ResiWater deliverable report D3.3: Integration of hydraulic information into the Event Detection Module

Olivier Piller, Denis Gilbert, Christian Kühnert, Thomas Bernard, Nicolas Cheifetz, Martin Wagner

RESIWATER

http://www.resiwater.eu

Deliverable Report D3.3

Written by Olivier Piller

INNOVATIVE SECURE SENSOR NETWORKS AND MODEL-BASED ASSESSMENT TOOLS FOR INCREASED RESILIENCE OF WATER INFRASTRUCTURES

Deliverable 3.3

Integration of hydraulic information into the Event Detection Module

Dissemination level: Public

WP3

Enhanced self-learning Monitoring and Event Detection Module

10th October 2018

Contact persons:
Olivier PILLER Olivier.Piller@irstea.fr
Fereshte SEDEHIZADE Fereshte.Sedehizade@bwb.de

Project reference for France & for Germany: ANR-14-PICS-0003 & BMBF-13N13690

WP 3 – Enhanced self-learning Monitoring and Event Detection Module

D3.3 Integration of hydraulic information into the Event Detection Module

List of Deliverable 3.3 contributors:

From IOSB

Thomas Bernard (thomas.bernard@iosb.fraunhofer.de)

Christian Kühnert (christian.kuehnert@iosb.fraunhofer.de)

From VEDIF

Nicolas Cheifetz (nicolas.cheifetz@veolia.com)

From Irstea

Olivier Piller (Olivier.Piller@irstea.fr)

Denis Gilbert (Denis.Gilbert@irstea.fr)

From TZW

Martin Wagner (martin.wagner@tzw.de)

Work package number	3	**Start date:**		01/02/2016
Contributors	IOSB	VEDIF	Irstea	TZW
Person-months per partner	2	8	6	6
Keywords				
Hydraulic and transport modelling – Water quality monitoring - Hybrid approach - Early warning detection system (EWDS) - Spatial segmentation - Sequential pattern mining – residence time - sensors				
Objectives				
Enhance the abnormal event classification by using information coming from the hydraulic and the transport models.				

TABLES OF CONTENTS

LIST OF FIGURES:

1 DELIVERABLE SUMMARY

The aim of work package 3 is the development of an enhanced self-learning monitoring and event detection module. In summary the principal items of work package 3 are:

1) Development of a data analysis platform for the integration of the heterogeneous sensor measurements
2) Development of self-learning monitoring and event detection algorithms. These algorithms will take into account the spatial distribution of the measurements.
3) Integration of online plausibility checks for the results of the algorithms
4) Deployment of tools for the launch of the enhanced event detection module

This deliverable describes the integration of hydraulic information into the event detection module with the aim to reduce the rate of false positive alarms.

Three main concepts are proposed:
- Use of backward transit time between two sensors;
- Knowledge of source provenance;
- Aggregation of clusters of similar water quality for spatial segmentation.

2 INTRODUCTION

In order to ensure a secure water distribution in the network, it is essential to detect any intentional and accidental contamination inside the water distribution system [1]. The design of sensor-based contaminant warning systems (CWS) is a promising approach for the mitigation of contamination risks in drinking water distribution systems [2]. Traditional detectors are based on data-driven techniques to analyse the collected signals at each monitoring station independently [3] or after synchronization [4] using statistical, heuristics or machine learning methods. Such event detection algorithms are formulated in deliverable 3.2 and require taking into account some spatio-temporal information in order to reduce significantly the rate of false positive alarms.

Furthermore, the hydraulic conditions of a water distribution system are hardly the same in operation (varying water sources, tank levels, etc.) which implies the emergence of changes in the water quality [5]. A quality monitoring system should not trigger alarms for such normal operating changes, by using hydraulic modelling for example [6]. It is classically assumed that the detector can discriminate the presence of a specific pollutant using some drinking quality parameters (*e.g.* free chlorine residual, conductivity, pH, turbidity, etc.) [7].

The deliverable is structured in two parts:

Initially, an approach is described which takes into account the water flow direction and pipe velocity to increase the robustness of the event detection module. As a use case, a part of the network from Eurométropole Strasbourg is investigated. It is additionally shown how the transit time between two sensors can be used backward to enhance the measure repeatability. Following, the importance of the provenance of sources for triggering alarm is revealed.

The second part tackles a more general event detection problem in Water Distribution Networks (WDNs) and proposes a new methodology to extract some prior knowledge which would enhance the performance of any detector. It can be seen as a pre-processing step easily usable by monitoring strategies like the event detection algorithms implemented in D3.2 or the freely available CANARY software [8] for instance.

3 VALIDATION OF EVENTS USING THE PIPE FLOW VELOCITY AND DIRECTION

Distortions in water quality (*e.g.* contaminations, chlorine peaks) travel through the water distribution network. Depending on the flow and the sensor network, several sensors located in the network on the same path will measure these distortions. On the contrary, the recalibration of a sensor or some maintenance work will only affect it at a specific time. In addition, maintenance and distortions resemble each other making it difficult to distinguish *e.g.* a chlorine peak from some sensor recalibration. Hence, by knowing only the measurements of one sensor leads to an increased rate of false positive alarms.

