A Hierarchy of Change-Point Methods for Estimating the Time Instant of Leakages in Water Distribution Networks

Giacomo Boracchi, Vicenç Puig, and Manuel Roveri*

Giacomo Boracchi and Manuel Roveri are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy; Vicenç Puig is with the Department of Automatic Control, Universitat Politècnica de Catalunya, Barcelona, Spain.
{giacomo.boracchi,manuel.roveri}@polimi.it
vicenc.puig@upc.edu

Abstract. Leakages are a relevant issue in water distribution networks with severe effects on costs and water savings. While there are several solutions for detecting leakages by analyzing of the minimum night flow and the pressure inside manageable areas of the network (DMAs), the problem of estimating the time-instant when the leak occurred has been much less considered. However, an estimate of the leakage time-instant is useful for the diagnosis operations, as it may clarify the leak causes. We here address this problem by combining two change-point methods (CPMs) in a hierarchy: at first, a CPM analyses the minimum night flow providing an estimate of the day when the leakage started. Such an estimate is then refined by a second CPM, which analyzes the residuals between the pressure measurements and a network model in a neighborhood of the estimated leakage day. The proposed approach was tested on data from a DMA of a big European city, both on artificially injected and real leakages. Results show the feasibility of the proposed solution, also when leakages are very small.

Keywords: Leakages in Water Distribution Networks, Change-Point Estimation, Nonstationarity Detection

1 Introduction

Water losses in distribution drinking water networks are an issue of great concern for water utilities, strongly linked with operational costs and water resources savings. Continuous improvements in water losses management are being applied and new technologies are developed to achieve higher levels of efficiency.

Among the wide range of water losses, we consider leakages, a specific type of hydraulic fault, that may be due to pipe breaks, loose joints and fittings, and overflowing water from storage tanks. Some of these problems are caused by the deterioration of the water delivery infrastructure, which is affected by ageing effects and high pressures. Leakages are classified by water utilities as *background* (small undetectable leaks for

* This research has been funded by the European Commission's 7th Framework Program, under grant Agreement INSFO-ICT-270428 (iSense).

which no action to repair is taken), *unreported* (moderate flow-rates which accumulate gradually and require eventual attention), and *reported* (high flow-rates which require immediate attention). In practice, there may be a significant time delay between the time instant when a leakage occurs, when the water utility detects the leakage, and when the leakage is located and repaired [1].

The traditional approach to leakage control is a passive one, whereby the leak is repaired only when it becomes visible. Recently developed acoustic instruments [2] allow to locate also invisible leaks, but unfortunately, their application over a large-scale water network is very expensive and time-consuming. A viable solution is to divide the network into District Metered Area (DMA), where the *flow* and the *pressure* are measured [3, 1], and to maintain a permanent leakage control-system: leakages in fact increase the flow and decrease the pressure measurements at the DMA entrance. Various empirical studies [4, 5] propose mathematical models to describe the leakage flow with respect to the pressure at the leakage location.

Best practice in the analysis of DMA flows consists in estimating the leakage when the flow is minimum. This typically occurs at night, when customers' demand is low and the leakage component is at its largest percentage over the flow [1]. Therefore, practitioners monitor the DMA or groups of DMAs for detecting (and then repairing) leakages by analyzing the minimum night flow, and also employ techniques to estimate the leakage level [1]. However, leakage detection may not be easy, because of unpredictable variations in consumer demands and measurement noise, as well as long-term trends and seasonal effects. Complementary to the minimum flow analysis, pressure loggers at the DMA entrance provide useful information for leak detection and isolation [6]. When a leakage appears in a DMA, the pressure at junctions typically changes, showing the key evidence for the leakage and providing information for its isolation.

In this paper, we address the problem of estimating the time-instant when a leakage has occurred within a DMA. Obtaining accurate estimates of leak time-instant is important, as this information improves the leak-diagnosis operations – including quantifying the leakages effect and understanding the leak causes – as well as the accommodation operations. Peculiarity of the proposed solution is to combine two change-point methods (CPMs) in a hierarchical manner. At first, a CPM analyzes the minimum night flow in the DMA to estimate the day when the leak has occurred. Then, within a range of this specific day, the residuals between the pressure measurements and a network model are analyzed by a second CPM, which estimates the time-instant when the leak has started. Such a coarse-to-fine analysis prevents the use of network models over large time-intervals, where these may be not accurate because of the large dimension of the network. The leak-detection problem is not addressed here, as this can be managed by specific techniques, such as [7]. To illustrate the feasibility of the proposed approach, real data – and a real leakage – coming from a DMA of a big European city are considered.

