Flowmeter Data Validation and Reconstruction Methodology to provide the Annual Efficiency of a Water Transport Network: the ATLL Case Study in Catalonia

J.Quevedo[1], J.Pascual[1], V. Puig[1], J. Saludes[1], R. Sarrate[1], S. Espin[2] and J. Roquet[2]

[1] {Advanced Control Systems Group of Technical University of Catalonia (SAC-UPC), Rambla S. Nebridi, 10, Terrassa (Barcelona, Spain)}

[2] {Aigues Ter Llobregat Company (ATLL), Barcelona (Spain)}

Correspondent to J.Quevedo (joseba.quevedo@upc.edu)

Abstract

The object of this paper is to provide a flowmeter data validation/reconstruction methodology that determines the annual economic and hydraulic efficiency of a water transport network. In this paper, the case of Aigües Ter Llobregat (ATLL) company, that is in charge of managing the 80% of the overall water transport network in Catalonia (Spain), will be used for illustrating purposes. The economic/hydraulic network efficiency is based on the daily data set collected by the company using about 200 flowmeters of the network. The data collected using these sensors are used by the remote control and information storage systems and they are stored in a relational database. All the information provided by ATLL is analyzed to detect inconsistent data using an automatic data validation method deployed in parallel with the evaluation of the network efficiency. As a result of the validation process, corrections of flow measurements and of the volume of billed water are introduced. The results of the ATLL water transport network obtained during year 2010 will be used to illustrate the approach proposed in this paper.

1 Introduction

The performance of the water network can be measured in two ways. First, the economic performance from the annual net income of the delivered water (VAF) is determined. Second, the hydraulic performance between the volume of water delivered which is measured by billing flowmeters (VAM) and the volume of water entering the network (VED) is also computed. The study presented in this paper covers the performance analysis of the 99 sectors composing the ATLL network, as well as of the 10 zones containing them and of the full network. This study identifies the sectors with the lowest economical and hydraulic performances. It also proposes where new flowmeters should be installed for a better assessment of the network performance by defining new zoning and sectorisation and it helps locating which flowmeters need to be recalibrated.

The main aim of this paper is to carefully analyze all raw data of the telemetry system using a set of validation tests. The invalidated data are reconstructed with the available models used for data validation.

In ATLL network, the telecontrol system acquires, stores and validates data from different sensors (collected at different sampling rates: 10min, 1hour, 1 day) to achieve accurate monitoring of the whole network. Frequent operating problems in the communication system between the set of the sensors and the data logger, or in the telecontrol itself, generate missing data during certain periods of time. The stored data are sometimes uncorrelated and of no use for historic records. Therefore, missing data must be replaced by a set of estimated data. A second common problem is the lack of data sensor reliability (offset, drift, breakdowns, etc.) leading to false measurements. Data sensors are used for several complex system management tasks such as planning, investment plans, operations, maintenance, security and operational control (Quevedo et al, 2010b). So wrong data must be detected and replaced by estimated data. Recorded data quality is a basic requirement to determine water network efficiency and further assess the non-revenue water of the network (Lambert, 2003).

2 Proposed methodology

In a previous work (Quevedo et al, 2009), a methodology was presented to compute the network efficiency taking into account raw flowmeter data and the network topology. Basically, consistency of raw flowmeter data was analyzed using spatial network models (mass balance of each sector). Wrong or missing data were removed and replaced by estimated data using models, and filtered data were analyzed to compute the performance (flowmeter inaccuracies, interval efficiency) of each sector. Finally, the economical and hydraulical efficiencies of each zone and of the overall network were derived and analyzed to generate new actions to improve the instrumentation (location of new sensors, recalibrations) and new plans for the network maintenance to locate leakages in the pipes. Further, in a second work (Quevedo et al., 2010a), a more general tool was developed to check the consistency of raw flowmeter and level sensor data of the water network taking into account not only spatial models but also temporal models (time series of each flowmeter) and internal models of several components in the local units (pumps, valves, flows, levels, etc.). This last proposal allows the robust isolation of wrong data that must be replaced by adequate estimated data. In this work, an integrated methodology of both previous proposals is presented (Figure 1).

