



Detection of Anomalous Patterns in Water Consumption: An Overview of Approaches

José Carlos Carrasco-Jiménez^{1(✉)}, Filippo Baldaro^{2(✉)},
and Fernando Cucchietti^{1(✉)}

¹ Barcelona Supercomputing Center, Barcelona, Spain
{jose.carrasco,fernando.cucchietti}@bsc.es

² CETAQUA Water Technology Center, Barcelona, Spain
fabaldaro@cetaqua.com

http://ris3catutilities.com/eng/index_ris3cat.html

Abstract. The water distribution system constantly aims at improving and efficiently distributing water to the city. Thus, understanding the nature of irregularities that may interrupt or exacerbate the service is at the core of their business model. The detection of technical and non-technical losses allows water companies to improve the sustainability and affordability of the service. Anomaly detection in water consumption is at present a challenging task. Manual inspection of data is tedious and requires a large workforce. Fortunately, the sector may benefit from automatized and intelligent workflows to reduce the amount of time required to identify abnormal water consumption. The aim of this research work is to develop a methodology to detect anomalies and irregular patterns of water consumption. We propose the use of algorithms of different nature that approach the problem of anomaly detection from different perspectives that go from searching deviations from typical behavior to identification of anomalous pattern changes in prolonged periods of time. The experiments reveal that different approaches to the problem of anomaly detection provide complementary clues to contextualize household water consumption. In addition, all the information extracted from each approach can be used in conjunction to provide insights for decision-making.

Keywords: Anomaly detection · Water consumption · Time series · Decision making

1 Introduction

Late payment of bills as well as non-technical losses (NTL) derived from commercial fraud has increased in most European countries, leading to a critical point in the financial management of operators in the utilities sector.

NTL in the utilities sector leads to the necessity of improving the efficiency of the process of detecting anomalous consumption and the location of non-technical losses. Identifying the origin of such losses allows companies to offer new payment solutions that are customized to the needs of the customers, guaranteeing the economic sustainability of the service as well as the affordability for those with economic difficulties.

Anomaly detection in water consumption is at present a challenging task. Data is generally dispersed and estimated readings and manual readings taken bimonthly or quarterly, usually result in difficult scenarios in which the task of classifying the type of anomaly is hard to predict. For the water company understanding the nature of the anomaly is of avid importance. For example, if an anomalous consumption is due to a broken pipe or a leakage, the result of classifying the anomaly as a technical loss leads to the quick repair of the equipment. On the other hand, if the anomaly is due to an intended manipulation at the water meter, a corrective action can be taken.

The deployment of smart-meters and the development of algorithms to process the data generated by these devices opens up new opportunities in the rapid identification of anomalous consumptions. This allows the operator to take action through the application of search criteria and the selection of possible cases of water pipe breakdown or irregular consumption, leading to savings in the cost of operations and an optimal, as well as, an efficient supply of water. As a consequence, the number of irregular consumptions is expected to decrease.

The manual identification of anomalous consumptions is conventionally done as follows:

1. identify water bills that have values out of the expected range given the region and the type of customer (e.g. industrial, residential);
2. prioritize the verification given a number of rules;
3. verify the service points that have been marked as possible anomalies due to identification of anomalous readings;
4. take a sample of the bills of a given user that is suspected of presenting anomalous consumption;
5. request a visual inspection of those pieces of equipment that are likely to show anomalous behavior (either due to technical or non-technical issues).

The process of determining anomalous behaviors in water consumption can be a tedious business given the large amount of data that needs to be processed and the large volume of clients. The cost of manually identifying possible anomalies can be diminished if the workflow can be automatized and supplied with intelligent algorithms. The development of new techniques to detect anomalous behaviors in water consumption has a direct impact in the business model of water companies. Among the benefits we can find targeted inspections with higher probabilities of identifying technical or non-technical issues, identification of metrological inefficiencies (e.g. manipulated meters), and identification of misclassified clients (i.e. incorrect type of tariff).

The aim of this research work is to develop a methodology to detect anomalies and irregular behavior from water consumption time series. This work is limited to the clues that can be exploited from a time series, considering relevant domain knowledge such as the types of clues that indicate that a consumption is potentially out of the normal behavior including point anomalies and pattern anomalies found in the time series.

