Uncovering Abnormal Water Consumption Patterns for Sustainability's Sake: A Statistical Approach

Ana Borges ⓘ**, Clara Cordeiro** ⓘ**, and M. Rosário Ramos** ⓘ

Abstract Monitoring domestic water usage may help the water utilities uncover abnormal water consumption. In this context, it is necessary to improve and develop tools based on data analysis of households' meter readings. This study contributes to this goal by using a statistical methodology that detects abnormal water consumption patterns, namely, significant increases or decreases. This approach relies on a combination of methods that analyse billed water consumption time series. The first step is to decompose the time series using Seasonal-Trend decomposition based on Loess. Next, breakpoint analysis is performed on the seasonally adjusted time series to look for changes in the pattern over time. Afterwards, the Mann–Kendall test and Sen's slope estimator are applied to assess whether there are significant increases or decreases in water consumption. The strategy is applied to water consumption data from the Algarve, Portugal, successfully detecting breakpoints associated with significant increasing or decreasing trends.

Keywords Breakpoints · Time series decomposition · Trend analysis · Water consumption

A. Borges (✉)
CIICESI, ESTG, Politécnico do Porto, Porto, Portugal
e-mail: aib@estg.ipp.pt

Department of Natural and Exact Sciences, ESTG, Politécnico do Porto, Porto, Portugal

C. Cordeiro
Department of Mathematics, Faculty of Science and Technology, Universidade do Algarve, Faro, Portugal

M. Rosário Ramos
Department of Science and Technology, Universidade Aberta, Lisbon, Portugal

C. Cordeiro · M. Rosário Ramos
CEAUL, Faculty of Science, Universidade de Lisboa, Lisbon, Portugal

M. Rosário Ramos
CEG, Universidade Aberta, Lisbon, Portugal

1   Introduction

In the last decades, water has been recognised as an essential resource for guaranteeing economic development and maintaining living standards. Water stress makes it indispensable to acknowledge water as a scarce resource and to enhance focus on managing demand [1]. In the context of water scarcity, [2] alert to the importance of assessing losses in water distribution systems since the compensation of water losses increases water demand. Enhancing water use efficiency and conservation are priorities to ensure, for example, universal access to drinking water and reduce the population suffering from its scarcity.

Water companies' awareness for the responsible use of water has gained importance, with climate changes emphasising this need. The analysis of urban water consumption patterns and the estimation of the corresponding water demand are expected to be among the top priorities for water companies in the near future [3]. In this sense, controlling domestic water usage can help reduce both water consumption and protect the environment [4]. Therefore, investigating water consumption patterns will provide a better understanding at a household level. This will promote water use efficiency and help to reduce non-revenue water (NRW). Detecting an anomalous increase will allow companies to take measures, such as alerting their consumers to have sustainable behaviours.

The Portuguese region of Algarve is known for registering the highest values of water consumption [5]. This region faces an enormous challenge in optimising water management and usage standards due to long periods of drought. Consequently, water utilities feel the need to develop mechanisms for water planning based on data analysis. Overall, they concentrate their efforts on addressing the consequences of climate change. Therefore, managing Portuguese water resources is likely to become challenging due to the potential decrease of water availability and the increase of the seasonal hydrological asymmetries [6].

This paper presents an application of a procedure capable of detecting significant changes in a time series anchored on statistical methods. The aim is the assessment of abnormal increasing and decreasing trends in water consumption. The methodology is an extension of the work developed by [7] that detects significant decreasing trends in water consumption time series. This approach can be synthesised in four steps: the first step consists of time series decomposition using Seasonal-Trend decomposition based on Loess [8]; on the second step, a breakpoint analysis is performed on the seasonally adjusted time series; the third step consists of the search for decreasing or increasing changes in the periods between breakpoints through the Mann–Kendall [9] test, and Sen's [10] slope estimator. In the end, an indicator for the magnitude of change is presented. Monthly time series of billed water consumption from Loulé Municipality, located in Algarve—the southern region of Portugal, is used.

