Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 49-28 (2016) 203–207

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.

2405-8963 © 2016, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.11.035

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.

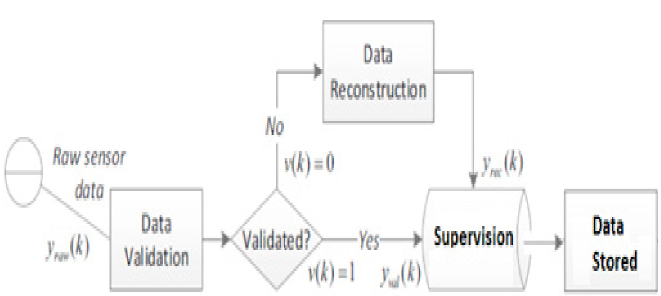

Figure 1. The three levels of the procedure

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