UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

UPCommons

Portal del coneixement obert de la UPC

http://upcommons.upc.edu/e-prints

Aquesta és una còpia de la versió *author's final draft* d'un article publicat a la revista *IFAC-PapersOnLine*.

URL d'aquest document a UPCommons E-prints:

http://hdl.handle.net/2117/102257

Article publicat / *Published paper*:

Quevedo, J., Pascual, J., Espin, S., Roquet, J. Data Validation and reconstruction for performance enhancement and maintenance of water networks. *IFAC-PapersOnLine*, 1 Gener 2016, vol. 49, núm. 28, p. 203-207. DOI10.1016/j.ifacol.2016.11.035

© 2017. Aquesta versió està disponible sota la llicència CC-BY-NCND 3.0 http://creativecommons.org/licenses/by-nc-nd/3.0/es/

Data Validation and Reconstruction for Performance Enhancement and Maintenance of Water Networks

J. Quevedo*, J. Pascual*. S. Espin and J. Roquet****

**Center of Supervision, Safety and Automatic Control (CS2AC),*
BarcelonaTech (UPC), Campus de Terrassa,
Sant Nebridi, 10 08222 Terrassa, Barcelona, Spain,
{joseba.quevedo, josep.pascual}@upc.edu
*** ATLL Concessionària de la Generalitat de Catalunya S.A.*
Sant Martí de l'Erm, 30. 08970 Sant Joan Despí, Barcelona, Spain

Abstract:

In a real water network, a telecontrol system must periodically acquire, store and validate data gathered by sensor measurements in order to achieve accurate monitoring of the whole network in real time. For each sensor measurement, data are usually represented by one-dimensional time series. These values, known as raw data, need to be validated before further use to assure the reliability of the results obtained when using them. In real operation, problems affecting the communication system, lack of reliability of sensors, or other inherent errors often arise, generating missing or false data during certain periods of time. These wrong data must be detected and replaced by estimated data. Thus, it is important to provide the data system with procedures that can detect such problems and assist the user in monitoring and processing the incoming data. Data validation is an essential step to improve data reliability. The validated data represent measurements of the variables in the required form where unnecessary information from raw data has been removed. In this paper, a methodology for data validation and reconstruction of sensor data in a water network is used to analyze the performance of the sectors of a water network. Finally, from this analysis several indicators of the components (sensors, actuators and pipes) and of the sectors themselves can be derived in order to organize useful plans for performance enhancement and maintenance. Nice practices have been developed during a large period in the water network of the company ATLL Concessionària de la Generalitat de Catalunya, S.A.

Keywords: Data Analytics, Data Validation, Data Estimation, Water Networks, Performance Index, Maintenance

1. INTRODUCTION

Data analytics is the science of examining raw data with the purpose of drawing conclusions about that information. Data analytics are used in many industries to allow companies and organization to make better business decisions. Data Validation and Reconstruction is a promised tool of Data Analytics, which allows testing if the raw data is reliable or not. In the positive case, this raw data is stored as a validated data and, in the negative case the raw data is rejected and replaced by an estimated or reconstructed data. Once all the data are validated useful information could be derived for system management tasks (e.g. maintenance, planning, investment plans, billing, security and operational control).

Critical Infrastructure Systems (CIS), including water networks, are complex large-scale systems geographically distributed and decentralised with a hierarchical structure. These systems require highly sophisticated supervisory and real-time control schemes to ensure high performance achievement and maintenance when conditions are non-favourable due to faults (e.g., sensor and/or actuator and/or pipes malfunctions) (Schutze, 2004).

In CIS, a telecontrol system is acquiring, storing and validating data gathered from different kind of sensors every given sampling time to accurately real-time monitor the whole system. Several problems can occur during the data acquisition process, as those related with the communication system, e.g., communication failure between sensors and data loggers or in the telecontrol system itself. These problems produce missing or corrupted data which may be of great concern in order to have valid historic records. When this is occurring, missing data should be replaced by a set of estimated data, which should be representative of the data lost. Since missing data may severely jeopardise further processes needing complete datasets in order to get meaningful conclusions or analysis.