The paper is organised as follows: the Methodology section describes the statistical methods that underlie the procedure and how they are connected; the following section presents the data set used to exemplify the procedure; the results are detailed in the next section and it ends with the Conclusion and Future Work section.

2　Methodology

A time series is a set of consecutive observations indexed in time t, $t = 1, \cdots n$, during regular intervals. Often time series exhibit seasonal behaviour, and adequate "control" for a seasonal component is essential before using any statistical model. Also, the time series may exhibit patterns such as an upward or downward movement (trend). The irregular component is the remaining time series behaviour that is not attributed to trend or seasonality. Both trend and seasonality components are potential confounding features in analysis, so identification and removal are important.

The methodology is organised into four steps described below.

In the first step, the Seasonal-Trend decomposition procedure based on Loess (STL) [8] is applied to decompose each time series into a trend (T_t), seasonal (S_t) and irregular or residual (I_t) components using nonparametric regression. Assuming the additive model, the time series is decomposed into

$$Y_t = T_t + S_t + I_t, \tag{1}$$

where $t = 1, \cdots n$, is the time period and n its length. This method was chosen over other decomposition methods in the literature because it has attractive modelling features, such as the seasonal component being allowed to change over time and being robust in the presence of outliers. This procedure is available in the ℝ software through function $\texttt{stl()}$ [11]. However, this procedure requires a subjective selection of two smoothing parameters: the seasonal (*s.window*) and trend (*t.window*) window widths. Therefore, the algorithm used was proposed by [12], named as $\texttt{stl.fit()}$ [13], which overcomes this drawback. The latter selects the best STL model with the smallest error measure achieved with a specific combination of the smoothing parameters. In this study, the Mean Absolute Error (MAE) is used. For more details, see [12].

The second step consists of the detection of breakpoints in the seasonally adjusted time series of water consumption given by

$$Y_t^* = Y_t - S_t \tag{2}$$

$t = 1, \cdots, n$. The ℝ package *strucchange* [14] is used to obtain the breakpoints. This package features methods from the generalised fluctuation and F-test (Chow test) frameworks. That includes methods to fit, plot and test fluctuation processes (e.g. CUSUM, MOSUM, recursive/moving estimates and F-statistics, respectively). This procedure tests for structural changes in linear regression models, estimating the number of segments (m) and the set of the breakpoints $bp = \{ t_1^*, t_2^*, \cdots, t_{m-1}^* \}$, minimising the Bayesian information criterion and the residual sum of squares [15]. The present study uses the two expressions proposed in [7] for obtaining the minimum length between consecutive breaks (*min.h*) and the maximum number of breaks (*max.breaks*).

In a third step, the change identified in the previous procedure is submitted to a nonparametric analysis through Mann–Kendall (MK) test [9], and Theil–Sen's (TS) Slope [10]. The choice of these methods is linked to the fact that they can handle situations where the segments correspond to short periods of asymmetric distributions and allow assess of the underlying increase or decrease through robust methods [16]. If the result is significant, positive or negative, then the breakpoint adjacent to the segment is considered relevant. To obtain these statistics, the function sen.slope(), available in the **R** package *trend* [17] is used.

In the last step, the magnitude of the change in water consumption before and after the significant break is obtained by the Relative Magnitude of the Change (RMC) proposed by [7]. This indicator is a ratio that compares the water consumption pattern before and after a breakpoint as follows:

$$RMC = \frac{slp_{after} - slp_{before}}{|slp_{before}|}, \tag{3}$$

where slp_{after} and slp_{before} are the nonparametric Sen's slopes in the neighbourhood of a specific breakpoint $t_k^* \in bp^*$. Higher negative values of RMC represent a higher decreasing change in water consumption after the considered breakpoint. While high positive values of RMC represent a higher increase in water consumption after the considered breakpoint.