The remainder of this paper is organized as follows. Section 2 describes the background information and Sect. 3 describes the related works. Sect. 4 details the methodology, including a description of the dataset and the data processing workflow. Sect. 5 describes the fundamental model in which manual inspection is based. Furthermore, Sect. 6 elaborates on the analysis of point anomalies found in time series of water consumption while Sect. 7 presents the result of anomalous behavior detection algorithms. We conclude with a discussion of the most notable findings and future work in Sect. 9.

2 Background

In this section we describe the background information required to contextualize the problem at hand. We start by defining in a concise manner how the water distribution network works. An explanation of definition and possible cause of anomalies in household water consumption.

2.1 Water Balance of a Drinking Water Distribution System

In the water sector, Unregistered Water (UW) indicates water that has been *produced*, i.e. processed in a water treatment plant and supplied into the distribution network, and is lost before reaching the end clients. Furthermore, Registered Water (RW) or Billed Water (BW) indicates the amount of water supplied and registered by the automated meter reading devices deployed at the residential level or municipal distribution point.

The International Water Association (IWA) proposes categories to evaluate the different components of UW [1]:

- real or physical losses corresponding mainly to leakage in the transportation network, the reservoir, the distribution network, or end users
- apparent or commercial losses including Authorized Consumption Unregistered (ACU), fraud or erroneous readings

The ACU consumptions correspond to the water used in the operations of network maintenance, or agreements with the public sector including firefighters or park irrigation. Figure 1 summarizes the water balance of a drinking water distribution system from the utilities perspective. The dotted-circle represents the part of the distribution system where technical and non-technical losses are to be inferred.

Frauds are defined as irregular connections to the distribution network, non-regulated discharge, or manipulation of smart-meters with the aim of benefiting

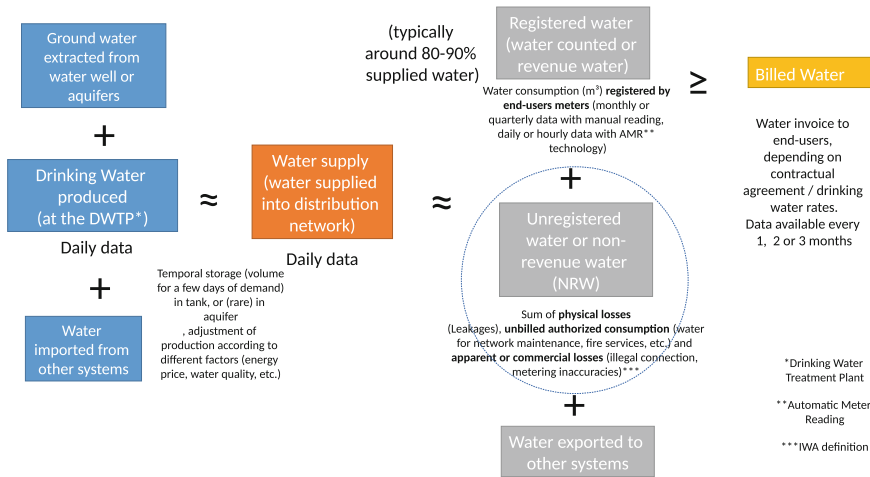


Fig. 1. Water balance of a drinking water distribution system (simplified, water utilities perspective).

with lower charges. Moreover, a distinct usage from that contracted fare, usually applying a lower fare than that required.

Lastly, lack of precision in the metering devices can be due to poor installations, underestimated volumes of water leading to smaller devices than needed, leakage, or aged devices. We can quantify the losses due to fraud or underestimation through the mean monetary value of the water service in Spain. The mean price in Spain is 2,24 €/m³ from which €1,23 correspond to the supply service, €0,32 stand for sewage system, and €0,69 for depuration. Using these reference values we can estimate the commercial loss of the operators due to fraud or underestimation, leading to a value of approximately 730 million €/year.

2.2 Anomalies

In [2], anomalies are defined as patterns that do not conform to the normal expected behavior. As a consequence, we need to identify which consumptions are to be defined as normal.

The anomalies associated to the consumption of water can be categorized as follows:

1. data transmission errors,
2. change in the behavior of water consumption,
3. fraud or manipulation of the smart-meters.

Furthermore, associating a given pattern to a deviation from the expected behavior of water consumption remains a challenging task. In other words, defining what the normal behavior looks like depend much on the context, not only of a single user but of the patterns exhibited in the region at hand.

