Data Setup for Water Distribution System Supervision

Ramon Pérez[1], Sergi Grau[2], Víctor Jiménez[1], Xavier Martínez[3], Ricard Tomàs[2]
[1] SAC-UPC, Rambla Sant Nebridi 10, Terrassa 08222, Catalunya.
[2] Aigües de Manresa, Plaça de la Plana de l'Om, 08241 Manresa, Catalunya.
[3] Fundació CTM Centre Tecnològic Plaça de la Ciència, 2 08243 Manresa, Catalunya.

[1]ramon.perezr@upc.edu

ABSTRACT

The availability of on-line data coming from the water distribution systems (WDS) allows the monitoring of such critical infrastructures. Nevertheless the huge amount of data that come to the control centre implies an enormous processing challenge [4]. The use of data avoids any physical theory and relies on statistical correlations and inferences. Nevertheless the previous efforts in modelling the networks encourage the companies to fusion information coming from both sources. Models help validating data and data update these models. This paper presents the first stage of an ongoing project focused in the integration of data and models. Data are collected, harmonised and validated using the models so that they will be used in following stages of the project for the supervision of the WDS (water balance and quality). A tool is being developed in R where the different modules will be integrated.

Keywords: Data Management, Calibration, Sensor Validation

1 INTRODUCTION AND PROBLEM STATEMENT

One of the main issues that motivate the use of water distribution models and on-line measurements are the leakage detection and localisation. There are methods based on data analysis [12] but most of the methods that use on-line measurements rely on models, either transient [1] or static [13, 9, 3]. The *Centre for Supervision, Safety and automatic Control* (CS2AC) at the UPC in Terrassa collaborated with the water utility of Manresa for the calibration of the models in order to use them in the leakage management procedure [8, 12]. This work was presented in the previous Conference in Computer Control for Water Industry in Leicester [10]. One main conclusion of this work was the confirmation of calibration benefits in the insight that models together with the data may produce and the need for an automated procedure. This automatization is mandatory as the number of district metered areas (DMA) in a network makes impossible a manual maintenance of all the models.

The water company, motivated by the preliminary results, has undertaken a new project with the *fundació CTM Centre Tecnològic* (CTM) that collaborated with CS2AC. This three-year project has ambitious objectives:

1. To create an automatic procedure and application for merging data coming from the SCADA with the hydraulic models used by the company. This procedure should update the models automatically inducing their intensive use.
2. To use the combination of modelling and on-line monitoring for the leakage management strategy, detecting and locating the leaks within the DMA.

3. To associate a chlorine decay model to the hydraulic model. The reliability of the hydraulic model comes from the first objective of the project while the quality model must be combined with an on-line calibration of its parameters.

This paper presents the results of the first phase of this project. The methodology applied is described in Section 2. Next, in Section 3, the tool developed and results obtained in a real network with real data are presented in order to illustrate the procedure and allow its analysis. Finally conclusions and future work are discussed in Section 4.

2 CASE STUDY

The transport network of Pineda de Bages (Figure 1), a sector of the Manresa network located in the centre of Catalunya, is used as illustrative example. The data of year 2015 were acquired and treated, the balances analysed so that conclusions about sensor reliability could be stated and finally the hydraulic model in EPANET was simulated.