For more details about each step, see [7].

3 Data

The empirical analysis uses billed water consumption data from residential households (**RH**) from a municipality located in the Portuguese region of Algarve. The municipality occupies about 200 km^2 and has an estimated population of around 5,000 inhabitants. It is characterised by an elderly population and an agricultural-based economy.

Two case studies will be presented to exemplify the procedure. Both cases refer to household's monthly water consumption (m^3) from February 2011 until December 2017, registered by two water meters: **RH1** and **RH2**.

On meter **RH1**, the higher values of water consumption (see Fig. 1a) were registered in the summer months of July and August. This is consistent with a strong seasonal behaviour, with the higher temperatures justifying the need for higher water consumption, related to Algarve's tourism period. Moreover, a decrease in the trend until 2016 was followed by an increase more pronounced during 2017, as seen in Fig. 1. The latter might be explained by its replacement on 8 November 2016. The consumption registered by this meter showed a noticeable abrupt increase in 2018, reaching a value higher than 30 m^3.
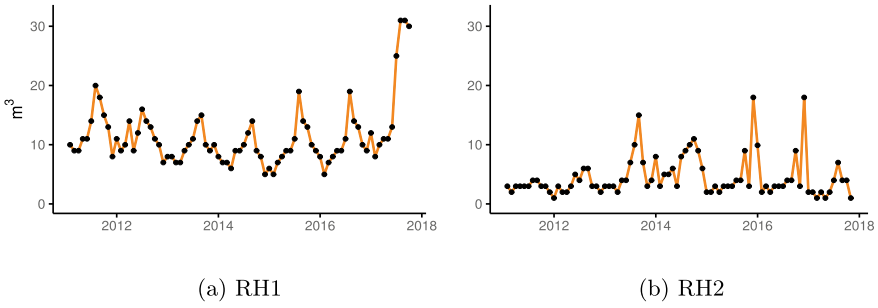
(a) RH1   (b) RH2

Fig. 1 Time series of water consumption

In contrast, meter **RH2** did not present a seasonal pattern as regular as in the previous case. In addition, it showed the highest values of water consumption in months such as October 2014, December 2016 and December 2017, as shown in Fig. 1b.

4   Results

The proposed strategy was applied to two case studies showing different water consumption patterns.

The first step was the decomposition of the water consumption time series into its components: trend, seasonal and remainder. STL [8] has already been successfully applied in studies of water consumption such as [18] and more recently [7]. Since STL is robust against outliers, the detection of these observations was done according to [8]. From Fig. 2, the robust approach of the STL was applied to **RH1** and **RH2**.

The `stl.fit()` proposed by [12] was applied, and the decomposition plots are shown in Fig. 3a and b. Note that this function searches for the best combination of the parameters (s.window and s.trend) minimising an error measure, which in this case
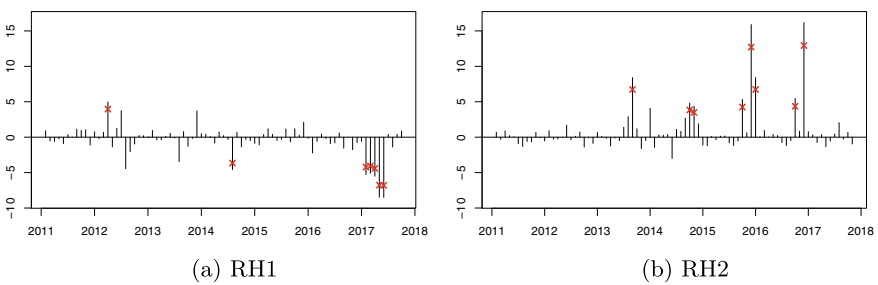

