Leak Detection in a DMA, a Real Application of Flow Modelling

Víctor Jiménez[1], Sergi Grau[2], Ramon Perez[1]
[1] SAC-UPC, Rambla Sant Nebridi 10, Terrassa 08222, Catalunya.
[3] Aigües de Manresa, Plaça de la Plana de l'Om, 08241 Manresa, Catalunya
[1] *ramon.perez@upc.edu*

ABSTRACT

This paper presents a versatile methodology to calculate parameters that characterise the demand of a DMA. This parameters are used for the leak modelling so that a predictive model is built and trained with historical data and used to detect on-line new leaks so that the repair time can be reduced applying proper leak localisation techniques. This methodology has been programmed using R where the modelling packages available provide assortment of predictive models easily to implement. It has been integrated in a Data analysis tool in order to utilise the great amount of information coming continuously from the WDN. Once the methodology and the tool are described the results applied to a real DMA are presented. This work has been carried out by the fundació CTM Centre Tecnològic (CTM) collaborating with Research Center for Supervision, Safety and Automatic Control (CS2AC) within a research project of Aigües de Manresa.

Keywords: Data Management, Leak Localisation, Demand Modelling

1 INTRODUCTION

The leakage management in a Water Distribution Network (WDN) is an issue in continuous improvement. It attains asset management, supervision and control. The supervision is based on the information that can be gathered from the WDN both on-line and off-line and uses models to convert this information into knowledge and decisions. The aim of supervision is to localise the leaks present in the WDN so that they can be fixed or reduced by means of control. The search in such huge systems is a real challenge and the literature available proves it [3, 2, 10, 6, 4, 8]. The sectorisation of the WDN in District Metered Areas (DMA) is a generalised procedure that help in this task. Once the problem is divided the effort has to be focused in the right part. The leak detection procedure helps the practitioners to decide which DMA's have to be analysed. The performance of the DMA's based on the quarterly information coming from billing has to low frequency [5]. Other approaches based on DMA classification [7] or the minimum night flow analysis [5] are related poorly with their on-line behaviour. The analysis of the flow entering a DMA is used in an on-line scheme [9] thanks to the availability of the input flow in the DMA's. This approach has been reproduced by the authors who missed flexibility in the modelling of the leaks based on flow information

2 METHODOLOGY

The modelling and leak detection are integrated in a data analysis tool. It has been programmed in R using, among others, the shiny library so that it was ready to become a web-tool. Its purpose is to carry out data analysis studies to study the water distribution networks hydraulic balances and validate the data and sensors in first instance. Once the data are reliable they are used for the leak detection based in the methodology presented in previous section. Figure 1 shows the appearance of its main menu. It can access the different functions via the pop-up menu. Down below, is a brief description of this functions.

Figure 1: classical web appearance, consisting of a pop-up menu at left side of the web and in the main panel it can see each of the functions pages

- **Extraction and data cleaning**: The application allows to extract the information from the generated SCADA files stored in a database for setting up the data for the next analysis functions. It allows define the kind of the sensor because it future treatment would be different and it allows define also spurious high values to delimiter the accept values and reconstruct spurious and empty data.
- **Consulting the data**: The sensors and the range data to use for the analysis can be selected. It allows to consulting the data at each step of the data analysis too. It allows also to import and export the data table at any time to save the progress and continue the work in future.
- **Data table operations**: The family define functions permit defining the kind of sensor of each data series; the *system define* function creates new data series with the inputs and outputs flowmeters that belong to each system; the delete data series function; the filter functions create new data series with a fill or empty reservoir condition; the spurious function deletes values out of the minimum-maximum range.
- **Summary of results**: The numerical indicators of the predefined systems are presented: the quantity of input or output water, the difference of two, the non-revenue for water (NRW), and its linear and quadratic approximations' coefficients. For the filtered data series, it shows a table with the amount of water input and output, the difference and the percentage that the balance represents.
- **Graphics representation**: It visualizes an interactive data series graph using *dygraph* package. It helps as a decision support system, exploration and interpretation of data.
- **Leak modelling and prediction**: The flow characterisation is done using the data of the flowmeter. The historical leaks registered can be added so that a model of leaks can be trained. This model is applied to fresh data and the prediction of leaks using the characteristics of the flow is presented numerically and graphically. This paper devotes to this functionality. The

methodology is described next, both the night flow estimation and the leak modelling and prediction. The results obtained using the tool are presented in next section.