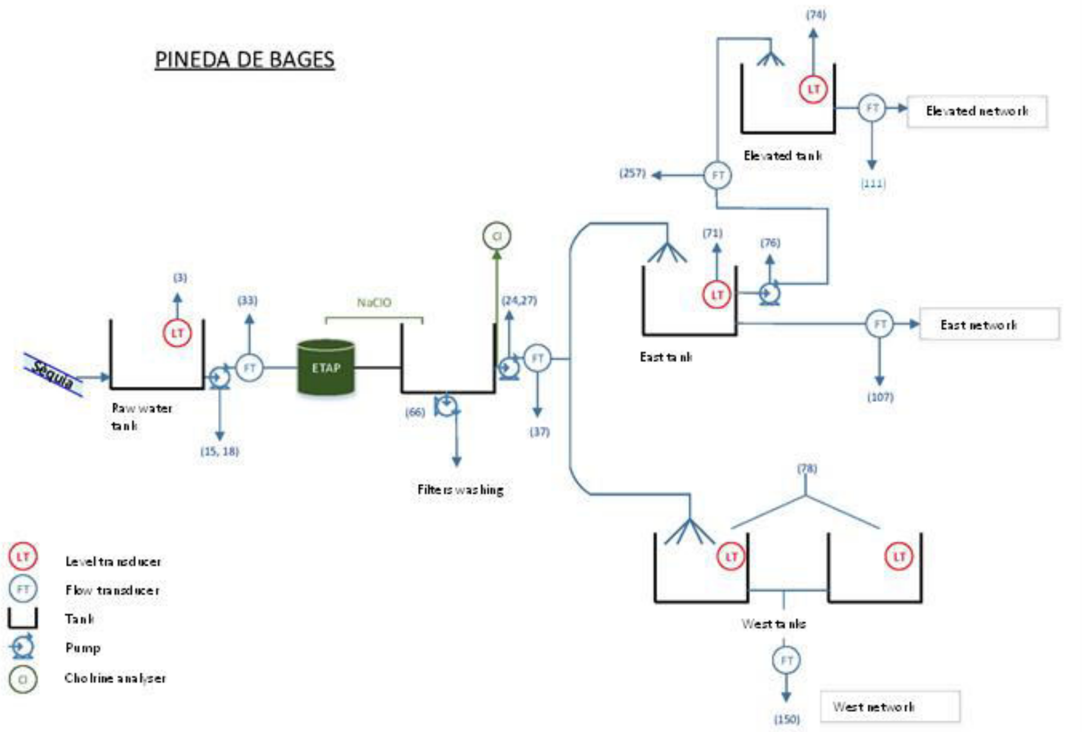

Figure 1: Case study WDN for Pineda de Bages

The system includes a treatment plant (ETAP) that provides the water to three subsystems with storage capacity. The outflows to the District Metered Areas (DMA) are monitored, so are the tank levels and their inflow. Nevertheless, two of the systems get the water from the same measured pumping. The water balances are not difficult to stablish but they are not straightforward, thus, an interface for their definition is of major interest for the company.

3 METHODOLOGY

The validation, resampling and homogenisation of the data are the first steps described. Once the data are available for the entire network they are organized by DMA using a simple water balance model. This model allows the generation of graphical and numerical analysis for sensor and parameter validation. Once the information of the water balance in the DMA becomes coherent the hydraulic simulation of the sector becomes reliable for further analysis. Figure 2 shows how the data go through the three stages becoming the knowledge expected by the experts. The on-line function of the system is represented by the white arrows.

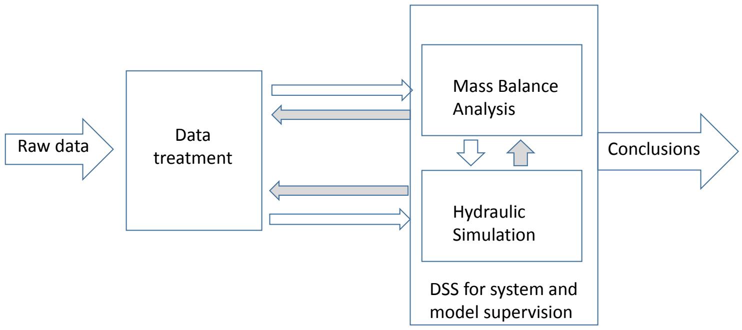

Figure 2: Data flux through the DSS for the system analysis

3.1 Data Treatment

The objective of the data treatment is to take the data provided by the SCADA system and adapt them to be used by the analysis tool. Data must be resampled in order to have a constant and synchronous sampling frequency. The company provides daily files with all the SCADA registers. The variables corresponding to the case study network are selected. Their relative time label (daily) is converted to the absolute time format. Interpolation is used for the resampling so that data are available both for the water balance analysis and the simulation.