Therefore, by integrating information from the flow direction and velocity into the event detection module, it is assumed, that the amount of false positive alarms can be considerably reduced. Figure 1 sketches this concept in this deliverable investigated approach. If an event has been detected by the data-driven module (for details deliverables 3.1 and 3.2), it is checked if the event is detected again with another sensor downstream. If the event is detected and the time between the two events corresponds to the flow velocity, the module raises an alarm, otherwise the event is rejected.

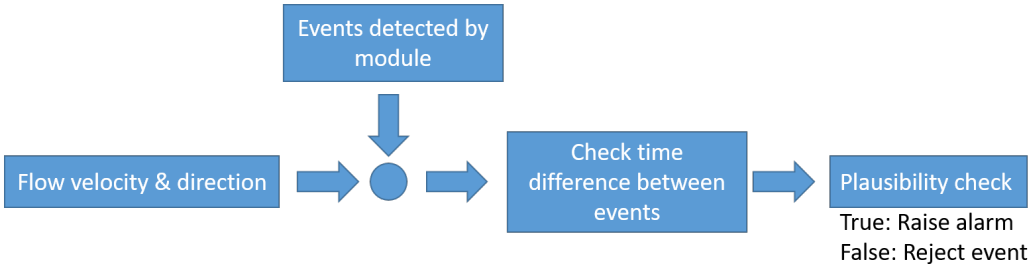


Figure 1: Concept to reduce the amount of false positive alarms by taking into account the pipe flow velocity.

3.1 Use case: Validation Chlorine Peaks at Eurométropole Strasbourg

As a use case, the prior described approach is tested on the propagation of a chlorine peak at the water distribution network of CUS (Eurométropole Strasbourg). Therefore, three chlorine sensors have been selected that are located close to each other and shown in Figure 2 left side. In that case, the sensors S139 and S138 are located in parallel pipes, while sensor S137 is located farer away. Since sensor S139 has the highest flow rate it detects the chlorine peak in the beginning followed by sensor S139. Finally, sensor S137 located farer away from the others will detect the peaks the last one.

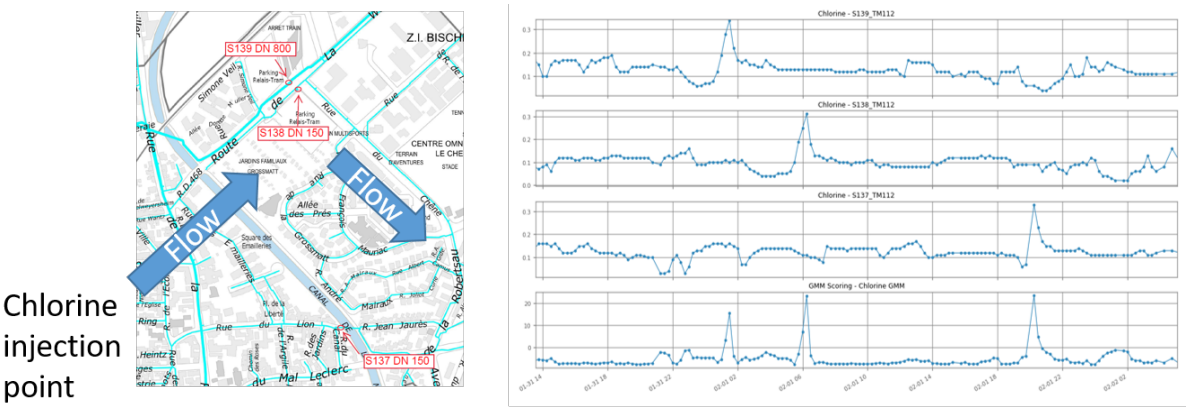


Figure 2: Chlorine sensor position at CUS (left); Chlorine detected by the three sensors, leading to three peaks in the event detection module last subfigure bottom right (right).

Figure 2, right side, shows the chlorine measurements for each station as well as the alarm index of the event detection module. Three distinguished peaks can be seen in the resulting alarm index generated from the event detection module. In this case each peak corresponds to the prior injected chlorine. By comparing the time difference between the peaks of the alarm index with the flow velocity in the pipes, it can be confirmed, that the peaks are due to an injection and not resulting from sensor maintenance works.

3.2 Two of ways of using the travel time between two sensors

The concept of integrating information from the pipe velocities can be further extended by calculating the transit time between two sensors. There are two ways: forward and backward. For the first one, it is illustrated in Figure 3, the water quality transport simulation can be used to anticipate and calculate which sensor will be reached and in in which time. It is here assumed that no action is taken to change the topology of the network (closing valves, flushing, etc.) so this is only possible when no health hazard is assumed. The second one, ***upstream backward*** is aimed to check the repeatability of the positive alarm and so ***will permit to limit the false positives*** if no detection further upstream. The principle is shown in Figure 4. The Irstea backtracking algorithm applied to sensor binary answers may be used as in [9].

