Prognosis of Quality Sensors in the Barcelona Drinking Water Network

Diego García[2], Ramon Creus[1], Meritxell Minoves[1], Xavier Pardo[1], Joseba Quevedo[2] and Vicenç Puig[2]

Abstract—One of the most important areas of the water utilities is the water quality management. This area is responsible of guaranteeing safety in the water supply to the citizens. The strategy to guarantee the safety is based on two principal elements: disinfection and monitoring. Disinfection techniques, such as chlorination, allow to prevent the growing of microorganisms present in the water. Moreover, in order to guarantee this safety in the whole water network and avoid any unexpected event, on-line sensors are required to monitor a set of quality parameters. The whole process is based on the assumption that the information retrieved from quality sensors is totally reliable. But due to the complexity of the calibration and maintenance of these chemical sensors, several factors affect the accuracy of the raw data collected. Consequently, any decision based on this raw data might be based on a non solid base. Therefore, this work presents a data analytics approach consisting in two modules: fault diagnosis and prognosis. The fault diagnosis module first discerns if a sensor is detecting a real change on water quality parameters or actually is providing inconsistent information due to some malfunction. The prognosis module aims to predict the fault instant due to a slow degradation, which is very common in chlorine sensors. This approach allows to apply a predictive maintenance strategy reducing corrective actions. The proposed methodology has been satisfactorily tested on the Barcelona Drinking Water Network.

I. Introduction

One of the main tasks of the water utilities (WU) is to transport and supply drinking water to the citizens throughout water distribution systems (WDS). Two of the WU's main areas concerned are: On the one hand, the operations department to manage the hydraulic infrastructure (e.g. pumping stations, reservoirs, pipes,...). On the other hand, the water quality control department to manage the drinking water safety. Furthermore, different legal frameworks regulates the quality of drinking water to supply.

Water quality monitoring and control management programmes involve several tasks. As detailed in [1], such tasks are monitoring network design (e.g. which parameters to be measured, how often, etc.), laboratory work (e.g. chemical analysis, laboratory tests, etc.) and analytical quality assurance (e.g. production of reliable data) among other elements.

This paper is focused on developing a methodology that guarantees the production of sensor reliable data.

[1]Ramon Creus Head of Operative Support, Meritxell Minoves Exploitation Support and Xavier Pardo Quality Management at Aigües de Barcelona {rcreusro, mminoves, xpardo}@aiguesdebarcelona.cat

[2]Diego García, Joseba Quevedo and Vicenç Puig are at Specific Center of Research CS2AC in the Universitat Politècnica de Catalunya (UPC), Terrassa (Barcelona), Catalonia. {diego.garcia,joseba.quevedo, vicenc.puig}@upc.edu

There are several techniques to treat the water in WDS and keep it healthy for human consumption. One common disinfection technique is the chlorination of water. This process consists in injecting chlorine or derivatives in the water. The injected chlorine is consumed (i.e. chemical reaction) in the WDS because of two main factors [2]. On the one hand, due to reactions in the bulk water: by the presence of organic content in the water, the initial chlorine concentration and physical conditions (e.g. temperature). On the other hand, the chlorine reacts at the pipe wall, known as *biofilm*.

The chlorine in the water drops exponentially as follows:

$$C(t) = C_0 \cdot e^{-kT} \tag{1}$$

where $C(t)$ is the chlorine concentration (mg/l) at the instant t, C_0 is the initial chlorine concentration and T the time interval since the injection.

Thus, in order to keep residual chlorine in the water distribution network after a certain time T, it is necessary to inject a certain chlorine dose C_0. The chlorine injection, usually done in the reservoirs, is regulated by an automatic controller, where a feedback control loop (typically based on a proportional-integral-derivative controller) injects a quantity of chlorine determined by the error between the concentration reference and the measured chlorine concentration.

The WU monitors the water quality parameters with on-line water quality sensors (multi-parametric and single-parametric) installed along the water transport and distribution networks. The most common water quality parameters monitored on-line are: conductivity, temperature, pH and chlorine. Other interesting parameters such as total organic carbon (TOC) are well-known indicators of water quality. Moreover, laboratory analyses of water samples taken from different points of the network are essential to analyze biological and chemical components unobserved by the on-line sensors, or even to contrast them against on-line observations.

