water Resour. Plan. Manag. 131 (6), 458–466. http://dx.doi.org/10.1061/(ASCE)0733-9496(2005)131:6(458).

Amoatey, P.K., Minke, R., Steinmetz, H., 2014. Leakage estimation in water networks based on two categories of night-time users: a case study of a developing country network. Water Sci. Technol.: Water Supply 14 (2), 329–336. http://dx.doi.org/10.2166/ws.2013.201.

Bragalli, C., Neri, M., Toth, E., 2019. Effectiveness of smart meter-based urban water loss assessment in a real network with synchronous and incomplete readings. Environ. Model. Softw. 112, 128–142. http://dx.doi.org/10.1016/j.envsoft.2018.10.010.

Casale, G., Cordeiro Ortigara, A.R. (Eds.), 2019. Water in the 2030 Agenda for Sustainable Development: How Can Europe Act? UNESCO and Water Europe, Paris, France and Brussels, Belgium, ISBN: 978-92-3-100367-7, URL https://unesdoc.unesco.org/ark:/48223/pf0000372496. Available in Open Access under the Attribution ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) license.

Chu, P.C., Fan, C., Huang, N., 2012. Compact empirical mode decomposition: an algorithm to reduce mode mixing, end effect, and detrend uncertainty. Adv. Adapt. Data Anal. 4 (3), 1250017. http://dx.doi.org/10.1142/S1793536912500173.

Covas, D., Jacob, A.C., Ramos, H.M., 2008. Water losses' assessment in an urban water network. Water Pr. Technol. 3 (3), http://dx.doi.org/10.2166/wpt.2008.061.

ERSAR, 2023. Relatório anual dos serviços de águas e resíduos em Portugal 2023 (RASARP2023) - volume 1 – caraterização do setor de águas e resíduos (in Portuguese).

Fu, G., Jin, Y., Sun, S., Yuan, Z., Butler, D., 2022. The role of deep learning in urban water management: A critical review. Water Res. (ISSN: 0043-1354) 223, 118973. http://dx.doi.org/10.1016/j.watres.2022.118973, URL https://www.sciencedirect.com/science/article/pii/S0043135422009204.

Fuller, W.A., 2009. Introduction to Statistical Time Series, second ed. John Wiley & Sons.

Hassani, H., 2007. Singular spectrum analysis: methodology and comparison. J. Data Sci. 5, 239–257.

Hong, L., 2011. Decomposition and forecast for financial time series with high-frequency based on empirical mode decomposition. Energy Procedia 5, 1333–1340. http://dx.doi.org/10.1016/j.egypro.2011.03.23.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 454, pp. 903–995. http://dx.doi.org/10.1098/rspa.1998.0193.

Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. John Wiley & Sons.

Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural Netw. 13 (4), 411–430. http://dx.doi.org/10.1016/S0893-6080(00)00026-5.

Jianwei, E., Shenggang, L., Jimin, Y., 2017. A new approach to gold price analysis based on variational mode decomposition and independent component analysis. Acta Phys. Pol. B 48 (11), http://dx.doi.org/10.5506/APhysPolB.48.2093.

Khandelwal, I., Satija, U., Adhikari, R., 2015. Efficient financial time series forecasting model using DWT decomposition. In: IEEE International Conference on Electronics, Computing and Communication Technologies. pp. 1–5. http://dx.doi.org/10.1109/CONECCT.2015.7383917.

Kume, K., Nose-Togawa, N., 2015. Spectral structure of singular spectrum decomposition for time series. arXiv.

Liemberger, R., Wyatt, A., 2019. Quantifying the global non-revenue water problem. Water Supply 19 (3), 831–837. http://dx.doi.org/10.2166/ws.2018.129.

McKenzie, R., Seago, C., 2005. Assessment of real losses in potable water distribution systems: some recent developments. Water Sci. Technol.: Water Supply 5 (1), 33–40. http://dx.doi.org/10.2166/ws.2005.0005.

Mijovic, B., De Vos, M., Gligorijevic, I., Taelman, J., Van Huffel, S., 2010. Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis. IEEE Trans. Biomed. Eng. 57 (9), 2188–2196. http://dx.doi.org/10.1109/TBME.2010.2051440.

Moore, K.J., Kurt, M., Eriten, M., McFarland, D.M., Bergman, L.A., Vakakis, A.F., 2018. Wavelet-bounded empirical mode decomposition for measured time series analysis. Mech. Syst. Signal Process. 99, 14–29. http://dx.doi.org/10.1016/j.ymssp.2017.06.005.

