

# METHODOLOGY OF A DATA VALIDATION AND RECONSTRUCTION TOOL TO IMPROVE THE RELIABILITY OF THE WATER NETWORK SUPERVISION

*J. Quevedo\*, J. Blanch\*, V. Puig\*, J. Saludes\*, S. Espin\*\* and J. Roquet\*\**

*\* Advanced Control Systems Research Group of Catalonia Technical University (SAC UPC)*

*\*\*Aigües Ter Llobregat Company (ATLL)*

## **ABSTRACT**

The drinking water network of *Aigües del Ter-Llobregat* (ATLL) has applied a methodology to obtain a reliable data. The methodology is based on a first phase to validate the data with six levels and a second phase that they analysis and reconstruct the non-validate data.

## **INTRODUCTION**

Every day, the telecontrol system supervised the whole network status. This process needs accurate and reliable sensors data. If we work with a wrong data, we do not able to control efficiently the network and we can not detect the anomalous situations as faults or leakages. So, we need to analyzes the received data and validate. If the data is non-validate, we have to replace by well forecast.

**The most important information, in a drinking water network, is captured by flowmeters.**

**The flowmeters measures the water consumption of the several towns and the introduced water in the network. The flow analysis is used to differentiate the correct and the wrong data. Moreover, the analysis is used to estimate the wrong data and their confidence intervals. Anyway, the method needs to know the other sensors values of the network, as valve position, pumps status and levels tank, to correlate with the flow. This methodology is been applied in the drinking water of Aigües del Ter-Llobregat (ATLL)**



Figure 1).

Recorded data quality is a basic requirement to determine the efficiency of a water network and further to assess the non revenue water (A.Lambert, 2003) and its components of real and apparent water loss of the network.

The analysis of the data quality recorded by the supervision system is a basic task to discriminate between correct and erroneous (or questionable) data. This task allows not only to detect an incorrect or questionable data but also to ignore and to replace it by a good estimation in order to maintain a complete and reliable data base of the network.

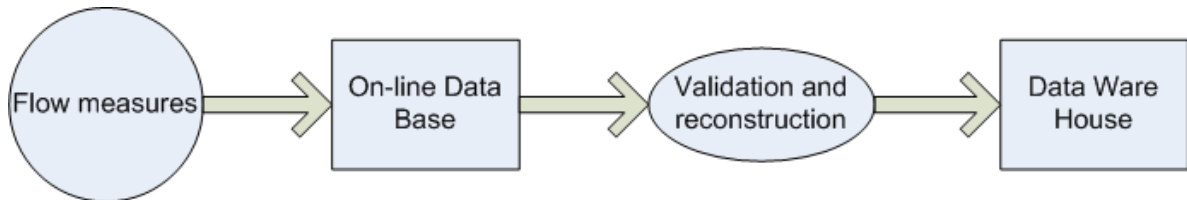


Figure 1: ATLL drinking water network.

The proposed method to detect the quality of the data apply until 6 level of tests, following the same principles of the Spanish UNE norm 500540 (UNE, 2004) used to analyse the data quality of meteorological stations.

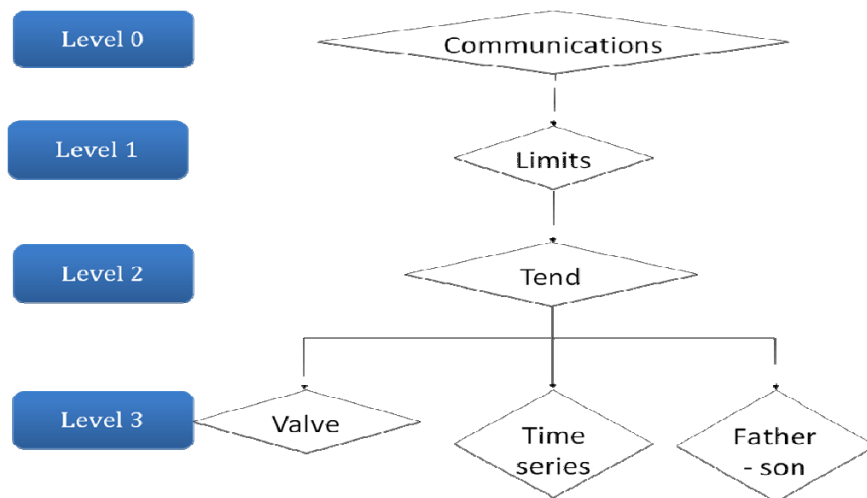
## METHODS

The proposed methodology is divided in two phases: validation and reconstruction. The procedure is added between the on-line database and the historical database (Figure 2). So, the procedure only affects the historical data and this validate data are going to make a better historical analysis.