[2] defines a few challenges that raise from the sheer definition of anomaly. The first thing the authors identify as challenging is defining what is the normal expected behavior. In the case of water consumption, normal behaviors tend to evolve and expose a cyclical tendency leading to different definitions of *normality*, understood as more frequent. In this research work, we deal with time series that exhibit patterns that may be suggested as deviation from the normal (i.e. more frequent) consumption when considered outside of its context. Nevertheless, when the patterns are analyzed inside their context, those punctual patterns can be categorized as normal behavior. In other words, anomalous behavior is determined within the limits of a specific context (i.e. region and fare).

3 Related Work

The notion of anomaly has several applications in the water sector. Given the nature of the data, different challenges arise and a variety of anomaly detection techniques apply in different scenarios. Anomaly detection serves different purposes that go from general water analytics to more specific applications such as water quality, identification of changes in household consumption, and detection of technical and non-technical losses in the water distribution network.

In [3], the authors propose a constrained-clustering-based approach for anomaly detection in real-world monitoring time series, a task that cannot be manually executed or analyzed given the large number of customers managed, even by small operators. Similar applications have been reported for hydrological time series in which abnormal patterns play an important role in the analysis and decision making [4].

With regard to water quality, operators are interested in understanding deviations from what they consider *normal* values of quality. The search for deviations is considered a problem of anomaly detection. Identifying anomalous points in drinking-water quality data may lead to improved quality. Recent works [5,6] seek to improve water quality by identifying anomalies that need to be eliminated or avoided.

Determining anomalous behaviors in water consumption is also a major task in water analytics. For example, in [7], changes in household water consumption are identified from water usage readings with the aim of detecting anomalies that may lead to the discovery of manipulated meters. Similarly, [8,9] propose two different approaches for determining anomalous behaviors in water consumption: punctual anomalous consumption and anomalous changes in behavior of water consumption, respectively.

Other interesting challenges in which the detection of anomalies may shed some light is the identification of technical and non-technical losses, which is of avid concern for water operators. For instance, [10] describes the implementation of a methodology for the automated detection of pipe bursts and other events that induce similar abnormal pressure/flow variations (technical losses). Similarly, in [11], authors use machine learning to identify malfunctioning meters and leaks (considered technical losses), as well as fraud (considered non-technical

loss). In this respect, [12] proposes the use of an anomaly detection approach based on unsupervised clustering algorithms employed to identify non-technical losses. A more recent work combines anomaly detection techniques with computer vision strategies to detect fraud in water meters, such is the case of [13], in which water meter seals are analyzed through morphological image processing. Then pattern recognition techniques serve to identify suspected irregularities and anomalies in water meters.

4 Methods

In this section we describe the methodology proposed to reach the objectives of this research work. We start by describing the dataset, followed by a description of the data processing.

4.1 Dataset Description

The main source of information for the detection of anomalous behavior in water consumption comes from two types of databases: 1) *Customer Information System (CIS)* and 2) *Customer Relationship Management (CRM)*, both are commercial management systems.

The smart-meters, also known as *Automated Meter Reading (AMR)*, collect and transmit data at different rates, although data is transmitted generally on a daily basis.

The dataset comprises over 1.460.000 records, totaling more than 4000 customers from two different regions over a period of one year. Customers are assigned a tariff based on the average volume of water consumed. Different regions may contain different types of fares depending on the kind of profiles present in each region. Part of the analysis of this work will make use of the region and type of tariff assigned to each client, in order to understand the customer’s behavior with respect to the typical consumption behavior of the region and tariff.

For this work, data has been previously anonymized and pre-processed by the water company. Each record collected contains three pieces of information as summarized in Table 1.

Table 1. Summary of data collected for each recording.

Data	Description	Example
CUSTOMER	Anonymized customer ID	AC11UA487082
DATE	Date in the format <i>yyyy-mm-dd</i>	2014-01-11
VOLUME	Amount of water consumed (liters)	176

Given the data structure described in Table 1, the sequence of consumption can be reconstructed to show the evolution through time. The construction of the time series is the foundation to discover anomalies.

4.2 Data Processing

The readings transmitted from the smart-meters follow three sequential phases:

1. the transmitter reads the datum from the automated meter reading every hour (or every six hours depending on the configuration of the AMR),
2. the data is sent to the CIS on a daily basis in chunks of 4 or 24 data per transmitter,
3. the data is centralized in the Master Data Management (MDM).

The data collected from the smart-meters are indices of consumption that are later transformed into consumption by the difference of two indices. In other words, the consumption is computed as the difference between the index registered at the beginning and the index registered at the end of the reading cycle.