The methodology presented in this paper has been applied to the water network of the company ATLL, which transport the 85% of Catalonia in Spain, around 240 hm^3 per year. The raw data analysis of around 200 flowmeters and 100 level meters of reservoirs allow to determine the network performance evolution and quantifying the effects of different actions (new instrumentation, maintenance plans, etc.) applied during these years in the overall network (Espin, 2012). The application of these analyses allows to identify sectors with the lowest economic performance with possible leakages in the network assets (Quevedo, 2011 and Quevedo, 2014). It also allows to identify which new flowmeters should be installed for a better assessment of the network performance by defining new zoning and sectorisation. Finally, it allows locating which flowmeters need to be recalibrated in a maintenance plan of the sensors. The core of this methodology of validation and reconstruction has been described recently in (MA. Cuguero, 2016) and the main focus of this paper is to present a new supervision process which it tries to confirm if the validated or reconstructed data of the previous methodology are reliable data to replace the raw data and to highlight the interest that may have of these results for an efficient plan of the sensors .

2. Methodology of Data Validation and Reconstruction

2.1 Introduction

This section details the proposed methodology, which is divided in three stages (Figure 1): data validation, invalid/missing data reconstruction and a supervision system. The input to this procedure is the raw data vector y_{raw} gathered from the sensors. At the first stage if the data $y_{raw}(k)$ at a certain sample time k is validated, flag v is set to 1 and data $y_{val}(k) = y_{raw}(k)$ is stored in an operational data base (DB) as validated data. Conversely, if data $y_{raw}(k)$ is invalidated, flag v is set to 0 and the data reconstruction process (second stage) is performed to provide a reconstructed estimation $y_{rec}(k)$ of the invalid/missing data $y_{raw}(k)$ to be stored in the DB. Finally, a third stage supervises the coherence of the results applying different rules to guarantee the quality of new data.

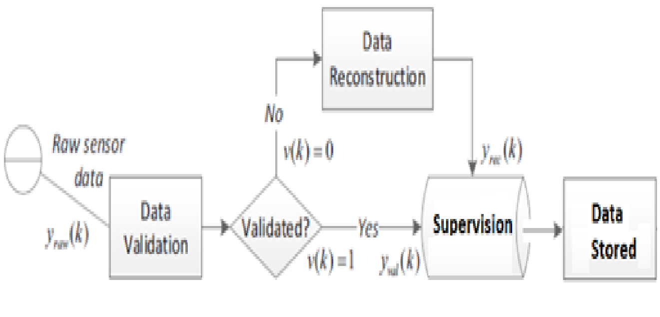

Figure 1. The three levels of the procedure

2.2 Data validation process

The data detection process is inspired by the Spanish AENOR-UNE norm 500540 developed for data validation in meteorological stations (UNE, 2004). The methodology presented here applies a set of consecutive detection tests to a given dataset to finally assign if the raw data is validated or not.

Level 0 (communications level) checks whether data are properly recorded at a regular sample rate by the acquisition system. If this is not fulfilled, there is some communication problem involving, e.g., the data transmission from the ground sensors to the operational database. Hence, this level allows detecting problems in the data acquisition or communication system.

Level 1 (physical range limits level) checks whether data are within the physical range of the sensor acquiring the corresponding measurement. The expected range of the measurements may be obtained from sensor specifications, expert knowledge or historical records of the data.

Level 2 (trend level) checks whether the data derivative, i.e., the magnitude change of the data among consecutive sample times, are within their expected rate. This allows detecting unexpected and possibly undesired sudden changes in the data, e.g. in a water network, tank water level sensors measurements cannot change more than several centimetres per minute. The expected range may be obtained from expert knowledge or historical records of the data.

Level 3 (equipment state level) allows to check the consistency of the variables in a given equipment unit, i.e. sensor or actuator. For example, in a water network system, in a pipe with a valve, a pump and a flowmeter installed, there is a relation between the valve and pump states and the flowmeter reading.

Level 4 checks the spatial consistency of the data collected by a certain sensor with other sensors installed in the network (Quevedo,2010a) i.e. the correlation between data coming from spatially-related sensors. This spatial model is obtained from the physical relations among these variables. In hydraulic systems, this relation is generally obtained from the mass balance relation of the element relating the different measured variables involved.

