European Water 58: 429-434, 2017.

Waterloss detection in streaming water flow timeseries using change-point anomaly methods

S.E. Christodoulou[1,2*], E. Kourti[1] and A. Agathokleous[1]
[1] *Nireas International Water Research Center, University of Cyprus, 1 Panepistimiou Avenue, Aglantzia, Cyprus*
[2] *Department of Civil and Environmental Engineering, University of Cyprus, 1 Panepistimiou Avenue, Aglantzia, Cyprus*
[*] *e-mail: schristo@ucy.ac.cy*

Abstract: The work described herein evaluates the utilization of a change-point method for the detection of anomalies in water consumption time series, and its applicability to waterloss detection in water distribution networks. The change-point detection method is based on the Relative unconstrained Least-Squares Importance Fitting (RuLSIF) algorithm and was implemented on a three-month-long hourly water consumption signal, successfully detecting unusual reductions and increases in the water consumption patterns, and classifying them as anomalies. The first water consumption anomaly type related to a discontinuity in the signal (a break in the consumer's water consumption patterns), whereas the second type related to an unusual increase in the signal (two waterloss incidents). Unlike traditional approaches that monitor the average current periodic consumption and compare this average to the one from a corresponding past period, the RuLSIF change-point detection approach is dynamic and sensitive to any change in the time series in study. Even though the proposed analysis does not predict future anomalies, it is suitable for past and near-real-time anomaly detection; an attribute that is sufficient for waterloss management as it allows for a timely detection of anomalies is streaming water flow data. Further, the method dynamically assigns anomaly scores to the detected changes in the signal, thus easing waterloss detection and appraising the severity of each detected incident. The change-point detection method is able to extract knowledge from past consumption patterns (the water flow signal over time) to detect anomalies in the signal, both in the past and in near-real-time, and for both physical and apparent water losses.

Key words: Waterloss detection; Water-flow anomaly detection; Real-time time-series analysis

1. INTRODUCTION

As water distribution networks (WDNs) worldwide are rapidly degrading and damages in them are increasing, the need for efficient assessment and real-time monitoring is increasingly becoming pressing. The damages in WDNs are most often manifested as pipe-failure incidents and are leading to significant levels of non-revenue water (typically in the range 20%-30%). While data on pipe-failures can, post-event, be analyzed and provide useful knowledge on the causalities of the failures, records of these incidents are usually hard to produce, are inconsistent and poorly recorded, and most often only span over short periods of time. Thus, the need for timely detection of any anomalies in the WDNs and the correlation of such anomalies to waterloss, has become of outmost importance to the efficient operations of WDNs.

Anomalies in WDN operations usually manifest themselves in the form of pipe-failure incidents (pipe breaks) and in the form of water loss. While pipe-failures are not a perfect indicator of WDNs, they are amongst the primary symptoms of network degradation and one of the few sources of information on the condition of WDNs. However, recording pipe-failure incidents is a relatively recent practice. Furthermore, knowledge of the type of pipes and their condition over an area is often unknown prior to a pipe-failure incident which requires repair.

Past efforts had led to the identification of several risk factors influencing the fragility of WDNs and devising decision support systems for sustainably managing them. For example, Christodoulou and Deligianni (2009) used data mining techniques and neurofuzzy systems in order to identify pipe-failure data clusters and convert these into decision support rules for waterloss management. Subsequently, Christodoulou and Agathokleous (2012) studied the effects of intermittent water supply on the vulnerability of WDNs, reporting that intermittent water supply policies resulted in increased stresses in the network they were applied to that eventually lead to an abnormally high number of pipe-failure incidents and higher volumes of water losses. The spatial distribution of pipe-failure incidents was studied by Gagatsis et al. (2012) who, first used spatial analysis techniques and a spatial clustering method to study the extend of pipe-failure clustering in the Limassol WDN and, then, transformed the deduced knowledge into a decision support system.

