Making Use of Continuous Measurements for Change Detection Purposes: an Application to Water Distribution Networks

Andrea Scozzari
Institute of Information Science and Technologies,
National Research Council of Italy,
Via Moruzzi 1,
56124 Pisa, Italy
a.scozzari@isti.cnr.it

Gianpiero Brozzo
ACAM Acque S.p.A.,
Via Picco 22,
19124 La Spezia; Italy
gianpiero.brozzo@acamspa.com

***Abstract*— The monitoring and control of industrial processes often requires the capability to identify particular patterns in a set of acquired measurements. One of the most recurrent applications consists in the detection of changes and the related minimization of false alarms. This requirement is true also when dealing with natural systems. The monitoring of a natural resource usually involves the identification of a set of parameters, which are considered as representative of its underlining processes, in order to extract useful information about its current status and its expected behavior. This work is focused on the water resources destined to the drinkable water distribution, paying attention to two particular aspects: i) the need for a suitable metric to detect anomalous values in the assessment of water quality indicators; ii) the experimentation of a simplified data-driven strategy to estimate natural variations of one or more indicators, in order to mitigate false alarms. This paper proposes a preliminary investigation and a selected case study, in order to exemplify one practical implementation of the proposed approach. The possible application to a context of low-cost distributed sensors is also briefly discussed.**

Keywords— Environmental measurements, data-driven modelling, drinking water

I. INTRODUCTION

There are currently many applicative frameworks among the Instrumentation & Measurement disciplines, in which the real-time or near-real-time detection of significant changes in a process under monitoring represents one of the most demanding exercises. The exploitation of a natural resource like water represents a particular case, due to the peculiarities of the natural process under observation, moreover to the critical aspects connected with the quality and quantity of the resource when destined to the human consumption.

Nowadays, in the European Union, in the United States and in many other countries, the majority of drinkable water supplies rely on groundwater, and this happens at an increasing rate [1, 2]. This fact introduces a new need for the deployment of in situ monitoring instrumentation, to perform a direct observation of the natural process at the catchment point and even upstream. In this case, the surveillance of the abstracted resource goes beyond the mere distribution network and deals directly with the natural and/or anthropic processes influencing the availability and quality of the water. Such interconnection between the relevant parameters characterizing the resource and its surrounding environment makes the definition of an optimal sensor distribution and the interpretation of the collected data a challenging problem, which generated much literature in the sector of hydrological measurements [3, 4] as well as in the context of water distribution systems [5, 6, 7, 8, 9], which is closer to the scope of this paper. It's also important to note that there's already a well-established literature concerning water consumption and the detection of flow anomalies in water distribution networks [10, 11]. To probe further the relatively limited literature listed here, the interested reader may find many additional references in the cited papers.

In this context, an important and recurrent topic is the development of efficient criteria for the detection of anomalies, which may be used for raising alarms and/or taking appropriate actions (e.g., pump switching, valve operations, automatic sampling, etc.). For what regards the threats to the quality of water, this concept is generally applicable to a wide range of accidents, criminal actions or natural contaminations, and it's also true in front of the recent discussions and projects about CBRN (Chemical, Biological, Radiological and Nuclear) threats [12, 13, 14].

In addition to the field instrumentation based on standard analytical methods, other promising sensing systems selective to particular chemical/biological threats have been proposed [15, 16, 17], and also non-specific sets of sensors have been investigated in the literature, in particular for what regards change-detection purposes [18, 19]. These devices are typically suggested in the perspective of low-cost distributed sensor networks, where the change-detection capability aims at covering the broadest possible range of potential threats, and the degree of pervasivity is intended to reach even the single water tap in a drinkable water network. Non-specific equipment may implement a wide variety of sensing techniques, which are characterized by different degrees of maturity. A non-exhaustive list may include conventional physico-chemical measurements, innovative electromagnetic sensing systems [20], and e-tongues based on electrochemical, optical and even electro-acoustic principles [21, 22, 23]. This aspect is also giving new value to traditional and very mature sensing techniques like in the case of conductimetry, ion-selective electrodes (e.g., pH, ORP etc.) and basic optical methods, which are applied to the context of new miniaturized sensors and proposed in applicative projects [24].