The in-flow of a DMA is often the only variable measured that has to do with a possible leaks existing in it. The normal strategy is to study the evolution of the night flow which increases as the leak evolves. The idea of the proposed methodology is to automatize this study. If the study is done automatically it can be more sophisticated without the limits of human resources. In order to enhance this study not only the night flow is taken into account but other characteristic variables of the in-flow.

In this work all the variables are calculated using the data of a single day. This approach avoids the temporal dependency of the predictions that are done daily. Nevertheless, new variables can be easily programmed and added to the leak models as it will be seen. Some experimental variables are developed and may be included in future up-dates of the tool. The variables currently available are:

- **Qmin**: Mean of the 5 minimal flow measurements.
- **Qnoct_mean:** Mean of the flow measurements of the night time (user defined).
- **Qdia_mean:** Mean of the flow measurements of the day time (user defined).
- **Qnoct_median:** Median of the flow measurements of the night time (user defined).
- **Qdia_median:** Median of the flow measurements of the day time (user defined).
- **Qnit/Qdia:** Ratio of **Qnoct_median** and **Qdia_median**.
- **Qnit_sd/Qnit:** Ratio of standard deviation of flow measurement of night time and **Qnoct_median**.
- **Qdia-Qnit:** Difference of **Qnoct_median** and **Qdia_median**.
- **Qmin/Qnit:** Ratio of **Qmin** and **Qnoct_median**.

The chosen model is Logistic Regression. It is particularly fit for predicting the probability of an event (dependent variable, leak) to occur from the observations (independent variables, ex. night flow). The probability function used is the logistic response:

$$P(y=1) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}}$$

Where x_i are the values of the independent variables and β_i are the parameters that have to be estimated. Manipulating the expression their estimation becomes easily a linear regression assuring the convergence and numerical efficiency. This model is estimated with the historical data including episodes of leaks detected and then it is used for the estimation of the probability of a leak to be present that day.

3 CASE STUDY AND RESULTS

In order to validate the methodology and the tool it has been applied to a real water network model of Pineda de Bages, a housing estate of Sant Fruitós de Bages counting about 1000 residents and provided by CTM (Centre Tecnològic de Manresa) and Aigües de Manresa, a water supplier company in Manresa near Barcelona in Spain. The water comes from a ditch and goes to a reservoir. Here it was a drinking water treatment station that loses water with the filter cleaning process. Once this has been done, the water goes to supply three neighbourhoods via other reservoirs, pumps and flowmeters.

The water balance study has been done to check the sensors reliability. Thanks to these results the water company has been able to found a model behaviour anomalies. In order to study the night flows, the data from 15 years about the validated X111 flow meter has been used, since 2000 to 2015. The mean night flow in this period is of and its mean day flow is about. The night flow and the other parameters associated with its behaviour are calculated for each day. The company provided the registered events that included the leak episodes that were reported and fixed. These events are associated with a day (Figure 2).

In order to include all the days when the leak was present it is necessary to evaluate each alarm event. Only the most clear leak events are selected. An expert analyses the graphical information about the night flow indicator of each day with the events overlapped and identify all the range of days that this event leakage reaches, the biggest leakages are easier to delimited, as always occurs. In Figure 2 this procedure is presented while in Figure 3 the all scenario of 15 years appears with the alternative visualisation that empathises the leaks.