Quality sensors require a specific calibration planning prescribed by the manufacturer depending on the sensor's model to guarantee the reliability of the observations. Moreover, a preventive maintenance planning (e.g. bi-monthly or quarterly) is also specified by the manufacturer to preserve data reliability.

Even though applying a preventive planning, these quality sensors could be affected by several problems such as the ones listed in Table I. Thus, a corrective planning is always required to solve these unexpected problems affecting the sensors reliability.

Cause	Consequence
Communications problem	Data gap
Loss of sensitivity	Flat signal or slow drift down
Electronic malfunction	Noise and peaks in the signal
Miscalibration of the sensor	Offsets affecting the real value

There is significant research to detect and avoid intended and unintended injection of hazardous substances in the water distribution network to guarantee the drinking water safety. Several works have studied this particular subject in order to detect water contamination events. In [3], different contaminants introduced in tap water are detected measuring pH, turbidity, conductivity, total organic carbon and chlorine and establishing as detection limits a threshold based on three time the standard deviation above the average of each magnitude. In [4], a probabilistic principal component analysis (PPCA) method using UV-Vis spectrometers is detailed to detect contaminant injection into WDS. In [5], a model-based approach considering the chlorine input injection is used to compute bounds to compare with the sensors measurements. In [6], a benchmark of a set of sensors from different manufacturers measuring distinct quality parameters is presented allowing to analyze and compare the sensitivity on the presence of various contaminants. In [7], operational data and water quality are combined to reduce false positives rate in the quality event detection. In [8], different change-point detection algorithms are applied to the residuals of an autoregressive model. Sensor placement is also an important topic to improve quality monitoring meanwhile reducing operational costs as discussed in [9].

Model-based approaches, such as [5], have the main drawback that the performance depend directly on the water network model's accuracy. Moreover, due to the complex behaviour of the water parameters, it is very difficult to develop a physical model to describe the water quality dynamics. Hence, data-driven approaches are very interesting in this case and therefore widely used.

The major drawback, in general, of the existing approaches to detect drinking water quality events is that are based on the assumption that data gathered from these sensors are accurate and precise. But as we have pointed out, raw data from quality sensors could not be ready to be analyzed or to extract solid conclusions. Unreliable water quality information is a serious problem for the WU in order to guarantee a water supply that assures the citizens health.

The most common fault is the loss of sensitivity of the sensor producing a slow drift (decays of 0.3-0.4 from 4 days to two weeks) of the chlorine signal. Due to the slowness of the incipient degradation is not straightforward to detect this fault. Hence, a prognosis approach is applied to predict the remaining useful life (RUL) of the sensor.

Prognosis is an important field in the fault community research [10], [11], [12], [13], [14].

Hence, the main motivation of this work is to provide a data analytics methodology for monitoring quality sensors and events applicable to drinking water networks, such as the mentioned before.

The contributions of this work are twofold. On the one hand, this work provides a procedure to get a solid information basis, discarding unreliable data, to improve decision making of the WU in the water quality management. On the other hand, a set of indicators are provided allowing to improve the preventive planning reducing the number of expensive corrective actions.

The proposed methodology has been satisfactorily tested on the Barcelona drinking water network.

The structure of the paper is the following: In SectionII, the considered case study to illustrate the proposed methodology is introduced. In Section III, the methodology is described. In Section IV, the results obtained in several real scenarios in the considered case study are presented and discussed. Finally, in Section V, the conclusions are provided as well a future research paths.

II. CASE STUDY

The case study, used to illustrate the proposed methodology for monitoring quality sensors and events, is based on the Barcelona drinking water network. The Barcelona drinking water network is a complex WDS of over 4,600 km that supplies supply drinking water to 218 demand sectors. In this WDS, there are installed thousands of sensors along the network to know with precision the hydraulic state of the network, to control and manage it efficiently. In addition, there are installed quality sensors and analysers to handle the water quality control.

For illustrative purposes, this paper is focused on the zone depicted in Figure 1. The water supply of this zone can come from two different water sources: the rivers Ter and Llobregat.

The tank collects water to satisfy three demand sectors. A chlorination process is continuously done in this tank based on an actuator (chlorine injection), a chlorine analyzer and some reference given by the WU's operators. At the entrance of each demand sector, a multi-parameter water quality sensor is installed to monitor and control the quality of the supplied water.

The WU collects hourly observations from multi-parameter sensors and 15-minutes observations from chlorine analyzers. The parameters observed are: temperature, conductivity, pH and chlorine. The single-parameter sensors measure chlorine.

