Control Engineering Practice 18 (2010) 640–651

Contents lists available at ScienceDirect

Control Engineering Practice

journal homepage: www.elsevier.com/locate/conengprac

Validation and reconstruction of flow meter data in the Barcelona water distribution network

J. Quevedo [a], V. Puig [a,*], G. Cembrano [a], J. Blanch [a], J. Aguilar [b], D. Saporta [c], G. Benito [d], M. Hedo [d], A. Molina [d]

[a] Automatic Control Department and Industrial Robotics Institute, Technical University of Catalonia (UPC) and CSIC, Rambla Sant Nebridi, 10, 08222 Terrassa, Spain
[b] LAAS-CNRS, 7 Avenue du Colonel Roche, 31077 Toulouse, France
[c] AGBAR Barcelona Water Company, C/Diputació, 351, 08009 Barcelona, Spain
[d] Adasa Sistemas, C/Pedrosa B, 30, 32, 08908 L'Hospitalet de Llobregat, Spain

A R T I C L E I N F O

Article history:
Received 4 March 2008
Accepted 9 March 2010
Available online 2 April 2010

Keywords:
Water distribution network
Telecontrol system
Flow meter
Fault detection
Sensor failure
Fuzzy logic classifier

A B S T R A C T

This paper presents a signal analysis methodology to validate (detect) and reconstruct the missing and false data of a large set of flow meters in the telecontrol system of a water distribution network. The proposed methodology is based on two time-scale forecasting models: a daily model based on a ARIMA time series, while the 10-min model is based on distributing the daily flow using a 10-min demand pattern. The demand patterns have been determined using two methods: correlation analysis and an unsupervised fuzzy logic classification, named LAMDA algorithm. Finally, the proposed methodology has been applied to the Barcelona water distribution network, providing very good results.

1. Introduction

In a complex water distribution network, such as the case of Barcelona city, a telecontrol system must acquire, store and validate data from many flow meters and other sensors every few minutes to achieve accurate monitoring of the whole network in real time. Frequent operation problems in the communication system between the set of the sensors and the data logger, or in the telecontrol itself, generate missing data during a certain periods of time. The stored data are sometimes uncorrelated and of no use for historic records. Missing data must, therefore, be replaced by a set of estimated data. A second common problem is the lack of reliability of the flow meters (offset, drift, breakdowns), producing false flow data readings. These false data must also be detected and replaced by estimated data, since flow data are used for several network water management tasks, namely: planning, investment plans, operations, maintenance and billing/consumer services and operational control (Cembrano, Wells, Quevedo, Pérez, & Argelaguet, 2000). The same type of problem can be found in gas or electricity networks (Matheson, Jing, & Monforte, 2004). Therefore, the methodology presented in this paper could also be applied to these networks.

According to the nature of the available knowledge, different kinds of data validation can be built, with varying degrees of sophistication. In general, one may distinguish between elementary signal based ("low-level") methods and model based ("higher level") methods (see, e.g. Denoeux et al., 1997; Mourad & Bertrand-Krajeswski, 2002).

Elementary signal based methods use simple heuristics and limited statistical information of a given sensor (Burnell, 2003; Jörgensen et al., 1998; Maul-Kötter & Einfalt, 1998). Typically, these methods are based on validating either signal values or signal variations. In the signal value-based approach, data are assessed as valid or invalid according to two thresholds (a high one and a low one); outside these thresholds data are assumed invalid. On the other hand, methods based on signal variations look for strong variations (peaks in the curve) as well as lacks of variation (flat curve).

Model-based methods rely on the use of models to check the consistency of sensor data (Tsang, 2003). This consistency check is based on computing the difference between the predicted value from the model and the real value measured by the sensors. Then, this difference, known as residual, will be compared with a threshold value (zero in the ideal case). When the residual is bigger than the threshold, it is determined that there is a fault in the system. Otherwise, it is considered that the system is working properly. Models are usually derived using either multivariate procedures exploiting the correlation or the analytical relations between several quantities obtained using first principles, sometimes measured at

* Corresponding author. Tel.: +34 9373986327; fax: +34 9373986328.
E-mail address: vicenc.puig@upc.edu (V. Puig).

0967-0661/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.conengprac.2010.03.003

different times ("*temporal redundancy*") and/or locations ("*spatial redundancy*"). The result of data validation may be either a binary variable indicating whether the data are considered valid or not, or a continuous validity index interpreted as a degree of confidence in the data. When the degree of confidence is too low, data can be either discarded or replaced by an estimate computed using a statistical or physical model (see, e.g. Petit-Renaud & Denoeux, 1998). Moreover, a subproduct of using model-based approaches for sensor data validation is that the prediction provided by the model can be used to reconstruct the faulty sensor. Some examples of these methods in literature applied to the water domain are:

- time-series analysis techniques (Bennis, Berrada, & Kang, 1997; Bennis & Kang, 2000; Crobeddu & Bennis, 2006; Lobanova & Lobanova, 2003; Prescott & Ulanicki, 2001);
- Kalman filters (Ciavatta et al., 2004; Pastres, Ciavatta, & Solidoro, 2003; Piatyszek, Voignier, & Graillot, 2000);
- parity equations (Boukhris, Giuliani, & Mourot, 2001; Hamioud, Joannis, & Ragot, 2005a, 2005b; Ragot & Maquin, 2006; Puig et al., 2003);
- pattern recognition methods (Valentin & Denoeux, 2001);
- principal component analysis (Arteaga & Ferrer, 2002; Harkat, Mourot, & Ragot, 2006; Nelson, Taylor, & MacGregor, 1996).