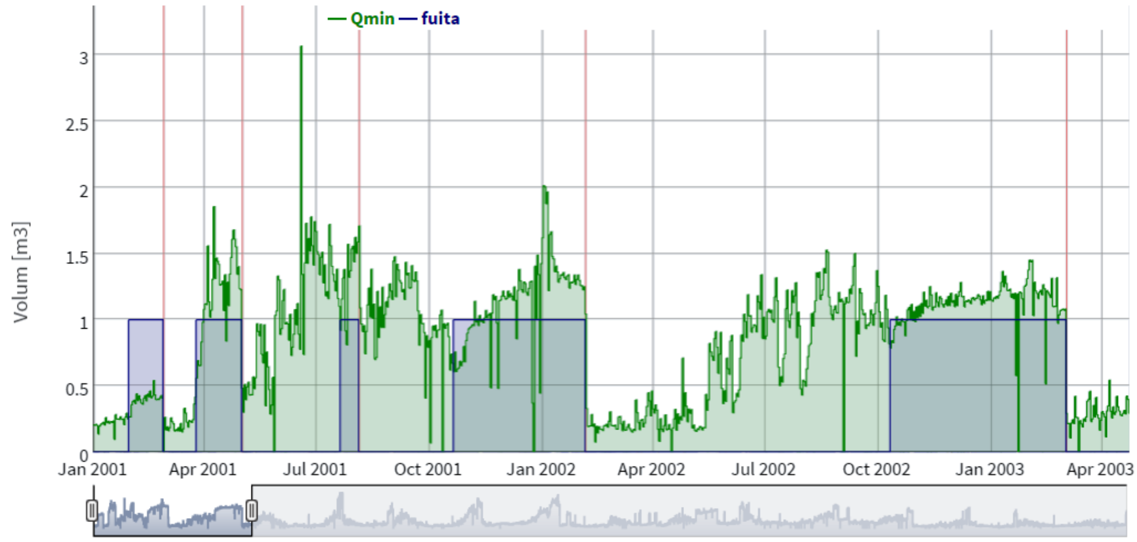

Figure 2: Night flow (green) values by days, the vertical red lines are the event alarm stored in the historical alarms company files, and in blue, a binary value with the expected duration of each leakage as the expert thinks.

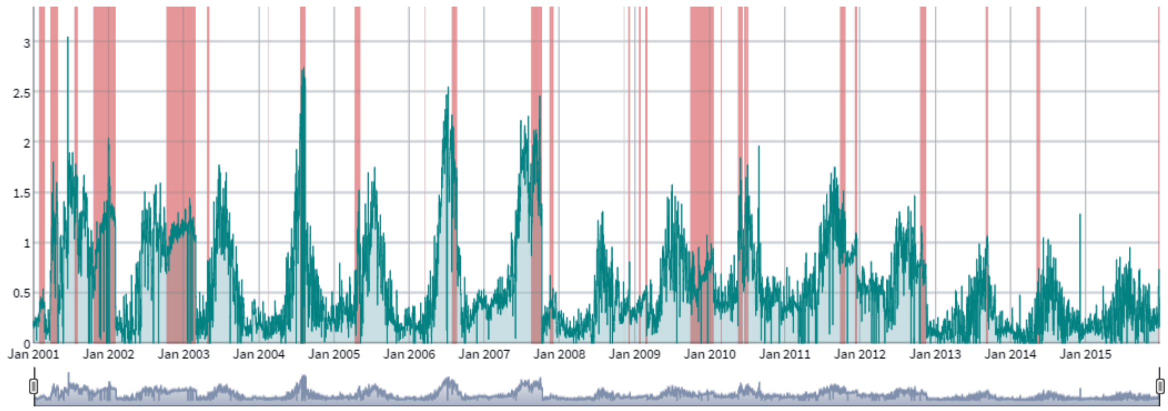

Figure 3: Night flow and selected leak episodes plotted in red for the whole scenario.

This case have a 17% of the days with allocated leakage value and the other 83% without them. The amount of water loss is calculated extracting the evolution associated with the season. Figure 4 exemplifies this procedure carried out by the expert off-line. The rate of water included in the addressed leaks is of 4.6% of the total consumption for 15 years that is 5.4% of the total consumption (around 106m3).