The water quality data collected are analyzed by the experts using a visualization software to check any existing quality event or sensor problem. Another software system allows the experts to contrast field samples analyzed in the laboratory against the data collected from the sensors.

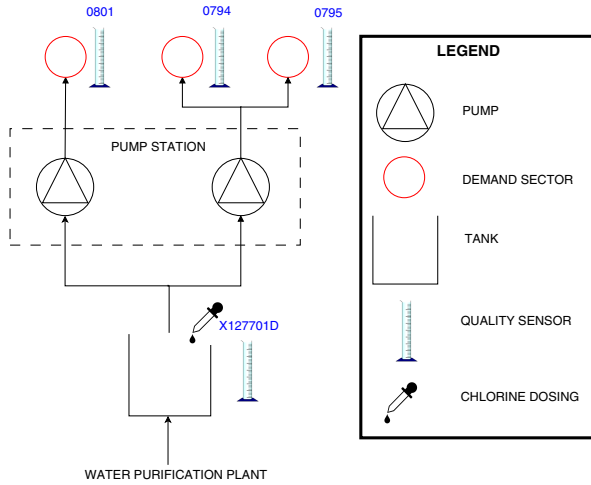The methodology presented next has been inspired on the knowledge of the experts used to analyze, check and

Fig. 1. *Water Demand Sector* from the Barcelona Water Network.

TABLE II
FAULT SIGNATURES BASED ON THE MODELS RESIDUALS.

PD	ANN	$\overline{PD} \wedge$ ANN	Cause
1	1	0	Distribution sensor fault
1	0	0	Distribution sensor fault
0	1	1	Quality event
0	0	0	Normal state

even forecast problems in the water quality sensors.

III. METHODOLOGY

The methodology presented consists in two modules: fault diagnosis and fault prognosis. Fault diagnosis detects a sensor fault, discerning if it has occurred a quality event. And the prognosis module tries to forecast the sensor remaining useful life (RUL) before the fault occurs.

This section describes and analyze the procedure followed by both modules. On the one hand, the diagnosis module, as we discussed before, has two objectives: the first objective is to detect changes in the water quality parameters that can compromise the safety of the water supplied, and the second objective is to discriminate if the problem detected is a real change in the water quality parameters or whether it has been generated by unreliable observations due to some of the problems presented in Table I. On the other hand, the prognosis module forecasts the RUL of the quality sensors, giving a valuable information to the WU to apply predictive maintenance.

A. Diagnosis

The detection of the event is done using a time series model (TSM) based on an Artificial Neural Network (ANN) and used to detect unexpected changes using historical knowledge. The ANN uses an autoregressive model using past observations as inputs with the predicted value as outcome.

This ANN model can detect abrupt changes in the pattern, but it is not able to distinguish if the change in the pattern is due to a sensor fault, i.e. unreliable data collected, or a quality event in the WDS.

Hence, spatial information is required to contrast the events detected against additional information provided by other sensors related.

Thus, the spatial model considered to provide this additional information is the predecessor (PD) model. The PD model measures the coherence between the sensor analyzed and the predecessor sensor. In our case study, the predecessor sensor is an on-line chlorine analyzer placed in the chlorination control loop.

Indeed, this is the procedure followed by the WU experts. First, they look for anomalous behaviours in the signals and next they validate their conclusions looking for information from other sensors hydraulically related to conclude if it is only a sensor problem or a real water quality problem.

Combining all this knowledge, a fault diagnosis scheme is developed to interpret the combination of the results and provide the key indicators to the WU to improve and anticipate the sensors maintenance operations.

Each model-based test generates a 1 if the residual is within the model threshold and 0 otherwise. The lower bound θ^{LB} and upper bound θ^{UB} are estimated based on the following expression:

$$\theta^{LB} = Q_1 - 3 \cdot IQR$$
$$\theta^{UB} = Q_3 + 3 \cdot IQR \tag{2}$$

where Q_1 and Q_3 are the first and third quartiles respectively, and IQR is the interquartile range (difference between the third and first quartiles) obtained from the residuals of the validation data set.

The fault diagnosis system can be formalized as a discrete-event system. Figure 2 presents the state diagram. From the normal state, there are two possible outcomes: a quality event or a sensor fault. When a sensor fault is detected, a maintenance operation is performed. A quality event can be caused by an intended action (e.g. hydraulic action, chlorine reference change) or by some unexpected infiltration.