2.1 Raw data validation and wrong data reconstruction (steps 1 and 2)

Raw flowmeter data validation is inspired in the Spanish norm (AENOR-UNE norm 500540). The methodology consists in assigning a quality level to data. Quality levels are assigned according to the number of tests that have been passed, as represented in Figure 2.

An explanation of each level is as follows:

- *Level 0*: The ***communications*** level simply monitors whether the data are recorded taking into account that the supervisory system is expected to collect data at a fixed sampling time (problems in the sensor or in the communication system).
- *Level* 1: The ***bounds*** level checks whether data are inside their physical range. For example, the maximum values expected for flowmeters will be determined by a simple analysis of the flow capacity of the pipes.
- *Level 2:* The ***trend*** level monitors the data rate. For example, level sensor data cannot change more than several cm by minute in a real tank.

- *Level 3:* The ***models*** level uses three parallel models:
    - ***Valve***: the valve model supervises the possible correlation that exists between the flow and the opening valve command in the same pipe or pump element.
    - ***Time series***: This model takes into account a data time series for each variable (Blanch et al., 2009). For example, analysing historical flow data 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.
    - ***Up-Downstream****:* the up-downstream model checks the correlation models between historical data of sensors located at different but near local stations in the same pipe (Quevedo et al., 2009). For example, data of flowmeters located at different points of the same pipe in a transport water network allows checking the sensor set reliability.

A decision tree method has been developed to invalidate data in level 3. This method detects invalid data from the result of the three models. From that, the ***Up-Downstream*** model is very useful not only to detect problems in sensor data but also to detect leakages in pipes and to compute the balance in transport network sectors.

Once data have passed all test levels, if data inconsistency is detected, next step is to isolate the fault by combining the previous tests. For instance, if the three tests detect an inconsistency in a set of two flowmeters, the system analyses the historical data and other features of both flowmeters to diagnose the cause of the problem and to identify the sensor in faulty operation. And then, all the data of this faulty sensor are replaced by the data of the healthy sensor of the same pipe.

Finally, the proposed method includes reconstructing erroneous data by completing database with estimated values that replaces bad data. For this task, the outputs of the models derived at level three are very useful to generate reconstructed data.

2.2 Network models and performances based on filtered data (steps 3, 4 and 5)

A water transport network can be divided into a set of interconnected sectors (see Figure 3). Inside each sector, there could be demand nodes, tanks and flowmeters. Flowmeters measure sector inputs and outputs. External demand is considered as an output. In this study, pipes are considered pressurized. Hence, it is assumed that there are no delays in the pipes. The sector model is based in mass balance equations and the following hypotheses should be assumed:

- Flowmeters are maintained and calibrated by the water management company following a maintenance program (confirmed in the case of ATLL company in Catalonia network).
- Flowmeters have been installed and operated fulfilling the manufacturer recommendations, thus avoiding systematic errors in the measurements ("unbiased").
- Random errors are normally distributed around the measured value (“normal”).
- Random errors between measurement instruments are uncorrelated (“independence”).

When a sector has several flowmeters at both input and output (Figure 4), the model is given by

$$\sum_{j=1}^{n_{in}} F_{in_j}(t) = K \sum_{l=1}^{n_{out}} F_{out_l}(t) + M \tag{1}$$

where $\sum_{j=1}^{n_{in}} F_{in_j}(t)$ and $\sum_{l=1}^{n_{out}} F_{out_l}(t)$ are the daily flows measured by the input and output sensors, respectively. Parameters K and M are determined using least squares and real data. In the ideal case, they should be equal to $K = 1$ and $M = 0$, respectively.

