A spatial and temporal analysis for long term renewal of water pipes

Youssef Tlili, Amir Nafi, Umberto Fratino

A spatial and temporal analysis for long term renewal of water pipes

Youssef TLILI*,** , Amir NAFI** and Umberto FRATINO***

*Technical University of Bari, Italy- Joint Research Unit in Water and Environmental Territory Management, Cemagref-Engees, Strasbourg, France
**Joint Research Unit in Water and Environmental Territory Management, Cemagref-Engees, Strasbourg,
***Technical University of Bari, Italy
(E-mail: youssef.tlili@engees.unistra.fr; amir.nafi@engees.unistra.fr)

Abstract

Drinking Water distribution systems as well as other technical networks represent an important accounting. Managing these networks aims to sustain people's access to drinking water quality and quantity according to a satisfactory level of service. Water networks are an assembly of technical devices, most of them are buried. This makes their diagnosis more difficult and requires the implementation of a policy, known by «asset management". This requires planned actions in the short, medium and long term. It aims to describe the state of the asset by developing specific tools in a first time and plan the actions needed to maintain its state in a second phase.

This current research deals with the problematic of establishing priorities for pipes that should be selected for rehabilitation. The developed approach is based on the discrimination of vulnerable pipes by a spatial and temporal analysis starting from a simple idea: the diagnosis is often unaware dimensions: topology and evolution of failures pattern. Thus, detailed approach provides a holistic vision of the network by identifying areas with high failures, called "hotspots" and their evolution over time. Therefore, this research allows a visual representation using the network topology; such representation provides a resolution of the network and can be used to evaluate, a priori, the effectiveness of the rehabilitation policies according to their incidence on hotspots' reducing.

Keywords

Breakage, cluster density, indicator, prioritization, pipe replacement.

INTRODUCTION

Asset management (AM) of drinking water systems is an approach that tracks the status of the buried heritage and anticipates works to be done for his continued service during its service lifetime. It is a managerial approach which aims to sustain the user's water service in a required quality and quantity, in the best technical and financial conditions. First of all, AM tends to improve the network knowledge by implementing inventories of equipments and checkups of components. It requires the establishment of the criteria and indicators for assessing the level of asset condition. By considering that it's almost impossible to rehabilitate the network at once, it is required to establish priorities in order to discriminate the more deteriorated parts and thus to carry out the appropriate actions to ensure the service continuity. Deterioration implies a decrease of the condition level that may be physical or functional. It depends on many parameters that can be environmental, technical, managerial and economic. The asset status is established in present time i.e. today, so current network state depends on the policies implemented earlier by the network manager.

The recurring questions that the manager arises concerning the renewal of the network are the following: What pipes do I renew? What are the criteria for renewal? When do I have to renew? How much is it going to cost? In general we have two ways to be mistaken: to replace a pipe that should not be or to unselect one that has to be. In practice, the renewal of water pipes is conditioned by the available financial resources. In most cases, the budget for renewal is fixed in advance and works to be performed must meet the financial constraint.

In response to these challenges, several decision support systems for prioritizing water pipes have been developed (Utilnets, Caree-W (Le Gauffre and Haidar, 2008)), Kanew, Riva, Sirocco, Prams (Nafi and Werey, 2009)). Most researches (Dridi et al, 2006) , (Halfawy et al, 2008) and (Giustolisi and Berardi, 2009) proposed a non-exhaustive list of various criteria and encouraged water utilities to better know their asset by collecting new data and by ensuring their update. Even if the proposed criteria are relevant, the spatial dimension and "network scale" are not enough taken into account. Indeed, the prioritization criteria identified in literature are assessed at the pipe scale and allow discriminating pipes between them. Some criteria are difficult to quantify or require too much input data, often unavailable. In reality, the most used criteria for pipes rehabilitation are the age of pipes; roadwork's or planned works on adjacent networks as gas, electricity and sewers.

However the pipe condition depends on both physical and functional deterioration. The importance of a pipe depends not only on its location in the network, but also on its connection with other pipes and other endogenous and exogenous factors. The analysis scale should therefore be both the pipe and the whole network. In addition, analysis across the network refers to a spatial analysis which was not considered in general. Often, the pipe is a recording in a database, identified by a key; we do not even know its location in the network during its selection. This induces a loss of important informations and can penalize the prioritization of pipes to be chosen for renewal.