The states are characterized in the Table II as a function of the activation of model-based tests, except the calibration state which is clearly known by the WU maintenance department.
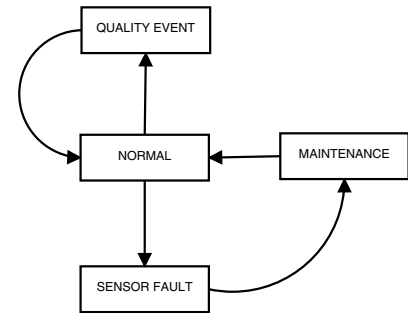

Fig. 2. State diagram of a quality sensor.

As detailed in Table II, a sensor is in normal state when all the tests are not active. A quality event is diagnosed when PD test is not active and ANN is active. When PD test is activated, a sensor fault is diagnosed, regardless of the ANN test.

B. Prognosis

This module forecast the Remaining Useful Life (RUL) based on a predetermined Failure Threshold (FT). As proposed in [11], the RUL is given by:

$$\text{RUL} \in \mathbb{N} \mid \hat{y}(t + \text{RUL}|t) = \text{FT} \qquad (3)$$

where $\hat{y}(t + \text{RUL}|t)$ is the RUL-step ahead forecast at time t of a given predictive model $\hat{y}$.

A data-driven approach is used to derive the predictive models from the data collected. Three different methods for multi-step forecast the chlorine decay are compared: Brown's double exponential smoothing, drift and Holt's linear filter.

The Brown's Double Exponential Smoothing can be expressed as follows

$$y_1(t) = \alpha y(t) + (1 - \alpha) y_1(t - 1) \qquad (4)$$

$$y_2(t) = \alpha y_1(t) + (1 - \alpha) y_2(t - 1) \qquad (5)$$

$$\hat{y}(t + h|t) = \left(2 + \alpha \frac{h}{1 - \alpha}\right) y_1(t) - \left(1 + \alpha \frac{h}{1 - \alpha}\right) y_2(t) \qquad (6)$$

where h is the forecast horizon and α the smoothing parameter. In the results section, the α parameter is arbitrary selected.

The drift model is a simple way to estimate the change over time from a set of observations. Indeed, it estimates the drift between the first observation and the last one. The drift's forecast expression is as follows

$$\hat{y}(t + h|t) = y_t + h \left(\frac{y_t - y_{t-m}}{m} \right) \qquad (7)$$

where m is the size of the window of observations.

Finally, the Holt's linear method is considered, but in the state-space model form. [16] presents a classification of exponential smoothing methods and the associated state space model for each class. They are classified using the triplet (E,T,S), where trend (T) and seasonal (S) are the components of a time series and the error (E) is the unpredictable component. Each component can be considered none (N), additive (A) or multiplicative (M). In this work we use the model ETS(A,A,N), with additive error, additive trend and none seasonality. Actually, this corresponds with the Holt's linear method with additive error.

The state-space general representation has the following form

$$y_t = w' \boldsymbol{x}_{t_1} + \epsilon_t, \qquad (8a)$$

$$\boldsymbol{x}_t = \boldsymbol{F} \boldsymbol{x}_{t-1} + g \epsilon_t, \qquad (8b)$$

where y_t is the observed value at time t and $\boldsymbol{x}_t$ is the state vector. The state vector $\boldsymbol{x}_t = [l_t \quad b_t]'$ is composed by the

level l_t and the growth rate b_t, $w = [1 \quad 1]'$, $F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $g = [\alpha \quad \beta]'$.

These parameters α and β are estimated using the maximum likelihood estimators.

The forecast of the state-space model ETS(A,A,N) is expressed as follows

$$\hat{y}(t + h|t) = \begin{bmatrix} 1 \\ h \end{bmatrix} \boldsymbol{x}_t + \epsilon_t \qquad (9)$$

IV. RESULTS

In this section, results based on the Barcelona case study described in Section II are presented next to show the performance of the methodology proposed in this work.

The data used to generate the results come from the multi-parametric (chlorine, pH, temperature and conductivity) sensors (0794, 0795 and 0801), the chlorine analyzer X127701D and the incidences reported by the WU experts to the maintenance department.