One fundamental aspect in the development of contamination warning systems is the need to define a metric for a proper assessment of anomalies, which has to minimize both false alarms and missing detections. This issue is even more severe when dealing with a network of inexpensive distributed sensors, like those mentioned above. In addition to the observed changes due to normal operations in the water distribution network, also those variations at source level which happen in the absence of contamination events (i.e., in the natural domain) must be taken into account and interpreted as baseline-changes. Thus, when dealing with parameters characterized by a significant degree of natural variability (e.g., due to meteo-climatic conditions or to different levels of exploitation), it is appropriate to define anomalous values in terms of distance from an expected value.

The practical application of this concept requires prior knowledge about the mechanisms underlining the quality and quantity of the resource, including its response to natural or anthropic inputs. Such knowledge is typically hard to determine with physical approaches, due to the complex nature of the processes and to the lack of exhaustive information (e.g. geological and geochemical data, etc.), in order to build an effective model for this purposes. In this case, system identification techniques may deliver models and data-driven strategies are often proposed [25]. A well-known example of an event detection system for the monitoring of water quality is represented by the US EPA CANARY software package [26], which bases its data analysis tools on both historical and simulated data from multiple sensors.

Despite the fact that the data-driven approach may look particularly attractive, practical operating conditions make it often complicated, due to incomplete datasets, missing data or too short time series to be representative of the natural variability (mostly seasonal) of the observed processes. Thus, the adopted approach should be determined on a case-by-case basis, according to the available data, the peculiarities of the resource, and the characteristics of the distribution network. Moreover, since historical data sets are often missing anomalous events, it may be reasonable to use a mixed strategy, in which contamination scenarios are simulated by simplified physical models, in order to create a statistically relevant set of contamination examples, suitable to feed a data-driven detection algorithm. A similar mechanism is described in [26].

In our investigation we made use of the measurements collected in few years of monitoring activity, in order to assess the preliminary feasibility of both a purely data-driven method and of a mixed one. For reasons of brevity, this paper limits its scope to one specific experiment, based on a data-driven approach. In particular, here we explore the possibility to embed a highly simplified model in the measurement strategy, to be potentially ported to a light computing system. After describing the selected case study and the modelling technique adopted, results are briefly presented and discussed, focusing on the possible industrial interest (i.e., water utilities) into the experimentation and usage of such techniques.

II. The Selected Case Study

The experimental dataset used for this work regards the alluvial aquifer of the Magra River (Italy). Two distinct abstraction sites, next to the villages of Battifollo and Fornola, are managed by the local drinkable water distribution company (ACAM Acque), which serves the area of La Spezia (about 150,000 people), on the Ligurian coast. The main clusters are in Fornola, which hosts 44 wells grouped in 7 clusters, totalizing an abstraction rate of about 900 liters per second.

Quantitative measurements about the abstraction activity are collected by the in situ instrumentation mounted on the wells' clusters and archived by a centralized SCADA. The available data comprise the following parameters: hydraulic head (level) in reference piezometers, status of the electrical pumps (i.e., percentage of running time), and total abstracted water flow. For the purpose of this work, daily averages have been taken, being the natural variations to be observed relatively slow with respect to a daily sampling rate. This work focuses on the longest available sequence of validated data, which covers the period between 14 December 2010 to 17 October 2016.

In addition to the archived data acquired by the process instrumentation, other field measurements were performed, partly with portable instrumentation and partly by laboratory analyses. The measured parameters regarded the assessment of water quality, including physico-chemical features (e.g., conductivity, pH and temperature) and the major dissolved ions. These measurements were carried out in various sampling points, both inside the city and in the proximity of the clusters, with an irregular sampling interval.

Groundwater in this area is characterized by a remarkable seasonal variability [27], connected with rainfall events. In particular, the delivered water exhibits a variable degree of mineralization, mostly involving two chemical species: SO_4 and Cl, thus impacting the electrical conductivity and the overall quality of the water, as it will be better discussed in section III. In addition, water does not undergo any particular treatment, which can potentially modify its main physico-chemical features before being distributed. Thus, the experimentation described in this paper is based on the hypothesis that observed (or predicted) variations at source level are significantly reflected by the quality of the water at the points of delivery.