The spatial analysis of the network shows clusters of breaks, which are more crowded in some areas of the water network. Empirically, it seems that the first break is an independent event but the following breaks events are dependent on the latter's and locate nearby (Goulter and Kazemi, 1988) .

Therefore, we explore the construction of pipes prioritization criteria based on the gathering of pipe breaks in "clusters" in order to discriminate areas of high density of breaks called "hotspots" surrounding areas characterized by low density of breaks or "noise breaks" that does not reflect a potential concentration of failures. These criteria will be used to identify hotspot zones and rank the pipes candidate for renewal within the same cluster. The following paper is arranged on 4 main sections; first the analysis of the context by literature review of existing criteria for pipes prioritization. The second section introduces the spatiotemporal approach by explaining the principal, objectives and used algorithm, the third section shows the implementation of the approach on a real case study and the significant results. A first evaluation of the approach, results and also a further work is presented in the last section.

PRIORITIZITION CRITERIA: CONSIDERATION OF THE SPATIAL AND TEMPORAL DIMENSIONS

Since the network status is uncertain and the opportunities to inspect it are rare (repair, work on other network), the establishment of a renewal program requires the implementation of a proactive approach for failure prediction and the definition of relevant criteria for the selection and prioritization of pipes to be renewed. Prioritization criteria can be classified into families as follows:

- Hydraulic criteria: Surplus or pressure deficit, demand satisfaction, hydraulic capacity, reliability, resilience.
- Economic criteria: replacement cost, opportunities and economy of scale, direct and indirect costs, social costs, cost of water loss, repair cost,
- Social criteria: customer dissatisfaction, number of complaints,
- Structural criteria: number of breaks or leakage, breakage rate, linear index of loss.

Beyond these criteria, we propose to build additional indices to characterize hotspots and their evolution over time. This implies a spatial analysis of failures through clustering assumption according to historical failures data. The knowledge of failure trend, occurrences and their spatial distribution within the network remains essential before any prioritization task. It will be inadequate to plan renewal works without knowing a priori the trends, the tendency and the breaks location. A discernment of failures concentrations grants to focus on the most vulnerable pipes and thus avoids an early renewal of pipes having just suffered from noise breaks. This helps to better target the renewal works and perform significant savings.

Distance analysis

In this section, we analyse how clustered or dispersed the breaks are. Based on the nearest neighbour approach, the analysis put us on track for identifying and understanding hotspots. We used distance analysis techniques to answer questions about the dispersion of breaks and hotspot analysis to identify areas where breaks concentrate.

Nearest Neighbour Analysis (NNA)

The *Nearest Neighbour Analysis* (*NNA*) measures the distance of each point to its nearest neighbour, determines the mean distance between neighbours, and compares the mean distance to what would have been expected in a random nearest neighbour distribution. We can also control whether to compare each point to its single nearest neighbour or to run the routine against the second-nearest, third-nearest, and so on.

NNA allows calculating the Nearest Neighbour Index (*NNI*). In the *NNI*, a score of 1 would indicate absolutely no discrepancy between the expected distances in a random distribution and the measured distances in the actual distribution. Scores lower than 1 indicate that incidents are more clustered than would be expected in a random distribution, and scores higher than 1 indicate the incidents are more dispersed than would be expected in a random distribution. The *NNI* calculation process is given by:

$$\overline{NND}_R = \frac{1}{2\sqrt{Density}} \qquad Density = \frac{n}{A} \tag{1}$$

$$R = \frac{\overline{NND}}{\overline{NND}_R} \qquad \overline{NND} = \frac{\sum_{i=1}^{n} NND}{n} \tag{2}$$

Where:

n = Number of failure points in the distribution
A= Size of the study area
R = Ratio of the observed distance to the expected distance
NND = Distance between each point and its nearest neighbour
$\overline{NND}$ = The observed mean distance between nearest neighbours
$\overline{NND}_R$ = Expected value of the nearest neighbour distance in a random pattern.

Spatial clustering: Finding patterns

Clustering is the process of creating a collection of similar data within the same group and dissimilar when they belong to different groups. Clustering is the unsupervised classification, which means without previously predefined classes. These methods are widely used in risk and crime spread analysis. A good clustering method ensures a high similarity «within-group» and allows dissimilarity "inter-group." It is through this characteristics that will be based the analysis of the spatial distribution of pipe breaks. There are four major clustering approaches. *Partitional Clustering Algorithms* constructs usually predefined clusters. *Hierarchical Clustering Algorithms* creates a hierarchical decomposition which can be represented as dendograms. *Density-based partitioning Algorithms* search for regions which are denser according to a predefined threshold. *Grid-based Algorithms* are based on a multiple-level granularity by quantizing the search space into a finite number of cells (Abdulvahit and Düzgün, 2006) .
We use in this study the *Neighbour Hierarchical Spatial Clustering* (*NNH*) for failures clustering points in the network based on two parameters *i)the maximum neighbourhood radius that separates a basic break point for other points of breakage "Eps" and ii) the minimum neighbourhood number of breaks "MinPts"*.