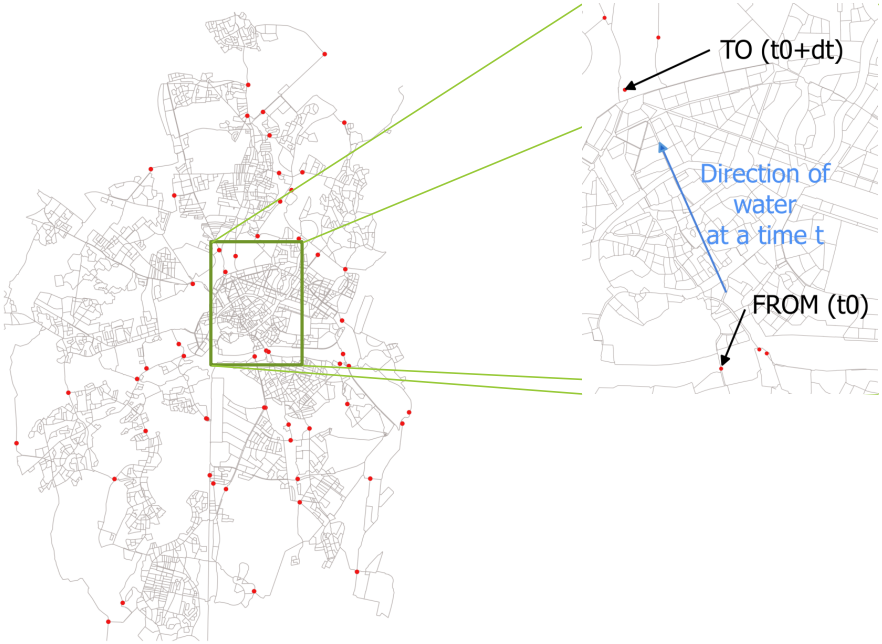


Figure 3: Forward transport model for normal water quality operation (*e.g.* a chlorine peak).

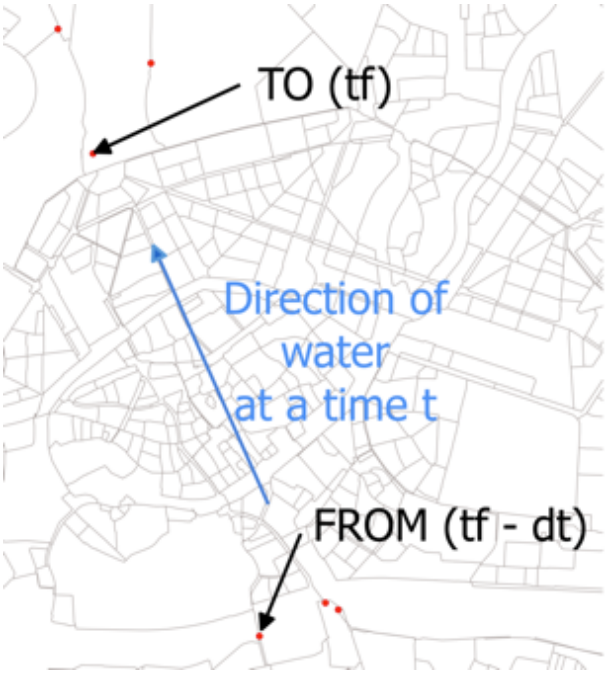


Figure 4: Inverse transport simulation for improving the specificity of the classification.

3.3 Impact of a burst on source provenance

It is also possible to use the hydraulic simulation to replay abnormal events and understand the different impact. A 400-mm nominal pipe (DN 400) has broken near the main station at 5:45. It was necessary to close four valves in the hydraulic segment of the broken pipe to isolate the leak. It was found that the consumers were not under the influence on the usual water source and there was a shift/move on the water quality barrier (a change of chlorine level). The software Porteau has been used to study the provenance of the source under different scenarios. ***The knowledge of water source provenance at sensors position is useful to identify false positive***.