(a) RH1   (b) RH2

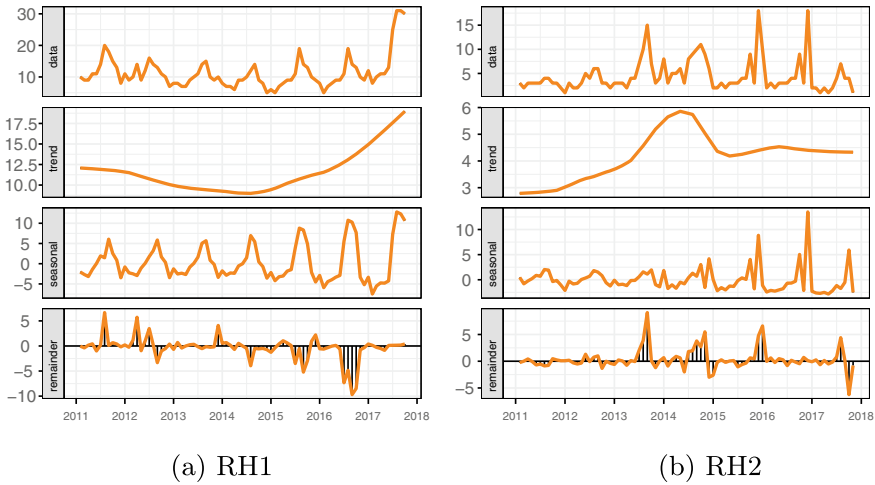**Fig. 2** Detecting outliers according to [8]

(a) RH1   (b) RH2

Fig. 3 Water consumption decomposition plots

Table 1 MAE results

RH	stl	stl.fit
1	1.33	1.25
2	1.57	1.09

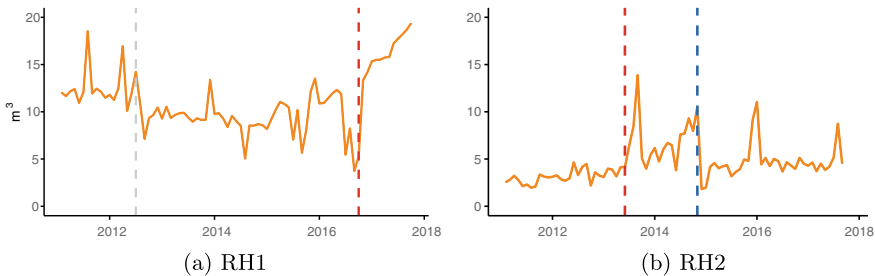(a) RH1   (b) RH2

Fig. 4 Breakpoints analysis

was the MAE. Table 2 presents the results of the `stl()`, with s.window="periodic"
(fixed seasonality), and the `stl.fit()` that search for the "best" combination, in
terms of MAE. Therefore, based on these results, the latter was chosen (Table 1).

The breakpoint algorithm was applied to the seasonally adjusted water consump-
tion, considering *min.h* = 0.15 (12.2 months) and *max.breaks* = 4. The relevance of
each breakpoint was detected through MK and TS methods that infer the significance
of the adjacent trends before and after the breakpoint.

For **RH1**, the procedure detected two breakpoints in water consumption in July
2012 and in October 2016 (Fig. 4a and Table 2). However, the estimated Theil–Sen's

Table 2 Breakpoints & Trend analysis

RH	Breakpoints	Theil–Sen's slopes			RMC
		Segment 1	Segment 2	Segment 3	
1	2012 (Jul), 2016 (Oct)	0.009 (0.880)	−0.016 (0.533)	**0.536 (<0.001)**	34.5
2	2013 (Jun), 2014 (Nov)	**0.052 (0.002)**	**0.236 (0.005)**	0.004 (0.744)	3.538, −0.983

Note 1 Statistically significant trend **slope (p-value)** in boldface
Note 2 All values are rounded with three decimal places

slopes of the segments, before and after the breakpoint in 2012, were not significant (p-values of 0.880 and 0.533, respectively). This means that the breakpoint was not considered as a relevant one by the procedure. Hence, the water utility should not be concerned with the pattern of water consumption of this household at that moment. Nonetheless, regarding the second breakpoint in October 2016, the segment slope estimated after it was positive (0.536) and statistically significant (p-value < 0.001). This represented an increased pattern in water consumption after that moment, with a magnitude RMC = 34.5.