MATERIAL & METHODS: CASE STUDY

In this section, we follow the spatial and temporal trend of pipes breaks to understand in which manner evolves the size of clusters during the time, which add time dimensionality to the study. The developed methodology includes 4 steps:

- Step1: partition the break history (1962-2003) in four equivalent intervals of time: [1962, 1973], [1974, 1984], [1985, 1995] and [1996, 2003].
- Step 2: plot on the network breaks corresponding to each interval of time.
- Step 3: for each horizon, a first way to detect the spatial distribution of breaks is to calculate the nearest neighbour index (NNI).
- Step 4: divide the network into clusters using *NNH* to identify areas of high failure from surrounding areas of low density of breakage (noise breaks), hence understand and discern the breaks pattern in time. To provide stable results regarding the number of clusters, quality of clustering output and to derive clusters that are meaningful to condition assessment and replacement optimization needs, many simulations where performed to fix the clustering parameters values. Ben-David and Simon (2002) and De Oliveira et al. (2010) demonstrated that the loss of clustering quality is precipitated by the increase of the "eps"- distance and with decrease of "MinPts" parameters. In another hand, while nearby pair of breaks might result from poor repair of previous breaks, spatial clusters with few breaks, e.g., clusters with pairs of breaks will not generate useful replacement candidates because they tend to have a very small extension [9], the "MinPts" parameter was chosen equal to values of three breaks and above in order to avoid the generation of numerous small spatial clusters and to reach

significant clusters in the same time. An upper bound for the "eps"- distance is defined by looking into the pipes length and considering that replacement will not be performed in arbitrarily large extensions of pipes but rather in a small length. For instance, taking in account the longest pipe (410 m) and the shortest one (37 m), we used an "eps" distance or search radius equal to the average value of pipe length equal to 223 m, this distance is not generated randomly and it is meaningful for decision makers. One of the interest of the analysis comparing to large clusters, was first to have small clusters since smaller ones at finer hierarchy tend to present more uniform distribution of breaks along pipes and also more homogenous in term of environmental factors (e.g., soil characteristics), pipe conditions (e.g., material , traffic..) and especially density of breaks along pipes. Additionally, this allows getting a large number of clusters which offer in principle, more choices of pipes candidate for renewal (De Oliveira et al, 2010) and (De Oliveira et al, 2011).

- Step 5: Define criteria for giving a first level of ranking and identifying which clusters exactly constitute hotspot zones. A possible criterion is the breakage rate of a cluster. Once we managed to set a classification of clusters, the pipes included will be automatically prioritized. The breakage rate of clusters will be compared to rates of groups of pipes based on non spatial criteria (i.e., pipe diameter and material).
- Step 6: Define prioritization ranking for each pipe or street into the clusters. Within each cluster, prioritization could not be only on the pipes' level but rather on the level of streets or roads level.

Clustering analysis

The methodology was applied to a part of a network consisting of 147 cast iron pipes with a diameter between 150 and 250 mm. For each interval time, we performed separately the cluster analysis (Nearest Neighbour hierarchical clustering algorithm with a fixed radius of 223 m). Additionally, one way to analyze and observe the temporal trend of breaks was by adding the clusters of the period n+1 to those of the previous period n in an incremental way of time. This analysis is illustrated in the figure 1. The figure below shows the obtained clusters for 4 periods of time considered separately. The Nearest Neighbour Index corresponding to periods is respectively 0.55, 0.53, 0.48 and 0.6 (<1) that denote the clustering behaviour of breaks. We note also that between these time periods, clusters have trend to occur in the same location as in the previous period, they can be superposed, with same deviation or also between two adjacent previous clusters. This may accentuate breaks and are more and more clustered in a precise location. We observe that there are three main annotations : i) appearance of new clusters, ii) evolution of clusters from one or more other clusters and iii) above all, overlapping of clusters of new clusters over old ones, particularly these clusters seems to be the most critical zones characterized by a higher density of breaks.