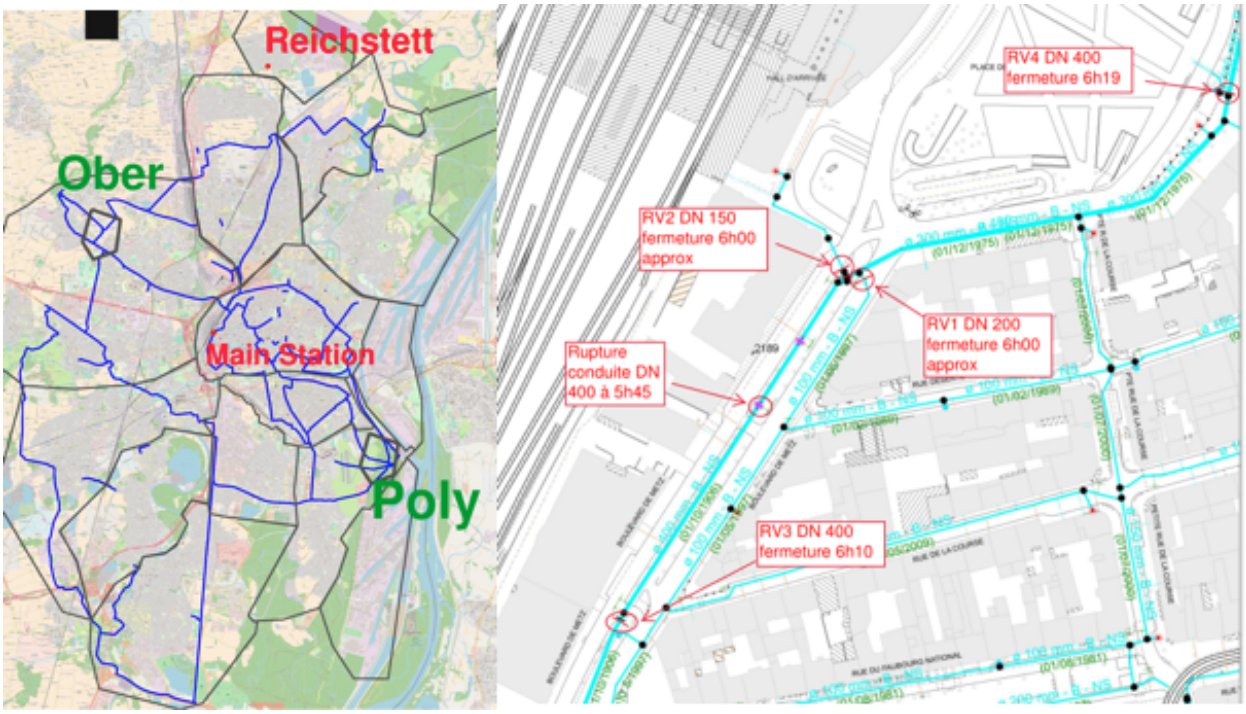


Figure 5: a drastic change in water quality after a pipe break near the Strasbourg main station.

4 SPATIO-TEMPORAL SEGMENTATION FOR WATER QUALITY EVENT DETECTION SYSTEMS

4.1 Problem Formulation

This section investigates how to deal with the variability of water quality signals for monitoring WDN based on an existing deployment of water quality sensors. *A WDS is a vulnerable infrastructure subject to deliberate or accidental contamination intrusions.* The proposed approach is a two-step methodology to run before any Early Detection System (EDS) for scaling the problem of water quality monitoring. The first step of the method can be seen as a fully data-driven method that identifies the most representative operational periods for a WDS based on the incoming and outgoing flows [10]. The second step relies in discriminating automatically zones in the WDN of different sizes in space and time for a specific operational configuration. In other words, the first step extracts macro behaviors of the WDS and the second step reveals a partition of zones of equal quality in the WDN.

The Figure 6 describes the global methodology to enhance any Event Detection Module monitoring water quality in WDN. The idea is to extract offline some prior knowledge about the spatial influence inside a WDN for a representative operating state of a WDS. Based on this existing spatio-temporal segmentation, the event detection system should trigger better alarms using online measurements of water quality parameters in WDN.

The aim of the "time segmentation" step is to identify automatically the most representative operating periods in terms of water flow incoming and outgoing a Water Distribution System (WDS). A multi-stage methodology is designed to address this initial problematic. The first substep consists in extracting elementary motifs from water flow time series. Each pattern characterizes a simplified hydraulic state defined by constant flow values, using a classical K-means algorithm. This multivariate discretization is used to compute a dedicated Levenshtein distance[1] that compares pairs of pattern sequence. The DBSCAN algorithm [11] is finally used to regroup similar sequences and their prototypes, called temporal patterns, are determined.

Then, the prototype of each operating configuration is used to design a specific hydraulic model with the modeling software Synergi™ Water. Based on such reference period, a hydraulic model is calibrated and used to trace conservative substances from sources to monitoring stations. The large amount of propagation simulations leads to multiple WDN partitions with zones under the influence of specific sources, mixing areas and "dead zones". We propose to summarize different topologies of the WDN in a single graph partition per operational configuration. It can be seen as a median graph [12] or consensus clustering [13] over a sequence of the resulting simulation graphs in the best possible manner. We propose a greedy like algorithm on a consensus matrix (co-occurrence of

[1] Initially, string metric or edit distance for measuring the difference between two sequences

vertices in clusters of the input partitions) which is particularly suitable to monitor the evolution of community structure in temporal networks [13].

