Modern Environmental Science and Engineering (ISSN 2333-2581)
June 2021, Volume 7, No. 6, pp. 588-594
Doi: 10.15341/mese(2333-2581)/06.07.2021/008
Academic Star Publishing Company, 2021
www.academicstar.us

Sensor Data Validation Methodology and Network Performance Assessment Applied to the Water Transport Network of Terrassa

Joseba Quevedo[1], Vicenç Puig[1], Pere Mora[2], and Adrian Oliva[2]

1. CS²AC Research Center, Polytechnic University of Catalonia, Spain

2. Taigua, Aigua Municipal de Terrassa Carrer de la Societat, 30, 08221 Terrassa, Spain

Abstract: Terrassa municipal water company (TAIGUA), through a collaboration research agreement with the CS²AC Research Center of the Polytechnic University of Catalonia (UPC), has applied a computer tool based on mathematical algorithms that allows validating the data from the flowmeters of the water transport network that supplies potable water to the city of Terrassa and other small cities surrounding. This water transport network is made up of 58 Km of pipes, 65 pumping stations and 13 reservoirs that supplies potable water to more than 218 thousand of inhabitants of Terrassa. Contains more than 28 flowmeters installed in pipes with a size ranging from 50 mm to 600 mm in diameter. The network is meshed and guarantees the continuity of the service through the contribution of the different production plants (ETAPs and IDAMs). Critical infrastructures (CIS), as in the case of the water transport network operated by TAIGUA, are complex systems of large scale, geographically distributed with a hierarchical structure. CIS require supervision and control schemes in real time (RTC) to guarantee high performance, but also demand adequate maintenance actions when malfunctions appear as, e.g., in sensors due to drifts, battery or communications problems. The development of the study presented in this paper involves the use of several mathematical models based on different algorithms that allow evaluating the automatic (daily) measurements and proposing the validity of the daily measurement or its possible error, leading in this case to a specific reconstruction of the value.

Key words: data analysis, data validation, data reconstruction, flowmeters

1. Description of the Water Network

TAIGUA operates water transport and distribution networks that supply the city of Terrassa and other small cities surrounding. It can be considered to be a real case of the so-called Critical Infrastructure Systems (CIS). These systems are complex because of the large scale, geographically distributed and decentralized with a hierarchical structure, requiring real-time control systems that guarantee high performance. However, they also demand adequate maintenance actions when malfunctions appear as, e.g., in sensors due to drifts, battery or communications problems.

The reliability of the information is the basis for making decisions that would allow to optimize energy consumption and to reduce water losses while guaranteeing an adequate supply to consumers in quantity and quality despite changing demands. The main objective of this methodology is to validate the raw data from the sensors (in this case flowmeters). And, if the data is not consistent, try to reconstruct it though estimation using mathematical models, maintaining a reliable, coherent and complete database system. This procedure allows to validate and to reconstruct the raw sensor data received and transform it into useful information. This will allow first to

Corresponding author: Joseba Quevedo, Dr.; research areas/interests: automatic control. E-mail: joseba.quevedo@upc.edu.

diagnose sensor malfunctions and then to calculate Key Performance Indicators (KPIs) for the management of the water transport and distribution system.

TAIGUA is a municipal company dedicated to the supply of drinking water for human consumption in the city of Terrassa and neighbouring cities (Ullastrell, Viladecavalls, Matadepera and San Quirze del Vallés) — It supplies water to more than 250 thousand inhabitants with an annual demand of more than 16 cubic hectometres. The network has more than 100 km of pipes with diameters of up to 700 mm. TAIGUA is in charge of managing this network guaranteeing the quality, quantity and continuity of the water supply. The water is extracted from Llobregat river at Abrera and from several wells. Water treament and delivery to the header tanks and the subsequent distribution of water to the consumers is done through a meshed network. The water service should be provided in an uninterrupted way 365 days a year, 24 hours a day, through an automated control centre.

TAIGUA network is subdivided into the transport and distribution network. The first network transports water either from previously captured in rivers and wells and after making it drinkable, or purchased from the ATL utility company, and then is driven by pumping stations to the head reservoirs. The distribution network supplies water from these tanks to the 8 pressure zones that include 43 DMAs. This work deals with the transport network that is schematized in Fig. 1, without considering water wells or supply to other cities different from Terrassa.