III. The Proposed Data-Driven Scheme

A. *Materials and methods*

Chemical and physico-chemical measurements often generate multiple variables, sometimes posing problems of dimensionality reduction. This is especially true when dealing with arrays of sensors to assess the quality of water [21] or electrochemical methods (e.g. voltammetry) generating high dimensionality signals [18]. The fundamental idea is to develop and test a data-driven modelling scheme based on some kind of machine learning, in order to approximate the output parameter(s) and determine the expected values, to be compared with the actual measurements in a change-detection framework. This study focuses on those relatively slow changes that are pertinent to the natural domain, thus generating a kind of ‘moving baseline’. This is a building-block of the whole detection system that is comparatively less covered in the literature. Other components of the system, like the detection of fast variations connected with sudden events and the mitigation of false alarms due to network activities (e.g., valves, tanks or pumps operations) are outside the scope of this paper.

The first simplifying action done in this investigation consisted in the selection of one single parameter, among those measured by the water utility, which may be considered as an overall indicator of the quality of water. Given the dataset described in section II, the parameter suitable for the purpose of this experiment has been individuated in the electrical conductivity of the analyzed water. The observed fluctuations in the chemistry of water can be briefly justified as follows: a highly mineralized (thus highly conductive), deep and almost-constant groundwater flow, is mixed with a shallow and highly variable component, mostly depending on rainfall events (poorly mineralized and low-conductivity water). The groundwater flow and the piezometric level are related by well-known hydraulic laws, thus, we can conceive to infer the electrical conductivity of the water by observing its level variations. Actually, the groundwater level reflects both seasonality and exploitation effects, carrying useful information about both. Field measurements were performed with a portable meter WTW model 340i, configured for simultaneous conductivity and pH measurements.

Electrical conductivity is thus treated as an output variable of an empirical I/O model. As a consequence, a piezometric level was used as an input variable. In particular, we chose the level measured at the wells cluster “C” by a standard down-well pressure transducer and recorded by the SCADA system. It is important to underline that the approach proposed in this study can be applicable also to one or more virtual parameters generated by extracting relevant features from a high-dimensionality space, such as the one generated by non-specific sensor systems.

B. The simplified implementation

As a general approach, an adaptive system is trained and validated on a dataset comprising input and output time series derived from archived data. Figure 1 shows a general flowchart, which includes multiple I/O parameters, in addition to the cited piezometric level h_w and electrical conductivity σ_w. Additional input parameters may include meteo-climatic measurements, water flow rate etc. Additional output parameters can be the estimated concentrations of specific chemical species, like Cl and SO_4 in our case. Figures 1a and 1b show respectively the data paths for the training and the simulation phases of a generic adaptive algorithm. Details will be given further in this section.

The modelling strategy used for this investigation undergoes few simplifying steps:

- Under a strong hypothesis of linearity and time-invariance (LTI), a phase relationship between input and output parameters is defined, and expressed by a constant delay
- The LTI assumption permits the definition of an impulse-response and to resort to transformed frequency domains for characterizing the system
- Non-linearity is introduced by means of a Multi-Layer Perceptron Neural Network (MLP-NN), used as a non-linear approximator
- The input and output datasets are reduced to a single variable, both at the input and at the output.

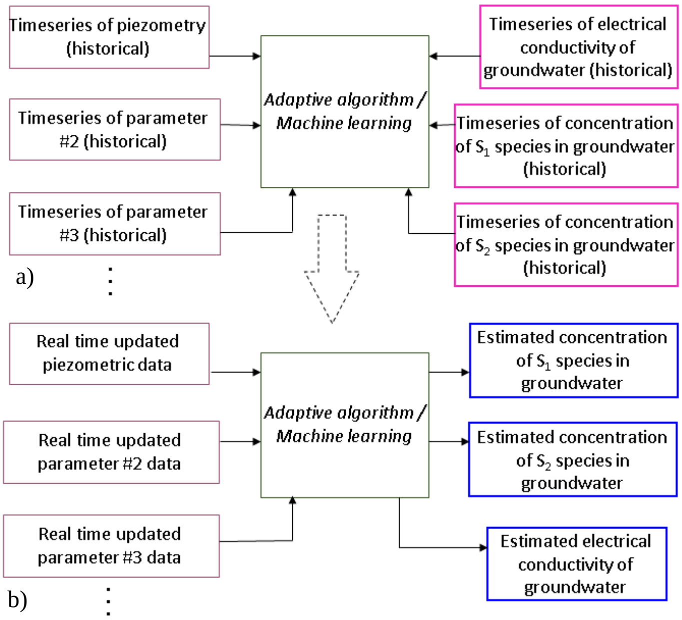

Fig. 1a, b. General scheme of the data-driven modelling approach.