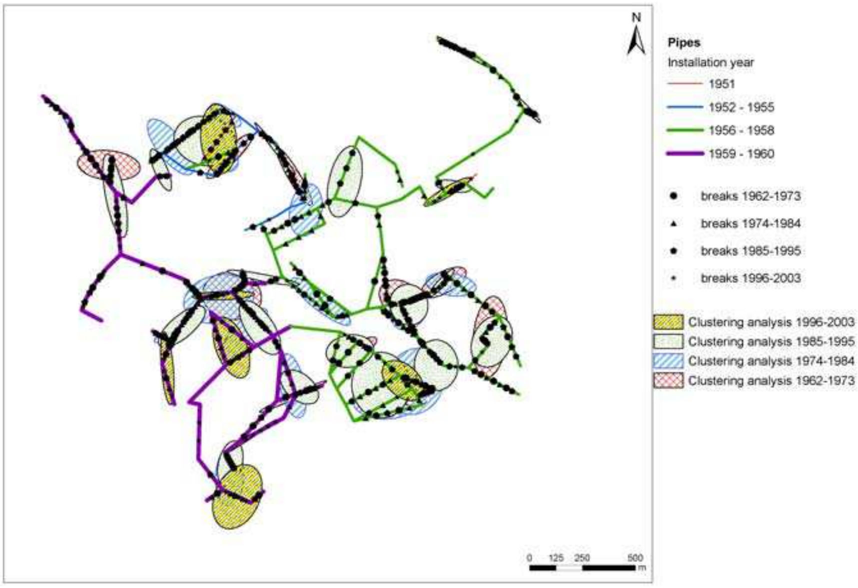

Figure 1. Clustering analysis since year 1962 to 2003.

Differently said, we observe that many scored clusters are close or have exactly the same location as those in the previous period, the thing that let us understand that more breaks were concentrated exactly in the same location of previous cycle and they are therefore encompassed.

To perform the reasoning concerning hotspots, the analysis was performed also using the kernel density estimation function, the figure 2 shows that there is practically the same hotspots and cold spots describing the breakage density in each raster. This make us understand that the zones A,B and C are the most critical zones in the study area, these zones are characterized by a high density of breakage and will be the sections where we will enhance our prioritization scheme .

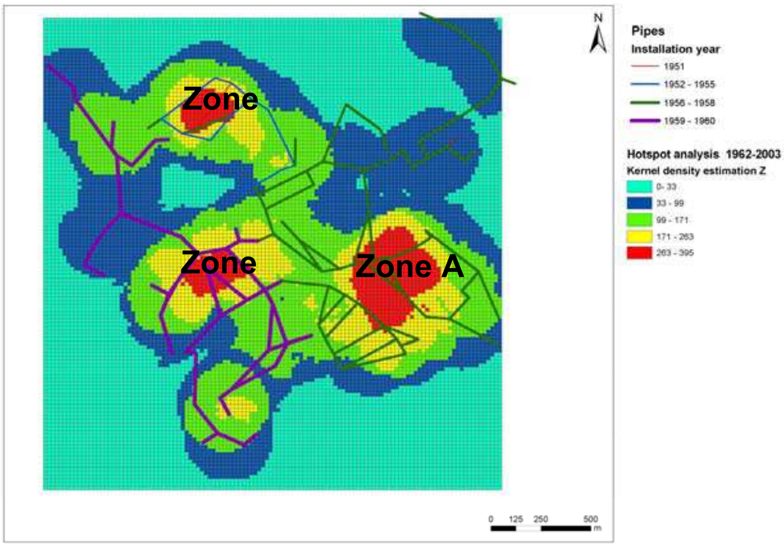

Figure 2. Hotspot analysis: Kernel density estimation

A second issue to be discussed here is that this kind of hierarchy, to move from smaller clusters to larger (dissolved) clusters was done in purpose, since a smaller clusters can provide an indicator of breakage that is higher that breakage density in vicinity at the level of small sections but not allow to work in a larger scale if we would like to concentrate our prioritization at the level of streets or roads. Thereby, merging small clusters into larger ones, will permit to obtain a second level of clusters through which it will more easy to catch streets to replace (figure 3).