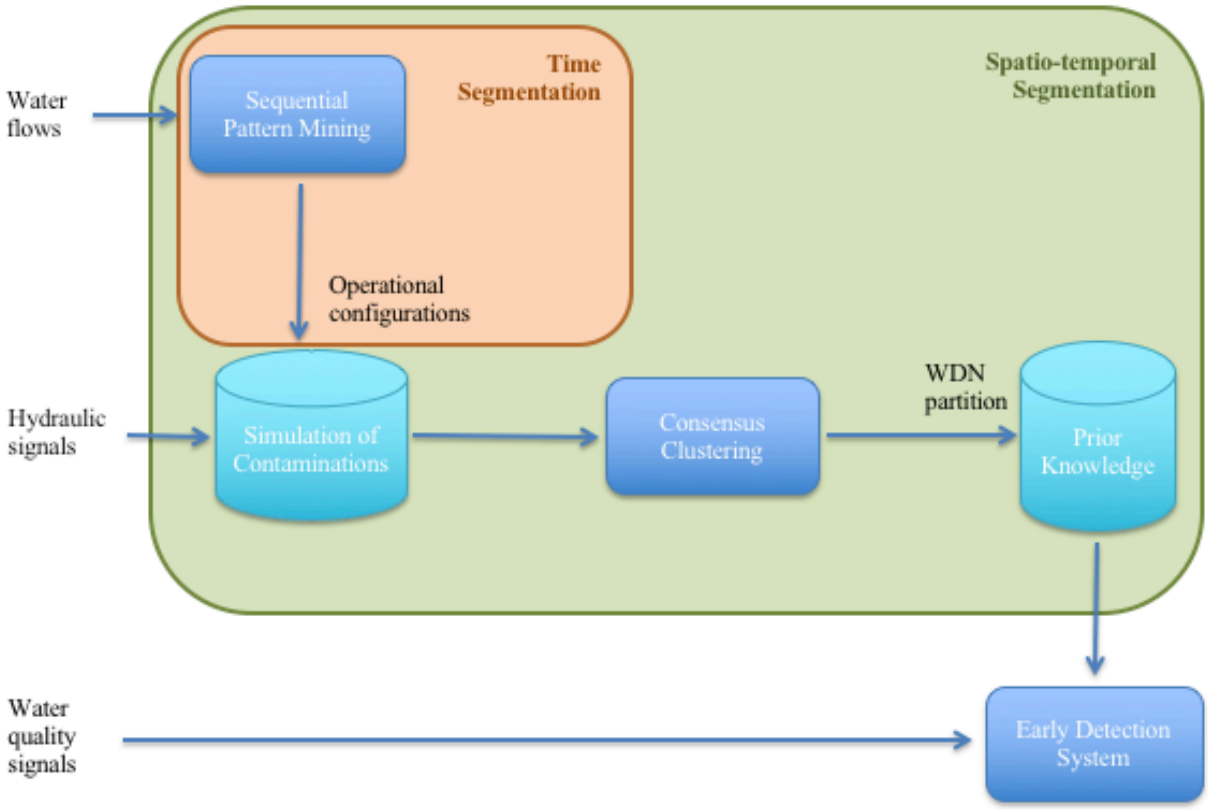


Figure 6: Global methodology to enhance the Event Detection Module.

4.2 The VEDIF Case Study

The proposed approach is illustrated on a large real-world network in France. The Syndicat des Eaux d'Ile-de-France (SEDIF) is an association including 150 municipalities that ensures the production and the distribution of drinking water to more than 4.5 million inhabitants of suburban Paris. The network of the SEDIF is the largest drinking WDN in France with about 8,600 km of pipes, almost 600,000 active connections and more than 750,000 m^3 of water produced each day. The water is produced in three large Drinking Water Treatment Plants (DWTP) located on the three main rivers of the Seine river basin, as shown in Figure 7. This paper is focused on a major part of the SEDIF network, mainly supplied by the Neuilly-sur-Marne DWTP and located on the Marne river. This subnetwork is depicted as the green area in Figure 7 and can be represented by a single hydraulic model including multiple sectors with different elevations. This hydraulic model is simplified as a system, only characterized by water flows collected in 2015. As the SEDIF network is fully interconnected (*e.g.*, large interconnections between the production plants, illustrated in dark blue), the various operational conditions are strongly impacting the water propagation into the entire WDN. Indeed, any point into the network can be under the influence of multiple sources depending on its location and time.

The next part presents the time segmentation, a procedure to give an insight on the recurrent operating periods over a year from a single water distribution system. Both the pattern extraction and pattern mining methods are briefly described and fully available in reference [10].

4.3 Time segmentation in a WDS

4.3.1 Extraction of elementary motifs

Let $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ denote a set of m time series, where each one of them $\boldsymbol{y}_j = \left(y_{1j}, \dots, y_{T_j j}\right)$ corresponds to water flows recorded at the border of the WDS. That is to say y_j is a univariate time series and $y_{tj} \in R$ is an incoming or outgoing flow. Note that no assumption is made about synchronization between time series and the production plant flow is omitted due to its value predominance and relative stability.

The original time series are recorded with a fine granularity where the time step is 2'30 and present classical issues like noisy data and missing values. The dataset representing more than 200,000 points in a year per water flow time series needs to be simplified using a piecewise approximation for instance. The TVD-MM algorithm [14] is used to denoise each time series independently while preserving the signal changes and aims to minimize the objective function:

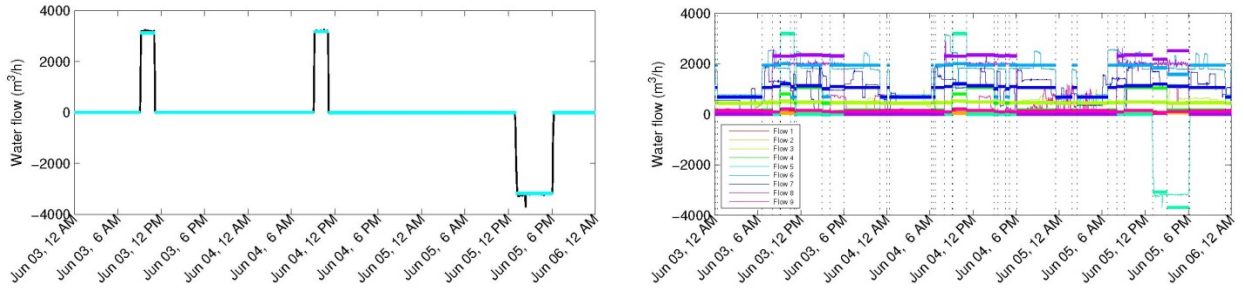
$$\sum_{t=1}^{T} |y_t - x_t|^2 + \lambda \sum_{t=2}^{T} |x_t - x_{t-1}| \,,$$

where $\lambda > 0$ is the regularization parameter, $(y_1, \dots, y_T)$ represents the original signal and $(x_1, \dots, x_T)$ is the smoothed signal. The higher λ, the smoothest the resulting signal. Note that the number of segments is not required, and the segment values are modeled as constants. The method is notably suitable when the water flow signal can be approximated by piecewise constant functions as illustrated in Figure 8a.