To date, pipe-failure incident records are usually the result of customer complaints (e.g. low pressure, no water, high bills, etc.) which are then investigated by Water Board technicians. However, all of these sources of information often require a certain amount of time before the source of a leak is discovered thus resulting in significant water losses until the damage is repaired. In order to resolve this problem, the use of automatic meter reading (AMR) is being advocated and is currently implemented widely.

Automatic meter readers are devices embedded in house connection meters which can provide readings in frequent intervals. These readings are transferred to a monitoring server. By examining unusual rises or drops in water consumption, water board technicians can be made aware of potential leaks and investigate them, thus resulting in a faster, more efficient leak detection method. Furthermore, the use of AMR can provide near real-time information on the water consumption within a district metered area (DMA) and compare it with the total water input in that DMA. A DMA is an area of the WDN where water originates and is distributed to the network from a single point of entry and where the network's water pressure is approximately uniform. Large differences between periodic readings indicate large leaks within the network.

In principle, the use of AMR devices permits Water Boards to detect water leaks before they are reported by customers. This is particularly useful in cases when customers are away on vacation or in cases when the leak is not detectable by the customers. The AMR devices send consumption information in frequent intervals to the central database. This data is then analyzed for detecting abnormal increases in water consumption readings sent from the AMR units. When such an abnormal increase is detected an alert is sent to the WDN operators of possible WDN malfunction, and then the WDN operators are tasked with linking the event with a possible water leakage.

This process is in many aspects subjective (in terms of how the anomaly thresholds are set) and manual (thus time consuming) with regards to the analysis of indicators (outliers) of water consumption anomalies. Further, the architecture shown in Figure 1 relies heavily on the wide use of AMR devices in the WDN (one for each house connection) and thus it is not cost-efficient.

A solution could be the selective location of AMR devices (or sensors) within the WDN, for which several sensor placement optimization techniques exist (Ostfeld et al. 2008, Christodoulou et al. 2013, 2015). The problem, though, of analyzing the sensed operational data of a WDN still remains, in terms of processing time-related data and sifting through this data for possible anomalies in the WDN's behavior.

A recent work on the dynamic analysis of time series related to water consumption was recently reported upon by Christodoulou et al. (2016), in which wavelet change-point detection classifiers were utilized for identifying anomalies in the consumption patterns. The wavelet change-point method utilizes the continuous wavelet transform (CWT) of time-series (signals) to analyze how the frequency content of a signal changes over time. In the case of water distribution networks the time-series relates to streaming water consumption data from automatic meter reading (AMR) devices, at either the individual consumers' level or at an aggregated district meter area (DMA) level. The wavelet change-point detection method analyzes the provided time-series to acquire inherent knowledge on water consumption under normal conditions at household or area-wide levels, to then make inferences about water consumption under abnormal conditions.

The paper proposes an approach to the timely identification of anomalies in WDN by use of a change-point anomaly detection algorithm and of streaming water consumption data from consumers and from district meter areas (DMA), in the form of 'automatic meter reading' devices (AMR). The anomaly detection approach is then coupled with spatio-temporal analysis to arrive at spatial decision support systems (DSS) that automate the process of waterloss detection, reduce water loss and increase the efficiency in the management WDN, and eventually positively contribute to the sustainability of such networks.

2. CHANGE-POINT DETECTION

Change-point detection refers to the identification of both whether or not a change (or several changes) has occurred in a time series (signal), and of the times of any such changes. Common change-point detection methods relate to changes in the mean, variance, correlation, density, or slope of the signal in study.

Among the several change-point detection methods found in literature, notable is the approach proposed by Yamada et al. (2013) and Liu et al. (2013), who presented a statistical change-point detection algorithm based on non-parametric divergence estimation between time-series samples from two retrospective segments. The algorithm uses the relative Pearson (PE) divergence as a divergence measure and a method of direct density-ratio estimation. Mathematically, the PE divergence from *p(x)* to *p'(x)* is defined as

$$PE[p(x), p'(x)] \coloneqq \frac{1}{2}\int \left(\frac{p(x)}{p'(x)} - 1\right)^2 p'(x)dx \tag{1}$$

