

Improving leakage management in urban water distribution networks through data analytics and hydraulic simulation

A. Candelieri¹, F. Archetti^{1,2} & E. Messina¹

¹*Department of Computer Science, Systems and Communications, University of Milano Bicocca, Italy*

²*Consorzio Milano Ricerche, Italy*

Abstract

Worldwide, water utilities are finding it increasingly difficult to meet the growing water demand. The problem, already acute in view of the urbanization trends, is compounded by the age of the infrastructure: one third of water utilities have 20% or more of their pipelines nearing the end of their useful life. This paper outlines an innovative approach for improving leakage management processes through the adoption of data analytics techniques and hydraulic simulation. More in detail, the aim is to provide analytical leaks localization and correspondent severity estimation in order to reduce time and costs for interventions and rehabilitation while improving asset management. The technical solution has been developed as a set of web services able to interoperate with other technological systems usually adopted by urban water distribution utilities, such as Supervisory Control And Data Acquisition (SCADA) system, Customer Information System (CIS), Geographical Information System (GIS) and Hydraulic Simulation tools. This paper presents some results obtained by the application of the approach to a real case study and also proves how its effectiveness may be improved in a sectorized water distribution network.

Keywords: leakage management, data analytics, water distribution network.

1 Introduction

Both social and economic development is strictly linked to the availability of fresh water sources and quality of the distribution service (Adams [1], Hoekstra



[2], Jung *et al.* [3]). In Europe, freshwater abstraction is allocated as follows: 45% for cooling in energy production, 22% for agriculture (50%–80% in the Southern Europe), 21% for public water supply and 12% for industry.

Urban water distribution networks are required to satisfy customer demand with a high quality of service. However, modifications in demand distribution are often unpredictable, both in space and time (in short and mid-long term). In order to satisfy these modifications the infrastructure of the water distribution network is generally modified (e.g., expansion/contraction) without following well defined and planned strategies, making system really complex to be monitored and managed.

According to the available data the growth rate of the water consumption with respect to the global population was doubled in the last century, domestic and industrial water consumption quadrupled from 1950 to 1995 and a further growth of 50% is estimated by 2025 as populations continue to increase (Rosegrant *et al.* [4]). Water utilities are finding increasingly difficult to meet the growing water demand and the problem, already acute in view of the urbanization trends, is compounded by the age of the infrastructure: one third of water utilities have 20% or more of their pipelines nearing the end of their useful life. Respect to this, breakages and leakages are highly probable within a water distribution network; according to data provided by the Italian National Institute of Statistics (ISTAT) during the World Water Day, on 22nd March 2011, in average, about 40% of water entered into the national water distribution networks is lost, with peaks greater than 80% in the South and the islands (data refer to water consumption in 2008).

Breakages and leakages produce both water resource and economical/financial losses mainly due to: extra energy needed to satisfy demand despite leaks, damages to infrastructure and third party, impacts on health due to affected water quality and diffusion of infections (Puust *et al.* [5]).

A detailed discussion on leakage management is given in (Preis *et al.* [6]). Summarizing, the traditional leakage management process should consist of three sequential steps: assessment, detection and (physical) localization. After the leakage management process, the administrator of water utility may define a proper rehabilitation strategy; the overall process is depicted in the following Figure 1.

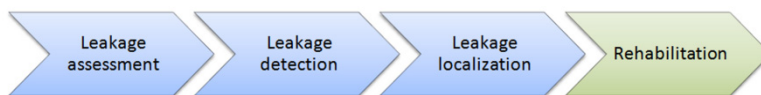


Figure 1: The three steps of the leakage management process before rehabilitation.

The first step, namely leakage assessment, is devoted to evaluate the entity of the losses by completely ignoring their possible location. Typically this step is performed by computing the overall water balance in the network (AHL [7]).