Figure 7: The SEDIF perimeter around Paris, the three main drinking water treatment plants in red and their interconnections in dark blue.

Obviously, segmenting independently each flow signal is simpler than tackling a multidimensional water flow. As the WDN is considered as a system, the segmented time series are then aggregated into a single matrix $\boldsymbol{x}$ sharing all the change-points which can be seen as multivariate time series $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_m)$ over the same time grid $\{1, \dots, n\}$. In other words, this matrix is composed by n multivariate segments where each segment is an m-dimensional vector revealing m constant flows for

a specific period, as shown in Figure 8b. In our case, this segmentation step allows to reduce significantly the size of the overall dataset by a factor 6.



(a) Univariate segmentation. *(b) Multidimensional time series.*

Figure 8: Univariate segmentation (8a) using the TVD-MM algorithm over three days in June 2015 – the raw data of water flow are painted in black and the piecewise constant functions are in cyan. The segmented multivariate time series (8b) with change points as vertical dotted lines.

A classical clustering method is performed on the multivariate time series $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)$. The well-known K-means algorithm [15] is applied using various random initializations and the partition with the lowest intra-cluster inertia is selected. The number of clusters K is usually assigned by minimizing some information criterion (*e.g.*, AIC or BIC) but here no clear minimum could be found due to the large size of the data. Then, the K-value is selected by minimizing a penalized and weighted version of the intra-cluster inertia defined by $C = D + \gamma \nu_K log(n(m+1))$, where D is a distance defined by the following equation, $\gamma > 0$ is the penalization parameter and $\nu_K = mK$ is the number of free parameters.

$$D = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \delta_i^2 \varphi\left(\boldsymbol{x}_i, \boldsymbol{\mu}_{z_i | z_i = k}\right)^2} \quad \text{with } \varphi(\boldsymbol{x}_i, \boldsymbol{\mu}_k) = \sqrt{\frac{1}{m}\sum_{j=1}^{m} (x_{ij} - \mu_{kj})^2}\,, \tag{1}$$

and δ_i is the duration (in hour) of the segment i, z_i is the label of segment i and $\boldsymbol{\mu}_k = (\mu_{k1}, \ldots, \mu_{km})$ is the mean centroid of cluster k. Note that D is linearly correlated to the inertia optimized by K-means.

The next part describes a strategy to extract meaningful "patterns" or subsequences in the time series $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n)$ where each temporal centroid $\boldsymbol{\mu}_i \in \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K\}$ is called an "elementary motif". The sequential pattern mining method is based on a dedicated Levenshtein distance.

4.3.2 Sequential pattern mining

Let us introduce a distance in order to quantify the difference between sequences of elementary motifs that represent the pattern instances. Moreover, a reformulation of the Levenshtein distance [16] is adopted due to its capacity to integrate each sequence order and the three single-character operations (insertion, deletion and substitution). Some other distances (e.g., Hamming distance) do not share these features. The distance noted L is based on the function φ defined in Eq. (1); let us note $\varphi(\mu_k, \mu_l) = \varphi_{k,l}$ and $\varphi(\mu_k, 0) = \varphi_k, \forall (k,l) \in \{1, \dots, K\}^2$. Considering two patterns u and v, the Levenshtein distance is defined as ($\forall i = 1, \dots, |u|, \forall j = 1, \dots, |v|$)

$$\begin{cases} L(0,0) = 0 \\ L(i,0) = L(i-1,0) + \delta_{u_i} \varphi_{z_{u_{i-1}}, z_{u_i}} \\ L(0,j) = L(0,j-1) + \delta_{v_j} \varphi_{z_{v_{j-1}}, z_{v_j}} \\ L(i,j) = min\left[L(i-1,j) + \delta_{u_i} \varphi_{z_{u_{i-1}}, z_{u_i}}, L(i,j-1) + \delta_{v_j} \varphi_{z_{v_{j-1}}, z_{v_j}}, L(i-1,j-1) + Sub(z_{u_i}, z_{v_j}) \right] \end{cases}$$

and the substitution cost is defined by

$$Sub\left(z_{u_i}, z_{v_j}\right) = \begin{cases} \delta_{u_i} \varphi_{z_{u_{i-1}}, z_{u_i}} + \delta_{v_j} \min\left(\varphi_{z_{v_{j-1}}, z_{v_j}}, \varphi_{z_{v_j}, z_{v_{j+1}}} \right) & \text{if } i = |u| \\ \delta_{u_i} \min\left(\varphi_{z_{u_{i-1}}, z_{u_i}}, \varphi_{z_{u_i}, z_{u_{i+1}}} \right) + \delta_{v_j} \varphi_{z_{v_{j-1}}, z_{v_j}} & \text{if } j = |v| \\ \delta_{u_i} \min\left(\varphi_{z_{u_{i-1}}, z_{u_i}}, \varphi_{z_{u_i}, z_{u_{i+1}}} \right) + \delta_{v_j} \min\left(\varphi_{z_{v_{j-1}}, z_{v_j}}, \varphi_{z_{v_j}, z_{v_{j+1}}} \right) & \text{otherwise} \end{cases}.$$