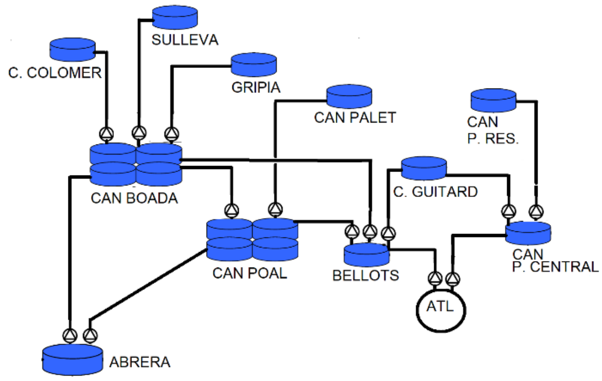

Fig. 1 Schematic diagram of the water transport network of Terrassa.

2. Data Validation and Reconstruction

The validation of the raw data of the flowmeters of the TAIGUA network is inspired by the UNE 500540 (2004) standard that defines a procedure to associate quality levels in the information obtained from automatic meteorological stations. This norm allows establishing minimum guidelines for the validation of meteorological data to know what manipulations they have been subjected to and the degree of validity of the information.

For each flowmeter, the validation system checks all the raw data collected by the remote control and certifies the validity of the daily data. This system labels as valid data those that have passed all the tests and as invalid data those that have failed at least one of the validation tests. The set of implemented tests that each of the raw data provided by the remote control system must pass is shown graphically in Fig. 2.

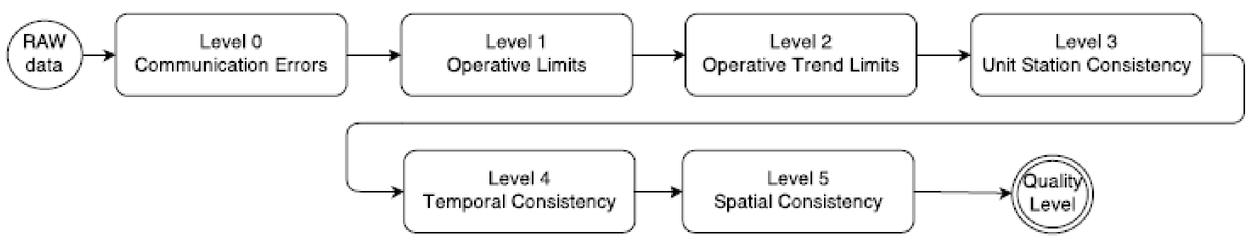

Fig. 2 Validation data flow diagram.

2.1 Level 0 Test (Communication Errors)

This level is intended to validate the structure of the data record. Thus, it is verified that all the expected information reaches the database adequately, and that none of the measured data is affected with any related alarm that could impair the quality of the measurement.

If any of the data cannot be correctly registered in the operational database because it has not arrived or because it has an alarm related to communication equipment, it will be considered as INVALID and it will become a candidate to being reconstructed.

2.2 Level 1 Test (Operative Limits)

Level 1 ensures that the data is within its physical working range. If the value is within the range (greater than a minimum and less than a maximum) the data is considered valid at this level. Otherwise, the data will be labelled as NOT VALID.

The upper and lower thresholds are the physical maximum and minimum values that are impossible to exceed by a consistent data. These limits are parameterized according to statistical analysis of their history.

2.3 Level 2 Test (Operative Trend Limits)

The level of trend or temporal coherence verifies if there is an abnormal variation of the measured flow in two consecutive sampling times, since the difference between two consecutive values must not exceed predetermined values.

The trend level validates the flow according to its growth or decrease. Therefore, if the difference in value absolute between the current value and the previous one does not exceed a threshold (maximum trend), the current flow is VALID. In otherwise, the data is considered NOT VALID.

2.4 Level 3 Test (Unit Station Consistency)

The unit station consistency monitors the possible correlation between the different variables in the same station remote (for example, the flow rate with the valve opening in the same pipeline).