The structure of the paper is the following: Section 2, states the problem of estimating the leak-time instant in a DMA, while Section 3 presents the hierarchy of CPMs to address this problem. Section 4 presents the results obtained on real data from a DMA. Finally, conclusions are drawn in Section 5.

2 Problem Statement

We focus on a single DMA, connected to the main water supply network through a reduced number of pipes, where the flows and pressure inlets, as well as the pressure of some internal nodes, are recorded. In fault-free conditions, the flow measurements represent the water consumption and follow a *trend* that is stationary, though unknown. After the leakage-time instant T^*, the flow follows a trend affected by a faulty profile. We assume that the leakage induces an abrupt and permanent fault, so that the measured flow at the inflows of the DMA becomes

$$f(t) = \begin{cases} f_0(t), & \text{if } t < T^* \\ f_0(t) + \Delta_f, & \text{if } t \geq T^* \end{cases}, \tag{1}$$

where $\Delta_f > 0$ is the offset corresponding to the leakage magnitude. Similarly, the behavior of the measured pressure can be defined:

$$p(t) = \begin{cases} p_0(t), & \text{if } t < T^* \\ p_0(t) - \Delta_p, & \text{if } t \geq T^* \end{cases}, \tag{2}$$

where $\Delta_p > 0$ is the offset representing the leakage affect the pressure measurements.

Let $\widehat{T}$ be the time instant in which the leakage is detected by a suitable detection method, our goal is to identify T^* by analyzing

$$F = \{f(t), 0 \leq t \leq \widehat{T}\} \text{ and } P = \{p(t), 0 \leq t \leq \widehat{T}\} \tag{3}$$

Thus, the leak is assumed to be constant within the time interval $[T^*, \widehat{T}]$, and we do not consider measurements after $\widehat{T}$ when the accommodation procedures start.

3 A Hierarchy of CPMs to Estimate Leak Time-Instant

Change-Point Methods are hypothesis tests designed to analyze, in an offline manner, whether a given data sequence X contains i.i.d. realizations of a random variable (i.e., null hypothesis) or a change-point that separates X in two subsequences generated from different distributions (alternative hypothesis). Interest reader can refer to [8–10].

In this paper, we illustrate the use of CPMs for estimating T^*, the time instant when a leakage occurs, and we show that, to this purpose, it is convenient to combine two CPMs in a hierarchical scheme yielding coarse-to-fine estimates. At first, a CPM analyses the statistical behavior of the minimum night flows: interestingly, when no leakage affects the network, the minimum night flows are expected to be stationary [7]. Thus, minimum night flow values can be (at least approximatively) modeled as i.i.d. realizations of a random variable. The analysis of the minimum night flow by means of a CPM provides us M_Φ, an estimate of the day when the leakage started, as described in Section 3.1. This is a coarse-grained estimate that can be refined by analyzing the pressure measurements in the few days before and after M_Φ. Unfortunately, CPMs cannot be straightforwardly used on pressure measurements, as these are time-dependent data,

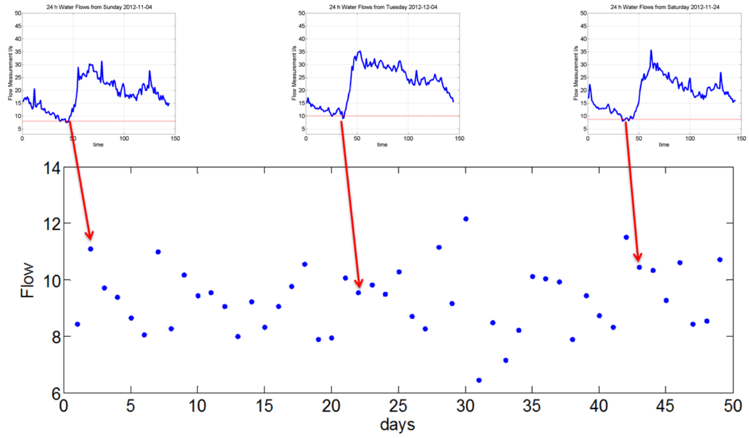

Fig. 1. An illustrative example of the sequence Φ. The values of Φ are displayed in the plot below, and correspond to the minimum night flow of each specific day. The small plots above represent the daily flows: the value of the minimum night flow is plot as an horizontal (red) line.