Level 5 checks for temporal consistency of a given sensor measurement, by means of a time series model obtained from sensor historical records under faultless assumption. A common method for time series signal forecasting is an autoregressive model approach (Quevedo, 2010b) because of its simplicity and low computational and storage requirements.

2.3 Data reconstruction process

This process is activated when a fault is detected at the validation stage and the corresponding data are voided, a reconstruction process is started until the sensor data are validated again. The output of the data validation process (Figure 1) is used to identify the invalidated data that should be reconstructed. SM, related with Level 4 and TSM, related with Level 5, are used for this purpose, depending on the performance of each model. The models accuracy is measured by the Mean Squared Error (MSE) of each model, evaluated in a moving horizon window

$$MSE(k)= \Sigma_j[e^2(j)]/m \qquad (1)$$

where m is the number of data samples considered in the window, $e(j)=y_{raw}(j)-y_{val}(j)$ is the error at instant j, $y_{raw}(j)$ is the raw value at instant j, $y_{val}(j)$ is the estimated value by the model (SM or TSM, respectively) at instant j and k is the actual time instant. The model having best MSE index before the fault is used to produce the reconstructed sensor signal. All the details of this methodology can be found in the recent paper (M.A. Cuguero, 2016)

2.4 Supervision process

The supervision process tries to confirm if the validated or reconstructed data are reliable data to replace the raw data. This process is important to assure the reliability of all the data because this information will be useful for system management tasks (e.g. maintenance, planning, investment plans, billing, security and operational control).

The supervision have several rules to be checked before to confirm the reliability of the new (validated and/or reconstructed) data. Some examples:

-If the daily accumulated new data are closed to the accumulated daily raw data, then the reconstruction will be supressed and the raw data will be take into account as the validated data.

-If the new data are outside the limits of the validation tests or practically are the same values than the raw data, then the reconstruction will be supressed and the raw data will be take into account as the validated data.

-If the new data are based on other data sensor (e.g. spatial models) and if these sensors have any non-validated data (e.g. multiple faults at the same time in two or more sensors), then the reconstruction will be supressed and the raw data will be into account as the validated data.

3. APPLICATION

The proposed methodology has been applied to ATLL network in the last 6 years, from 2008 to 2014. ATLL network supplies drinking water to 4.5 million inhabitants in Catalonia (Spain) with an approximate yearly demand of 240 cubic hectometres through 829 km of piping with diameters up to 3000 mm and its responsibility ends at municipal head tanks.

Figure 2. ATLL water network

During the considered period, 7 annual reports have been developed (analysing all the daily data per year of more than 200 flowmeters and 115 level sensors in the tanks) to provide the hydraulic and economic efficiency of more than 90 sectors, 10 zones and the whole ATLL network. The concept of network hydraulic efficiency analysed in this study is calculated as the ratio between the the volume of authorized consumption (CA) and the volume of water entering the network (VED). The CA includes the sum of consumption measured or not, but which have been authorized. On the other hand, the economic efficiency is calculated as the ratio between the volume of water billed division (VAF) and the volume of water entering the network (VED).

All the raw data of each sector, zone or whole network are validated and reconstructed allowing finally obtaining several index of performances: interval hydraulic efficiency, imprecision of the sensors, quality of the raw data regarding the number of non-validated raw data. Examples of the raw and validated data of two sectors during a year are presented Figure 3 and 4.