The SCADA reads all the measurements induced by a change in values of the sensors. This irregular sampling is overcome by a high frequency (0.2Hz) reconstruction where in absence of data the previous value is kept. This reconstruction avoids the different number of samples depending on the values (high flows variability are more likely to induce measurements than lower values in the night) and assures that all the switches of the pumps are detected. Thresholds for spurious are defined so that they are removed. Values are averaged in a reasonable sample (10 minutes). Finally each type of measurement has to be treated differently in order to harmonise the units (for example levels are given in percentage of the total).

3.2 Water Balance

The first evaluation of the model adjustment is based on the water balance between the available data. Four levels of analysis are defined. Each of these levels allows to explain with further detail the inaccuracies detected in the previous level. In a well-adjusted system the first level would suffice.

1. Total integral of those systems with known input and output.

2. Evolution of the input and output volume (integration of flows). Linear and quadratic models are adjusted to the non-revenue for water (NRW) in order to estimate the possible leak/mismeasurement .
3. Comparison of the tank volume and the flow balance (with different pumping status). A form factor is calculated in order to explain the difference between both measurements in terms either of the volume or flow measurement.
4. Study of the pairs pumping status and flow measurements.

In a first version matlab was used for carrying out the calculations and the results were stored in a spread sheet so that the expert may analyse both the graphics and the data themselves. Afterward a web-application was developed in R using the library Shiny. In Section 4 the screens and graphics of this tool are presented using the illustrative example. The conclusions will help improving the hydraulic model or detecting faults in the system (mainly sensors but even leaks).

4 RESULTS

An application has been developed to interpret contour data, to automatize the repetitive processes, to give decision support to the expert user and to macro-calibrate the model. This application was made under R programming language because its use has many advantages: most used language for Analytics, Data Mining and Data Science asked about a poll on a representative sample of the professional in these areas [7]; it is for free, it is a collaborative project where the users are part of the continuous improvement of the Language and the Package creation (it has more than 8100 libraries) including many for the job analysis of BigData like sqlf, DBI, RSQLite, filehash, bigmemory, ff, other for communication environments as EPANET as epanetReader library and others for powerful graphic representations like dygraphs, ggplot2, lattice, rCharts, googleVis, ggvis, shiny. The latter, Shiny, has been able to create this tool like a web page with all the advantages of use and visual environment involved, as well as access to it through web servers.

This APP allows to do several functions:

- Load the data from a .csv file, which identifies, first, the type of variable before read all the data table.
- Define the family of each data set, to identify from which kind of sensor has been reading for each data series, as will different treatment.
- Accumulate the data series values to see trends over time and identify behaviours.
- Subsystems creation, to make more intuitive the work isolating the cases and assigning relations between readings, filtering data sets to study concrete cases (for a different pump status cases)
- Graphical data representation.
- Save the table results and the procedures in an external file.

Figure 3 presents the subsystem Elevated Network (AE) of the case study (Figure 1), formed by the X257 flowmeter as the water input and the X111 flowmeter as water output and the X74 as level from the deposit. The analysis presented here is for 15 days:

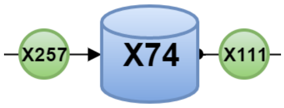

Figure 3: Scheme from AE sector

Figure 6 shows how the evolution of the accumulated flowmeter X257 follows a phased manner due to the on/off status of the pumps providing water to the tank.