while CPMs operates on i.i.d. realizations of a random variable. To address this problem we run a CPM on residuals of approximating models, as in [11]. Specifically, in Section 3.2 we apply a CPM on pressure residuals measuring the discrepancy between the measurements and estimates provided by a DMA model, see [6]. The use of the ensemble of CPMs [11] instead of conventional CPMs to compensate temporal dependencies in the minimum night flow and in pressure residuals is discussed in Section 3.3

3.1 CPM on the Minimum Daily Flow

Let us denote by $\phi(\tau)$ the minimum night flow of the day τ, which may be computed as the minimum value or the average flow in a neighborhood of the minimum. Given the flow measurements F in (3) we compute

$$\Phi = \{\phi(\tau), 0 \leq \tau \leq \widehat{\tau}\}, \tag{4}$$

the sequence of the minimum night flows, being $\widehat{\tau}$ the day containing $\widehat{T}$. We say that Φ contains a change-point at τ^* if $\phi(\tau)$ is distributed as

$$\phi(\tau) \sim \begin{cases} \mathcal{P}_0, & \text{if } 0 \leq \tau < \tau^* \\ \mathcal{P}_1, & \text{if } \tau^* \leq \tau \leq \widehat{\tau} \end{cases}, \tag{5}$$

where $\mathcal{P}_0$ and $\mathcal{P}_1$ represent the distribution of the minimum night flow without and with a leakage, respectively, and τ^* and $\widehat{\tau}$ are the day when the leak occurred and when the leakage has been detected, respectively.

Within the CPM framework, the null hypothesis consists in assuming that all data in Φ are i.i.d., and when the null hypothesis is rejected the CPM provides also an estimate of the change point τ^*, which here corresponds to an estimate of day when the leak has occurred. From the practical point of view, when running a CPM, each time instant $S \in \{1, \dots, \widehat{\tau}\}$ is considered as a candidate change point of Φ, which is accordingly partitioned in two non-overlapping sets

$$\mathcal{A}_S = \{\phi(\tau),\ \tau = 1, \dots, S\},\ \text{and}\ \mathcal{B}_S = \{\phi(\tau),\ \tau = S+1, \dots, \widehat{\tau}\},$$

that are then contrasted by means of a suitable test statistic $\mathcal{T}$. The test statistic

$$\mathcal{T}_S = \mathcal{T}(\mathcal{A}_S, \mathcal{B}_S), \tag{6}$$

measures the degree of dissimilarity between A_S and B_S. Among test statistics commonly used in CPMs, we mention the Mann-Withney [8] (to compare the mean over A_S and B_S), the Mood [12] (to compare the variance over A_S and B_S) and the Lepage (to compare both the mean and variance over A_S and B_S). Other statistics in [13, 14, 10, 15].

The values of $\mathcal{T}_S$ are computed for all the possible partitioning of Φ, yielding $\{\mathcal{T}_S, S = 1, \dots, \widehat{T}\}$; let $\mathcal{T}_{M_\Phi}$ denote the maximum value of the test statistic, i.e.,

$$\mathcal{T}_{M_\Phi} = \max_{S=1,\dots,\widehat{T}} (\mathcal{T}_S)\,. \tag{7}$$

Then, $\mathcal{T}_{M_\Phi}$ is compared with a predefined threshold $h_{l,\alpha}$, which depends on the statistic $\mathcal{T}$, the cardinality l of Φ, and a defined confidence level α that sets the percentage of type I errors (i.e., false positives) of the hypothesis test. When $\mathcal{T}_{M_\Phi}$ exceeds $h_{l,\alpha}$, the CPM rejects the null hypothesis, and Φ is claimed to contain a change point at

$$M_\Phi = \underset{S=1,\dots,\widehat{T}}{\operatorname{argmax}} (\mathcal{T}_S)\,. \tag{8}$$