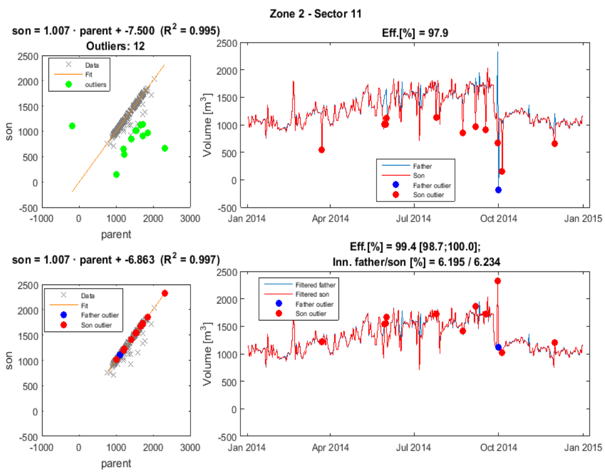

Figure 3. Sector 11 Zone 2 daily raw and validated data

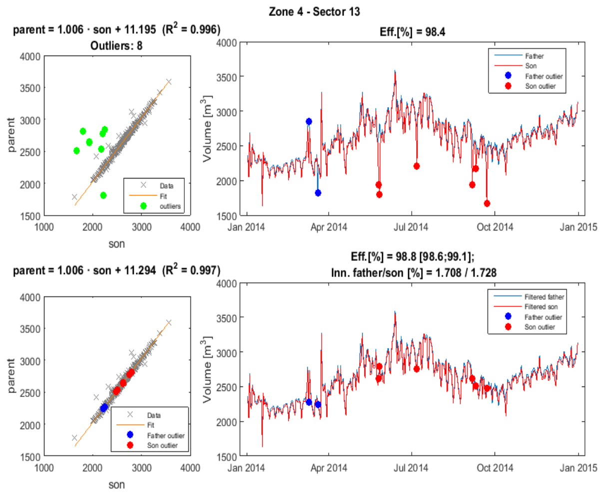

Figure 4. Sector 13 Zone 4 daily raw and validated data

The annual report also contains several ranking of all the 90 sectors ordered from the larger to the smaller volume, by efficiency, by sensor imprecision and by the quality of the data. Figure 5 shows a piece of ranking table according to the volume per sector. These rankings are very useful to extract recommendations for ATLL Company and a list of recommendations is proposed every year as a report. Moreover, the actions developed by the ATLL Company from the previous recommendations are also studied in this report.

Sector	Vol.Parent [Mm3]	Vol.Son [Mm3]	Imprecision Parent	Imprecision Son	Efficinecy	Filtered Eff.
Z5a	92.717	95.639	1,2	1,2	101,1	101,1
Z5aS1	92.668	93.440	1,4	1,4	100,1	99,8
Z3S1	84.739	82.429	1,6	1,7	97,02	97,27
Z3	84.737	81.851	1,6	1,7	96,40	96,59
Z4	68.468	66.502	0,6	0,6	97,69	97,13
Z4S1	68.385	66.671	1,1	1,2	97,95	97,49
Z9S1	29.371	28.936	0,9	0,9	98,56	98,52
Z9	29.369	28.653	0,8	0,9	97,59	97,56
Z5aS2	25.735	27.020	3,5	3,3	101,2	100,9
Z6	24.936	24.629	0,8	0,8	98,77	98,77

Figure 5. Ranking of the sector per volume

Among the set of recommendations of this analysis, some important issues are:

- Sectors of the network with lowest hydraulic performance are selected to check possible leakages or unmeasured consumptions by maintenance plans.
- Sectors with suspicious high performance are selected to re-calibrate the sensors by another maintenance plan, and
- Sectors with a suspicious high imprecision are selected to re-calibrate their sensors and/or to re-calibrate the area of their tanks in another maintenance plan to improve the performance of the network.

Finally, the annual report provides the economic and hydraulic interval efficiencies of the whole network as well as the comparison with the results in previous years. As, it can be seen in the Figure 6 for hydraulic efficiency, this indicator has been improved from 2008 to 2013 more than 2% (Quevedo 2014).

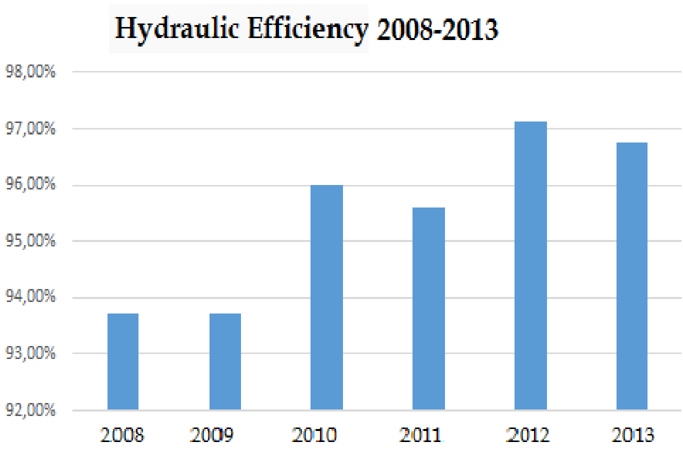

Figure 6. Historic evolution of the hydraulic efficiency of ATLL network (2008-2013)