Considering that input and output flowmeters have errors, named respectively e_{in} and e_{out}, Eq. (1) is rewritten as follows:

$$\sum_{j=1}^{n_{in}} (F_{in_j}(t) + e_{in_j}(t)) = K \sum_{j=1}^{n_{out}} (F_{out_l}(t) + e_{out_l}(t)) + M \tag{2}$$

and model residuals are given by

$$e(t) = \sum_{j=1}^{n_{in}} F_{in_j}(t) - K \sum_{l=1}^{n_{out}} F_{out_l}(t) - M = K \sum_{l=1}^{n_{out}} e_{out_l}(t) - \sum_{j=1}^{n_{in}} e_{in_j}(t) \sim N(0, K^2 n_{out} \sigma_{out}^2 + n_{in} \sigma_{in}^2) \tag{3}$$

Consider that input and output sensors have the same characteristics, i.e, it is assumed that $\sigma_{in} = \sigma_{out} = \sigma$. If the main sectors are close to the ideal case (K=1), then the residual error $e(t)$ is normally distributed ($N(0,\sigma_{fit}^2)$ with $\sigma_{fit}^2 = (n_{in} + n_{out})\sigma^2$) and the variance of the error can estimated as follows

$$\sigma = \frac{\sigma_{fit}}{\sqrt{n_{in} + n_{out}}} \tag{4}$$

If a confidence interval α is considered with an standard deviation radius λ(α), the relative error is given by

$$Flowmeter\ error\ (\%) = 100 \frac{\lambda(\alpha)\sigma}{mean\,(flowmeter)} \tag{5}$$

The network efficiency calculation is the ratio between the network output flow V_{out} and the network input flow V_{in},

$$R = \frac{V_{out}}{V_{in}} \tag{6}$$

.

As these two quantities are affected by flowmeter errors, the network efficiency calculation has an uncertainty that can be quantified by means of the following interval

$$\left[R_{\min}(n), R_{\max}(n)\right] = \left[\frac{V_{out} - \lambda(\alpha)\sqrt{n \cdot n_{out}} \cdot \sigma_{out}}{V_{in} + \lambda(\alpha)\sqrt{n \cdot n_{in}} \cdot \sigma_{in}}, \frac{V_{out} + \lambda(\alpha)\sqrt{n \cdot n_{out}} \cdot \sigma_{out}}{V_{in} - \lambda(\alpha)\sqrt{n \cdot n_{in}} \cdot \sigma_{in}}\right] \tag{7}$$

where n is the number of days taken into account in the efficiency calculation horizon (e.g. n = 365 for a year).

This analysis is very useful to detect problems in the sensors and leakages in the sectors of the network. The efficiency interval [R_{min}, R_{max}], the flowmeter imprecision (%) and the parameters K and M provide the following logic rules:

- If $K \approx 1$, $M \approx 0$ and the flowmeter imprecision is of the order of manufacturer sensor imprecision, the sensors are working well. While, if the flowmeter imprecision is larger than the manufacturer sensor imprecision, this can be caused by an operation fault of the sensors.
- If $K >> 1$ and $M >> 0$ the input flowmeters measure a greater flow than output flowmeters and consequently [R_{min}, R_{max}]<<1. This can be caused by leakage or bad-calibration of the sensors.
- If $K << 1$ and $M << 0$, the inputs flowmeters measure less flow than outputs flowmeters and consequently [R_{min}, R_{max}]>>1. This also can be caused by a bad calibration of the sensors.

3 ATLL network results

The methodology described above has been applied to ATLL network (Figure 5) continuously every year from 2007 until now to determine the annual economic and hydraulic efficiency. This has allowed to analyse the evolution of the network efficiency and to quantify the effects of different actions (new instrumentation, maintenance plans, etc.) in the overall network. This methodology has been applied firstly in a sector by sector basis in order to distinguish the real efficiency of all the components of the network. In this section, the results of year 2010 in two sample sectors will be presented. The first sample sector is composed of one input flowmeter and three output flowmeters.

Figure 6a presents the upstream and downstream flowmeter daily time series. It shows with a circle the outliers which have been detected and isolated by the time series models of upstream and downstream flowmeters.

Figure 6b is a scatterplot of raw input data versus raw output data corresponding to flowmeters and its linear approximation.

Figure 6c presents again the upstream and downstream flowmeter daily time series. It shows with a circle the outliers which have been replaced by estimated data obtained from time series models of upstream and downstream flowmeters.

Figure 6d is a scatterplot of the filtered input data versus filtered output data corresponding to flowmeters and its linear approximation. The linear approximation fits well because the linear coefficients are K≈1 and M≈0 and the Pearson's coefficient is 0.997.

The upstream flowmeter error is close to 0.5% whereas the downstream flowmeter error is close to 1.0%. The confidence interval of the hydraulic efficiency corresponding to this sector is [98.7%, 98.9%].