and, in their work, Yamada et al. (2013) and Liu et al. (2013) proposed an approach to distribution comparison called α-relative divergence estimation, by which they estimated the α-relative divergence (defined as the divergence from *p(x)* to the α-mixture density $\alpha p(x) + [1 - ap'(x)]$ for $0 \leq \alpha < 1$. By use of Eq. 1, the α-relative Pearson (PE) divergence is given by

$$PE_a[p(x), p'(x)] \coloneqq \frac{1}{2}\int \left(\frac{p(x)}{ap(x)+(1-a)p'(x)} - 1\right)^2 [ap(x) + (1-\alpha)p'(x)]dx \tag{2}$$

and the α-relative divergence was estimated by the researchers by direct approximation of the α-relative density ratio given by

$$r_\alpha(x) \coloneqq \frac{p(x)}{ap(x)+(1-a)p'(x)} \tag{3}$$

Yamada et al. (2013) also showed that using an estimator of the α-relative density-ratio $r_\alpha(x)$, we can construct several estimators of the α-relative PE divergence, of which the following one is easy to compute:

$$\widetilde{PE}_a \coloneqq \frac{1}{2n}\sum_{i=1}^{n} \hat{r}\,(x_i) - \frac{1}{2} \tag{4}$$

In its final form, and by using an estimator $\hat{g}(Y)$ of the α-relative density ratio, the RuLSIF method approximates the α-relative PE divergence by use of the following equation (Yamada et al., 2013):

$$\widetilde{PE}_a \coloneqq -\frac{\alpha}{2n}\sum_{i=1}^{n} \hat{g}\,(Y_i)^2 - \frac{1-a}{2n}\sum_{i=1}^{n} \hat{g}\left(Y'_j\right)^2 + \frac{1}{n}\sum_{i=1}^{n} \hat{g}\left(Y'_j\right) - \frac{1}{2} \tag{5}$$

where, *y(t)* is the time-series sample at time *t*, *Y(t)* is the subsequence of *k* timeseries samples at time *t*, and *n* is the number of points in the analysis window.

3. CASE STUDY IMPLEMENTATION, ANALYSIS AND DISCUSSION

The RuLSIF change-point detection method is tested on a three-month-long hourly timeseries of a water flow signal (approximately 2,200 hours), depicting the water consumption of a 3-person household, with an average total daily water consumption of about 419 L (based on records from a local Water Board), distributed hourly across the day as per Figure 1. Water consumption starts in the early morning hours (around 06:00), increases peaking up at around 10:00, then drops until the early afternoon hours (16:00), peaking up again in the late afternoon and early evening hours (18:00 - 21:00), before dying down at night (21:00 - 06:00). It should be also noted that the household's daily water consumption does not, at any point in time, zero out. This is in agreement with the 'minimum night flow (MNF)' concept, commonly used in WDN operations. MNF is a common method used to evaluate water loss in a water network, and refers to the water volume flowing through the network even when all true water demand is zero (typically in the time band of 02:00 - 04:30).

Further to the profiled daily water consumption, three induced anomalies in the household's consumption are recorded (as shown in Figure 1): (1) a three-day-long drop in consumption ($t \in [680, 720]$ hours); (2) a 1-L/hour water loss for a duration of 48 hours ($t \in [510, 558]$ hours); and (3) a 2-L/hour water loss for a duration of 48 hours ($t \in [990, 1038]$ hours). The water consumption's time series (Figure 1) is first processed macroscopically to identify the time periods of concern, and then microscopically to zoom in on possible consumption anomalies.