On the contrary, when $\mathcal{T}_{M_\Phi} < h_{l,\alpha}$, there is not enough statistical evidence to reject the null hypothesis, and the sequence is considered to be stationary. Summarizing, the outcome of a CPM to estimate the day when the leak occurred is

$$\begin{cases} \text{The leak occurred at day } M_\Phi & \text{if } \mathcal{T}_{M_\Phi} \geq h_{l,\alpha} \\ \text{No leak can be found in } \Phi, & \text{if } \mathcal{T}_{M_\Phi} < h_{l,\alpha} \end{cases}. \tag{9}$$

3.2 CPM on the Residuals of the Pressure Measurements

A CPM executed on Φ provides an estimate M_Φ of $\widehat{\tau}$, the day when the leakage started. To provide an fine-grained estimate of the leakage time-instant, we analyze the pressure measurements in a neighborhood of M_Φ (e.g., one or two days before and after M_Φ). Let $P_{M_\Phi} \subset P$ collects such pressure measurements. It is worth noting that, differently from values in Φ that can be assumed to be i.i.d., the pressure measurements follows their own dynamics, hence CPMs cannot be directly applied to P_{M_Φ}. We address this issue as in [11], and we compute the residuals between the measured pressure and its estimates using the DMA mathematical model f_θ

$$\widehat{p}(t) = f_\theta(p(t-1), \dots, p(t-n_x), u(t), u(t-1), \dots u(t-n_u))\ , \tag{10}$$

where θ is the model parameter vector, $p(t) \in \mathbb{R}$ and $u(t) \in \mathbb{R}^m$ represent the output and input of the DMA model, respectively. The parameters $n_x \geq 0$ and $n_u \geq 0$ set the order of the output and input, respectively. The DMA model is based on the hydraulic laws describing the flow balance in the DMA nodes and the pressure drop in the pipes. This model leads to a set of non-linear equations with non-explicit solution that must be solved numerically using a water network simulator (EPANET), as it is done in [6], to obtain $\widehat{p}(t)$.

The second CPM of the hierarchy is executed on the residual sequence that refers to pressure measurements in P_{M_Φ}:

$$\mathcal{R}_P = \{p(t) - \widehat{p}(t), T_{\text{init}} \leq t \leq T_{\text{end}}\}, \tag{11}$$

where T_{init} and T_{end} are the initial and final time instant of P_{M_Φ}, respectively. Following the CPM formulation in Section 3.1, the CPM on $\mathcal{R}_P$ is immediately obtained by replacing Φ with $\mathcal{R}_P$. We denote by M_P the leakage time-instant estimated form $\mathcal{R}_P$.

3.3 Ensemble of CPMs

Unfortunately, the minimum night flow might suffer from seasonalities or nonstationarities that are difficult to compensate or address, thus, the sequence Φ may not contain truly i.i.d. observations. Moreover, approximating models are never exact, and, because of model bias, sequence $\mathcal{R}_P$ is far from being i.i.d. both before and after T^*, where a large degree of dependency among the residuals is expected. These circumstances violate the hypothesis required by the CPM and explain why the CPMs are not able to properly estimate the change-point on residuals from approximating models, [11].

To reduce the effect of time-dependency in the analyzed sequence, it is possible to use the ensemble of CPMs, which is detailed in [11]. Because of space limitation here, we briefly describe its peculiarities. The ensemble $\mathcal{E}_d$ aggregates d individual estimates provided by CPMs executed on subsequences obtained by randomly sampling the original sequence (either Φ or $\mathcal{R}_P$). Such a random sampling is meant to reduce the temporal dependencies. Experiments in [11] on residuals of ARMA processes show that the ensemble provides better performance in locating the change-point than a single CPM executed on the whole residual sequence.

4 Experimental Results

To illustrate the feasibility of using CPM to estimate the leak time-instant, we consider data from the DMA of a big European city. The DMA is characterized by two inlets where flow an pressure are measured as well as five pressure monitoring sensors right inside. Real records have been collected from 11th November to 22nd December 2012.

