Urban Water Distribution Network Asset Management Using Spatio-Temporal Analysis of Pipe-Failure Data

S. Christodoulou, A. Gagatsis, A. Agathokleous, S. Xanthos & S. Kranioti
University of Cyprus, NIREAS International Water Research Center, Cyprus

Abstract

The work presented herein investigates the use of spatio-temporal analysis as a decision support tool for increasing the efficiency of maintenance strategies ("repair or replace" decisions) related to urban water distribution networks (UWDN). The analyses used utilize both classical statistical tools and neurofuzzy systems, as well as GIS-based spatio-temporal clustering and visualization techniques. The study investigates UWDN in an urban locale, under both normal and abnormal operating conditions as manifested by the effects of intermittent water supply on the performance and fragility of such networks. The spatio-temporal analysis examines the resulting water losses, the inefficiencies and the overall maintenance cost during the aforementioned operating conditions and it is based on a six-year dataset (2003-2010) including thousands of pipe failure incidents. The spatio-temporal analysis, further to showing an increase in the number of water-leak incidents during and immediately after the years of enacting intermittent water supply policies, it allows for spatio-temporal clustering and pattern recognition, it helps devise repair-or-replace strategies and it reinforces the belief that intermittent supply increases the vulnerability of UWDN.

Keywords: spatio-temporal analysis, UWDN, GIS.

1 Introduction

Sustainable management of urban water distribution networks should include not only new methodologies for monitoring, repairing or replacing aging infrastructure, but more importantly it should facilitate the modeling of the state of each piping network and the predictive analysis of its behavior over time. Among the primary goals should be the arrival at intelligent 'repair or repair' decisions through the devising of effective assessment/management strategies and intelligent decision support systems. Since the decision to repair or replace failing pipe segments relates to the tolerance a water agency exhibits towards risk and the socioeconomic implications from a possible pipe failure, UWDN managers need to be able to dynamically assess the reliability of their piped networks and to arrive at intelligent management decisions through the monitoring and evaluation of a plethora of operational and risk-of-failure factors.

Research to-date has helped identify several of the potential risk factors that contribute to pipe breaks, such as: the pipe diameter, material and length, the operating pressure in the network, a pipe's age, the number of previously observed breaks for each pipe segment, the soil conditions, and the external loads to the underground piping network. Some of the factors are time-invariant and some are time-dependent, but all have been shown to be contributing factors to the overall risk-of-failure level.

Interestingly enough, though, the various analyses to-date are primarily based on one-dimensional snapshots of UWDN performance data, lacking the spatial and temporal dimensions and the associated analyses to go beyond the current state of affairs. The work reported herein introduces such spatio-temporal analyses of pipe-failure data which have been developed to complement analytical tools for evaluating the performance of UWDN, to improve on the detection of data clustering and to increase the efficiency of associated decision support systems.

2 State of knowledge

To-date a number of studies on infrastructure assessment and deterioration modeling of urban water distribution networks have been undertaken. The intent of such studies has traditionally been to assist UWDN operators in improving their understanding of the network's behavior over time, its deterioration rate and its reliability with respect to presumed risk factors. A primary goal has been the arrival at "repair-or-replace" decisions on a more scientific basis. The studies, thus, usually attempt to identify statistical relationships between water main break rates and influential risk factors such as a pipe's age, diameter and material, the corrosiveness of the soil, the operating pressure and temperature, possible external loads and recorded history of pipe breaks.

Most studies in literature show a relationship between failure rates and time of failure (age of pipes), and some of them suggest a methodology to optimize pipe replacement times. Shamir and Howard (1979) reported an exponential relationship, and Clark (1982) developed a linear multivariate equation to characterize the time from pipe installation to the first break and a multivariate exponential equation to determine the breakage rate after the first break. A study by Andreou et al. (1987) suggested a probabilistic approach consisting of a proportional hazards model to predict failure at an early age, and a Poisson-type model for the later stages, and further asserted that stratification of data (based on specific parameters) would increase the accuracy of the model. A non-homogeneous Poisson distribution model was later proposed by Goulter and Kazemi (1987) to predict the probability of subsequent breaks given that at least one break had already occurred. Finally, Kleiner et al. (1998) developed a framework to assess future rehabilitation needs using limited and incomplete data on pipe conditions. Christodoulou et al. (2009, 2010a, 2010b) reported on several conclusions derived from New York City and Cyprus's datasets and outlined a framework for integrated GIS-based management, risk assessment and prioritization of water leakage actions. Of special interest is the work reported by Christodoulou and Deligianni (2009) in which data mining techniques and neurofuzzy systems are employed in identifying pipe-failure data clusters and then in converting these clusters into decision support rules (figure 1).