This model takes advantage of the elements adjacent to the flow meter (valves, pumps, etc.) to monitor the system coherence. A flowmeter should not measure flow, if the associated valve is closed. On the contrary, the measure performed by a flowmeter is considered doubtful if its valve is open and it has not recorded any measure. Otherwise this raw data is considered NOT VALID

2.5 Level 4 Test (Spatial Model Consistency)

The spatial coherence model verifies that the values of the same variable measured at the same instant of time in nearby stations cannot differ too much from each other. It allows to know if there is coherence between physically separated flowmeters but that measure the same flow.

That is, the model estimates the inlet flow measurement of an inlet flowmeter according to the coherence of the sum of the outlet flow measurements located in the same pipeline or branch pipes close to each other.

Between the flowmeters that measure the inlet flow and the flowmeters that they measure the outlet flow we can find water tanks. Thus, between them there may be a volume variation that corresponds to warer stored in the tank. Then, the mass balance of this spatial model compares the inlet flows substracting the volume in the tank stored in the last period of time equals to outlet flowmeters. If the balance model is not consistency the raw data involved in the model are NOT VALID.

2.6 Level 5 Test (Time Series Model Consistency)

The flow that passes through the pipes of a water transport network is used for water consumption of a population that generally presents a daily repetitive behaviour and, therefore, can be estimated using a time series model.

The time series model takes advantage of the redundancy of the historical data of the same variable with periodic behaviour to estimate a present value. Later, it will compare the prediction to one step with the actual data received, if this difference in absolute value is greater than a preset threshold, the data is considered NOT VALID.

2.7 Data Reconstruction

When a raw data does not pass any of these five test levels, the data must be replaced by an estimation. Our methodology proposes to use either the spatial or time series models to reconstruct the not validated data. Also, other prediction techniques or a simple interpolation could be applied to reconstruct the wrong data.

3. Application to the water Transport Network of Terrassa

In order to illustrate how the methodology of data validation and reconstruction works in the water transport network of Terrassa, the application to several sectors of this network will be presented in this section.

3.1 From Can Boada to Sulleva reservoir

Sulleva in one the pressure zones of the distribution network and the header tank of Sulleva is fed from Can Boada reservoir by means of a pump station. Two flowmeters have been considered: the inflow of this sector is located in the pump station of Can Boada and the outflow of this sector is located at the output of the Sulleva reservoir which measure the volume of the water supplied to the Sulleva distribution subnetwork (Fig. 3).

±5% of the maximum head loss throughout the system; 95% of them within ±0.75 m or ±7.5% and 100% within ±2 m or ±15%.

A set of daily raw data of both flowmeters and the measurements of the Sulleva reservoir level are analyzed for data validation. Fig. 4 shows the time evolution of the inflow flowmeter substracting the stored water in the reservoir and outflow flowmeter during the year 2019. Fig. 5 shows the same signals in a X-Y plot showing clearly two red circled set of inconsistency data far from the straight line. And,

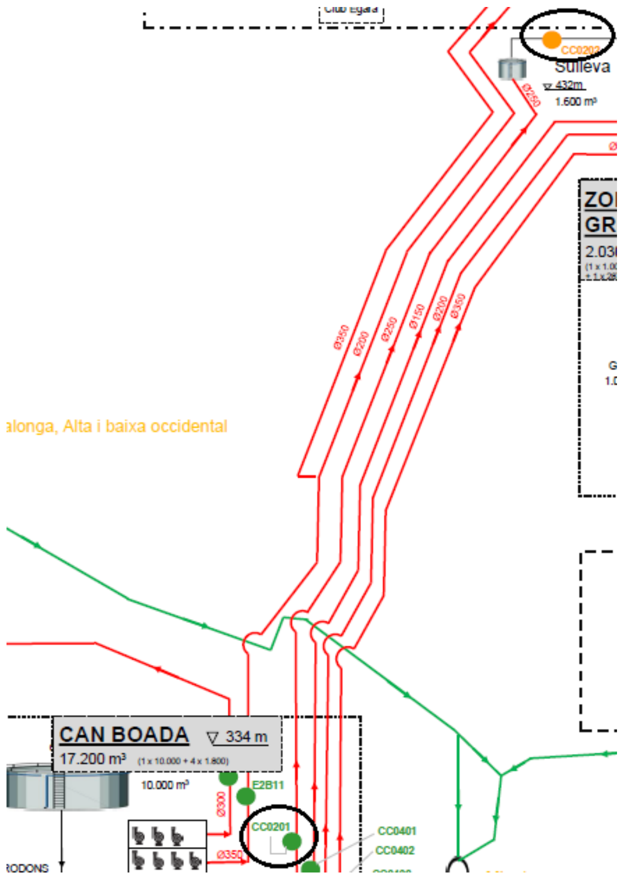

Fig. 3 A schema of the pressure zone Can Boada-Sulleva.