A clustering algorithm exploiting the previous distance is used to aggregate similar patterns among the overall sequence of elementary motifs. It is worth noting that successive motifs with identical labels are merged. First, a sequence of p candidate patterns are enumerated according to some prior knowledge relative to the addressed problem. Then, the DBSCAN algorithm [10] groups candidate patterns in high density regions; that is to say, a similar pattern has a distance less than a given threshold $\varepsilon > 0$. This algorithm has a worst case complexity of $O(p^2)$ and does not require setting the number of clusters (unlike K-means). The estimation of the ε-value is needed and a greedy-like procedure is performed to identify few potential clusters, where each iteration is defined such as

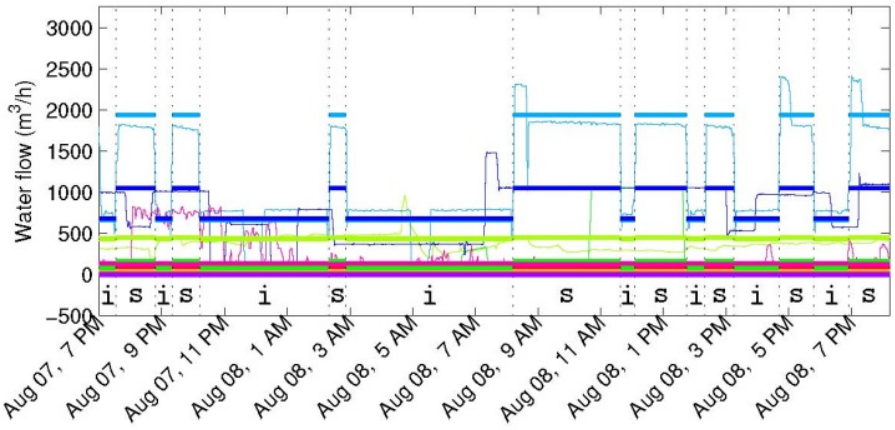
1. Selection of the best pattern (depending on the addressed problem: most frequent, etc.);
2. Aggregation of candidate patterns similar to the best pattern instances (distance $< \varepsilon$).

Then, the DBSCAN algorithm is used on all the patterns identified by the greedy clusters. The final threshold ε is chosen such as the DBSCAN rate of good classification is maximized while its ε-value is the lowest. Note that pattern overlapping can occur between patterns of different clusters but not inside each cluster. Finally, the most meaningful operational periods are identified as the medoid pattern per cluster.

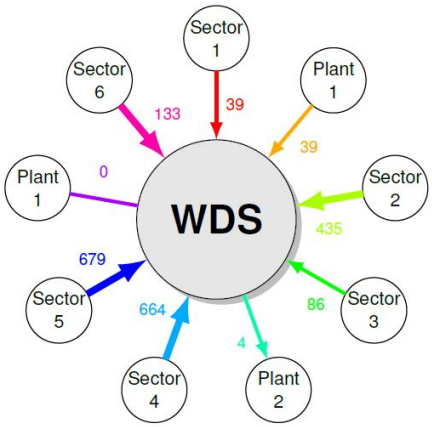
4.4 Spatial segmentation in a WDS

4.4.1 Contamination simulation

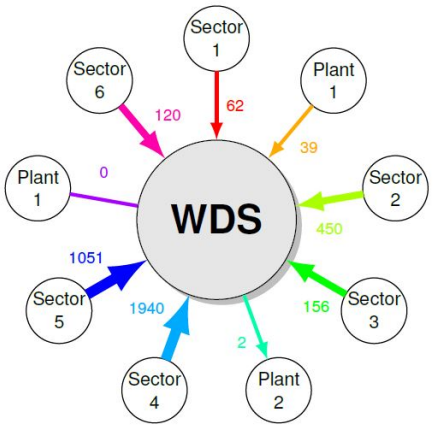
Following the description of the case study in subsection 4.2, nine time series of water flow collected in 2015 are used for characterizing successive hydraulic states of a WDS with respect to water exchange with its adjacent hydraulic systems (two plants and six sectors). The medoid of the first representative pattern is illustrated by Figure 9. The SubFigure 9a draws the medoid which is defined by the sequence of labels '`isisisisisisisis`', a succession of two elementary motifs '`i`' and '`s`'. It lasts 25h12min early from the 7th to the 8th of August and its belonging cluster represent a cumulated duration of 42% in 2015. The motif `s` marks the intermittent significant flows that come from sectors 4 and 5. For brevity, a short description of the two elementary motifs occurring in the first medoid is given in SubFigure 9b and SubFigure 9c. The motif '`i`' shows an incoming flow at about 700 m^3/h from sector 4 and 5, while the period '`s`' displays higher flows of about 2,000 m^3/h and 1,000 m^3/h respectively.



(a) Medoid of the first most representative operating period.



(b) Elementary motif 'i'.



(c) Elementary motif 's'.

Figure 9: Description of the elementary motifs appearing in the first two pattern medoids: arrows represent water flows (m^3/h) between the WDS and other adjacent hydraulic systems (two plants and six sectors). Each color is associated to a specific water flow: flow 1 in red, flow 2 in orange…up to flow 9 in mauve. Note that most of the differences between the four elementary motifs can be seen through water coming from the sector 3, 4, 5 and plant 1 (via the flow 8).