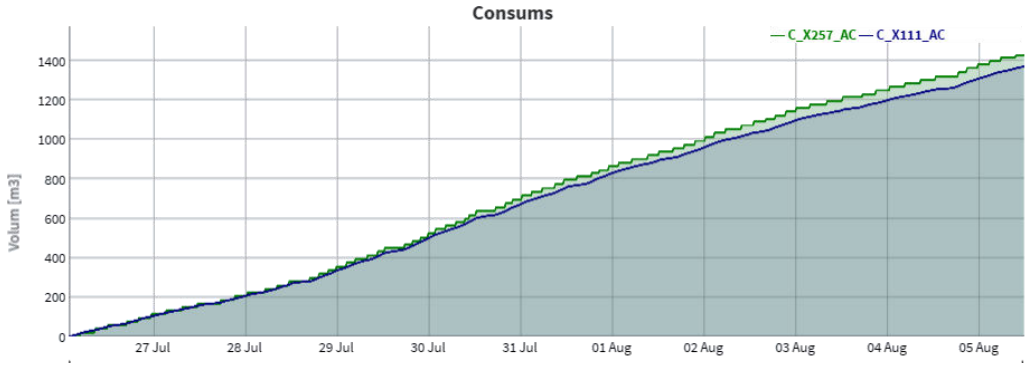

Figure 6: Water input and output evolution in AE sector

With the system defined, the accumulated balance is displayed (the difference between the input and output accumulated data set). Figure 7 shows the balance of the AE sector in green, and it shows, in blue, the balance considering the information on the tank, removing the dynamic balance. At Table 1, we can observe the total values of the input, output, difference (NRW) and the NRW factor as well as linear and quadratic regression coefficients, AL R2, for both system definitions. An improvement is observed in the regression when the dynamic is removed. If the quadratic regression R2 AR were much higher than the linear regression would get evidence of water leakage. In this case, indicates that the error is due to poor calibration of some sensors:

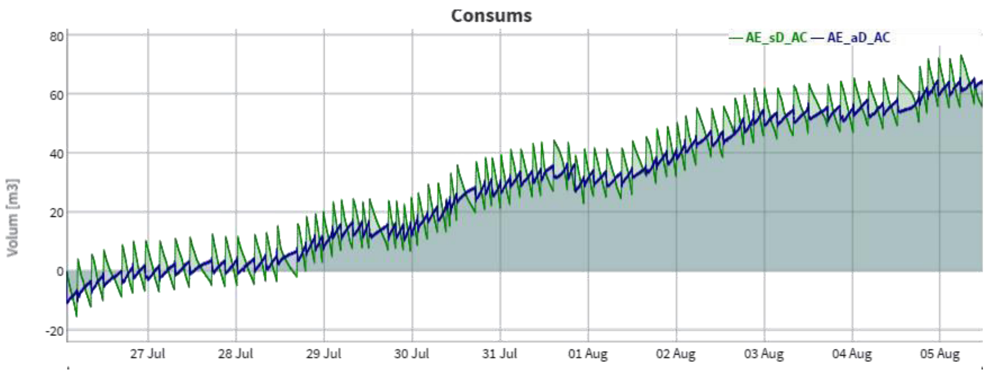

Figure 7: Shows the accumulating balance of the AE sector in green and, in blue, the balance considering the information of the tank

Table 1. Informative indicators of the AE sector

	In (m3)	Out (m3)	NRW(m3)	NRW Rel%	AL origent	AL coef. lineal	AL R2	AQ origent	AQ coef. lienal	AQ coef. quadratic	AR R2
AE_sD	1429.91	1368.76	61.14	4.28%	-7.8668	0.3015	0.9372	-7.8892	0.3021	-0.0000	0.9372
AE_aD	1429.91	1366.62	63.28	4.43%	-9.3500	0.2979	0.9825	-8.9172	0.2875	0.0000	0.9826

One of the main functions of this tool is to apply filters in datasets to study the different cases isolating the possible mismatch cause. Utilising the status of the pump two cases arise, The the filling up tank case is compared with the emptying tank case. Figure 8 shows the resulting graph for the AE sector of emptying case, when there isn't any water input by the X257 flowmeter, in other words, when the status of the pump that actives the tank filling is switched off (=0). It shows the integral of output flow (AE_B_D_AC measured with sensor 111) and the integral of volume variation (AE_B_Cf_AC measured by level sensor 74) and the adjustment of one curve to the other using a form factor (equation 4.1) applied at the level measurement (equation 4.2) and its inverse applied at the flowmeter (equation 4.3). Both adjustments work perfectly as the form factor is calculated for this situation. The application of this form factor when the tank is being filled may enlighten which of the sensors is more likely not to work properly.

$$ff_{AE} = \frac{AE\_B\_C\_AC}{AE\_B\_D\_AC} = \frac{\sum_{j=0}^{k}(-q_{X111}(j)) * \Delta t}{\sum_{j=1}^{k}(V_{X74}(j) - V_{X74}(j-1))} \quad \text{(equation 4.1)}$$

$$AE\_B\_Df\_AC = ff_{AE} * AE\_B\_Df\_AC \quad \text{(equation 4.2)}$$

$$AE\_B\_Cf\_AC = \frac{AE\_B\_Df\_AC}{ff_{AE}} \quad \text{(equation 4.3)}$$

All these calculus are subjected to the condition of emptying, i.e. $q_{X257} = 0$.