The second step, namely leakage detection, is devoted to identify the possible area where a loss occurs and may be performed through analytical or equipment-based methods; also a combination of the two strategies is possible. A discussion on well known equipment-based approaches is given in (Puust *et al.* [5]), in particular according to Leak Reflection Method (LRM), Inverse Transient Analysis (ITA), Impulse Response Analysis (IRA), Transient Dumping Method (TDM) and Frequency Response Method (FRM).

Minimum Night Flow (MNF) analysis 24 Hours Zone Measurements (HZM) are two of the most widely adopted analytical methods for leakage detection. The main difference between the two methods is that HZM requires a temporary isolation of some network areas. MNF consists of analyzing possible trends, over time, in flow and pressure values acquired during the early morning, usually 02:00 to 04:00 (Liemberger and Farley [8]), when customer's consumption can be considered near to zero and, therefore, acquired measures are mainly related to leaks.

Furthermore, effectiveness of MNF analysis in detecting leakages (Covas *et al.* [9], Garcia *et al.* [10]) resulted substantially higher for sectorized water distribution networks (Behzadian *et al.* [11], Izquierdo *et al.* [12]). Sectorization consists of defining sections of the network, usually known as District Metering Areas (DMA), easier to be monitored and controlled. Sectors are usually defined by a set of valves through which is possible to isolate, temporary or permanently, each single sector from the rest of the network.

Last step of the leakage management process, namely the (physical) localization, may be performed through different approaches and devices, such as vibration sensors and hydrophones (Muggleton and Brennan [13]), ground penetrating radar (GPR) (Farley [14]), noise correlators (Muggleton *et al.* [15]) and gas injection (Farley and Trow [16]) and aims at identifying the pipelines affected by the leak.

Leakage localization has become more and more relevant in the last 30 years and, according to this, efficiency in water distribution network management has been recently better defined through the IWA performance indicators for water supply systems (Alegre *et al.* [17]).

Many cost optimization models have been already proposed to support replace/repair decisions in asset management according to the ensuing costs, due to repair, replacement and expansion. Only recently the issues of computational leak detection and localization have been brought to the forefront leading to the proposal of several approaches, from traditional statistical analysis to machine learning, such as Artificial Neural Networks, Bayesian inference and clustering. (Puust [18], Buchberger *et al.* [19], Mashford *et al.* [20], Shinozuka and Liang [21], Polukais *et al.* [22], Candelieri and Messina [23], Candelieri *et al.* [24, 25]). The basis of all methods is that leakages can be detected/localized correlating changes in flow and pressure values within the real water distribution network to the output of a simulation model whose parameters are related to location and severity of the leak.

This paper presents the results obtained by the application of the approach proposed by the authors on a real case study and aimed at improving

effectiveness of the leakage management process by introducing a further step of analytical leakages localization and characterization (severity estimation), as depicted in Figure 2.

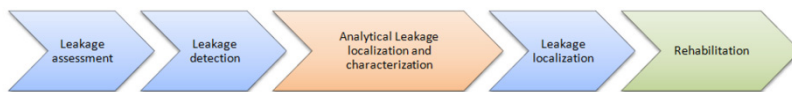


Figure 2: Improving leakage management process through the proposed approach.

The solution consists of a combination between hydraulic simulation of leakage “scenarios” and the association of the actual pressure and flow values with the most similar simulated scenario, as well as the estimation of the correspondent severity in terms of discharge coefficient. This permits to identify a limited set of pipelines, rather than an area in the network, probably affected by a leak – according to the simulation of leakage scenarios – to physically check, reducing time and costs for intervention and rehabilitation.

The rest of the paper is organized as follows: the approach is presented and results are provided. Then the application of the approach on a sectorized water distribution network, in combination with the traditional MNF analysis, is presented: the main result is a further improvement in localizing leaks. Finally, a discussion and planned future works are presented.