regarding the evolution of the input and output signals, we can determine that all the non-validated data are due to the inflow flowmeter and these data have been replaced from the values of the outlet flowmeter. Finally, Fig. 6 shows all the validated data in a new X-Y plot and an efficiency performance (KPI) of this pressure zone of Can Boada-Sulleva of about 98% and has been computed from the filtered data.

3.2 From Abrera to Can Boada Reservoir

One of the most important pressure zones of the upstream network of Terrassa is the interconnection between the Abrera pump station and the four head reservoirs of Can Boada with a pipe of 700 mm of diameter 27 Km of length and a pump station of 8 units to thrust the 250 meters of elevation from Abrera to Can Boada (Fig. 7). In this pressure zone, the inlet flowmeter is near to the pump station of Abrera and the

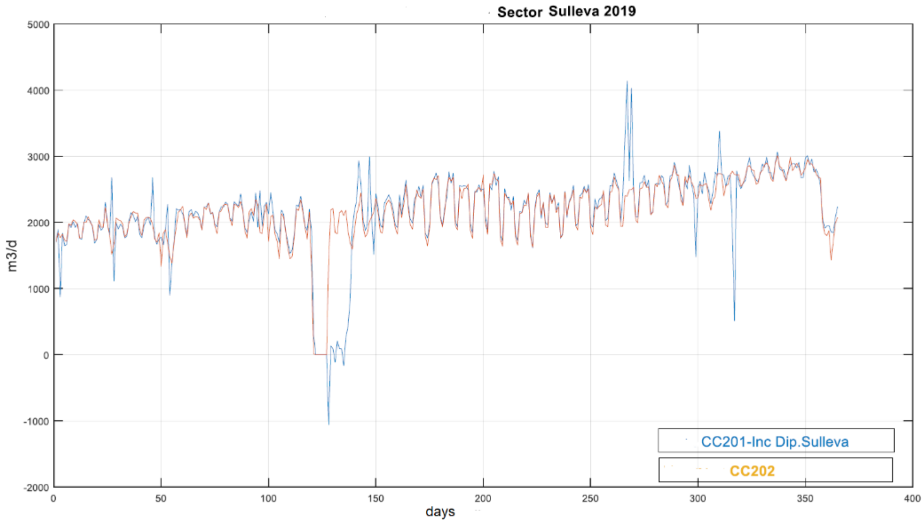

Fig. 4 Time evolution of the daily input and output of the Can Boada-Sulleva sector in the year 2019.

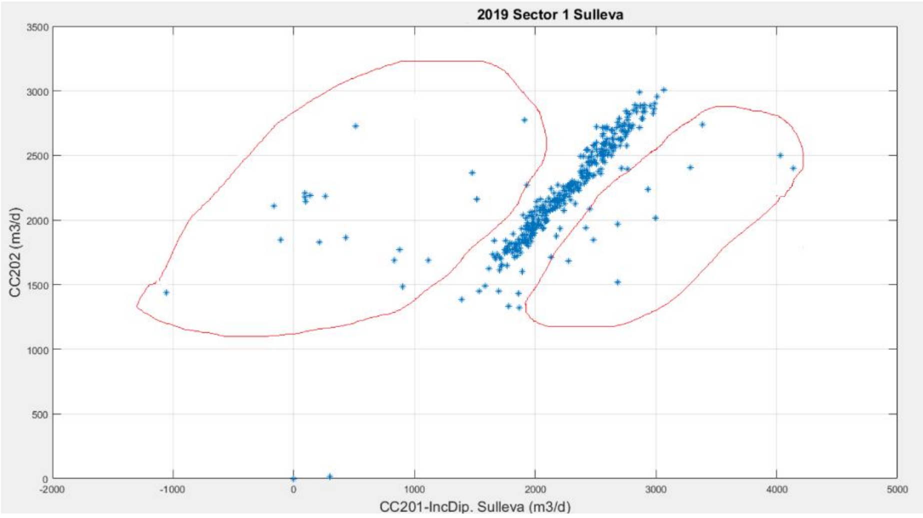

Fig. 5 X-Y plot of daily raw data input and output of the Can Boada-Sulleva pressure zone in the year 2019.