The prototype of the most representative operating period is used as a reference period to calibrate a hydraulic model based on real water signals. The hydraulic modelling implements a major part of the SEDIF network represented by a single calibrated model including multiple sectors and containing about 30,000 nodes and 40,000 pipes. Using the hydraulic modeling software Synergi™ Water, several contamination scenarios are set up including various times, durations and locations of injections. The resulting dataset of contamination simulations leads to multiple WDN partitions of water quality zones. The next part describes briefly how to extract a unique network partition for each operating period.

4.4.2 Consensus clustering

Consensus clustering, also called cluster ensemble, has received considerable attention in the statistics and machine learning communities. Different cluster ensemble approaches are considered in the literature, including graph partitioning, Voting approach, Mutual information algorithms and Co-association based functions [17]. The Figure 10 illustrates the time series of the water distribution system. The consensus clustering aims to get a unique graph partitioning among the various WDN partitions obtained with the different scenarios of contamination distribution.

Figure 10: Time series of a real water distribution network.

Most of the methods are based on the computation of a consensus matrix. Let us suppose that we wish to combine n_P partitions found by a clustering algorithm on a network with n vertices. The consensus matrix D is an $n \times n$ matrix, whose entry D_{ij} indicates the number of partitions in which vertices i and j of the network were assigned to the same cluster, divided by the number of partitions n_P. The matrix D is usually much denser than the adjacency matrix A of the original network, because in the consensus matrix there is an edge between any two vertices which have co-occurred in the same cluster at least once. On the other hand, the weights are large only for those vertices which are most frequently co-clustered, whereas low weights indicate that the vertices are probably at the boundary between different (real) clusters, so their classification in the same cluster is unlikely and essentially due to noise [13].

5 REFERENCES

[1] N. Sankary and A. Ostfeld. Inline mobile sensors for contaminant early warning enhancement in water distribution systems. *Journal of Water Resources Planning and Management*, 143(2):04016073, 2016.

[2] W. E. Hart and R. Murray. Review of sensor placement strategies for contamination warning systems in drinking water distribution systems. *Journal of Water Resources Planning and Management*, 136(6):611–619, 2010.

[3] X. Yang and D. L. Boccelli. Bayesian approach for real-time probabilistic contamination source identification. *Journal of Water Resources Planning and Management*, 140(8):04014019, 2013.

[4] C. Kühnert, M. Baruthio, T. Bernard, C. Steinmetz, and J.-M. Weber. Cloud-based event detection platform for water distribution networks using machine-learning algorithms. *Procedia Engineering*, 119:901–907, 2015.

[5] M. Housh and Z. Ohar. Integrating physically based simulators with event detection systems: Multi-site detection approach. *Water research*, 110:180–191, 2017.

[6] N. Oliker, Z. Ohar, and A. Ostfeld. Spatial event classification using simulated water quality data. *Environmental Modelling & Software*, 77:71–80, 2016.

[7] D. G. Eliades, D. Stavrou, S. G. Vrachimis, C. G. Panayiotou, and M. M. Polycarpou. Contamination event detection using multi-level thresholds. *Procedia Engineering*, 119:1429–1438, 2015.

[8] A. Leow, J. Burkhardt, W. E. Platten III, B. Zimmerman, N. E. Brinkman, A. Turner, R.Murray, G. Sorial, and J. Garland. Application of the canary event detection software for real-time performance monitoring of decentralized water reuse systems. *Environmental Science: Water Research & Technology*, 3(2):224–234, 2017.

[9] H. Ung, O. Piller, D. Gilbert, and I. Mortazavi, "Inverse Transport Method for Determination of Potential Contamination Sources with a Stochastic Framework," in World Environmental and Water Resources Congress 2013, C. L. Patterson, S. D. Struck, and D. J. Murray, Eds. Cincinnatti (Ohio), USA: ASCE, 2013, pp. 798-812.

[10] N. Cheifetz, S. Kraiem, P. Mandel, C. Féliers, and V. Heim. Extracting temporal patterns for contamination event detection in a large water distribution system. In *the 15th International Computing and Control for Water Industry conference (CCWI 2017)*, 2017.

[11] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *the Second International Conference on Knowledge Discovery and Data Mining*, volume 96, pages 226–231, 1996.

[12] X. Jiang, A. Munger, and H. Bunke. On median graphs: properties, algorithms, and applications. *IEEE Transactions on pattern analysis and machine intelligence*, 23(10): 1144–1151, 2001.
[13] A. Lancichinetti and S. Fortunato. Consensus clustering in complex networks. *Scientific Reports (Nature Publisher Group)*, 2:336, 2012.
[14] M. A. Figueiredo, J. B. Dias, J. P. Oliveira, and R. D. Nowak, "On total variation denoising: A new majorization-minimization algorithm and an experimental comparison with wavelet denoising," in *IEEE International Conference on Image Processing*, 2006, pp. 2633–2636.
[15] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Symposium on Mathematical Statistics and Probability*. Univ. of California Press, 1967.
[16] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," in *Soviet physics doklady*, vol. 10, no. 8, 1966, pp. 707–710.
[17] H. Elghazel, K. Benabdeslem, and F. Hamdi. Consensus clustering by graph based approach. In *the 18th European Symposium on Artificial Neural Networks (ESANN)*, 2010.