Continuous measurements of electrical conductivity of water at selected taps in the city and in the proximity of the clusters would represent an ideal dataset for this experiment. The dataset used, based on sparse measurements conducted with portable instrumentation, is a sub-optimal alternative, which requires some pre-processing as described in Figure 2.

For what regards the conductimetry data, multiple measurements performed on the same day are averaged and outliers removed. Then, data are interpolated according to a piecewise cubic Hermite interpolating polynomial, in order to preserve sampled values and the continuity of the first derivative. Thus, the time series $\sigma_w(\mathrm{i})^{\mathrm{RS}}$ is obtained after regular re-sampling of the interpolated signal. The piezometric time series undergoes a bandwidth limitation stage, in order to mitigate possible overfitting due to training the NN with the $\sigma_w(\mathrm{i})^{\mathrm{RS}}$ signal as a target. In particular, the bandwidth of the piezometric signal has been limited to the frequency interval containing 99% of the conductimetric signal power. Bandwidth limitation is done in the DCT (Discrete Cosine Transform) domain [28], a real transform suitable for real and causal signals, which exhibits high compaction properties. Filtering in the transformed domain is done by using a rectangular window. The bandwidth-limited time series is named $h_w(\mathrm{i})^{\mathrm{BL}}$. Signals are then split into Low-Pass (LP) and High-Pass components, again by selecting the low-frequency components in the DCT domain with a rectangular window, followed by inverse transformation. The residual signals in the time domain represent the High-Pass components, as shown in Figure 2.

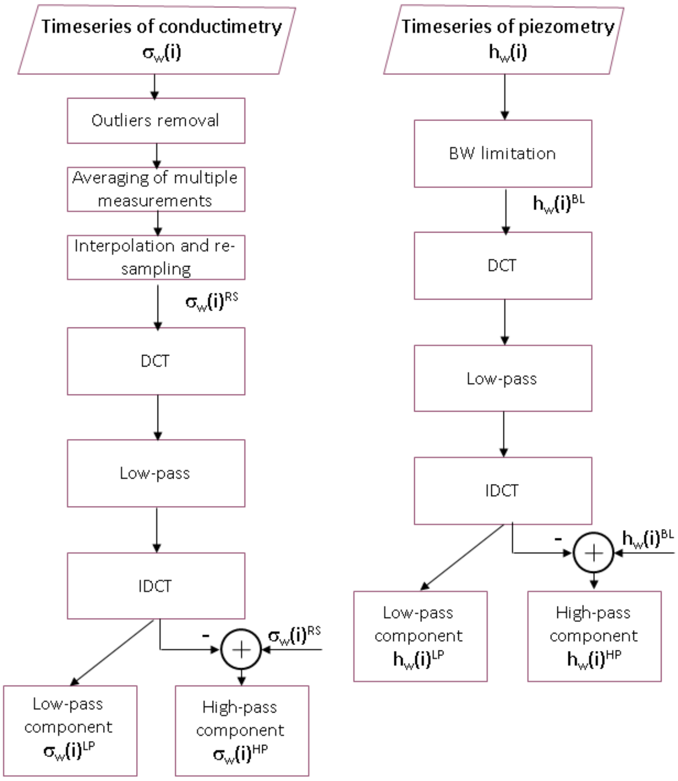

Fig. 2. Data pre-processing.

Resulting signals ($\sigma_w(i)^{LP}$, $\sigma_w(i)^{HP,}$ $h_w(i)^{LP}$, $h_w(i)^{HP}$) are then fed to two dedicated neural networks, for the LP and HP components (Figure 3). In particular, the non-linear approximation is based on a classical Multi-Layer Perceptron Neural Network (MLP-NN) architecture [29, 30], with one hidden layer. A non-linear (*tansig*) firing function was assigned to the neurons in the hidden layer and a linear firing function was used for the output neuron. The MLP-NN was implemented by using the Neural Networks toolbox in MATLAB 2014a. The NN input (piezometry data) can be seen in a vectorial form, having dimensionality equal to the number of past samples of the input parameter plus one:

$$\boldsymbol{H_w}(i) = [h_w(i), h_w(i-1) \dots, h_w(i-PS)] \qquad (1)$$

where *PS* is the number of past samples fed to the input, in addition to the current one (the *i-th*). Two different sets of vectors $\boldsymbol{H_w}$ are formed for the LP and HP components (Figure 3). The output of each MLP-NN is a scalar, representing the respective component of the current estimated value of the electrical conductivity of the water ($\sigma_w^{EST}(i)$ in Figure 3b).