The models have been calibrated with data sets different to the faults scenarios. The Drift model has only one parameter, m, that is set to 48 samples (hours) for all the scenarios. The Brown's double exponential smoothing has the α parameter, that is set to 0.05 for all the scenarios. This value has been estimated such that good forecasts of the RUL are provided. Finally, the ETS model has the parameters: α, β, h, and the initial state $\boldsymbol{x}_0 = [l_0 \quad b_0]'$, where l_0 is set to be the intercept of the linear trend computed on the first 10 observations, and b_0 is set to the slope of the trend. Hence, α, β and the initial state $\boldsymbol{x}_0$ are refined estimating the maximum likelihood, as expressed in [16]. The horizon h is set to 1 during the calibration of the three models.

The following results correspond to three scenarios (named A, B and C) based on similar types of faults: incipient decay of chlorine observed due to a loss of sensitivity of the sensor. The value $FT = 0.2$ is fixed as the minimum chlorine limit that corresponds with $RUL = 0$. Hence, at a given time instant using the prediction models presented before, the methodology determines when the chlorine sensor will arrive to FT.

To evaluate the prognosis models performance the Prognostic Horizon (PH), as proposed in [17], is used

$$PH = t_{\text{EOL}} - i \qquad (10)$$

that is the difference between the discrete time instant time index of actual end of life t_{EOL} and the one computed as follows

$$i = \min \left(j | (j \in h) \wedge (R_j \subset R_k^*) \quad \forall k \in [j, t_{\text{EOL}}] \right) \quad (11)$$

where h is the set of all time instants when a forecast is made, R_j is the predicted RUL according to (3) at time j, and $R_k^* = [t_{\text{EOL}}(1 - \varepsilon) - t(k), \quad t_{\text{EOL}}(1 + \varepsilon) - t(k)]$ are the ε-bounds of the actual RUL, i.e. the allowable prediction error. In this work, we set $\varepsilon = 0.05$ to allow a $\pm 5\%$ of error in the predicted RUL.

Figure 3 shows the chlorine signals of the multi-parametric sensor V0794 and the chlorine analyzer VX127701D. As we can see, from $t = 0$ until $t = 350$ the chlorine signal V0794 follows the reference VX127701D, with some offset due to the physical distance between them. But, from $t = 350$, the sensor V0794 starts to decay arriving to 0.2 in about 14 days. Figure 4 shows the actual RUL (black solid line) and the forecast RUL using the prognosis module using three forecasting models proposed in Section III-B. The Brown model has a $PH = 8$, Drift model with $PH = 19$ and ETS with $PH = 9$.
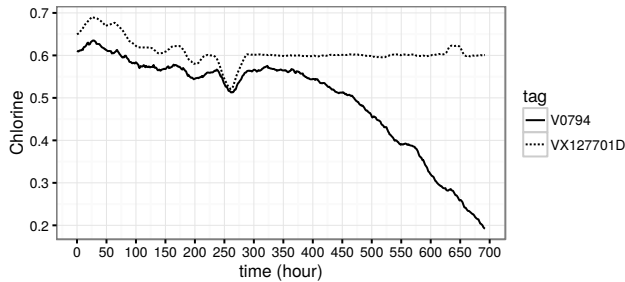
Fig. 3. Scenario A. Chlorine signal from V0794 sensor and VX127701D.
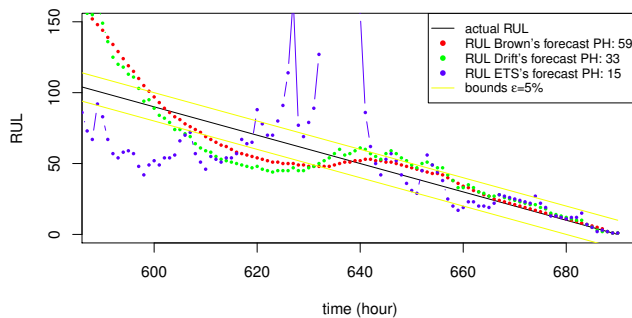
Fig. 4. Scenario A. RUL of the Chlorine signal from V0794 sensor and VX127701D.