4. CONCLUSIONS

In this work, a methodology for data validation and reconstruction of raw data has been described as a basic step to discover reliable information of the process for useful decision for multiple tasks, such as maintenance, planning, investment plans, billing, security and/or operational control.

The data validation consists to apply sequentially a set of six tests to all the raw data. It the raw data is not passing any of these tests, a reconstruction by estimation of several temporal and spatial models is applied to complete the data base with reliable information.

This methodology has already used for assessing the economic and hydraulic efficiency of water networks. The proposed methodology has been applied to ATLL water network in the last 7 years, from 2008 to 2014 with satisfactory results for performance enhancement and maintenance plans.

In particular, in this application the hydraulic efficiency has been improved from 2008 to 2014 more than 2% as a result of the application of the proposed methodology and derived actions, corresponding to a substantial improvement in a transport water network of more than 200 million of m3 per year.

ACKNOLEDGMENTS

This work was supported by Spanish Government (MINISTERIO ECNONOMIA Y COMPETITIVIDAD) and FEDER under project DPI2014-58104-R (HARCRICS). The authors also wish to thank the support received by the company ATLL in the development of this work.

REFERENCES

Cuguero M.A., Diego García, Joseba Quevedo, Vicenç Puig, Santiago Espin and Jaume Roquet (2016), "A methodology and a software tool for sensor data validation/reconstruction: Application to the Catalonia regional water network", Control Engineering Practice Volume 49, April 2016, Pages 159–172.

Espin S. and Roquet J. (2012) "Systematic control of efficiency and water losses reduction for Barcelona supramunicipal distribution network of 4.5 millions of inhabitants", New Developments in IT & Water Conference, Amsterdam. Holland.

Quevedo, J., Cuguerò, M.A., Pérez, R., Nejjari, F., Puig, V., Mirats, J.M.(2011). Leakage location in water distribution networks based on correlation measurement of pressure sensors. In: 8th IWA Symposium on System Analysis and Integrated Assessment (WATERMATEX 2011). San Sebastian, Spain.

Quevedo, J., Puig, V., Cembrano, G., Blanch, J., Aguilar, J. Saporta, D., Benito, G., Hedo, M., Molina, A. (2010a). Validation and reconstruction of flow meter data in the Barcelona water distribution network. Control Engineering Practice 18(6), pages 640 – 651. DOI http://dx.doi.org/10.1016/j.conengprac.2010.03.003

Quevedo, J., Blanch, J., Puig, V., Saludes, J., Espin, S. & Roquet, J.(2010b). Methodology of a data validation and reconstruction tool to improve the reliability of the water network supervision. In Water Loss Conference 2010, Sao Paulo, Brazil.

Quevedo, J., Blanch, J., Saludes, J., Puig, V. & Espin, S. (2009). Methodology to determine the drinking water transport network efficiency based on interval computation of annual performance. In Water Loss Conference 2009, Cape Town, South Africa .

Quevedo J., J. Pascual, V. Puig, J. Saludes, R. Sarrate, A. Escobet, S. Espin and J. Roquet (2014). "Flowmeter data validation and reconstruction methodology to provide the annual efficiency of a water transport network". Water Science & Technology: Water Supply. Vol. 14.2. Pages 337-346. DOI 10.2166/ws.2013.203

Schutze, M., Campisano, A., Colas, H., Schilling, W., Vanrolleghem, P.A. (2004): Real time control of urban wastewater systems – where do we stand today? Journal of Hydrology 299(34), pages 335–348. DOI http://dx.doi.org/10.1016/j.jhydrol.2004.08.010

UNE 500540 (2004) Redes de estaciones meteorológicas automáticas: directrices para la validación de registros meteorológicos procedentes de redes de estaciones automáticas: validación en tiempo real. AENOR.