The following step after the transformation from indices into consumption is the construction of time series, which are later fed into the algorithms developed in this research work.

The anomaly detection algorithms depend on the context to which the labels are available. In Sect. 1 we defined the steps executed in a manual inspection of potential customers who present anomalous behaviors in water consumption. As a consequence some customers have been manually analyzed but the number of labeled samples remains insufficient to use supervised anomaly detection.

On the other hand, unsupervised anomaly detection algorithms are widely applicable to situations where we lack labeled instances. Unsupervised techniques assume that those instances that are more frequent in a dataset can be considered as normal instances, leading to the identification of instances that deviate from normality, also called anomalies.

5 Deviation from Expected Normal Water Consumption Levels

In Sect. 4.2, we described how the data is transformed from its raw format into time series of water consumption. The sequence of consumption points is the input of the algorithms analyzed in this work.

An anomaly can be defined as a pattern that does not conform to the expected normal behavior [2]. In the scenario at hand, an anomaly is a deviation from the expected normal water consumption levels. A first approach to understanding anomalous levels of water consumption is to consider the time series of water consumption of a client and analyze it with respect to the context of the customer in accordance with the definitions in Sect. 2.2.

In other words, a segmentation of the dataset based on similar environmental and social characteristics, as well as similar tariffs, is needed. In our first approach we characterize the detection of deviation from expected normal water consumption levels as described in Algorithm 1.

The pseudocode described starts by splitting the dataset by region and tariff. The next step requires the estimation of the mean consumption of each group

Algorithm 1: Procedure to detect deviations from expected normal water consumption levels.

Input: threshold=3
Output: anomalies
Data: waterConsumptionData

```

1 groups ← splitByRegionAndFare(Data)
2 for group ← groups do
3   groupMean ← mean(group)
4   groupSd ← sd(group)
5   for c ← group do
6     upperThreshold ← groupMean + threshold * groupSd
7     lowerThreshold ← groupMean - threshold * groupSd
8     anomalies ← which( c ≥ upperThreshold || c ≤ lowerThreshold )

```

of customers as well as the standard deviation of each group. Assuming we have established a threshold to discriminate typical from anomalous values, which is commonly three standard deviations from the mean to be considered anomalous, we proceed to identify which points lie above the upper threshold and below the lower threshold. The aforementioned algorithm was applied to every customer in the database used for this research work. The resulting anomalies for a sample of customers is shown in Fig. 2.

For instance, in Fig. 2, the vertical bars show the deviations from the mean water consumption for the region and tariff to which the customer belongs. The vertical bars that lie above the zero line are the instances of time in which the water consumption is above the mean consumption of the region and those that lie below are the instances in which the consumption is below the mean consumption.

The light green rectangle drawn on each of the plots are interpreted as the maximum and minimum permitted deviations from the expected normal water consumption levels. In other words, vertical bars that lie outside the light green rectangle are to be considered as anomalous consumptions.

Consider the customer shown in Fig. 2 (c) which belongs to *region 2* and is assigned a tariff of the type *city*. As it is observed, most of the instances of time, the customer showed what would be classified as normal behavior of water consumption. On the other hand, the rest of customer in the sample exhibit many instances in which the consumption was anomalous.

In the next sections we will propose other complementary methods to detect anomalies. The information that can be extracted from this type of anomaly detection algorithm will serve to feed, and complement other sources of information, a meta-model proposed as future work.