2 Localizing leaks and characterizing them through data analytics

The proposed analytical leaks localization and characterization approach is based on a quite simple idea: hydraulic simulation software can be used to simulate the behaviour of a water distribution network according to several kinds of leaks added into the network hydraulic model, namely leakage scenarios. Simulation software is then able to provide all the hydraulic measurements related to each specific leakage scenario and, in particular, pressure and flow values in correspondence of the monitoring points in the real water distribution network. In the following subsections the approach is detailed.

2.1 Generating and simulating leakage scenarios

The first step consists of retrieving all the available data related to: *i*) network infrastructure (from Geographical Information System, GIS) and *ii*) customers water consumption (from Customer Information System, CIS), respectively.

These data are needed to configure the hydraulic model of the water distribution network and, in particular, build different leakage scenarios.

To implement our approach we have adopted the open-source software EPANET 2.0 (<http://www.epa.gov/nrmrl/wswrd/dw/epanet.html>) freely provided

by the Environmental Protection Agency (EPA). After a hydraulic simulation is performed, the hydraulic measurements for each network component are available: pressure at the nodes, flows on the pipelines, power of pumping systems, and so on. In order to realize the analytical leaks localization and characterization, the most significant data from simulations of leakage scenarios are pressure and flow values in correspondence of the actual monitoring points.

In EPANET a leak (as well as nozzles, sprinkler heads or irrigations systems) may be added within a network model by using the "emitter" component. The flow rate through the emitter varies according to the pressure according to the function $q = C p^\gamma$, where q is the flow rate, p is the pressure, C is the discharge coefficient and γ is the pressure exponent. The pressure coefficient depends on geometry of the orifice (Greyvstein and Van Zyl [26]) and material of the pipelines (Van Zyl and Clayton [27]) and its typical value is 0.5 for circular orifices.

A leak is applied, in turn, on each pipe with different values of discharge coefficient. After the simulation of all the correspondent leakage scenarios is performed a dataset can be built in order to perform two different types of analysis: grouping leakage scenarios and estimating discharge coefficient (severity of the leak) both according to the pressure and flow values at the monitoring points. An example of the resulting dataset is reported in the following Figure 3. In detail, L is the identifier of the pipe on which the leak is placed, C is the discharge coefficient, P_i and F_i are pressure and flow at the i -th monitoring point. Leakage Localization and Leakage Characterization are the two different tasks are performed on this dataset.

L	C	P_0	F_0	P_N	F_N
14	0.1	0.4	1.3	0.3	0.5
...
...

Leakage Localization
Clustering by ignoring L and C features

Leakage Characterization
Regression on C based on the values of the other features (L is ignored)

Figure 3: An example of dataset obtained through leakage scenarios simulation.

2.2 Clustering leakage scenarios for leaks localization

This step consists of grouping together leakage scenarios that are similar in terms of pressure and flow distributions at the monitoring points. According to the state of the art different strategies are possible. Authors already presented some results about the comparison between different techniques and proposed a modified version of the Induced Bisecting algorithm (Archetti *et al.* [28], Fersini *et al.* [29]), namely the In-deep Bisecting, with the aim to improve localization performances (Candelieri and Messina [23], Candelieri *et al.* [24, 25]).

In particular, after leakages scenarios are grouped together without considering information about the pipe on which leak is placed and the

correspondent discharge coefficient, the number of pipes belonging to each cluster respect to the overall number of mains of the network is taken into account as performance index. In particular the following four localizing level have been defined:

- Highly Localizing cluster: contains less than 25% of pipes;
- Average Localizing cluster: contains more than 25% and less than 50% of pipes;
- Poorly Localizing cluster: contains more than 50% and less than 75% of pipes;
- No Localizing cluster: contains more than 75% of pipes.

The proposed In-deep algorithm proved to be the best one, providing the highest number of highly localizing clusters and the lowest number of no localizing ones (Candelieri and Messina [23], Candelieri *et al.* [24, 25]).