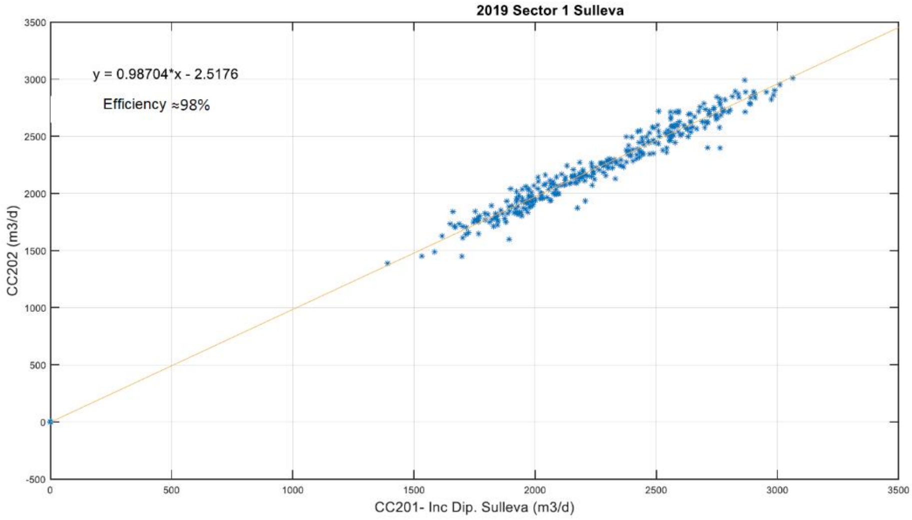

Fig. 6 X-Y plot of daily filtered data input and output of the Can Boada-Sulleva pressure zone in the year 2019.

outlet flowmeters are two sensors of the water supplied to Viladecavalls and one sensor measuring the inflow to the Can Boada head reservoirs.

Fig. 8 shows the time evolution of the inlet flowmeter and outlet flowmeters during the year 2019. This plot shows clearly four red circled sets of inconsistency of data, in the first circle the invalidated data are due to spurious values of the inlet flowmeter and the last three circles are not validated data of the outlet flowmeter. Once these wrong data are replaced by validated data, Fig. 9 shows the X-Y filtered data near the straight line which allows to determine a performance efficiency of 99.37%.