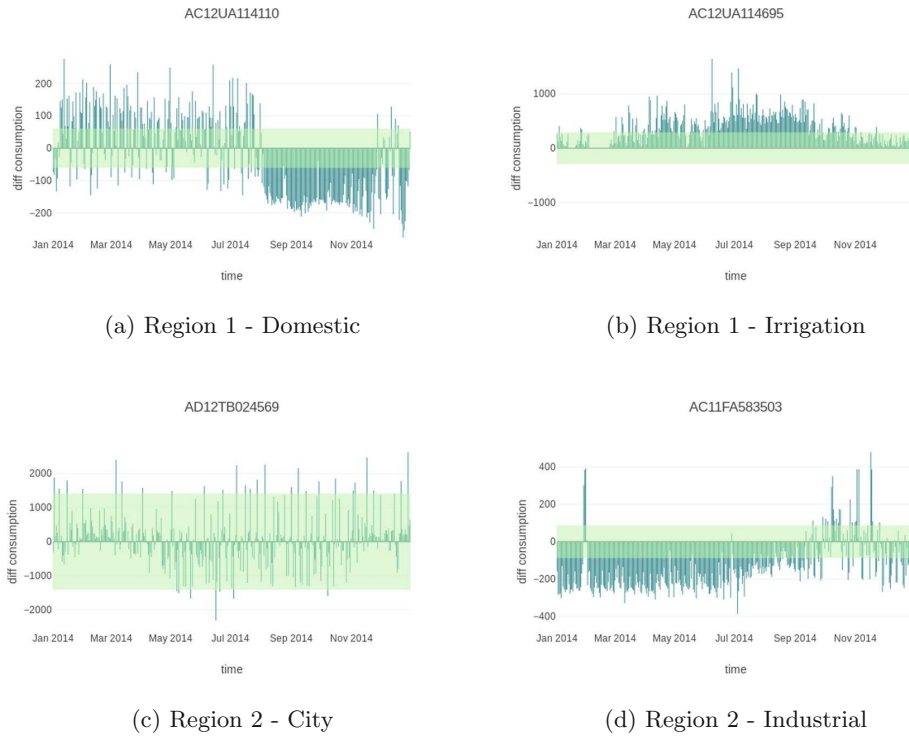


Fig. 2. Difference from mean consumption. It shows the comparison between the consumption of a customer with respect to the mean consumption of the region and fare to which the customer belongs. A sample of customers from different regions and fares are shown.

6 Punctual Anomalous Consumption

Punctual anomalies refer to those deviations that happen at a given timestamp regardless of the prior and posterior values. This type of anomaly refers to values that are either too high or too low with respect to the rest of the time series, and occur at single points in time.

In this part of the work we analyzed three different methods for anomaly detection in a time series, two of which are based on STL decomposition (mean and median) [14] and one based on the Seasonal Hybrid ESD Test [15].

The three techniques provide different number of anomalies. Figure 3 shows the Mean-STL Decomposition, Median-STL Decomposition and Seasonal Hybrid ESD Test, from top to bottom respectively, of customer identified as AC11UA487082. The results show that Median-STL Decomposition approach identifies too many points as anomalous, a tendency shown by many customers, while Seasonal Hybrid ESD Test exhibits a more moderate solution as can be visually inspected in the aforementioned figure.

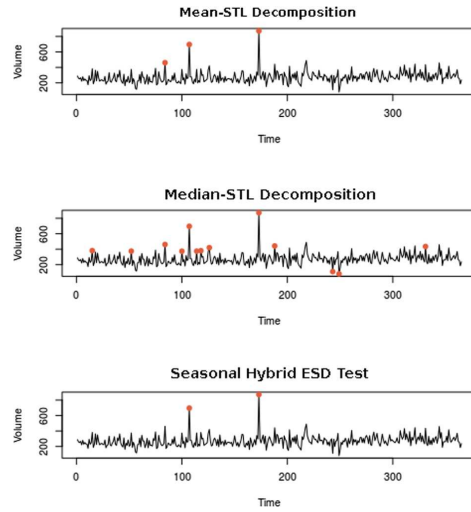


Fig. 3. Three different methods for point anomaly detection have been tested. Different techniques identify different number of anomalies for customer AC11UA487082.

On the other hand, in Fig. 4, Seasonal Hybrid ESD Test shows the worst viable solution given the large amount of anomalous points detected. Note that the Mean-STL Decomposition remains the most trustworthy.

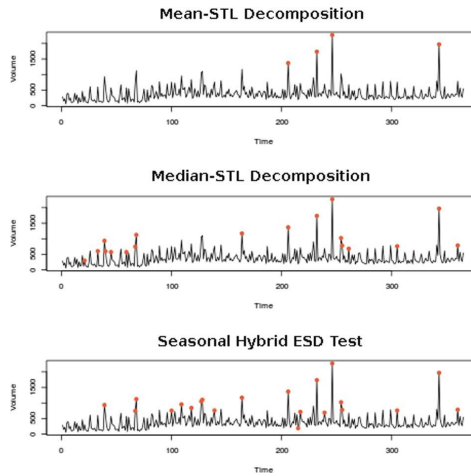


Fig. 4. Three different methods for point anomaly detection have been tested. Different techniques identify different number of anomalies for customer AC12UA114190.