Multiple NNs have been generated, with a variable number of neurons in the hidden layer from 1 to *NNMAX*. In addition, *N*=10 different nets have been generated per each number of neurons, with randomly selected initial biases, weights and validation sub-sets. In fact, the dataset was randomly split for each run into a training and a validation sub-sets, containing 85% and 15% of the whole data, respectively. Moreover, the last 547 days (approximately one year and half) of the whole dataset have been kept out of the training exercise as a completely independent test set. In this study, the bandwidth of the LP signal has been limited to (6 months)$^{-1}$, *NNMAX* has been set to 10, and a total of *N*· *NNMAX* = 100 neural networks have been generated for both the LP and the HP sub-domains.

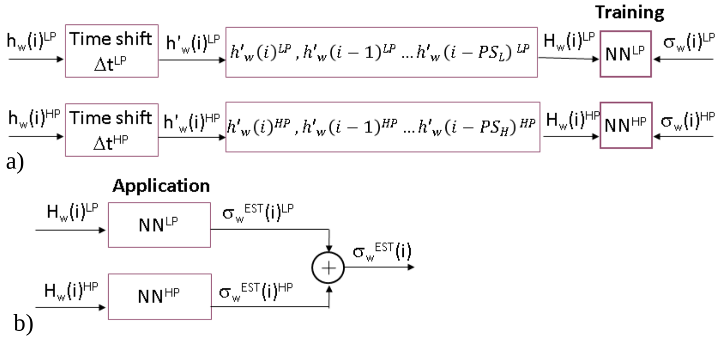

Fig. 3a, b. Implementation of the MLP-NN to the processed signals.

Determination of the number of past samples and of the applied time shift is based on assumptions derived by observing the cross-covariance between the input and output signals, i.e.:

$$\varphi(m) = \sum_n [(\sigma_w(n+m) - \overline{\sigma_w}) \cdot (h_w(n) - \overline{h_w})] \qquad (2)$$

where *m* is the discrete lag and *n* spans all the time series.

Figure 4 shows a plot of cross-covariance relating to the low-pass components, where *m* is expressed in days. Here, estimations are made by exploiting cross-covariance lobes. In particular, the constant time delay is associated to the lag corresponding to a maximum of $\phi(m)$, and the number of past samples is calculated as a half width of the lobe across *m*=0 (red arrows in Figure 4). In the absence of a rigorous metric for this matter, this approach is essentially intended to take into account the aliasing of periodic components in the two time series and choose long enough sequences for carrying sufficient information to the NNs.

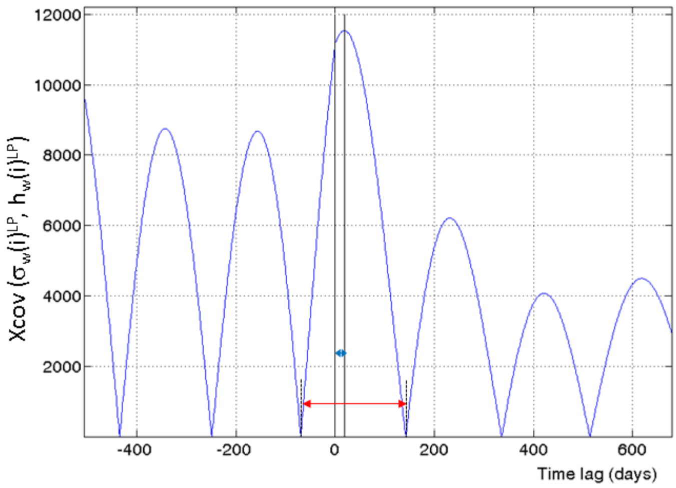

Fig. 4. Cross-covariance function calculated for the low-pass components.

In our case study, the time delay for the low-frequency components has been forced positive (1st maximum at *m*>0, vertical line in the picture) due to prior knowledge about the behavior of the natural domain, i.e., the seasonal effect on the σ_w signal is delayed with respect to h_w variations. The time lag for the high frequency components, instead, has been left free in the absence of prior knowledge. Each NN has been trained and validated (as briefly shown in Figure 3a) with randomly selected sub-sets, stopping each training phase upon decay of the validation performance (meaning loss of generalization) or when reaching a maximum number of training epochs [31]. This strategy generated 10000 combinations between LP and HP nets. A selection has been done by applying each pair to the independent test set (in the way described in Figure 3b), in order to assess and privilege the generalization capabilities of the nets. The best net was chosen by minimizing a performance index (*PI*), defined as the Euclidean distance between two vectors:

$$PI_{j,k} = \sum_{i=1}^{Ntest} \left[NET_{j,k}(H_w(i)) - \sigma_w(i)\right]^2 \tag{3}$$

where j and k are indexes to the j-th and k-th neural networks generated in the LP and HP sub-domains, respectively; $NET_{j,k}$ is a scalar (σ_w^{EST}(i) in Figure 3b), obtained by applying the j-th and k-th MLP-NNs to the input vector $\boldsymbol{H_w}(i)$; N_{test} is the number of samples in the test set.

IV. Results

The results shown in this section were obtained according to the modelling strategy and under the strong simplifications described in the section above.

A performance matrix was generated after applying the 100x100 network combinations to the independent test set, producing the results shown in Figure 5. Acronyms *NNLP* and *NNHP* used in the picture are equivalent to the indexes j and k in eq. (3). The absolute maximum (that is, the minimum of the *PI* matrix) is shown in the picture inside the dashed circle.

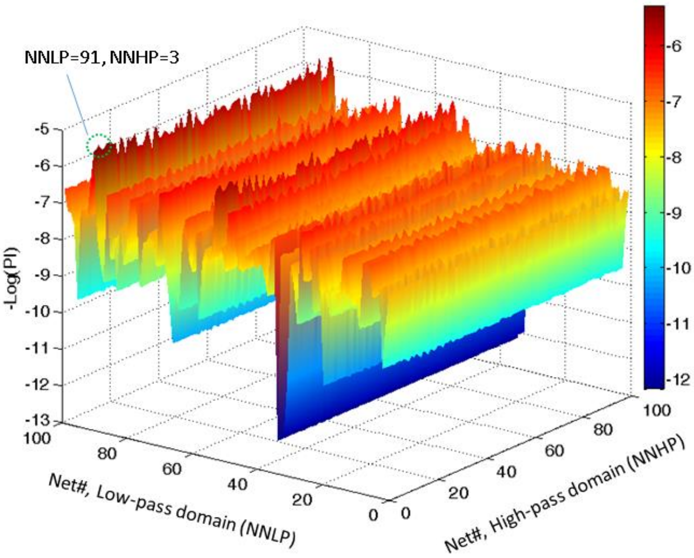

Fig. 5. 3D plot of the performance matrix. Note the –Log scale for representation purposes.

The best performing combination according to the assessed metric has been chosen, and applied to the independent test set. Figure 5 illustrates the model outputs (red line) versus conductivity measurements at their sampling times (black crosses). The x-axis represents the sequential day number inside the test set. This data sub-set has undergone the same pre-processing of the whole dataset as shown in Figure 2, and resulting $h_w(i)$ time series have been fed to the selected combination of networks.

As specified in the past sections of this paper, the water conductivity dataset was irregularly sampled in the time domain. Further uncertainty is introduced by the sparse and almost random measurement locations, which can reflect the effect of some particular features of the network (e.g. tanks operations and different water ages at different taps). In addition to instantaneous alterations of water quality, network operations may alter those delays that, in this first simplified approach, have been assumed as constant. Despite that, if compared to the performance of a change-detection scheme based on the distance of each sample σ_w(i) to the average value in the test set $\overline{\sigma_w}^{test}$, the approach defined in this study is still advantageous. In fact, by applying a distance threshold equal to one standard deviation of the σ_w(i) series, the number of "false alarms" given by this detection scheme would go down to 7 from the 13 given by the mere distance from the average value. In case of using twice the standard deviation as a threshold, false alarms would be completely eliminated. All the considerations above are referred to the independent test set.