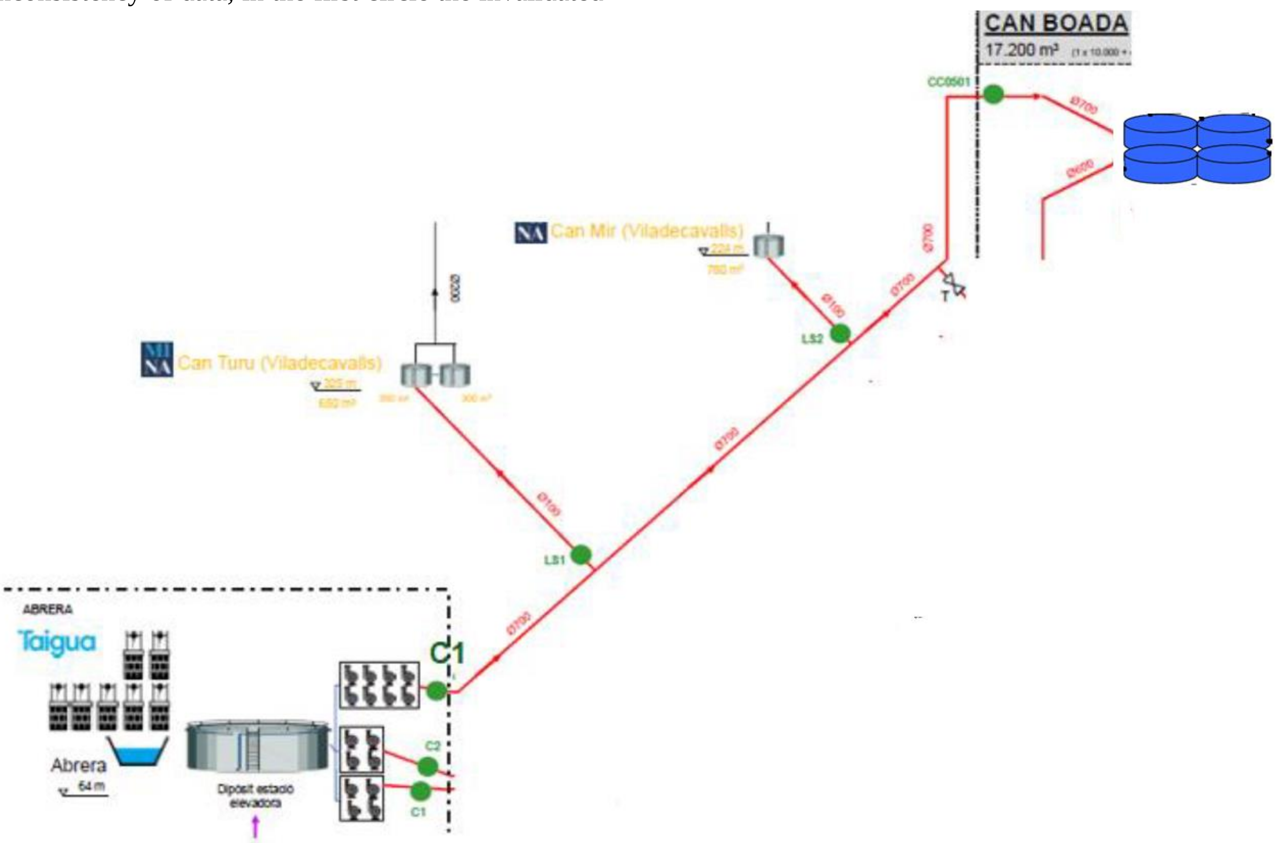

Fig. 7　A schema of the Abrera – Can Boada pressure zone.

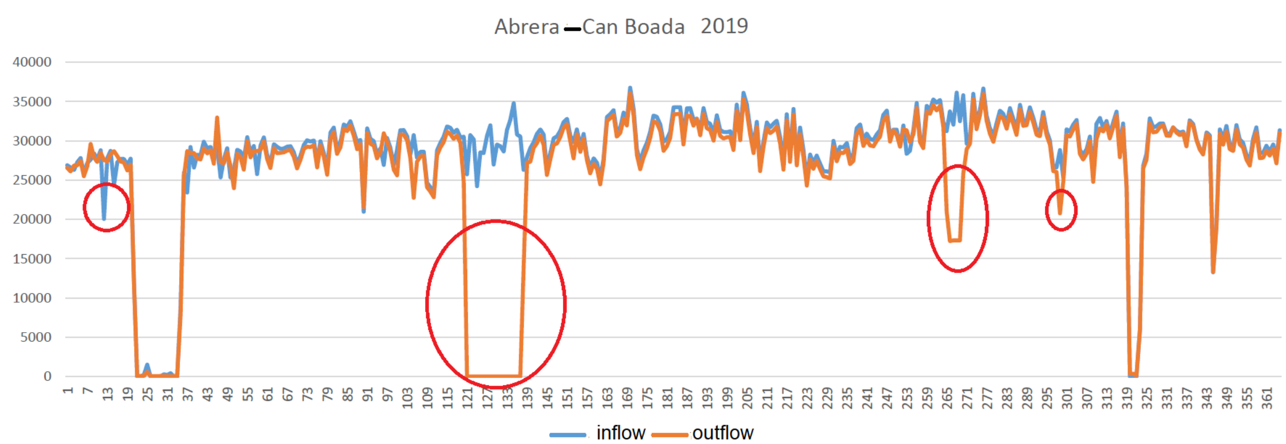

Fig. 8　Time evolution of the daily input and output of the Abrera-Can Boada pressure zone in the year 2019.