In order to provide a better approximation to the most viable anomaly detection technique, an evaluation on all the customers was performed. Figure 5 shows the performance of each algorithm in terms of the number of anomalies detected

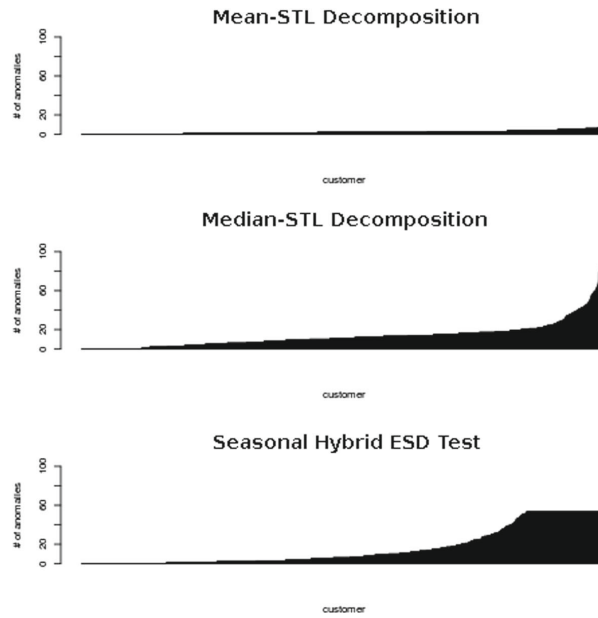


Fig. 5. The three different methods studied exhibit different sensitivity. Mean-STL Decomposition is more robust to small changes as defined by the smaller number of anomalies detected.

for each customer. If we look carefully at the distribution of anomalies we note that Mean-STL Decomposition has a lower number of anomalies in general.

Although, in general, Mean-STL Decomposition shows a more feasible solution, it would be interesting to know in which situations each technique is more viable.

7 Anomalous Pattern Changes in Household Water Consumption

As it was described in Sect. 3, changes in household water consumption are of avid importance for water operators. Point anomaly detection provides information about single instances of time in which an abnormal value is identified. But that is not the whole story. Another approach to anomaly detection is the identification of portentous changes in patterns of household water consumption. That is, changes in the patterns of consumption as opposed to single point anomalies. This approach is rather concerned with significant changes in prolonged periods of time, avoiding the raise of alarms when a single point has gone out of range. Although the approaches described in Sects. 5 and 6 are of different nature, the three approaches described in this paper do not compete against each other,

rather they can be used as complementary methods that perform on different scenarios or provide information about anomalies from different perspectives.

By understanding how medium to long-term behavior of water consumption, some non-technical losses can be detected. Take for example the case of an individual whose mean consumption drops drastically and remains stable for prolonged periods of time. In that case, point anomalies could not have detected a possible fraud. On the other hand, significant changes in the tendency may also indicate a technical loss when the mean consumption increases significantly.

In order to detect anomalous changes in household water consumption, we propose the use of algorithms for the detection of multiple change points in multivariate time series. Change point detection can be seen as a model selection problem that consists of three elements, as defined in [16]:

1. cost function (e.g. maximum likelihood estimation, kernel-based detection),
2. search method (optimal, approximate),
3. constrains (i.e. penalty functions).

The most crucial part is the selection of search method. The method depends largely on two pieces of information: precision desired and whether or not the number of change points are known. With this in mind, we selected the Pruned Exact Linear Time (PELT) algorithm [17] which works for unknown number of change points and has better performance than the naive approach which proves to be computational cumbersome because of the quadratic complexity. Furthermore, the cost function is related to the type of change to detect. In our case, we used a non-parametric cost function, namely Mahalanobis metric.

Figure 6 shows the resulting segments extracted after applying the change point detection algorithm. The left side of the figure, i.e. Fig. 6 (a) and (c) display the temporal evolution of water consumption for two customer from different regions and types of tariffs. The bottom bars with alternating colors (gray and red) show the segments with different behaviors.

The corresponding figures on the right hand side of Fig. 6 show the mean consumption of a given customer at each segment. The order of the bars correspond to the order of the bottom bars of the left hand side of the figure. As we can see, the segments exhibit different behaviors, showing changes in household water consumption. To measure the significance of the changes observed, we need to consider two pieces of information: 1) mean consumption difference among the segments detected by the algorithm and 2) length of each segment. As part of the future work we will measure statistical significance from one segment to the other. This will allow us to identify possible cases with higher accuracy. Future work will also include a study to measure the minimum length of each segment to be defined as independent from the neighboring segments.