The second sample sector is only composed of one upstream flowmeter and one downstream flowmeter, but the quality of the time series corresponding to raw data are worse than in the first sample sector (Figures 7a and 7b). In this case, the time series of the upstream flowmeter had an operating problem during almost half a year.

The validation method detects, isolates and adequately reconstructs wrong flowmeter data using the downstream flowmeter data. Filtered data are shown in Figures 7c and 7d. The coefficients of the linear approximation are K=0.653 and M=382 m^3. The Pearson's coefficient is 0.48. The upstream and downstream flowmeter inaccuracies are close to 17% and the confidence interval of the hydraulic efficiency corresponding to this sector is [104.7%, 108.5%].

The same procedure has been applied to all the sectors of ATLL water network allowing ranking them form the best to the worst taking into account several performance indices: hydraulic efficiency, sensor error, data quality (% of estimated data), etc. Finally, this work has addressed the 10 zones and the overall network in order to obtain global performances of the ATLL network.

4 Conclusions

In this work, a combined methodology to evaluate the annual economic and hydraulic efficiency corresponding to all sectors of a water network is proposed. It is based on checking raw flowmeter data consistency using several tests and models, and replacing wrong data by model estimations. Moreover, the proposed methodology evaluates the efficiency of all sectors, zones and complete network taking into account sensor inaccuracies and providing a confidence interval. This confidence interval collects the network misbehaviours either due to leakage or sensor bad-calibration. A tight confidence interval is indicative that the network is behaving well. Otherwise, a wide confidence interval corresponds to the existence of some leakage or bad-calibrated sensor.

Acknowledgements

This works belongs to a research applied project granted by ATLL Company. The authors wish also to thank the support received by CICYT (SHERECS, DPI-2011-26243) of Spanish Ministry.

References

J. Blanch, V. Puig, J.Saludes, Joseba Q.(2009): "ARIMA Models for data consistency of Flowmeters in water distribution networks", 7th IFAC Symposium on Fault detection, Supervision and Safety of technical processes. pp 480-485, 2009.

Burnell D.(2003): "Auto-validation of district meter data" Advances in Water Supply Management- Maksimovic, Butler, Memon eds., Swets & Zeitlinger P., Netherlands.

A.Lambert (2003): "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(2009): "Methodology to determine the drinking water transport network efficiency based on interval computation of annual performance", Water Loss Conference 2009, Cape Town, May 2009.

Quevedo, J., Puig, V., Cembrano, G., Blanch. J. (2010a): "Validation and reconstruction of flowmeter data in the Barcelona water distribution network". Control Engineering Practice Journal, 18 (6), pp. 640-651.

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

Tsang, K.(2003). "Sensor data validation using gray models" ISA Trans. 42, 9–17.

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.