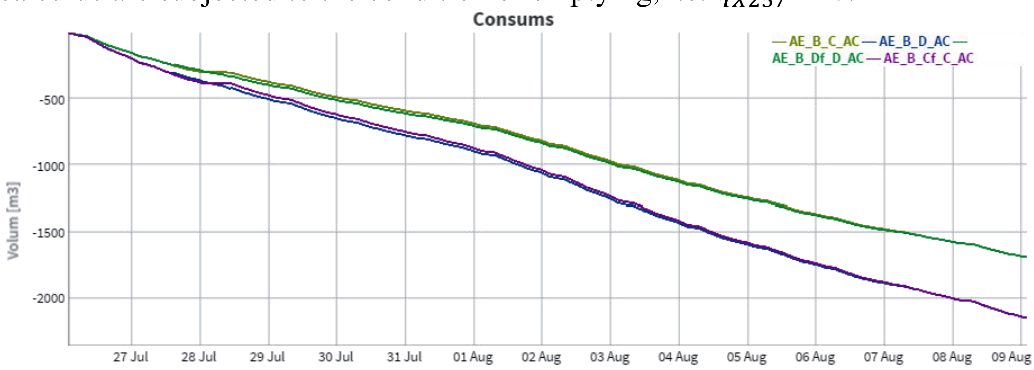

Figure 8: Example of the water consumption, in the emptying case, measured by the flowmeter (in ochre) or by the tank (in blue) with their respective calibrations (in green for the X74 level sensor and purple for the X111 flowmeter calibrated)

The following is a study of all the cases when the status pump of the water input flowmeter is switched on, different than 0 and positive, the fill up case. Figure 9 shows the flowmeters and the tank measurements as well as these ones with the emptying form factor (equation 4.1) applied on the output flowmeter (equation 4.4) and the last case with the emptying form factor (equation 4.1) applied at the level sensor of the tank (.

$$AE\_E\_Cf\_AC = \sum_{j=0}^{k}\left(q_{X257}(j) - \frac{q_{X111}(j)}{ff_{AE}}\right) * \Delta t \quad \text{(equation 4.4)}$$

$$AE\_E\_Df\_AC = ff_{AE} * AE\_E\_Df\_AC \quad \text{(equation 4.4)}$$

All these calculus are subjected to the condition of filling, i.e. $q_{x257} \neq 0$.

It can see, in the graph, how the tank calibrated case is better than the other (water output sensor calibrated) being more similar in this evolution. $AE\_E\_C\_AC \approx AE\_E\_Df\_AC$ que $AE\_E\_D\_AC \approx$

$AE\_E\_Cf\_AC$. The method signals the level sensor as the faulty one. Most likely the total capacity of the tank has been updated in the sensor without updating it in the SCADA and the percentage of the level is misunderstood by the system. This habitual fault provides the name to the calibration parameter (form factor of the tank).

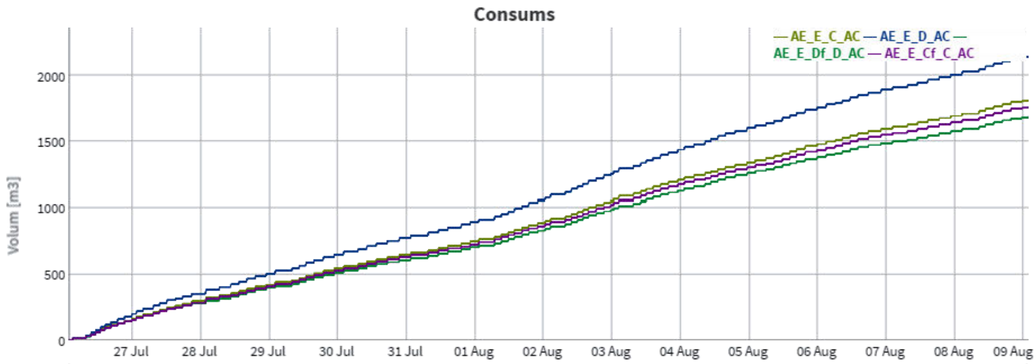

Figure 9: Evolution of the AE sector for the fill up cases. Flowmeters (X257-X111) measurements (AE_E_C_AC in ochre), tank level measurements (X74) (AE_E_D_AC in blue) with the tank sensor (X74) calibrated (AE_E_Df_D_AC in green) and with the water output flowmeter sensor (X111) calibrated (AE_E_Cf_AC in purple)