Regarding **RH2**, the strategy implemented was able to detect two relevant breakpoints in June 2013 and November 2014 (Fig. 4b and Table 1). The estimated Theil–Sen's slope of the segment before the first breakpoint was positive with the low value of 0.052 (p-value = 0.0017), and the slope of the segment after the breakpoint was higher with the value of 0.236 (p-value = 0.0489). This represents an increase in water consumption pattern, with a magnitude of RMC = 3.538 (lower than the change in consumption of water meter **RH1**). In a deeper inspection, we can deduce that this result may be associated with a change in the mean value between the two periods (before and after the break). For the second breakpoint detected in November 2014, a decrease in water consumption was detected since the estimated segment slope before it was 0.236 and the estimated slope after the breakpoint was 0.004 and non-significant. Thus, the indicator RMC = –0.983, i.e. negative, as expected.

5 Conclusion

Trend and breakpoint analysis of water consumption—at an individual level—are important for decision-making in a broad sense, including in environmental, health and sustainability concerns.

This study presents an integrated statistical approach to analyse water consumption time series within the framework of water sustainability. The goal is to detect the moment (or moments) when a significant increase or decrease in consumption occurs.

Several studies have adopted time series decomposition and breakpoint detection methodologies, such as [7, 18–21]. The classical methods of decomposition of time

series allow identifying the trend, the seasonality and the irregular components. However, these methodologies do not allow for a flexible specification of the seasonal component, and the trend component is generally represented by a deterministic time function, which is easily affected by the existence of outliers. The nonparametric Seasonal-Trend decomposition by Loess can identify a seasonal component that changes over time, a non-linear trend, and it can be robust in the presence of outliers. Similar to all nonparametric regression methods, STL requires the subjective selection of smoothing parameters. The two main parameters are the seasonal (*s.window*) and trend (*t.window*) window widths. Therefore, to overcome this limitation, the `stl.fit()` procedure [12] was used since it allows an "objective" choice of STL smoothing parameters. This procedure has been developed to obtain "automatically" the seasonal and trend smoothing parameters by minimising an error measure.

After estimating the components using STL, the seasonality was removed. Afterwards, a combination of methods was applied on the seasonally adjusted time series to detect breakpoints, using the algorithm implemented in the ⓡ package *strucchange* [14]. Subsequently, statistically significant decreasing and increasing segments of water consumption were analysed using robust nonparametric methods such as MK and TS.

Additionally, a water consumption change indicator, the RMC, was calculated. The RMC is unitless, allowing to compare different types of consumption. It allows the water utility to understand and compare consumption patterns between households (or other buildings) and between different periods. The water company can also choose threshold values for RMC, at which the consumption is considered problematic. Overall, the idea is to quantify the change of the decrease or increase in water consumption for each consumer and identify which ones the water company should investigate.

This strategy was applied to real data of billed water consumption from households located in the municipality of Loulé, characterised as an agriculturally based economy region with tourism activity. The methodology successfully detected breakpoints linked to a significant increase or decrease in water consumption. Moreover, the difference in household water consumption patterns justifies the importance of implementing a procedure at an individual level able to capture consumption specificities.

The detection of an abnormal increase will allow the water utility to alert its consumers of less environmentally sustainable behaviour. This is important since the impacts of climate change on water demand may be particularly relevant in the case of agricultural water use. In fact, the water needs for crop production increase as a consequence [6].

In conclusion, this integrated strategy may also contribute to the assessment of losses in water distribution systems as well as apparent losses and NRW. Furthermore, the application of the methodology is not limited to the time series of water consumption. The flexibility of the procedure allows, in each step, to regulate parameters such as the seasonal and trend windows in STL decomposition, the minimum length (*min.h*) of the segment and the maximum number of breaks (*max.breaks*).