**Figure 2:** The data capture diagram.

The validation methodology is inspired in the AENOR-UNE norm 500540. The methodology consist in associate each data a quality restrictive level. The progressive quality levels are represented in Figure 3.



**Figure 3:** Validation levels.

An explanation of each level are:

- *Level 0*: this level takes into account simply if the data is recorded or not in the case than the supervised system expect to collect data at a fixed sampling time (problems in the sensor or in the communication system).
- *Level 1*: The limits level checks that the data are inside the physical range. For example, the maximum values expected of the flowmeters will be determined by a simple analysis of the flow capacity limit of the pipes.

- *Level 2*: The tend level will take into account the time changes of the data in two or more sampling times. For example, the data of the level sensors in a tank cannot change more than several cm by minute in a real tank.
- *Level 3*: The models level is a multilevel and it checks three parallel models:
  - *Valve*: the valve model takes into account the possible correlation existing between the flow and the opening valve in the same pipe.
  - *Time series*: This model take into account the time series of the data (Blanch, 2009) of each variable. For example, analysing the historic data of the flows in a pipe a time series model can be derived and the output of the model is used to compare and to validate the recorded data.
  - *Father-Son*: the Father-Son model checks the correlation models (Quevedo, 2009) between historical data of the same sensors located in different but near local stations. For example, the data of the flowmeters located in different points of the same pipe in a transport water network allows checking the reliability of the sensor set.

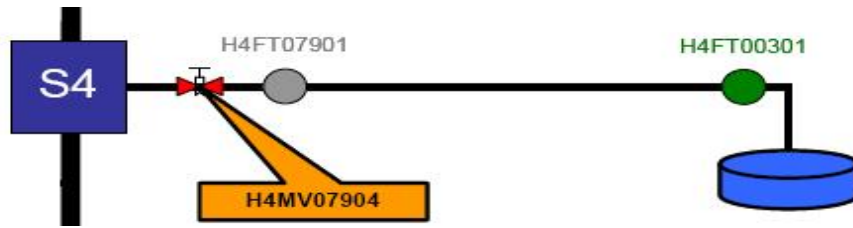
We develop logical method to invalidate the data in the level 3. This logic method detect the invalidate data from the result of the three models. Aside, the Father-Son model is very useful not only to detect problems in the data of the sensors but also to detect leakages in the pipe and to compute the balance of the sectors in a transport network.

Once the data pass all the test levels if the system detects misbehaviour, the next step is to isolate the fault combining the previous tests. For instance, if the three tests detect misbehaviour in a set of two flowmeters, the system analyzes the historical data and other features of both flowmeters to isolate the origin of the problem.

And finally, the proposed method includes a reconstruction task of the erroneous data in order to complete the data base with estimated values to replace the bad data. For this task, the outputs of the derived models of the level three are very useful as reconstructed data.

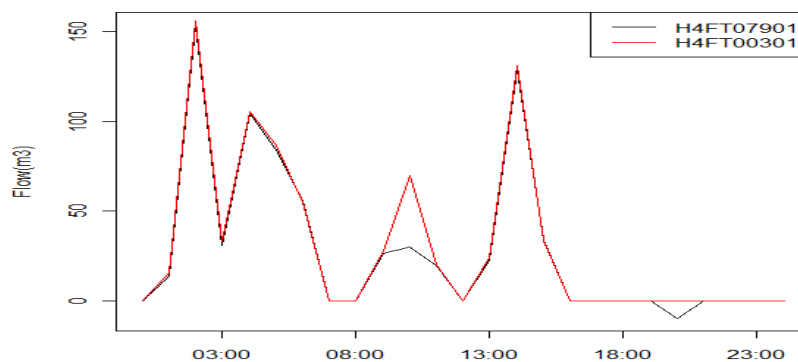
## **EXAMPLE**

The proposed methodology is applied in a studied zone of the drinking water network of ATLL. We present one pipe of this studied zone. This pipe contains two flowmeters (H4FT07901 and H4FT00301) and one valve (H4MV07904). The Figure 4 presents a pipe diagram.