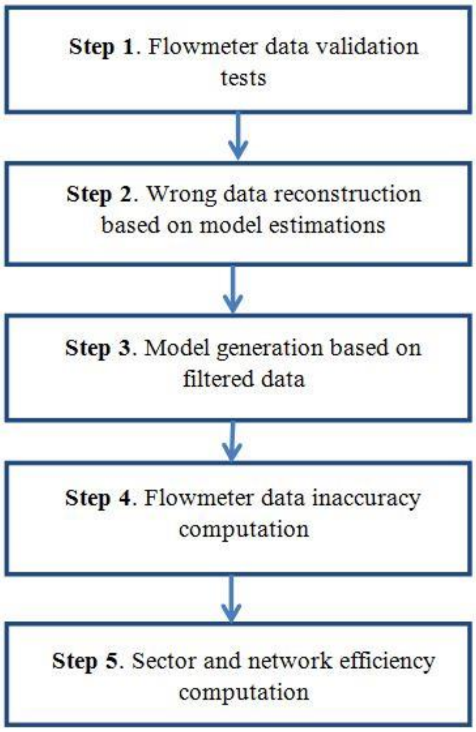

Figure 1. The integrated methodology

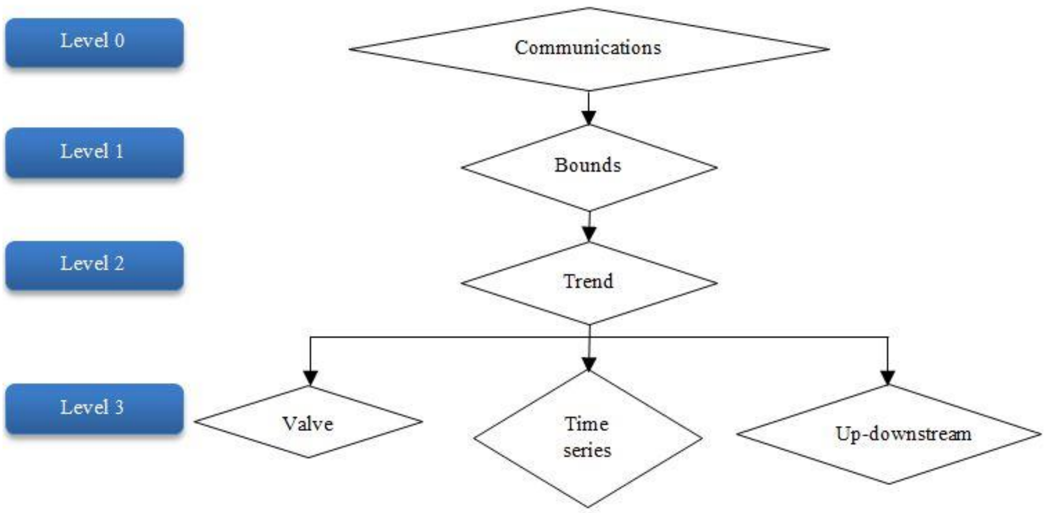

Figure 2. Raw flowmeter data validation tests

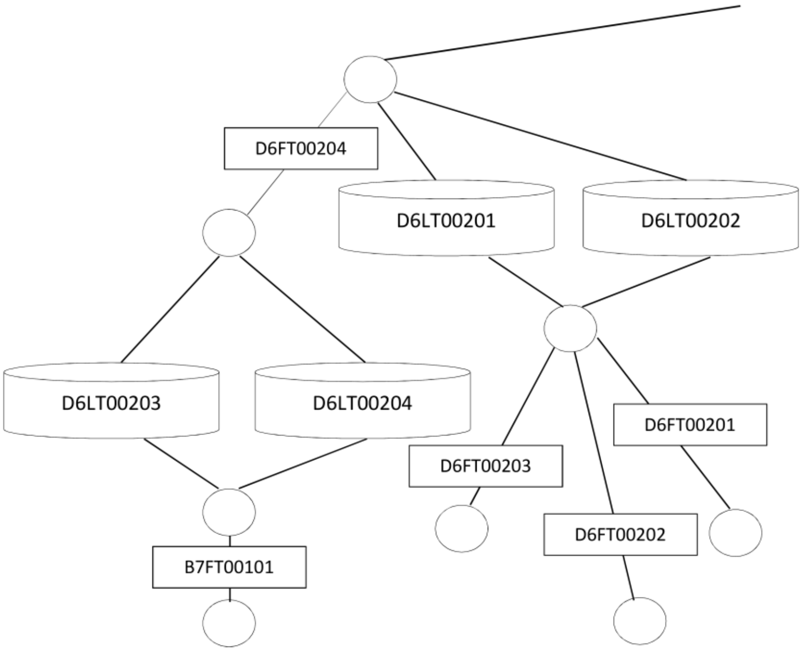