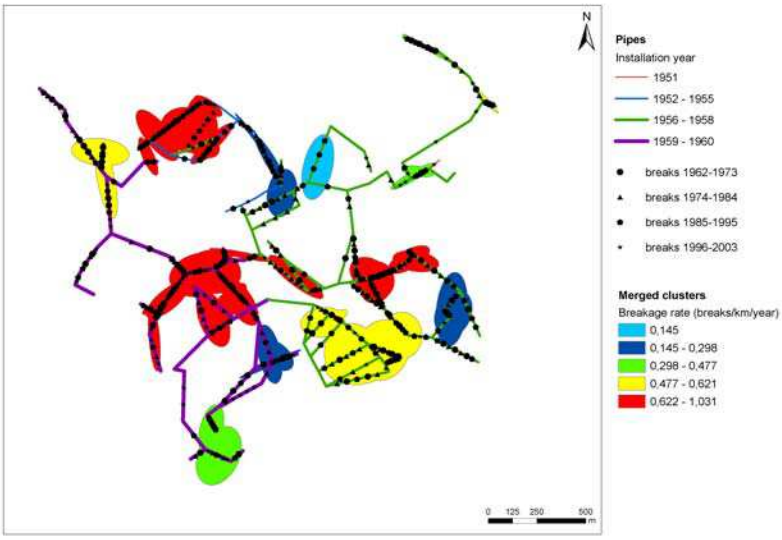

Figure 3. Merged clusters: Breakage rate

PRIORITIZATION SCHEME

Spatial clusters results are now analyzed in order to provide insight on what issues spatial clustering can address to the infrastructure management decision. In this section we aim to compare the information provided by the definition of (non spatial) groups of pipe with similar age to the information provided by spatial clusters of pipes. The figure 3 shows that basing on breakage number per cluster, zones A and B are the most critical despite they contain newer pipes (installation year respectively 1956-1958 and 1959-1960). Basing on the breakage rate (breaks/km/year/cluster), we observe that the most critical zones are not those installed before 1955 but rather where most breaks are concentrated. Furthermore, a same age of pipe may enclose different levels of breakage density which is one of the advantages of clustering (Fig.3).

Right now, the prioritization scheme within each cluster will be based only on breakage rate of pipes as pipes included in clusters are automatically classified. Besides, to compare the current methodology with the classic one based only on breakage rate of pipes, we sort pipes network according to the breakage rate (high to low) and we plot the curve describing the avoided number of breaks versus the cumulative percentage of pipes length renewed in the network (Fig.4). To get validation period, we share the time horizon since 1962 to 2003 into two periods: analysis [1962, 1995] and validation [1996, 2003].

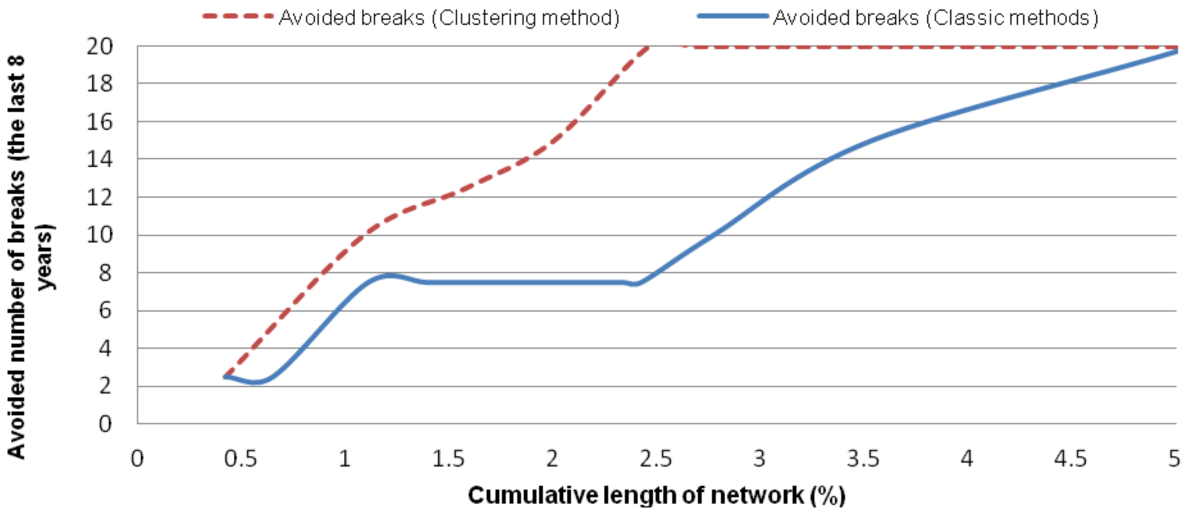

Figure 4. Avoided breaks versus renewed length

The figure shows that from 25% of the renewed length, the clustering approach is still advantageous over the classic one. This may justify the importance of the clustering concept as a long- term prioritizing method. It seems that the clustering method is very suitable for a long term rehabilitation planning where failures can form significant clusters during a meaningful period of time.