**Figure 4:** Studied pipe diagram.

The Figure 5 presents a raw data of one day for both flowmeters. It can be observed, the major part of the day, which the two flows are equal. But there are two hours that they do not measure the same.



**Figure 5:** Initial data for the two flowmeters.

When the procedure analysis the data, it detects this two hours as invalidate data. The hour 10h is invalidate in the level 3 (models) and for the both flowmeters. The hour 20 for the H4FT007901 only pass the level 0 (communications) and for the H4FT00301 pass the whole levels.

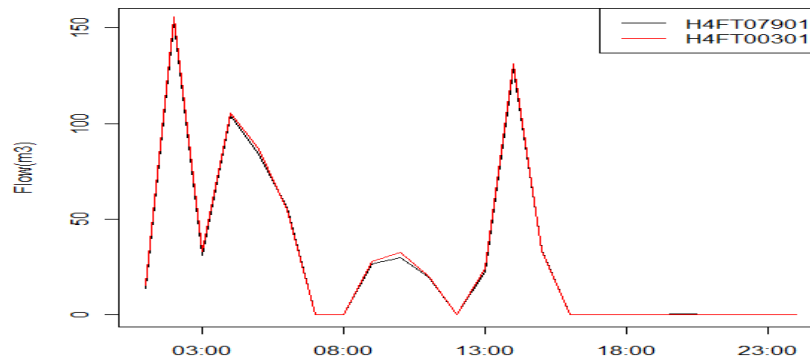
The reconstruction is the mean of the three estimations (valve, time series and Father-Son):

$$H4FT00301(20h) = 1/3 \cdot 0.9262 + 1/3 \cdot 0 + 1/3 \cdot 0 = 0.3087329 \text{ m}^3$$

Before to reconstruct the 11h, the procedure has identified which flowmeter is the invalid. The procedure checks that the H4FT00301 is coherent with the models of valve and the time series. So, the invalidate flowmeter is H4FT07901 and their reconstructed data is:

$$H4FT07901(10h) = 1/3 \cdot 35.0637 + 1/3 \cdot 31.2414 + 1/3 \cdot 30 = 32.1017 \text{ m}^3$$

The validate data with the reconstructed data is presented in Figure 6.



**Figure 6:** Validate data.

## CONCLUSIONS

In this paper, we present a methodology to validate and reconstruct incoming water network data. At the present, this procedure is running over an experimental DMA in ATLL drinking water network with satisfactory results.

The proposed methodology is able to detect non-valid data and propose a good reconstruction. The best properties are that the procedure only needs information from the database and the test levels are simple to verify. The major problem is that the validation and reconstruction is based on other sensors. So, these sensors also need to be validated.

The proposed methodology can be applied to other similar network in an easy way. Also, this methodology is able to apply in other types. It is only required to modify the parameters and change the valve model for other sensors models.

## Acknowledgements

This article is part of applied investigation project between ATLL Company and SAC group. The authors wish to thank also the support received by the SAC research group by the Comissió de Recerca de la Generalitat de Catalunya (Grup SAC ref. 2009SGR) and the CICYT (ref. HYFA DPI-2008-1999) from the Ministry of Education the Spanish State.

## REFERENCES

- A.Lambert, "Assessing non-revenue water and its components: a practical approach", Water 21, International Water Association, August 2003.
- J. Quevedo, J. Blanch, J. Saludes, V. Puig, S. Espin, "Methodology to determine the drinking water transport network efficiency based on interval computation of annual performance", Water Loss Conference 2009, Cape Town, May 2009.

UNE, 2004 "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.

J. Blanch, V. Puig, Jordi S. Joseba Q., "ARIMA Models for data consistency of Flowmeters in water distribution networks", 7<sup>th</sup> IFAC Symposium on Fault detection, Supervision and Safety of technical processes. pp 480-485.

Deltour, J.L., Canivet, E., Sanfilippo, F., Sau, J. (2005) Data Reconciliation on the Complex Hydraulic System of Canal de Provence. Journal of Irrigation and Drainage Engineering, mayo/junio, págs. 291-297.

Lozano Miguel A. y Remiro Jesus A. (2002), Clasificación de variables y reconciliación de datos en ingeniería de procesos, Revista Métodos numéricos para cálculo y diseño en ingeniería: Revista internacional, ISSN 0213-1315, Vol. 18, N.º 3, págs. 401-414.

Tsang, K.M. (2003). Sensor data validation using gray models. ISA-Transactions, N.º 42, págs. 9-17.