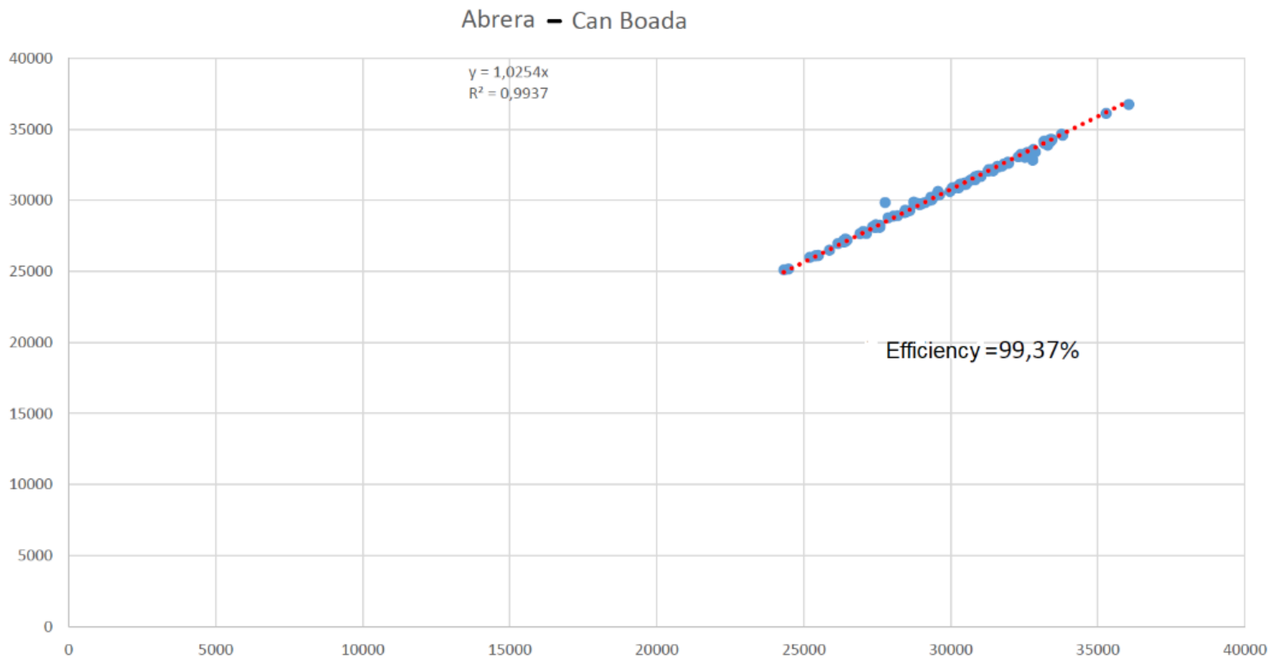

Fig. 9 X-Y plot of daily filtered data input and output of the Abrera-Can Boada pressure zone in the year 2019.

4. Conclusion

This paper has presented a methodology for the validation and reconstruction of raw sensor data in water networks to improve the reliability of the data and to extract interesting information such as the efficiency performance of the network. This methodology has been successfully applied to the water transport network of Terrassa and in the future could also be applied to the distributed water network.

Acknowledgment

This work has been partially funded by TAIGUA through the collaboration project with CS^{2}AC (ref. UPC C-11660) and data have been kindly provided by TAIGUA.

References

[1] A. Ostfeld and E. Salomons, Optimal layout of early warning detection situations for water distribution systems security, *J. Water Resour Plann Management* 130 (2004) (5) 37785, doi: 10.1061/(ASCE)0733-9496(2004)130: 5(377).

[2] C. Palau, F. Arregui and M. Carlos, Burst detection in water networks using principal component analysis, *J. Water Resour Plann Manage* 138 (2011) (1) 47-54, doi: doi: 10.1061/(ASCE)WR.1943-5452.0000147.

[3] J. Quevedo, H. Chen, M. A. Cuguero, P. Tino, V. Puig, D. García, R. Sarrate and X. Yao, Combining learning in model space fault diagnosis with data validation/reconstruction: Application to the barcelona water network, *Engineering Applications of Artificial Intelligence* 30 (2014) (0) 18-29, available online at: http://www.sciencedirect.com/science/article/pii/S0952197614000153.

[4] J. Quevedo, V. Puig, G. Cembrano, J. Blanch, J. Aguilar, D. Saporta, G. Benito, M. Hedo and A. Molina, Validation and reconstruction of flow meter data in the Barcelona water distribution network, *Control Engineering Practice* 18 (2010) (6) 640-651, doi: 10.1016/j.conengprac.2010.03.003.