Figure 5 shows another scenario corresponding to an incipient fault of the sensor V0795. The decay is faster than in the previous scenarios, arriving to FT in 5 days. Figure 6 shows the actual and predicted RULs of the three models. The Brown's PH is 8, Drift's PH is 18 and ETS's PH is 9.
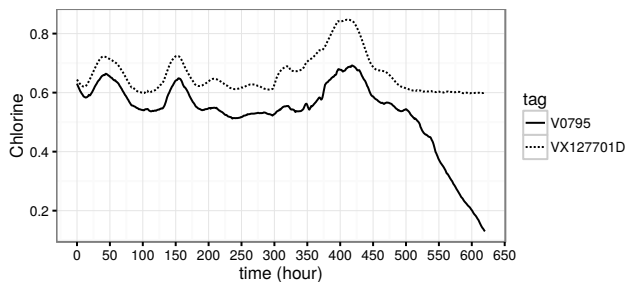

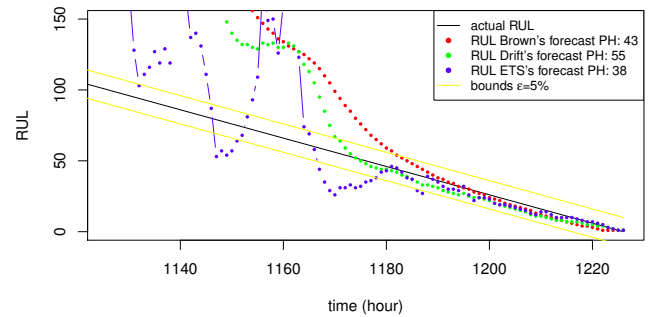Fig. 5. Scenario B. Chlorine signal from V0795 sensor and VX127701D.

Fig. 6. Scenario B. RUL of the Chlorine signal from 0795 sensor and X127701D.

Figure 7 shows a similar scenario as the previous one, an incipient fault in the sensor V0795. The chlorine decay is fast as the previous scenario, arriving to FT in around 5 days. And Figure 8 shows the actual and forecast RULs, where the Brown model has $PH = 11$, Drift model with $PH = 14$ and ETS with $PH = 12$.
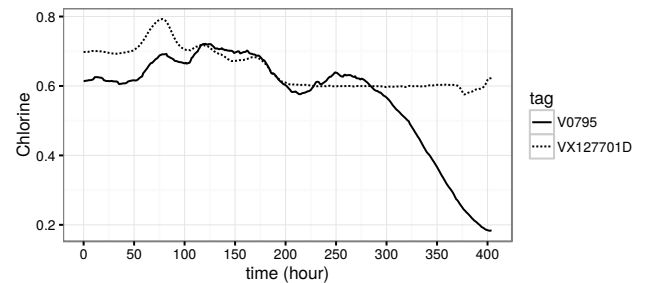

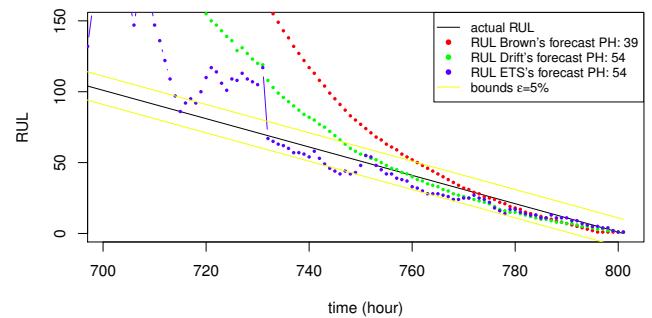Fig. 7. Scenario C. Chlorine signal from 0795 sensor and X127701D.

Fig. 8. Scenario C. RUL of the Chlorine signal from 0795 sensor and X127701D.

Table III lists the 4 scenarios and the PH results detailed previously to compare the forecast models' performance.

V. Conclusions

This paper has presented an approach to fault diagnosis and prognosis faults in water quality sensors. Different approaches to forecast the RUL of a water quality sensor have been analyzed and compared. A fault diagnosis algorithm has been developed able to distinguish between water

TABLE III
PH RESULTS OF THE PROGNOSIS MODELS

Scenario	Brown PH	Drift PH	ETS PH
A	59	33	15
B	43	55	38
C	39	54	54

quality events and problems affecting the sensors such as loss of sensitivity.

A prognosis algorithm, comparing three different forecasting methods, has been developed to predict the RUL of water quality sensors suffering loss of sensitivity.

This approach has been applied to the Barcelona Water Network and the results obtained show that the methodology detailed is able to anticipate the detection of future problems in chlorine sensors compared to the visual analysis applied by WU experts.