However, none of the existing methods satisfies data validation and reconstruction specifications established by the Barcelona water company. In particular, the company specifies that every flow meter at water distribution level should be validated and reconstructed using only their own data by exploiting the "*temporal redundancy*" that exists since flow readings follow the consumer demand patterns. No mass balances between different sensors can be used since they involve relating the information systems of the Barcelona transport and distribution networks. This is not currently possible because these systems are not integrated. The transport network is in charge of the right choice of water sources that enter the system at each moment (quality and quantity of the water supply) and of the bulk water transferences between sources and reservoirs, while the distribution network is responsible for delivering water from the reservoirs to the consumers.

To address these problem specifications, this paper proposes a model based methodology for data validation (detection) and replacement of faulty/missing flow meter measurements in a water distribution network with district metering areas (DMAs), based on the use of time series analysis in combination with short-term consumption patterns. A number of consumption patterns have been studied using correlation analysis and the fuzzy classification algorithm LAMDA. The proposed data validation and replacement algorithm exploits temporal redundancy existing in the demands because of the existence of weekly periodic behaviors associated with human habits (Brdys & Ulanicki, 1994). This algorithm is inspired on previously developed algorithms to predict water flow demands (Quevedo, Cembrano, Valls, & Serra, 1988).

The structure of the paper is as follows. In Section 2, the problem of flow meter data validation and reconstruction in Barcelona water distribution network information system is presented and the need for an algorithm to do it automatically is justified. In Section 3, the proposed methodology is introduced and motivated. Section 4 presents the two-level model used for such methodology that uses at the first level a time series and flow patterns at the second level to describe the flow time series recorded at each flow meter. Section 5 describes the proposed methodology of data validation and reconstruction. Results of application of the proposed methodology in several real scenarios coming from the Barcelona water distribution network are presented in Section 6. Finally, in Section 7, conclusions and further developments are presented.

2. Problem description

2.1. Barcelona water distribution network

The Barcelona water distribution network is comprised of some 200 sectors (DMAs) with approximately 400 control points (Fig. 1). At present, the Barcelona information system receives, in real time, data from 200 control points, mainly flow meters and a few pressure sensors. Most of these flow meter control points correspond to the single supply point of a DMA, so that their evolution is closely related to that of the water demand in that DMA.

Every 10 min, a data logger stores data from the sector sensors and, once a day, these data are sent to an operational database of the Telecontrol Information System implemented using TOPKAPI SCADA system (http://www.areal.fr/) via telephone XTC network or GSM radio using the ModBus communication protocol (Fig. 2). Communication with each of the data loggers is ensured daily through 20 XTC lines managed simultaneously by the two servers. The data are recovered using the TOPKAPI SCADA facilities and transferred to an ORACLE database.

The sensor data used by the telecontrol system must follow two functional procedures, previous to their integration for use in the Management System: raw data insertion process and data validation and replacement.

2.2. Raw data insertion process

This process consists of data acquisition from the data logger towards the operational database in the TOPKAPI telecontrol system. In this process, missing sensor or data-logger data, as well as communication failures occur, which must be recovered using artificial data, for further use in statistic and hydraulic balance studies (Fig. 3). In this case, failure detection is trivial: There is a gap in the data and an attached error message. A more interesting problem is the replacement of missing data with a virtual value, which approximates the real value of the missed reading.

2.3. Data validation and reconstruction

At present, there is an automatic data validation procedure implemented at each control point. This validation checks each reading value against a validity range (daytime and nighttime minimum and maximum values). Whenever a reading value lies outside this range, the reading is invalidated. Additionally, the previous reading and the next one are also invalidated. However, these invalid data have been, so far, left intact and propagated for further use in hydraulic balances and statistics.

Invalid data should be replaced by virtual data (software or virtual sensor) before further computations are performed. A procedure for estimating missing or invalid data is necessary for the complete water network, this being the aim of this paper. This procedure, described in the next sections, can also be used to improve the actual implemented data validation method as described.

3. Motivation and overview of the proposed methodology

3.1. Motivation

The proposed methodology for flow data validation and reconstruction is based on an analytical redundancy approach. A model is developed for each flow meter in the Barcelona water distribution distribution network based only on the actual and