Respect to our previous studies we have further validated our approach by taking a part, in turn, a limited set of scenarios as test while all the others have been used for clustering. Then each test case has been associated to the correspondent most similar scenario; more than 75% of the cases have been associated to a highly localizing cluster, always identifying, among others, the right pipelines.

2.3 Estimating severity of the leakage

This step consists of adopting supervised machine learning techniques to estimate leakage severity (i.e., discharge coefficient) according to pressure and flow values at the monitoring points. A simple Least Median Squared Linear Regression proved to be effective in order to perform this task. Reliability of this

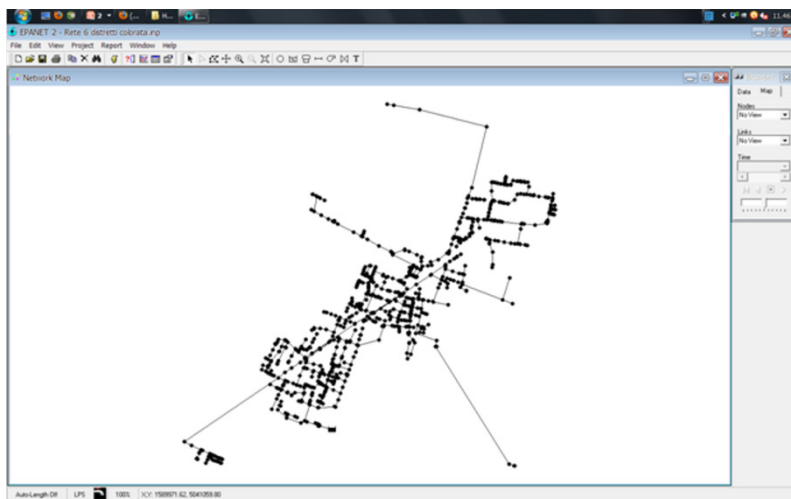


Figure 4: A case study water supply network in a little town in the North of Italy.

model has been evaluated through 10-fold cross validation on the real case study water distribution network in Figure 4 and provided a significant correlation coefficient (0.09997) and a really low error (Relative Mean Absolute Error = 0.8764% and Root Relative Mean Squared Errors = 2.5368%) (Candelieri *et al.* [24, 25]).

3 Defining an effective and efficient intervention plan

As final step, results provided by the two previous analytical task can be suitable combined in order to support the definition of an effective and efficient intervention plan. More in detail, after a leakage is assessed and possibly detected through traditional methods (e.g., MNF analysis), all the steps described in previous section 2 are performed. Finally, the real pressure and flow value are used to *i*) identify the cluster related to the scenarios most similar to the actual condition and *ii*) estimate discharge coefficient of leaks. These two information are then combined to define an ordered list of pipelines to check according as depicted in Figure 5.

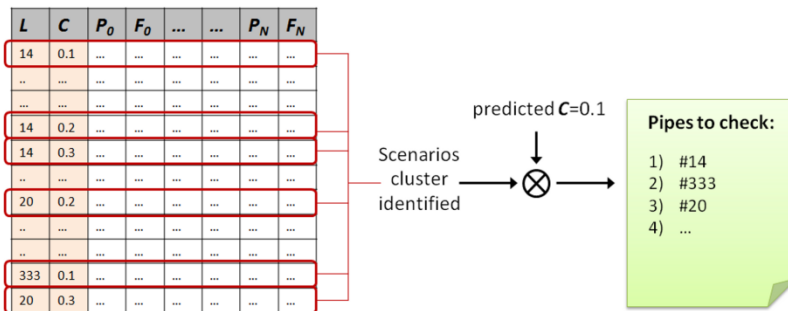


Figure 5: Building a plan of intervention by combining suggestions from leakage localization and estimation of leakage severity.