4.1 CPM Configuration

Since we expect the leak to induce an abrupt and permanent shift in Φ and $\mathcal{R}_P$, we exploit a nonparametric CPM based on the Mann-Whitney [16] statistics, $\mathcal{U}$, and we

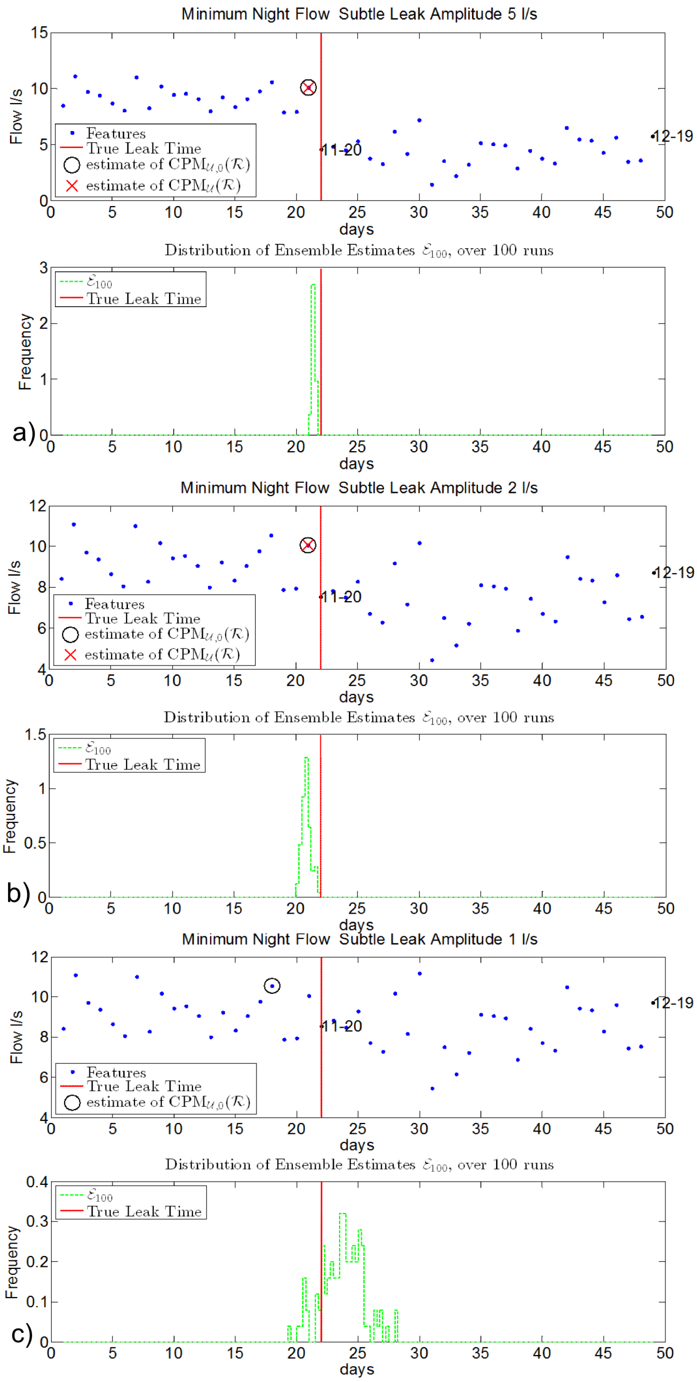

Fig. 2. Analysis of the minimum night flow to estimate the day when a leakage of 5 l/s (a), 2 l/s (b) and 1 l/s (c), appeared. The leak has been artificially injected on November, 20th.

analyze the performance of three solutions. At first, $\mathrm{CPM}_{\mathcal{U}}$, which is the CPM with the proper threshold $h_{\alpha,l}$ provided by the `CPM` package [17] in `R` statistical software. Then, we consider $\mathrm{CPM}_{\mathcal{U},0}$, the same as $\mathrm{CPM}_{\mathcal{U}}$ with threshold $h_{n,\alpha} = 0$; $\mathrm{CPM}_{\mathcal{U},0}$ locates a change-point where the partitioning yields the two most dissimilar sets. Finally, we consider the ensemble $\mathcal{E}_{100}$, which aggregates 100 individual estimates from random sampling. In all the CPM, we set $\alpha = 0.05$. The minimum night flow is computed by averaging measurements in 1 hour before and after the minimum of one day flow.