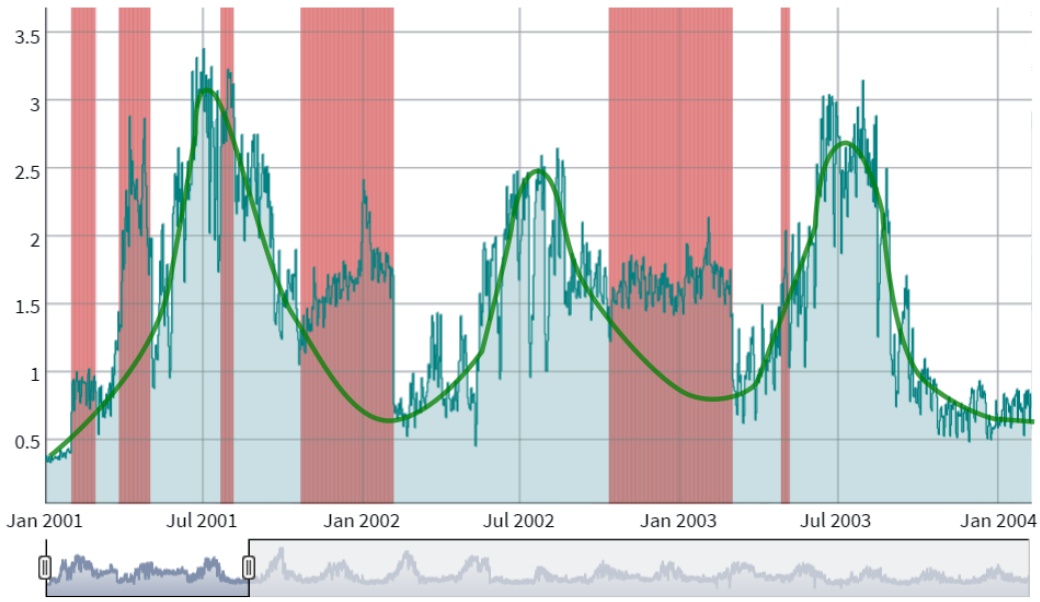

Figure 4: Estimation of the extra flow due to the leak. The green line represents the expected night flow due to the seasonal evolution.

This data table is divided in two: the first one, calls train represents the 75% of all data and the second one represents the 25% of the data. Both of them with the same proportional leakage representation. This is the first option that presents the tool (Figure 5). Afterwards the technician select the parameters that want to include in the model in this case they are the night flow, its standard deviation divided by itself, the rate between night and day flow and some parameters that characterise the distribution of the flows during the they. This indicators and the leakage state allows to create a predictive model using *glm*() R function. After creating this logistic regression model, it can be checked, to know the quality results, using the data table test and choosing the threshold. The mean of the probability of leakage for leaky and non-leaky days is displayed too.