It is possible to note that the pipe most probably affected by a leak is the L=14, because it appears in three different scenarios and also with discharge coefficient equal to the predicted one. Then, pipe L=20 appears twice but with discharge coefficients different from the predicted one, while the pipe L=333 appears only once but with exactly a discharge coefficient equal to the predicted one. With respect to this, a higher priority is given to L=333 than L=20 in order to be physically checked, generating the sorted list depicted in Figure 5.

4 Deployment into a sectorized water distribution network

As final result, this section presents how reliability of the leakage localization and characterization may be relevantly more effective in the case of a sectorized water distribution network. In this case, monitoring devices are usually placed at the entry points of each sector in order to monitor and control the sub-network

and perform traditional analysis such as MNF. The same set of sensors is used in order to perform the analytical tasks proposed by the authors. As significant example, a five sectors water distribution network is depicted in Figure 6, with the correspondent monitoring sensors.

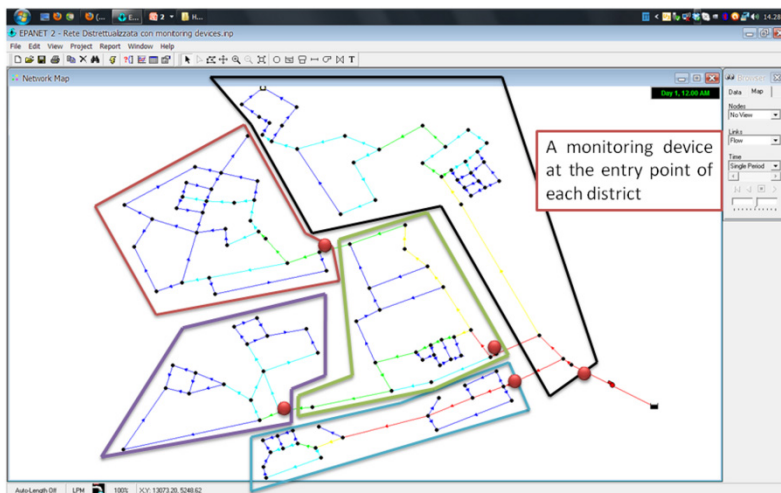


Figure 6: A water distribution network with 5 sectors.

Leakage scenarios have been generated and then grouped via In-deep clustering as described in section 2. As already mentioned, this algorithm provides a great number of highly localizing clusters, which are groups of scenarios associated to a limited set of the overall pipelines. In many cases, this set of pipelines is widely spread on the entire network: in the following Figure 7 is reported a highly localizing cluster associated to a specific pressure and flow distribution generated by a leak.

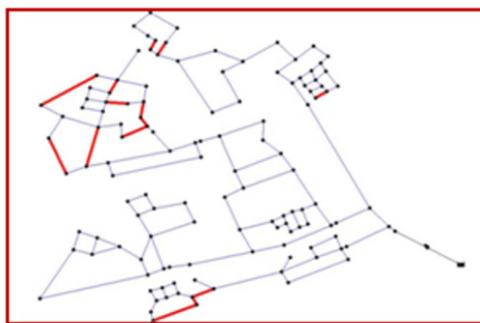


Figure 7: An example of a highly localizing cluster. Pipes belonging to this cluster, and probably affected by a leak (bold line) are spread on different sectors.

Since traditional (analytical) leakage detection methods, such as MNF, permit to identify network areas where the leakage probably occurs, this information may be combined with that related to the set of probably leaky pipelines identified through clustering. For example, according to the Figures 6 and 7, and supposing that MNF analysis has indicated the north-eastern sector as the area with a loss, only two pipes should be checked of the overall 14 belonging to the cluster.

5 Discussion and future works

This paper proposes a set of data analytics based approaches aimed at improving the leakage management process in urban water distribution networks. More in detail, a strategy to move from traditional leakage detection towards analytical localization and characterization of possible leaks is presented.

All the tasks presented have been developed as a set of web services in order to facilitate the interoperability with other systems traditionally used by water utilities, such as SCADA, GIS, CIS and hydraulic simulation software.