4.2 Performance on Artificially Induced Leakages

Since leakages induce a permanent offset in the minimum night flow, we simulate a leakage in Φ (see Eq. 4) by subtracting an amount proportional to the leakage amplitude (in liters per second) in $\phi(t)$ after T^*. To this purpose, we considered different leakage amplitudes, namely a reported leakage (5 l/s), an unreported leakage (2 l/s) and a background leakage (1 l/s) referring to the leak range set by water management company [6]. The leak has been artificially injected on November 20$^{\text{th}}$, and only the first level of the CPM hierarchy was tested here.

Fig. 2 shows the performance of the considered CPMs: while the outputs of $\mathrm{CPM}_{\mathcal{U}}$ and $\mathrm{CPM}_{\mathcal{U},0}$ are deterministic, i.e., given an input sequence they always provide the same result, the output of the ensemble is stochastic due to the random sampling phase. Therefore, we provide the empirical distribution of $\mathcal{E}_{100}$ estimates computed over 100 iterations. The plots show that the first level of the hierarchy estimates rather successfully the day when the leakage has occurred, though when the leak has a very small magnitude (background leakage) there is not enough statistical evidence for $\mathrm{CPM}_{\mathcal{U}}$ to assess the leak.

4.3 Performance on Real Leakages

A real reported leakage of magnitude 5.6 l/s occurred on 20$^{\text{th}}$ of December at 00h30 and lasted 30 hours. Fig. 3 a) shows the performance of the first level of the CPM hierarchy, and only $\mathrm{CPM}_{\mathcal{U},0}$ is effective in estimating the leakage day. The ensemble $\mathcal{E}_{100}$ is not accurate here since too few samples after T^* are provided, thus often subsequences obtained by random sampling may not contain measurements after the leak. However, the $\mathcal{U}$ statistic was able to point out the leak day when used on the whole dataset, though there is not enough statistical evidence for $\mathrm{CPM}_{\mathcal{U}}$ to estimate the leak day.

On the real leakage scenario, we assess the hierarchy of CPM to refine M_{Φ} by analyzing hourly pressure measurements from 00h00 of 19$^{\text{th}}$ till 23h00 of 21$^{\text{st}}$ of December, that were compared with those estimated using the DMA model, yielding to the residual sequence $\mathcal{R}_P$ (11). Fig. 3 b) shows that both $\mathcal{E}_{100}$ and $\mathrm{CPM}_{\mathcal{U},0}$ are rather effective in estimating the leak time-instant; while on the contrary, probably because the residuals are not i.i.d. (not even before and after T^*), $\mathrm{CPM}_{\mathcal{U}}$ was not able estimate the leak time-instant.

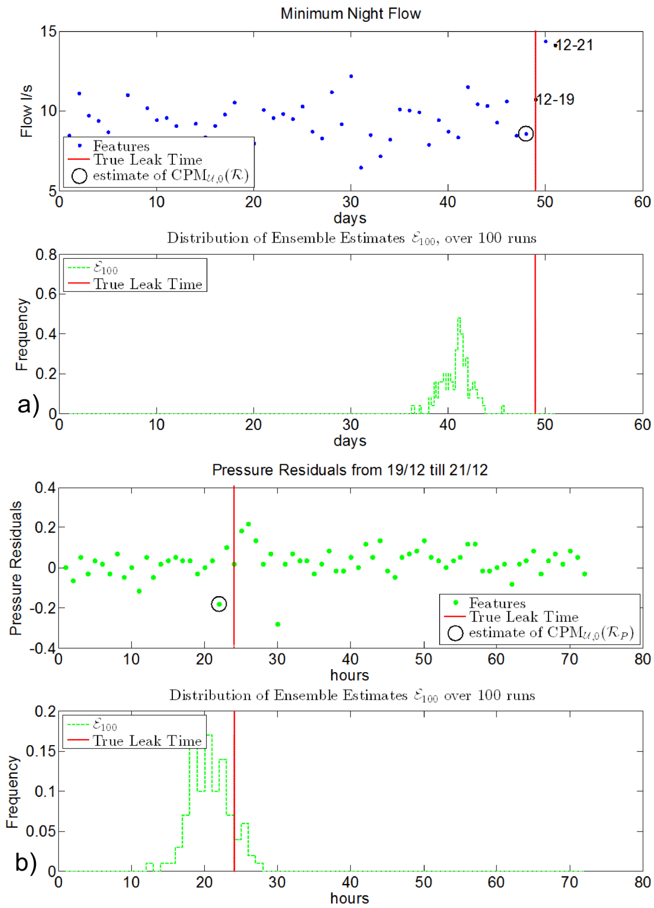

Fig. 3. Analysis of the minimum night flow to estimate the time instant when a leakage appeared in the network. The leakage occurred at 20th of December and lasted for two days.