Figure 3. A piece of the ATLL network with several sectors

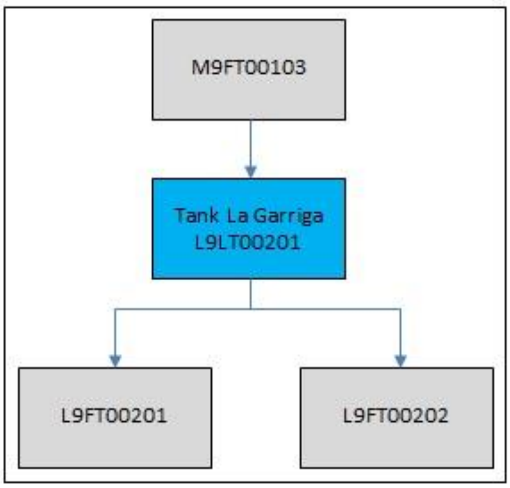

Figure 4. A sample sector with one input and two output flowmeters and one tank

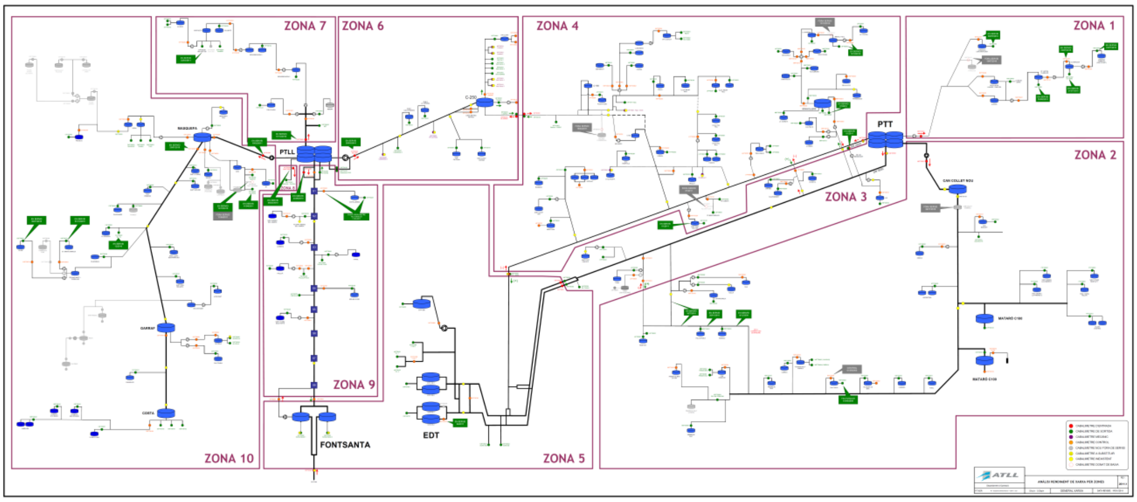

Figure 5. Sectorisation of ATLL transport water network