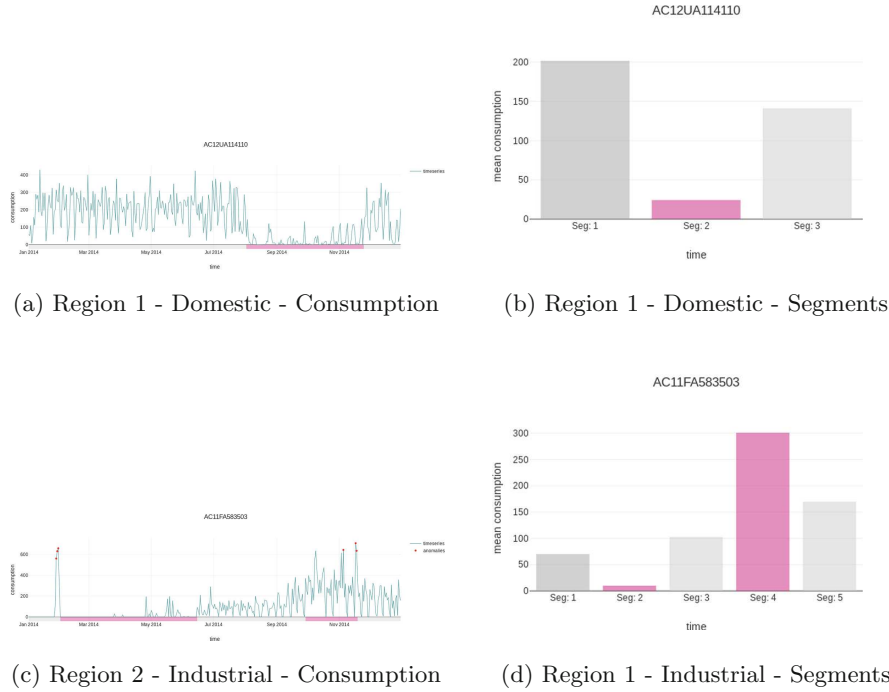


Fig. 6. Water consumption of customers. The red dots indicate days in which anomalous consumptions were identified.

8 Discussion and Limitations

The algorithms presented in this work are of different nature and have distinct capabilities. Each method approaches the contextualization of water consumption from different perspectives, emphasizing diverse aspects of the consumption. Although distinct, they offer complementary clues to contextualize household water consumption, leading to two broad categories of anomaly definitions, namely, those that are considered anomalous relative to the patterns of the region and tariff, and those that are tagged as anomalous when compared against their own consumption patterns. Those anomalies that are defined as such when a consumption exhibits deviations from expected normal water consumption levels of a given region and tariff fall in the first category. In this case, the context is defined by the behavior of water consumption of other customers. On the other hand, in punctual anomalies and anomalous pattern changes in household consumption, the context of the anomalies is dictated by past and future consumption of the same user, and does not depend on the behavior of others.

As it is explained in Sect. 9, the different approaches provide complementary clues to contextualize household water consumption and serve as the basis for a framework that is currently being designed. In fact, the methodology of this work sets the foundation of a broader goal, which consists in identifying households suspected of anomalous water use either due to technical or non-technical issues.

The scope of this work is limited to finding anomalous consumptions. It does not aim at interpreting the cause of each anomaly. Furthermore, each method provides different clues that may suggest whether the household presents anomalous patterns of water consumption due to technical issues or non-technical flaws, but identifying specific causes will be studied as part of the future work.

9 Conclusion and Future Work

Technical and non-technical problems derive in the loss of large quantities of resources that have a direct impact on the water supply chain. As it was stated in Sect. 2.2 the commercial loss can be estimated as much as 730 million €/year. Avoiding technical and non-technical losses remains a challenging task. Manual inspection of individual cases is a tedious task due to the large volumes of data. As a consequence, the water sector can benefit from an automatized workflow leading to targeted inspections with higher probabilities of identifying technical or non-technical issues.

In this work we developed a methodology to detect anomalous household water consumption and irregular patterns of consumption that may fall into fraudulent activities. The methodology consists of three approaches of different nature and capabilities. In the first part we implemented an algorithm to detect deviations from typical behavior. Although this piece of information can be useful, different perspectives can provide a richer set of conclusions. The second approach is defined as point anomaly detection in which punctual anomalies are detected regardless of the prolongation in time. Furthermore, to complement the previous approaches, a method to identify anomalous pattern changes in household water consumption was considered.