Mutikanga, H.E., Sharma, S.K., Vairavamoorthy, K., 2011. Assessment of apparent losses in urban water systems. Water Environ. J. 25 (3), 327–335. http://dx.doi.org/10.1111/j.1747-6593.2010.00225.x.

Mutikanga, H.E., Sharma, S.K., Vairavamoorthy, K., 2013. Methods and tools for managing losses in water distribution systems. J. Water Resour. Plan. Manag. 139 (2), 166–174. http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000245.

Nadaraya, E.A., 1964. On estimating regression. Theory Probab. Appl. 9 (1), 141–142. http://dx.doi.org/10.1137/1109020.

Nava, N., Di Matteo, T., Aste, T., 2018. Financial time series forecasting using empirical mode decomposition and support vector regression. Risks 6 (1), 7. http://dx.doi.org/10.3390/risks6010007.

Palau, C.V., Arregui, F.J., Carlos, M., 2012. Burst detection in water networks using principal component analysis. J. Water Resour. Plan. Manag. 138 (1), 47–54. http://dx.doi.org/10.1061/(ASCE)WR.1942-5452.0000147.

Palau, C.V., Arregui, F., Ferrer, A., 2004. Using multivariate principal component analysis of injected water flows to detect anomalous behaviors in a water supply system - a case study. Water Supply 4 (3), 169–182. http://dx.doi.org/10.2166/ws.2004.0052.

Qiu, X., Suganthan, P.N., Amaratunga, G.A.J., 2016. Electricity load demant time series forecasting with empirical mode decomposition based random vector functional link network. In: IEEE International Conference on Systems, Man and Cybernetics. http://dx.doi.org/10.1109/SMC.2016.7844431.

Ribeiro, A.S., Loureiro, D., Martins, L., Sousa, J.A., Batista, E., Dias, L., Soares, A.C., 2018. Measurement uncertainty of turbine flow meter calibration used in conformity assessment for water management. J. Phys.: Conf. Ser. 1065 (9), 092006. http://dx.doi.org/10.1088/1742-6596/1065/9/092006.

Silva, M.A., Amado, C., Loureiro, D., 2024. Inferring demand in drinking water distribution systems through stratified sampling of billing data for smart meter installation. J. Water Resour. Plan. Manag. (ISSN: 0733-9496) 150 (8), http://dx.doi.org/10.1061/JWRMD5.WRENG-6404.

Simões, C., Ribeiro, A.S., Almeida, M.C., Covas, D., 2024. Traceability and measurement uncertainty of non-removable field flowmeters using clamp-on ultrasonic flowmeters as reference. ACTA IMEKO (ISSN: 2221-870X) 13 (3), 1–6. http://dx.doi.org/10.21014/actaimeko.v13i3.1777.

Singh, P., Joshi, S.D., Patney, R.K., Saha, K., 2017. The Fourier decomposition method for nonlinear and non-stationary time series analysis. In: Proceedings of the Royal Society London A: Mathematical, Physical and Engineering Sciences, vol. 473, 20160871. http://dx.doi.org/10.1098/rspa.2016.0871.

Watson, G.S., 1964. Smooth regression analysis. Sankhyā: Indian J. Stat. Ser. A 359–372.

Wu, Z., Huang, N.E., 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv. Adapt. Data Anal. 1 (01), 1–41. http://dx.doi.org/10.1142/S1793536909000047.

Xian, L., He, K., Lai, K.K., 2016. Gold price analysis based on ensemble empirical mode decomposition and independent component analysis. Phys. A 454, 11–23. http://dx.doi.org/10.1016/j.physa.2016.02.055.

Yang, H., Jiang, Z., Lu, H., 2017. A hybrid wind speed forecasting system based on a 'decomposition and ensemble' strategy and fuzzy time series. Energies 10 (9), 1422. http://dx.doi.org/10.3390/en10091422.

Yaslan, Y., Bican, B., 2017. Empirical mode decomposition based denoising method with support vector regression for time series prediction: a case study for electricity load forecasting. Measurement 103, 52–61. http://dx.doi.org/10.1016/j.measurement.2017.02.007.

Yu, H., Guo, J., Kong, Q., Chen, X., 2018. Gravity tides extracted from relative gravimeter data by combining empirical mode decomposition and independent component analysis. Pure Appl. Geophys. 175, 1683–1697. http://dx.doi.org/10.1007/s00024-018-1864-3.

Zhang, X., Zhang, Q., Zhang, G., Nie, Z., Gui, Z., 2018. A hybrid model for annual runoff time series forecasting using Elman neural network with ensemble empirical mode decomposition. Water 10 (4), 416. http://dx.doi.org/10.3390/w10040416.