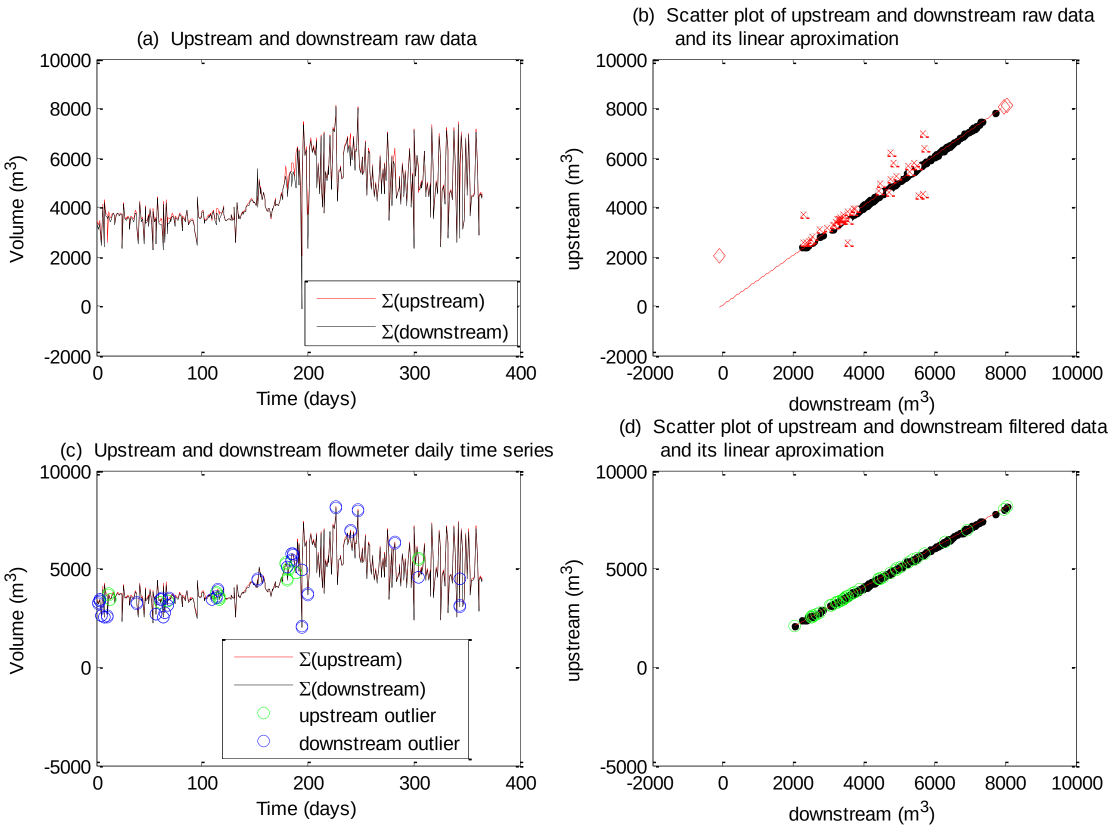

Figure 6. Graphical results corresponding to sector 1 zone 1

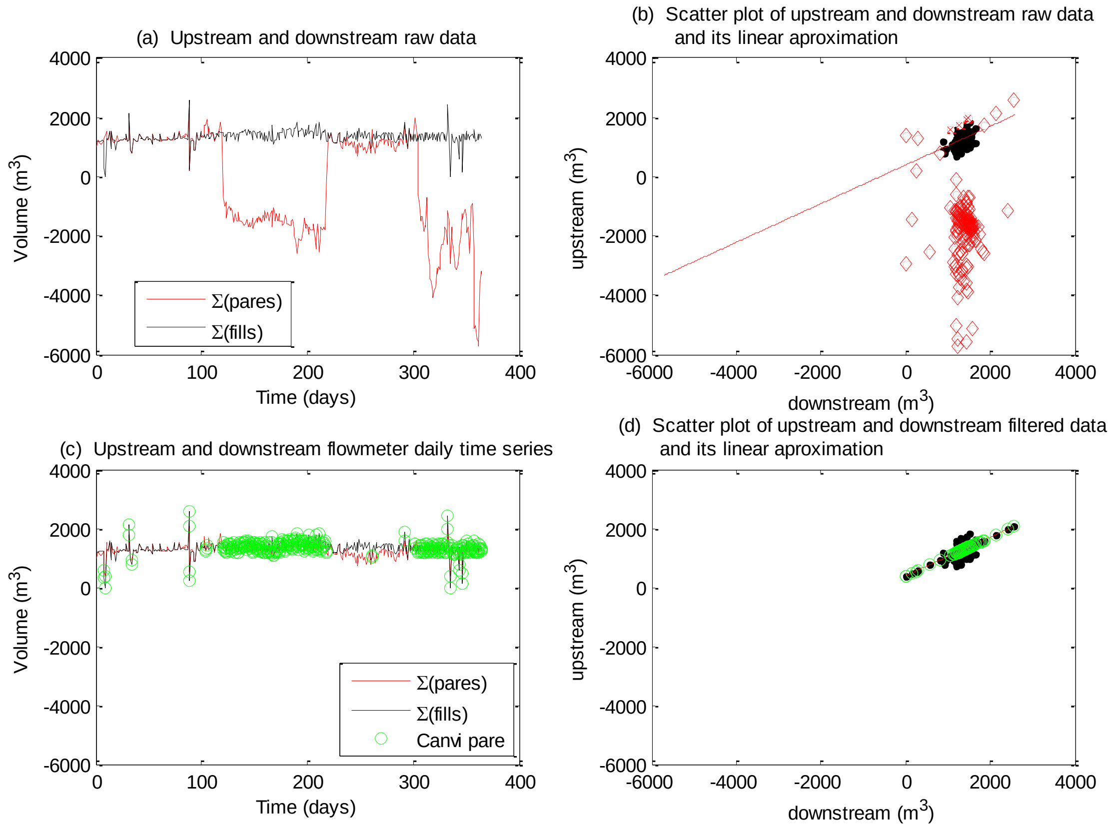

Figure 7. Graphical results corresponding to sector 2 zone 4