The different approaches provide complementary clues to contextualize household water consumption and serve as the basis for a framework that is currently being designed. Future work includes the development of a meta-model capable of classifying the type of anomaly, thus identifying whether the anomaly is due to a technical flaw or a non-technical issue. Other data sources will be incorporated including trend and seasonality of the demand of water in the region, weather information, holidays, reports of manual inspection of smart-meters, the state of the water distribution network, and authorized non-registered consumption (such as park irrigation).

Acknowledgment. This research work is cofunded by the European Regional Development Fund (FEDER) under the FEDER Catalonia Operative Programme 2014–2020 as part of the R+D Project from RIS3CAT Utilities 4.0 Community with reference code COMRDI16-1-0057.

References

1. Lambert, A., Hirner, W.: Losses from water supply systems: standard terminology and recommended performance measures, Report P-13, International Water Association (2000)

2. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. *ACM Comput. Surv.* **41**, 1–58 (2009)
3. Vercruyssen, V., Meert, W., Verbruggen, G., Maes, K., Baumer, R., Davis, J.: Semi-supervised anomaly detection with an application to water analytics. In: *IEEE International Conference on Data Mining (ICDM)*, pp. 527–536, November 2018
4. Sun, J., Lou, Y., Ye, F.: Research on anomaly pattern detection in hydrological time series. In: *14th Web Information Systems and Applications Conference (WISA)*, pp. 38–43, November 2017
5. Dogo, E.M., Nwulu, N.I., Twala, B., Aigbavboa, C.: A survey of machine learning methods applied to anomaly detection on drinking-water quality data. *Urban Water J.* **16**(3), 235–248 (2019)
6. Muharemi, F., Logofătu, D., Leon, F.: Machine learning approaches for anomaly detection of water quality on a real-world data set. *J. Inf. Telecommun.* **3**(3), 294–307 (2019)
7. Quinn, S., Murphy, N., Smeaton, A.F.: Tracking human behavioural consistency by analysing periodicity of household water consumption. In: *Proceedings of the 2019 2nd International Conference on Sensors, Signal and Image Processing, SSIP 2019*, New York, NY, USA, pp. 1–5. Association for Computing Machinery (2019)
8. González-Vidal, A., Cuenca-Jara, J., Skarmeta, A.F.: IoT for water management: towards intelligent anomaly detection. In: *IEEE 5th World Forum on Internet of Things (WF-IoT)*, pp. 858–863, April 2019
9. Christodoulou, S.E., Kourti, E., Agathokleous, A.: Waterloss detection in water distribution networks using wavelet change-point detection. *Water Resources Manage.* **31**, 979–994 (2017)
10. Romano, M., Kapelan, Z., Savić, D.A.: Automated detection of pipe bursts and other events in water distribution systems. *J. Water Resour. Plann. Manage.* **140**(4), 457–467 (2014)
11. Kermany, E., Mazzawi, H., Baras, D., Naveh, Y., Michaelis, H.: Analysis of advanced meter infrastructure data of water consumption in apartment buildings. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013*, New York, NY, USA, pp. 1159–1167. Association for Computing Machinery (2013)
12. Júnior, L.A.P., Ramos, C.C.O., Rodrigues, D., Pereira, D.R., de Souza, A.N., da Costa, K.A.P., Papa, J.P.: Unsupervised non-technical losses identification through optimum-path forest. *Electr. Power Syst. Res.* **140**, 413–423 (2016)
13. Detroz, J.P., da Silva, A.T.: Fraud detection in water meters using pattern recognition techniques. In: *Proceedings of the Symposium on Applied Computing, SAC 2017*, New York, NY, USA, pp. 77–82. Association for Computing Machinery (2017)
14. Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I.: STL: a seasonal-trend decomposition procedure based on loess. *J. Off. Stat.* **6**, 3–37 (1990)
15. Vallis, O., Hochenbaum, J., Kejariwal, A.: A novel technique for long-term anomaly detection in the cloud. In: *Proceedings of the 6th USENIX Conference on Hot Topics in Cloud Computing, HotCloud 2014*, USA, p. 15. USENIX Association (2014)
16. Truong, C., Oudre, L., Vayatis, N.: Selective review of offline change point detection methods. *Sig. Process.* **167**, 107299 (2020)
17. Wambui, G.D., Waititu, G.A., Wanjoya, A.K.: The power of the pruned exact linear time(pelt) test in multiple changepoint detection. *Am. J. Theor. Appl. Stat.* **4**(6), 581–586 (2015)