The proposed integration among hydraulic simulation, clustering, regression learning and traditional approaches, such as sectorization and MNF, proved to be successful in order to improve effectiveness and efficiency of the intervention and rehabilitation activities, by reducing both time and costs.

Nevertheless some further issues should be addressed. In particular, impact of possible noise on sensors and errors on consumption data will be investigated. Moreover, consumption data adopted in this study are daily averaged form accounted water consumption acquired through the billing process (twice a year) – respected to this daily averaged pressure and flow values have been computed for data retrieved from SCADA system. As future works authors plan to apply and test new versions of the approaches on case studies equipped with automatic metering readers (e.g. with half an hour sampling rate) in order to improve reliability of the leakage localization and characterization.

The usually unknown initial overall “health condition” of the water distribution networks is another relevant criticality to be investigated. Aged systems, in fact, may present a high probability to have multiple leaks; according to this case a computationally efficient techniques for building leakage scenarios having multiple losses will be developed to make more reliable the leaks localization and characterization process.

References

- [1] Adams, W.M., The Future of Sustainability: Re-thinking Environment and Development in the Twenty-first Century, *Report of the IUCN Renowned Thinkers Meeting*, 29-31 January 2006.
- [2] Hoekstra, A.Y., The global dimension of water governance: Nine reasons for global arrangements in order to cope with local water problems, *VALUE OF WATER RESEARCH REPORT SERIES NO. 20*, UNESCO-IHE Institute for Water Education, Delft, the Netherlands, 2006.



- [3] Jung, N.C., Popescu, I., Price, R.K., Solomatine D., Kelderman P., Shin J.K., The use of the A.G.P. test for determining the phytoplankton production and distribution in the thermally stratified reservoirs: The case of the Yongdam reservoir in Korea, *Environmental Engineering and Management Journal*, 10, 1647-1657, 2011.
- [4] Rosegrant, M.W., Cai, X., Cline, S.A., Global Water Outlook to 2025: Averting an Impending Crisis, published by the *International Food Policy Research Institute*, 2002.
- [5] Puust, R., Kapelan, Z., Savic, D. A. and Koppel, T., A review of methods for leakage management in pipe networks, *Urban Water Journal*, 7 (1), 25-45, 2010.
- [6] Preis, A., Allen, M., Whittle, A.J., On-Line hydraulic modeling of a Water Distribution System in Singapore. *Water Distribution System Modeling Issues* 1336-1348, 2010.
- [7] AHL, Water Management, In: *8th Report of Session 2005, London: Science and Technology Committee, Authority of the House of Lords*, 2006.
- [8] Liemberger, R., Farley, M., Developing a nonrevenue water reduction strategy Part 1: Investigating and assessing water losses. In: *Proceeding of IWA WWC 2004 Conference*. Marrakech, Morocco, 2004.
- [9] Covas, D., Ramos, H., Lopes, N., and Almeida, A.B., Water pipe system diagnosis by transient pressure signals. In: *Proceedings of the 8th Annual Water Distribution Systems Analysis Symposium*. Cincinnati, USA, 2006.
- [10] Garcia, V.J., Cabrera, E. and Cabrera, E. Jr, The minimum night flow method revisited. In: *Proceedings of the 8th Annual Water Distribution Systems Analysis Symposium*. Cincinnati, USA, 2006.
- [11] Behzadian, K., Kapelan, Z., Savic, D.A., Ardeshir, A., Stochastic sampling design using multiobjective genetic algorithm and adaptive neural networks. *Environmental Modelling and Software*, 24 (4), 530–541, 2009.
- [12] Izquierdo, J., Herrera, M., Montalvo, I., Pérez-García, R., Division of Water Supply Systems into District Metered Areas Using a Multi-agent Based Approach. In *Software and Data Technologies - Communications in Computer and Information Science*. Eds. Cordeiro, J., Ranchordas, A., Shishkov, B. Springer Berlin Heidelberg. 50: 167-180, 2011.
- [13] Muggleton J.M., Brennan M.J., (2005), Ax symmetric wave propagation in buried, fluid-filled pipes: effects of wall discontinuities, *Journal of Sound and Vibration*, 281, 849–867.
- [14] Farley, M., Finding ‘difficult’ leaks. *International Water Association Specialist Group – Efficient Operation and Management*, 2008.
- [15] Muggleton, J.M., Brennan, M.J., Linford, P.W., Asymmetric wave propagation in fluid-filled pipes: wave number measurements in vacuo and buried pipes, *Journal of Sound and Vibration*, 270, 171–190, 2004.
- [16] Farley, M., Trow, S., Losses in water distribution networks. London: *IWA Publishing*, 2003.
- [17] Alegre, H., Baptista, J.M., Cabrera, E., Jr., Cubillo, F., Duarte, P., Hirner, W., Merkel, W., Parena, R., *Performance Indicators for Water Supply Services*, Second Edition, IWA Publishing, 2006.