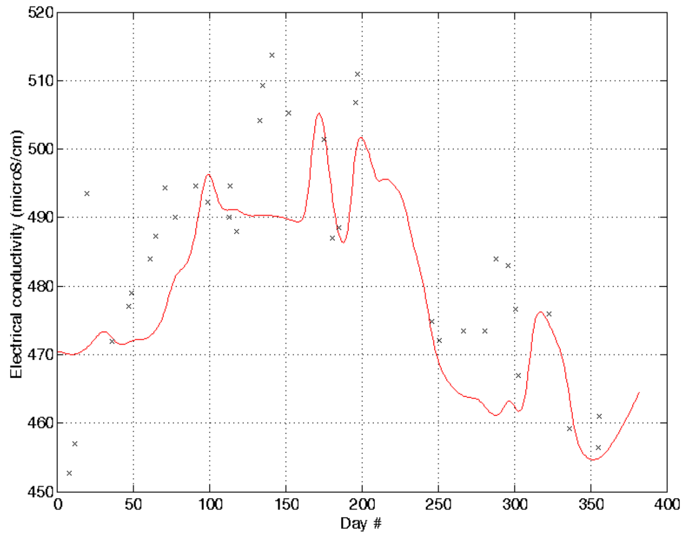

Fig. 5. Model output vs. conductivity measurements (independent test set). The red line shows the model output, the black crosses indicate the target measurements at their sampling times.

A significant improvement to the method may consist in installing online tap-mounted sensors, making periodic calibrations to them and embedding the measurements taken at calibration times in the training datasets. This would enhance the adaptation of the non-linear approximators to the process.

V. Conclusions

This paper describes the development of an experimental technique to process water quality measurements in a change-detection scheme. In this particular context, modelling exercises are essential tools in order to support the detection of those process alterations that can be symptoms of potential threats. Data-driven techniques are already offered in the literature for a wide range of applications, such as the identification and prediction of possible scenarios, the optimization of measurement schemes, and the estimation of output values. In this context, this work discusses a preliminary experience in the application of a simple embeddable building-block for a change-detection scheme, in the context of a water distribution network fed by a groundwater resource.

The encouraging results obtained in the simplest configuration (single input and output parameters) gives the perspective to use the same approach with more complete configurations, where multiple input variables may take into account both network operations and the peculiarities of each selected tap due to its position in the network.

Next step may consist in the experimentation with online sensors distributed on the net, calibrated and validated on a regular basis, in order to enhance and better evaluate the performance of the proposed approach.

ACKNOWLEDGEMENTS

Part of this activity was funded by the Italian national project ACQUASENSE (funding mechanism “INDUSTRIA2015”).

REFERENCES

[1] Zektser IS, Everett LG. Groundwater resources of the world and their use, vol 6, IHPVI, series on groundwater, UNESCO, Paris, 2004

[2] Margat J, van der Gun J. Groundwater around the world: a geographic synopsis, CRC/Balkema, Boca Raton, pp 376, ISBN 9781138000346, 2013

[3] Alfonso L, Chacon J, Solomatine D. Optimal design of hydrometric monitoring networks with dynamic components based on Information Theory, Geophysical Research Abstracts vol. 18, EGU2016-15822, 2016

[4] Park C, Kim SH, Telci IT and Aral MM. Designing optimal water quality monitoring network for river systems and application to a hypothetical river, IEEE Simulation Conference (WSC), Proceedings of the 2010 Winter, Baltimore, MD, USA, 2010, pp 3506-3513, doi:10.1109/WSC.2010.5679039

[5] Du R, Fischione C, Xiao M. Flowing with the water: on optimal monitoring of water distribution networks by mobile sensors, IEEE International Conference on Computer Communications (IEEE INFOCOM 16), San Francisco, CA, USA, 2016, doi:10.1109/INFOCOM.2016.7524480

[6] Palleti VR, Narasimhan S and Rengasamy R. Optimal sensor network design for early intrusion detection and identification in water distribution networks, 2016 IEEE Indian Control Conference (ICC), Hyderabad, 2016, pp 134-139, doi:10.1109/INDIANCC.2016.7441118

[7] Perelman L, Arad J, Oliker N, Ostfeld A and Housh M. Water distribution systems event detection, IEEE Conference on Complexity in Engineering (COMPENG), Aachen, Germany, 2012, pp 1-3, doi:10.1109/CompEng.2012.6242956

[8] Mounce SR, Mounce RB, Jackson T, Austin J and Boxall JB. Pattern matching and associative artificial neural networks for water distribution system time series data analysis, Journal of HydroInformatics,16(3):617-632, doi:http://www.iwaponline.com/jh/016/jh0160617.htm, 2014

[9] Mounce SR, Gaffney JW, Boult S and Boxall JB. Automated data driven approaches to evaluating and interpreting water quality time series data from water distribution system,. ASCE Journal of Water Resources Planning and Management, 141(11):1-11, doi:http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000533, 2015

[10] Laucelli D, Romano M, Savic D, Giustolisi O. Detecting anomalies in water distribution networks using EPR modelling paradigm, Journal of Hydroinformatics, 18(3) 409-427, doi:10.2166/hydro.2015.113, 2016