In particular, the Drift and Brown methods have shown in two of three scenarios slightly better performance than the ETS method. These two methods allow to anticipate, around 50 hours ahead, the loss of total sensitivity of the chlorine sensors.

Hence, the proposed approach improves the water quality control management with predictive maintenance thus reducing corrective maintenance actions on the Barcelona Water Network.

As a future research, it is planned to extend the proposed methodology to bound the uncertainty of the prognosis models.

REFERENCES

[1] J. Bartram, R. Ballance, W. H. Organization, and U. N. E. Programme, "Water quality monitoring : a practical guide to the design and implementation of freshwater quality studies and monitoring programs / edited by Jamie Bartram and Richard Ballance," 1996. [Online]. Available: http://www.who.int/iris/handle/10665/41851

[2] J. C. Powell, N. B. Hallam, J. R. West, C. F. Forster, and J. Simms, "Factors which control bulk chlorine decay rates," *Water Research*, vol. 34, no. 1, pp. 117–126, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0043135499000974

[3] D. Byer and K. H. Carlson, "Expanded Summary: Real-time detection of intentional chemical contamination in the distribution system," *Journal / American Water Works Association*, vol. 97, no. 7, pp. 130–133, 2005.

[4] D. Hou, S. Liu, J. Zhang, F. Chen, P. Huang, and G. Zhang, "Online Monitoring of Water-Quality Anomaly in Water Distribution Systems Based on Probabilistic Principal Component Analysis by UV-Vis Absorption Spectroscopy," *Journal of Spectroscopy*, vol. 2014, 2014.

[5] D. Eliades, T. Lambrou, C. Panayiotou, and M. Polycarpou, "Contamination Event Detection in Water Distribution Systems Using a Model-based Approach," *Procedia Engineering*, vol. 89, pp. 1089–1096, 2014. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1877705814023443

[6] B. Y. J. Hall, A. D. Zaffiro, and J. G. Herrmann, "System Contamination," no. January, pp. 66–77, 2007.

[7] D. B. Hart, S. a. McKenna, R. Murray, and T. Haxton, "Combining Water Quality and Operational Data for Improved Event Detection," *Water Distribution Systems Analysis 2010*, pp. 287–295, 2011. [Online]. Available: http://ascelibrary.org/doi/abs/10.1061/41203%28425%2926

[8] A. Ba and S. a. McKenna, "Water quality monitoring with online change-point detection methods," *Journal of Hydroinformatics*, vol. 17, no. 1, p. 7, 2015. [Online]. Available: http://www.iwaponline.com/jh/017/jh0170007.htm

[9] S. Rathi and R. Gupta, "Sensor Placement Methods for Contamination Detection in Water Distribution Networks: A Review," *Procedia Engineering*, vol. 89, pp. 181–188, 2014. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1877705814022905

[10] T. Escobet, J. Quevedo, V. Puig, and F. Nejjari, "Combining Health Monitoring and Control," *Diagnostics and Prognostics of Engineering Systems: Methods and Techniques*, no. 2002, pp. 230–255, 2012.

[11] T. Escobet, J. Quevedo, and V. Puig, "A Fault / Anomaly System Prognosis using a Data- driven Approach considering Uncertainty," *IEEE World Congress on Computational Intelligence*, pp. 10–15, 2012.

[12] T. Escobet, V. Puig, J. Quevedo, and D. Garcia, "A methodology for incipient fault detection," *2014 IEEE Conference on Control Applications, CCA 2014*, pp. 104–109, 2014.

[13] I. Roychoudhury, G. Biswas, and X. Koutsoukos, "A Bayesian approach to efficient diagnosis of incipient faults," *Proceedings of the 17th . . .*, vol. 1, no. Dx, 2006.

[14] Y. Zhang and J. Jiang, "Bibliographical review on reconfigurable fault-tolerant control systems," *Annual Reviews in Control*, vol. 32, no. 2, pp. 229–252, 2008.

[15] R. P. Brent, *Algorithms for Minimization Without Derivatives*, N. Englewood Cliffs, Ed. Prentice Hall, 1973.

[16] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose, "A state space framework for automatic forecasting using exponential smoothing methods," *International Journal of Forecasting*, vol. 18, no. 3, pp. 439–454, jul 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169207001001108

[17] A. Saxena, J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, and M. Schwabacher, "Metrics for evaluating performance of prognostic techniques," *2008 International Conference on Prognostics and Health Management, PHM 2008*, 2008.