5 Conclusion

We propose a hierarchy of CPMs to estimate the time instant when a leak has appeared within a DMA. Two CPMs are combined to estimate, first, the day, and then, the exact time when the leakage has occurred. The proposed hierarchy prevents the use of models approximating the pressure measurements over a large time interval, where these could be highly affected by model bias. Application results in a real DMA of a big European city have shown the feasibility of the proposed approach to estimate the leak time-instant also for subtle leaks. Ongoing works concern the use of multivariate CPM to analyze simultaneously the pressure, flow and different indicators that may be affected by the leak (e.g., water billed volumes), as well as providing a complete methodology – including leak detection and isolation – for leak diagnosis based on CPM.

References

1. Puust, R., Kapelan, Z., Savic, D.A., Koppel, T.: A review of methods for leakage management in pipe networks. Urban Water Journal **7**(1) (2010) 25–45
2. Khulief, Y., Khalifa, A., Mansour, R., Habib, M.: Acoustic detection of leaks in water pipelines using measurements inside pipe. Journal of Pipeline Systems Engineering and Practice **3**(2) (2012) 47–54
3. Lambert, M., Simpson, A., Vtkovsk, J., Wang, X.J., Lee, P.: A review of leading-edge leak detection techniques for water distribution systems. In: 20th AWA Convention, Perth, Australia. (2003)
4. Lambert, A.: What do we know about pressure:leakage relationships in distribution systems? In: IWA Conference System Approach to leakage control and water distribution system management. Brno, Czech Rebublic. (2001)
5. Thornton, J., Lambert, A.: Progress in practical prediction of pressure: leakage, pressure: burst frequency and pressure: consumption relationships. In: eakage 2005 Conference Proceedings. Halifax, Canada. (2005)
6. Pérez, R., Puig, V., Pascual, J., Quevedo, J., Landeros, E., Peralta, A.: Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. Control Engineering Practice **19**(10) (2011) 1157–1167
7. Eliades, D.G., Polycarpou, M.M.: Leakage fault detection in district metered areas of water distribution systems. Journal of Hydroinformatics **14**(4) (2012) 992–1002
8. Pettitt, A.N.: A Non-Parametric Approach to the Change-Point Problem. Applied Statistics **28**(2) (1979) 126–135
9. Hawkins, D.M., Qiu, P., Kang, C.W.: The changepoint model for statistical process control. Journal of Quality Technology **Vol. 35, No. 4** (2003) 355–366
10. Ross, G.J., Tasoulis, D.K., Adams, N.M.: Nonparametric monitoring of data streams for changes in location and scale. Technometrics **53**(4) (2011) 379–389
11. Alippi, C., Boracchi, G., Puig, V., Roveri, M.: An ensemble approach to estimate the fault-time instant. In: Proceedings of the 4th International Conference on Intelligent Controland Information Processing (ICICIP 2013). (2013)
12. Mood, A.M.: On the asymptotic efficiency of certain nonparametric two-sample tests. The Annals of Mathematical Statistics **Vol. 25, No. 3** (September 1954) 514–522
13. Hawkins, D.M., Zamba, K.D.: A change-point model for a shift in variance. Journal of Quality Technology **37**(1) (2005) 21–31
14. Zamba, K.D., Hawkins, D.M.: A multivariate change-point model for statistical process control. Technometrics **48**(4) (2006) pp. 539–549
15. Ross, G., Adams, N.M.: Two nonparametric control charts for detecting arbitrary distribution changes. Journal of Quality Technology **Vol 44, No. 22** (2012) 102–116
16. Mann, H.B., Whitney, D.R.: On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. The Annals of Mathematical Statistics **18**(1) (1947) 50–60
17. Ross, G.J.: Parametric and nonparametric sequential change detection in R: The cpm package. Journal of Statistical Software (Forthcoming)