[11] Giuliani M, Cominola A, Alshaf A, Castelletti A and Anda M. Data-driven behavioural modelling of residential water consumption to inform water demand management strategies, Geophysical Research Abstracts vol. 18, EGU2016-12007, 2016

[12] Lundberg S, Paffenrot HR, Yosinski J. Analysis of CBRN sensor fusion methods, IEEE 13th International Conference on sensor fusion (FUSION 2010), IEEE Conference Publications, 2010

[13] EU FP7 SAFEWATER project, retrieved from: www.safewater-project.eu, last accessed on 9 November 2016

[14] Borchers U, Gray J, Thompson KC (eds), Water Contamination Emergencies: Managing the Threats, RSC Publishing, Cambridge, UK, 2013

[15] Neeley A, Khan SA, Beqa L, Zhen F, Singh AK, Wentong L, Senapati D, Arbneshi T, Eumin L, Anderson Y, Banerjee S, Jinghe M, Dubey M., Amirtharaj P, Ray P. Selective detection of chemical and biological toxins using gold-nanoparticle-based two-photon scattering assay, in IEEE Transactions on Nanotechnology, 10(1):26-34, doi:10.1109/TNANO.2010.2076340, 2011

[16] Santonico M, Scozzari A, Brozzo G, Marini L, D’Amico A, Filippini D, Lundström I, Di Natale C. Detection of natural Cr (VI) with computer screen photo-assisted technology, Procedia Chemistry, 1(1):317-320, Elsevier, 2009

[17] Rabe J, Buttgenbach S, Schroder J, Hauptmann P. Monolithic miniaturized quartz microbalance array and its application to chemical sensor systems for liquids, IEEE Sensors Journal, 3(4):361 - 368, doi:10.1109/JSEN.2003.815783, 2003

[18] Scozzari A, Acito N, Corsini G. A novel method based on voltammetry for the qualitative analysis of water. IEEE Trans Instrum Meas 56(6):2688-2697 doi:10.1109/TIM.2007.903600, 2007

[19] Lvova L, Martinelli E, Mazzone E, Pede A, Paolesse R, Di Natale C, D’Amico A. Electronic tongue based on an array of metallic potentiometric sensors, Talanta, 70:833–839, Elsevier, 2006

[20] Gennarelli G, Soldovieri F. Analysis of water pollution by a microwave cavity sensor, Sensor Letters 11(9):1561-1566, American Scientific Publishers, 2013

[21] Di Natale C, Dini F, Scozzari A. Non-conventional electrochemical and optical sensor systems, Hdb Env Chem. doi:10.1007/698_2013_254, Springer, 2013

[22] Cole M, Sehra G, Gardner JW, and Varadan VK. Development of smart tongue devices for measurement of liquid properties, IEEE Sensors J., 4(5):543–550, 2004

[23] Krantz-Rulcker C, Stenberg M, Winquist F and Lundstrom I. Electronic tongues for environmental monitoring based on sensor arrays and pattern recognition: a review. Anal. Chim. Acta, 426(2):217–226, 2001

[24] EU H2020 PROTEUS project, retrieved from: http://www.proteus-sensor.eu/, last accessed on 9 November 2016

[25] Laurain V, Gilson M, Benoıt M. Data-driven modeling for water resource quality over long term trends, 7th International Congress on Environmental Modelling and Software (iEMS), San Diego, California, USA, 2014

[26] CANARY Users’ Manual Version 4.3.2, EPA 600/R-08/040B, United States Environmental Protection Agency, 2012

[27] Menichini M, Doveri M, El Mansouri B, El Mezouary L, Lelli M, Raco B, Scozzari A. Groundwater flow controlled by rivers and implications with future climate scenarios: a predictive modelling for the Magra valley aquifer (southeast Liguria, Italy), 42nd IAH Congress, Rome, Italy, 2015

[28] Rao KR, Yip P. Discrete cosine transform: algorithms, advantages, applications, Academic Press Professional, San Diego, CA, USA, 1990

[29] Luo FL and Unbehauen R. Applied neural networks for signal processing, Cambridge Press, UK, 1997

[30] Hu YH. and Hwang JN. Handbook of neural network signal processing, CRC Press, 2001

[31] Beale MH, Hagan MT, Demuth HB. MATLAB Neural Network Toolbox User's Guide, The Mathworks, 2016