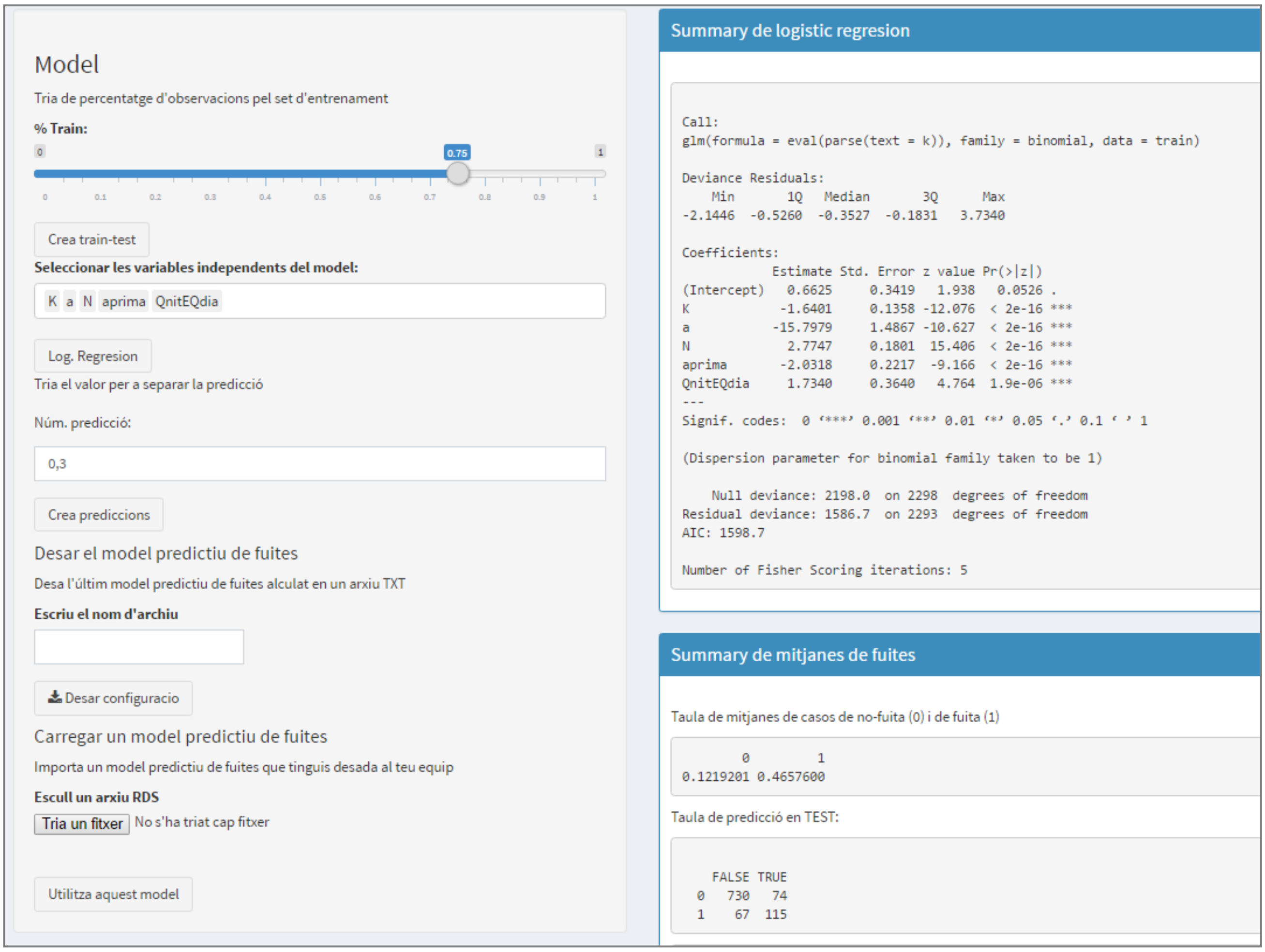

Figure 5: Application screen for the leak model generation.

Since the leakage days in the test data table is known, the application shows a table with the false-true positive-negative indicators and two graphics using the ggvis package for R, Figure 6 shows the evolution of the true positive versus false positive relation in order to the predicted cut value and Figure 7 shows the prediction value for each day and in highlight points shows the leakage days known.