In the future, an alternative approach could be a prior application of time series clustering on households' water consumption to group the consumers by pattern similarity. Afterwards, the proposed strategy would be implemented only to the most problematic consumer profiles.

References

1. Cámara, Á., Llop, M.: Defining sustainability in an input-output model: an application to spanish water use. Water **13**(1), 1 (2021)
2. Oviedo-Ocaña, E., Dominguez, I., Celis, J., Blanco, L., Cotes, I., et al.: Water-loss management under data scarcity: case study in a small municipality in a developing country. J. Water Resour. Plan. Man. **146**(3), 05020001 (2020)
3. Ioannou, A.E., Creaco, E.F., Laspidou, C.S.: Exploring the effectiveness of clustering algorithms for capturing water consumption behavior at household level. Sustainability **13**(5), 2603 (2021)
4. Abid, I., Khattak, H.A., Khan, R.W.A.: Water conservation and environmental sustainability approach. In: Collaboration and Integration in Construction, Engineering, Management and Technology, pp. 485–490. Springer, Berlin (2021)
5. Reynaud, A.: Modelling household water demand in Europe. Insights From a Cross-country Econometric Analysis of EU, vol. 28 (2015)
6. Ferreira, J.P.L., Vieira, J.M., et al.: Water in Celtic Countries: Quantity, Quality and Climate Variability. IAHS Press, Wallingford (2007)
7. Cordeiro, C., Borges, A., Ramos, M.R.: A strategy to assess water meter performance. J. Water Resour. Plan. Manag. **148**(2), 05021027 (2022)
8. Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I.: STL: a seasonal-trend decomposition. J. Off. Stat. **6**(1), 3–73 (1990)
9. Kendall, M.: Rank Correlation Methods. Griffin, London (1975)
10. Sen, P.K.: Estimates of the regression coefficient based on Kendall's tau. J. Amer. Stat. Assoc. **63**(324), 1379–1389 (1968)
11. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2021)
12. Cristina, S., Cordeiro, C., Lavender, S., Costa Goela, P., Icely, J., et al.: Meris phytoplankton time series products from the SW Iberian Peninsula (Sagres) using seasonal-trend decomposition based on loess. Remote Sens. **8**(6), 449 (2016)
13. Cordeiro, C.: `stl.fit()`: Function developed in Cristina et al. (2016)
14. Zeileis, A., Leisch, F., Hornik, K., Kleiber, C.: `strucchange`: an R package for testing for structural change in linear regression models. J. Stat. Softw. **7**(2), 1–38 (2002)
15. Zeileis, A., Kleiber, C., Krämer, W., Hornik, K.: Testing and dating of structural changes in practice. Comput. Stat. & Data Anal. **44**, 109–123 (2003)
16. Helsel, D.R., Hirsch, R.M.: Statistical Methods in Water Resources, vol. 323. US Geological Survey Reston, VA (2002)
17. Pohlert, T.: trend: Non-Parametric Trend Tests and Change-Point Detection (2020). R package version 1.1.4

18. Hester, C.M., Larson, K.L.: Time-series analysis of water demands in three North Carolina cities. J. Water Resour. Plan. Manag. **142**(8), 05016005 (2016)
19. Gelažanskas, L., Gamage, K.A.: Forecasting hot water consumption in residential houses. Energies **8**(11), 12702–12717 (2015)
20. Quesnel, K.J., Ajami, N.K.: Changes in water consumption linked to heavy news media coverage of extreme climatic events. Sci. Adv. **3**(10), e1700784 (2017)
21. Ohana-Levi, N., Munitz, S., Ben-Gal, A., Schwartz, A., Peeters, A., et al.: Multiseasonal grapevine water consumption-drivers and forecasting. Agric. For. Meteorol. **280**, 107796 (2020)