Fig.4 shows that a long-term rehabilitation program considering 25% of the network for renewal allows us to avoid 50% of breaks while replacing 50% of the network grant us to avoid 100% of failures and 90% with rehabilitation techniques not based on spatial and temporal clustering. A second issue to underline here is that the replacement based on spatial analysis of clusters allows us to have fewer replacements, less linear to replace (10.9 km against 11.3km for classic method) that's equivalent to 1.5% of the whole network but sparing the same number of future breaks (avoided breaks). Obviously, avoiding unnecessary replacements for 1.5% of the network will allow us to save about 99,650 € for the last seven years which corresponds to an average amount of 14,230 € per year.

CONCLUSION

This research includes two main analyses using data on water pipe failures to complete the study started by (Nafi and Kleiner, 2010). Geostatistical analyses through the use of scanning algorithm were addressed to discover not only the distribution of water pipes failures in space but also indicate various spatial and temporal trends.

Through the spatial, temporal and scanning process based on the nearest neighbour hierarchical clustering, the breakage rate is calculated for each cluster as a first level of prioritization. This study indicates that a significant number of failures appear in geographic clusters. Notably, it shows point distribution with a strong concentration in some areas and low in another where it will be interesting to know the causes and types of these failures. In conclusion, analyses conducted via the point mapping cluster identification and spatial analyses techniques constitute powerful tools that can help for understanding the water pipe failure pattern problem and the prioritization process of pipes that should be renewed. The current methodology offers economic savings by avoiding unnecessary renewal.

REFERENCES

- Le Gauffre P and Haidar H. (2008). Assessing the benefits of rehabilitation programmes defined with the CARE-W decision support system. *Water Asset Management International 4.1: 15-18*.
- Nafi A and Werey C (2009). Modèles d'aide à la décision et modèles de prévision de défaillances : Etat de l'art et enseignements méthodologiques. Research report. Strasbourg. France. 56 p.
- Dridi L, Mailhot A, Parizeau M and Villeneuve J-P (2006). Planification optimale du renouvellement des conduites dans un réseau de distribution d'eau. Conference INFRA: 12th Annual Urban Infrastructure Week, Quebec Under Construction: Optimizing Infrastructure Resources, Québec City, Nov.22, 2006, pp.1-14.
- Halfawy M.R, Dridi, L and Baker S. (2008). Integrated decision support system for optimal renewal planning of sewer networks. Journal of Computing in Civil Engineering, v. 22, no. 6, pp 360-372.
- Giustolisi O and Berardi L (2009). Prioritizing pipe replacement: From multiobjective genetic algorithms to operational decision support. Journal of water resources planning and management. ASCE. 135(6): 484- 492.
- Goulter I.C and Kazemi A. (1988). Spatial and temporal groupings of water main pipe breakage in Winnipeg. Journal of Civil Engineering, 15(1): 91–97.
- Abdulvahit Torun A and Düzgün S (2006). Using spatial data mining techniques to reveal vulnerability of people and places due to oil transportation and accidents: A case study of Istanbul Strait. ISPRS Technical Commission II Symposium, Vienna, 12 – 14 July 2006.
- S. Ben-David, N. Eiron, and H.U. Simon. The computational complexity of densest region detection. Journal of Computer and System Sciences, 64(1): 22- 47, 2002.
- De Oliveira D.P, Garrett .J.H and Soibelman L (2010). A density-based spatial clustering approach for defining local indicators of drinking water distribution pipe breakage. Journal of Advanced Engineering Informatics.
- De Oliveira D.P, Neill D.B Garrett .J.H and Soibelman L (2011). Detection of Patterns in Water Distribution Pipe Breakage Using Spatial Scan Statistics for Point Events in a Physical Network. Journal of Computing in Civil Engineering, Vol. 25, No. 1
- Nafi A and Kleiner Y. (2010). Scheduling renewal of water pipes while considering adjacency of infrastructure works and economics of scale. ASCE Journal of Water Resources Planning and Management, 136, (5), pp. 519-603, October 01, 2010.

14-WWW-YES-2012-Tlili-Paper-2012-06-26-DT.doc