- [18] Puust, R., Probabilistic leak detection in pipe networks using the SCEM-UA. *Thesis (PhD)*. Tallinn University of Technology, 2007.
- [19] Buchberger, S.G., Nadimpalli, G., Leak estimation in water distribution systems by statistical analysis of flow readings. *Journal of Water Resources Planning and Management*, 130(4), 321–329, 2004.
- [20] Mashford, J., De Silva, D., Marney, D., Burn, S., An approach to leak detection in pipe networks using analysis of monitored pressure values by support vector machine. *Network and System Security, Third International Conference on*. 534-539, 2009.
- [21] Shinozuka, M., Liang, J., On-line damage identification of water delivery systems. *Proceedings of Engng Mech Conf.*, 1999.
- [22] Polukais, Z., Valougeorgis, D., Papadimitriou, C., Leakage detection in water pipe networks using a Bayesian probabilistic framework, *Probabilistic Engineering Mechanics*, 18, 315–327, 2003.
- [23] Candelieri, A., Messina, E., Sectorization and analytical leaks localization in the H2OLEAK project: Clustering-based services for supporting water distribution networks management. *Environmental and Engineering Management Journal*, Vol.11, No. 5, 953-96, 2012.
- [24] Candelieri, A., Messina, E., Malacrida, A., Simulation and Unsupervised Learning Strategies for Enabling Integrated Water Resource Management: H2OLEAK Project, in *Proceedings of the International Environmental Modelling and Software Society (iEMSs) 2012 International Congress on Environmental Modelling and Software Managing Resources of a Limited Planet*, 2012.
- [25] Candelieri, A., Archetti, F., Messina E., Analytics for Supporting Urban Water Management, in *Proceedings of ECOIMPULSE 2012* [in press].
- [26] Greyvnstein, B., Van Zyl, J.E., An experimental investigation into the pressure leakage relationship of some failed water pipes, In: *Proceedings of Leakage 2005*, Halifax, California, 2005.
- [27] Van Zyl, J.E., Clayton, C.R.I., The effect of pressure on leakage in water distribution systems, *Water management*, 160, 109-114, 2007.
- [28] Archetti, F., Campanelli, P., Fersini, E., Messina, E., A Hierarchical Document Clustering Environment Based on the Induced Bisecting k-Means, In: H. Legind Larsen *et al.* (Eds.) *FQAS 2006, LNAI 4027*, Springer-Verlag Berlin Heidelberg, 257-269, 2001.
- [29] Fersini, E., Messina, E., Archetti, F., Multimedia Summarization in Law Courts: A Clustering-Based Environment for Browsing and Consulting Judicial Folders, In: P. Perner (Ed.): *ICDM 2010, LNAI 6171*, Springer-Verlag Berlin Heidelberg, 237-247, 2010.