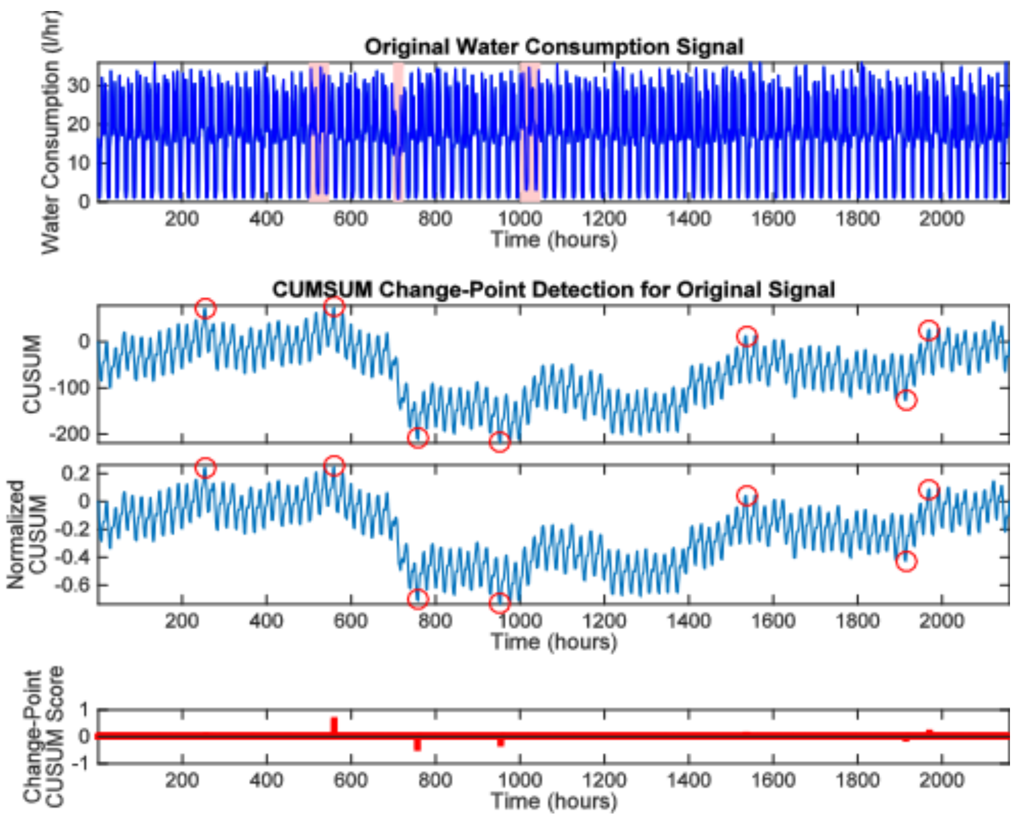


Figure 1. Water consumption signal for the period in study (approximately 2200 hours) and the CUSUM-based anomaly detection.

As a point of reference and subsequent assessment of the RuLSIF approach, the water consumption signal is first analyzed for anomalies by use of the cumulative sum (CUSUM) method, which is a well-known anomaly detection method. CUSUM is a statistical control chart used to track the variation of a process, able to detect small shifts in the process' mean. The CUSUM anomaly detection method is a statistical method for anomaly detection in univariate time series, using the out of control signals of the CUSUM charts to locate anomalous points. In effect, CUSUM works by tracking the individual cumulative sums of the negative and positive deviations from the mean (the high and low sums respectively), and by thresholding them against a predefined level of tolerance.

The water consumption signal with the three induced anomalies highlighted and the CUSUM-based anomaly detection for this signal are as depicted in Figure 1. As shown, the CUSUM method correctly identifies the time periods of anomalies in the signal, but it also gives several false-positive alarms.

We then turn our attention to the RuLSIF method and its suitability to water-flow anomaly detection, first investigating the abnormal drop (Figure 2) and then the abnormal increase in water flow (Figures 3 and 4).

Figure 2 shows the signal for the period in vicinity to the observed water-flow drop (t ∈ [680, 720] hours), and the corresponding RuLSIF-based anomaly detection. As depicted, the observed drop in water flow is correctly identified by the RuLSIF change-point detection method (with a slight time-shift in the detection as a result of the parameters used in the computation), and high anomaly scores are assigned to the approximate start and end points of the corresponding period.