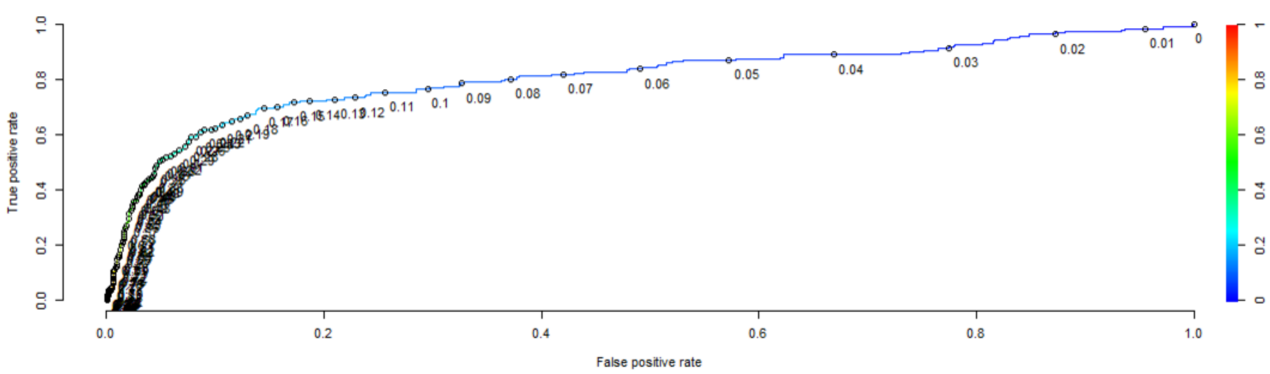

Figure 6: ROCR curve that aid in the threshold selection

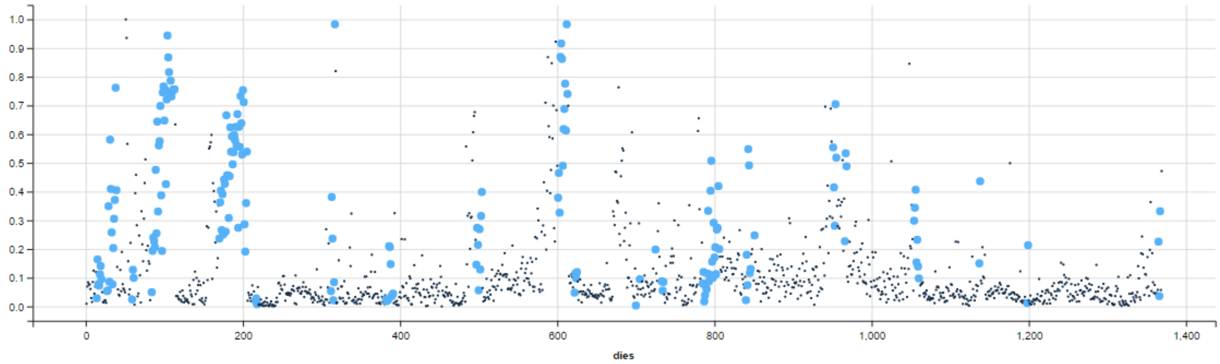

Figure 7: Probability of leak for each day of the test set (chronological order) in blue the reported leaky days.

The application allows also select the totally of the days like train dataset to create the predictive model and save it to use it in others sessions to others datasets, for example the daily data like online tool. In next example, for the X111 flow meter it has been created the predictive model with the days between 200 and 2010 to check the prediction success with data compressed between 2010 and 2015. Figure 8 presents the prediction results using a threshold of 0.8.

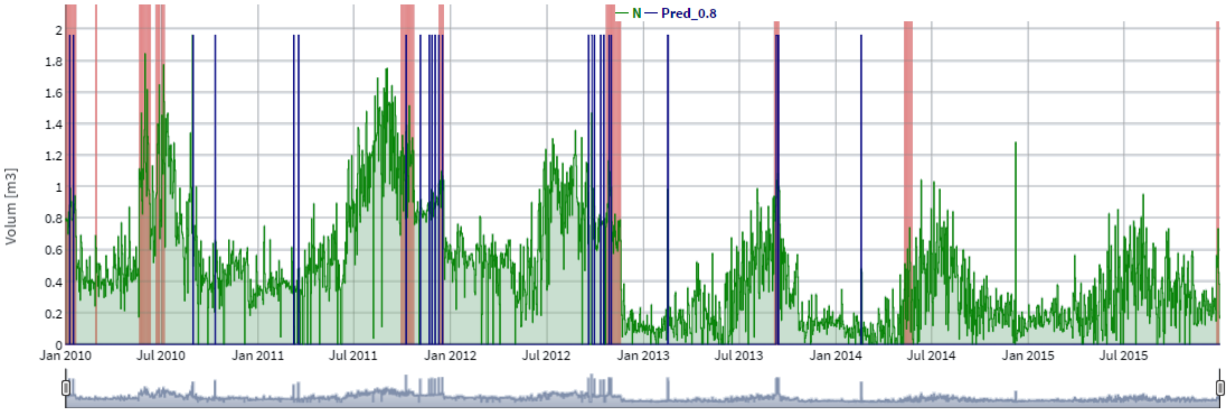

Figure 8: Alarms (blue lines) produced by the model with a threshold of 0.8, real events (red)and night flows (green)

4 CONCLUSIONS

This application aims to be a good decision support system for the preventive maintenance of the network sensors and to analyse the DMA data to calibrate the sensors as well as detect a leakage event. The predictive model can be customised by the selection of the independent variables. New independent variables can be defined easily. Furthermore, logistic regression is one of the models available in R, with little programming effort new models can be added with no change in the structure of the tool. The results obtained are promising despite of having poor registration of the leak events. The use of the tool should motivate the user to improve the registration protocols.

5 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] https://cs2ac.upc.edu/en

[2] 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.

[3] Farley, M., & Trow, S. (2003). Losses in Water Distribution Networks. London: IWA publishing.

[4] 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.

[5] Tabesh, M., Yekta, A. H. A., & Burrows, R. (2009). An Integrated Model to Evaluate Losses in Water Distribution Systems. Water Resources Management, 23(3), 477–492. http://doi.org/10.1007/s11269-008-9284-2.

[6] 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.

[7] Pérez, R., de las Heras, S., Aguilar, J., Pascual, J., Peralta, A., & Landeros, E. (2009). District management areas characterisation in water network based on clustering. Water Science & Technology: Water Supply, 9(5), 591.

[8] 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.

[9] van Thiennen, P. & Montiel, T. (2014). Flow Analysis and Leak Detection With the Cfpd Method in the Paris Drinking Water Distribution System. International Conference on Hydroinformatics. New York August 2014.

[10] 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