5 CONCLUSIONS

The tool presented in this paper has been applied to the case study detecting faults in two flow sensors and two tanks including the illustrative case presented in Section 4. The updated version of the tool is being introduced in the company for analysing the over thirty DMA data. Once the data are validated, the sensors suspicious of significant errors are tested on field and repaired or substituted if necessary.

Once these data are reliable the leak detection methods can be applied using the flows delivered to the DMA. Reliable pressure and flow measurements are required for the hydraulic and quality simulation aimed in the last stage of the project.

6 ACKNOWLEDGEMENTS

This work has been funded by the Spanish Ministry of Economy and Competitiveness (MINECO) and FEDER through the projects ECOCIS (ref. DPI2013-48243-C2-1-R) and HARCRICS (ref. DPI2014-58104-R).

References

[1] Colombo, A.F., Lee, P. & Karney, B.W., 2009. A selective literature review of transient-based leak detection methods. Journal of Hydro-environment Research, 2(4), pp.212–227

[2] Gangl, G., Kölbl, J., Haas, G., & Hassler, E. (2007). Influence of Measurements Inaccuracies at a Storage Tank on Water Losses. IWA Conference on Water Loss. Bucharest March 2007, 474–484.

[3] Goulet, J.-A., Coutu, S. & Smith, I.F.C., 2013. Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks. Advanced Engineering Informatics, 27(2), pp.261–269.
[4] Harford, T., Statistical, R., & International, S. (2014). Big Data: Are We Making a Big Mistake, 14–19.
[5] Jiménez V. Estudi i disseny d'una eina de calibració de models de xarxes d'aigua. Master thesis. http://upcommons.upc.edu/handle/2117/102475
[6] Mounce, S. R., Khan, A., Wood, A. S., Day, A. J., Widdop, P. D., & Machell, J. (2003). Sensor-fusion of hydraulic data for burst detection and location in a treated water distribution system. Information Fusion, 4(3), 217–229.
[7] Piatetsky-Shapiro, G. (s.f.). R, Python Duel As Top Analytics, Data Science software – KDnuggets 2016 Software Poll Results.
[8] Pérez, R. et al., 2014. Leak Localization in Water Networks: A Model-Based Methodology Using Pressure Sensors Applied to a Real Network in Barcelona [Applications of Control]. IEEE Control Systems, 34(4), pp.24–36.
[9] Pérez, R. et al., 2011. Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. Control Engineering Practice, 19(10), pp.1157–1167.
[10] Roma, J., Pérez, R., Sanz, G., & Grau, S. (2015). Model Calibration and Leakage Assessment Applied to a Real Water Distribution Network. Procedia Engineering, 119, 603–612.
[11] Sanz, G., Pérez, R., Kapelan, Z., & Savic, D. (2015b). Leak Detection and Localization through Demand Components Calibration. Journal of Water Resources Planning and Management, 04015057. doi:10.1061/(ASCE)WR.1943-5452.0000592
[12] van Thiennen, P. & Montiel, T. (2014). Flow Analysis and Leak Detection With the Cfpd 4Method in the Paris Drinking Water Distribution System. International Conference on Hydroinformatics. New York August 2014.
[13] Wu, Z.Y., Sage, P. & Turtle, D., 2010. Pressure-Dependent Leak Detection Model and Its Application to a District Water System. Journal of Water Resources Planning and Management, 136(1), pp.116–128.P. Snickety and M. Ticulous, “Abstracts: an Abstract Form,” J. Self. Ref. vol. 2, pp.11-132, 2011.