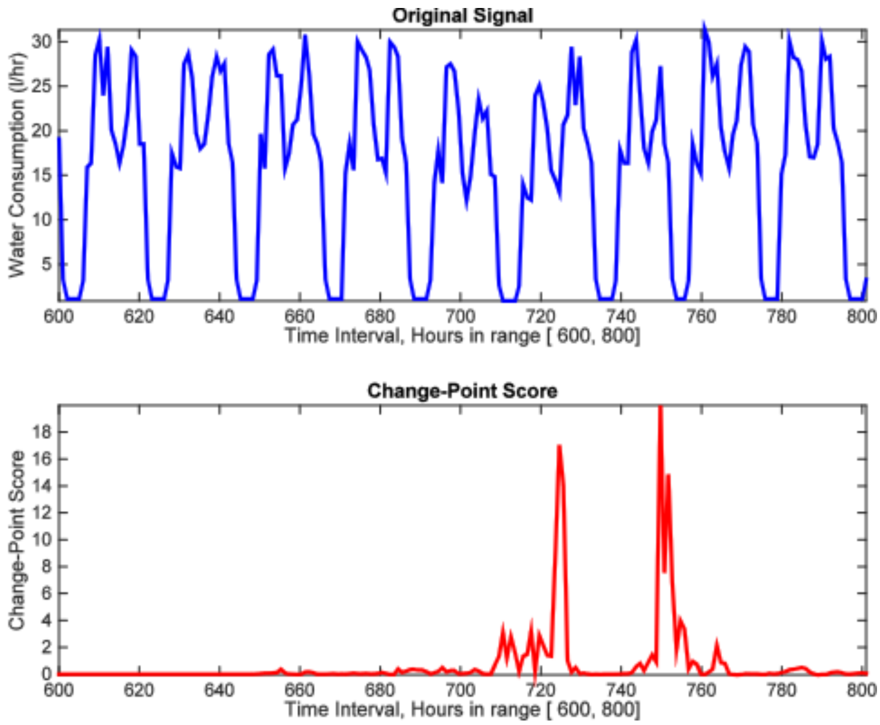
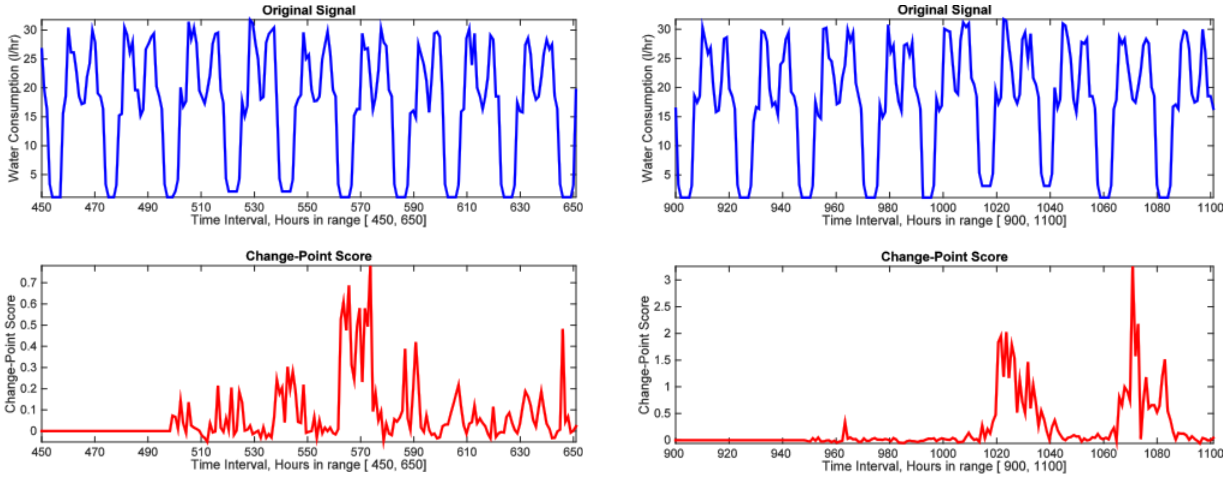


Figure 2. Water consumption signal for the period t ∈ [600, 800] hours, and the RuLSIF-based change-point detection of a drop in water consumption.

The method performs equally well in detecting the sudden increase in water flow values (case of water loss), for both the 1-L/hour and the 2-L/hour anomalies (Figures 3a and 3b, respectively). Further, the computed anomaly scores can be used as a measure of the extent of the observed anomalies. In our case study, the change-point score for the 1-L/hour water loss is lower than the corresponding score for the 2-L/hour water loss (~0.75 vs. ~3.2, respectively).

4. CONCLUSION

The work described herein evaluates the utilization of the relative unconstrained least-squares importance fitting (RuLSIF) change-point method for the detection of anomalies in water consumption time series, and its applicability to water-loss detection in water distribution networks. The RuLSIF change-point detection method was implemented on an hourly water consumption signal of about 2200 hours duration, successfully detecting unusual reductions and increases in the water consumption patterns, and classifying them as anomalies. The first water consumption anomaly type related to a discontinuity in the signal (a break in the consumer's water consumption patterns), whereas the second type related to an unusual increase in the signal (two water-loss incidents).



(a) (b)

Figure 3. Water consumption signal for the periods t ∈ [450, 650] and t ∈ [900, 1100] hours, and the RuLSIF-based change-point detection of an increase in water consumption; Case of (a) 1-L/hour water loss and (b) 2-L/hour water loss.

Unlike traditional approaches that monitor the average current periodic consumption and compare this average to the one from a corresponding past period, the RuLSIF change-point detection approach is dynamic and sensitive to any change in the time series in study. Even though the proposed analysis does not predict future anomalies, it is suitable for past and near-real-time anomaly detection; an attribute that is sufficient for water-loss management as it allows for a timely detection of anomalies is streaming water flow data. Further, the method dynamically assigns anomaly scores to the detected changes in the signal, thus easing water-loss detection and appraising the severity of each detected incident.

Ongoing research work on wavelet change-point detection will address spatial mapping and heat-map generation for the identification of areas of concern in a distribution network, as well as the blending of automatic meter reading (AMR) technologies and water auditing at a district meter area (DMA) level, for the dynamic (and near-real-time) detection of water leaks in an urban water distribution network.

REFERENCES

Christodoulou, S., 2015. Smarting up water distribution networks with an entropy-based optimal sensor placement strategy. Journal of Smart Cities; 1(1): 47-58.

Christodoulou, S., Agathokleous, A., 2012. A study on the effects of intermittent water supply on the vulnerability of urban water distribution networks. Journal of Water Science and Technology; 12(4): 523-530

Christodoulou, S., Deligianni, A., 2009. A neurofuzzy decision framework for the management of water distribution networks. Water Resources and Management; 24(1): 139-156.

Christodoulou, S., Gagatsis, A., Xanthos, S., Kranioti, S., Agathokleous, A., Fragiadakis, M., 2013. Entropy-based sensor placement optimization for waterloss detection in water distribution networks. Water Resources Management; 27(13): 4443-4468.

Christodoulou, S.E., Kourti, E., Agathokleous, A., 2016. Water Resources Management; doi:10.1007/s11269-016-1558-5.

Gagatsis, A., Kranioti, S., Christodoulou, S., Agathokleous, A., Xanthos, S., 2012. An integrated software solution for identifying, monitoring and visualizing water leak incidents in Water Distribution Networks. Proc., IWA's Intern. Conf. on New Developments in IT & Water, 4-6 November 2012, Amsterdam, Netherlands.

Liu, S., Yamada, M., Collier, N., Sugiyama, M., 2013. Change-point detection in time-series data by relative density-ratio estimation. Neural Networks; 43: 72-83.

Ostfeld, A., Uber, J.G., Salomons, E., Berry, J.W., Hart, W.E., Phillips, C.A., et al. 2008. The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms. Water Resources Planning and Management; 134: 556-569.

Yamada, M., Suzuki, T., Kanamori, T., Hachiya, H., Sugiyama, M., 2013. Relative density-ratio estimation for robust distribution comparison. Neural Computation; 25(5): 1324–1370.