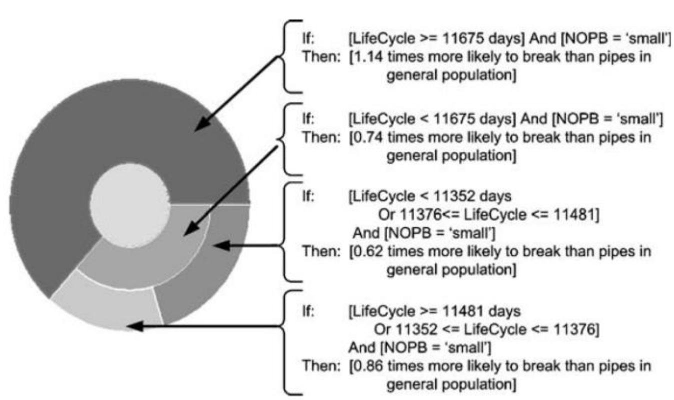

Figure 1. Sample DSS rule derived from applying a neurofuzzy classification. Adapted from Christodoulou and Deligianni (2009).

3 Risk-of-failure analysis under normal and abnormal operating conditions

3.1 Normal operating conditions

As aforementioned, previous studies by Christodoulou et al. (2009, 2010a, 2010b) have reported on the effect of several potential risk factors on the fragility and probability of failure for UWDN pipes. The studies initially identified as most influential risk-of-failure factors to be the number of previously observed breaks (NOPB), the material type (MAT), the length (L) and the diameter (D) of each pipe (Christodoulou et. al., 2006), and then deduced rules about the relative importance of such factors. The deduced knowledge was based on specific datasets and several analytical or artificial intelligence techniques such as pattern recognition, neurofuzzy systems and data-mining methods.

For example, Christodoulou et al. (2010b) report that by using several data stratifications on a Limassol dataset spanning a 5-year period (2002–2007) and about 2,000 break incidents, they were able to deduce that:

- The hazard rate of black medium density polyethylene (MDPE-Black) pipes is higher than galvanized (GI), asbestos cement (AC) and MDPE-Blue pipes.
- The hazard rate related to pipe deterioration greatly outpaces the hazard rate of the other incident types (corrosion, interference by others, tree roots, connection hose, other). This is an indication that aging pipes should be replaced at smaller time-intervals (at about 11,000 days ~ 30 years), otherwise the risk of failure increases substantially.
- The hazard rate related to corrosion accelerates in time and surpasses the rate of increase of interference-related, tree-roots and connection-hose incidents. Similarly, incidents related to the existence of tree roots in the vicinity of pipes and their effect on pipe failure have a hazard rate which is accelerating over time. This is an indication that pipes in the vicinity of trees should be monitored more closely and/or replaced in smaller time intervals than other pipes.
- Medium and large diameter pipes have approximately the same hazard rate over time but small-diameter pipes have an increasing hazard rate. This is an indication that small-diameter pipes should be replaced at smaller time intervals than medium and large-diameter pipes.
- The 'number of previously observed breaks (NOPB)' risk factor is very critical in the overall repair-or-replace management strategy. The hazard rate for the NOPB factor was subsequently shown to be a dominating parameter, especially for the range of $4 \leq NOPB \leq 8$, wherein if NOPB is increased by 1 incident then the hazard rate increases by 20.49%.

3.2 Abnormal operating conditions

The aforementioned risk-of-failure metrics seem to hold true, at least in terms of contributing factors, in a variety of UWDN operating under normal conditions. Recent studies seem to also uphold the metrics in the case of UWDN under abnormal operating conditions (Christodoulou at al. 2011). Such an abnormal operating condition is the case of intermittent water supply and the case study reported by Christodoulou at al. (2011), in which an UWDN has been studied for a two-year period during which intermittent water supply was forced upon the network operators in order to mediate the effects of extensive and severe water drought.

The UWDN in study (Water Board of Limassol) had been the focus of several related studies in the past and documented in literature (Christodoulou et al., 2006, 2009, 2010a, 2010b). The network is over 50 years of age and serves approximately 170,000 residents through approximately 64,000 consumer meters in an area of 70 km^2. The annual volume of potable water distributed through the network of pipes, of approximate length 795km, is about 13.7x106 m^3 and of value €7.0 million. The intermittent water supply lasted for about 2 years (Mar. 2008 – Oct. 2010) and was presumed to have reduced the overall water consumption on the island by about 20% (by volume of water) compared to previous years. This savings estimate, even though calculated based on the total water consumed, was actually the direct result of the reduction in the water supplied by the Republic's Water Development Department. Furthermore, the presumed water savings did not take into consideration the side-effects of the intermittent water supply policy on the piping networks and the resulting water loss thereafter.

Statistical analysis of the pipe-failure data during this two-year period, complemented by pipe-failure data from a second UWDN spanning a seven-year period (2003-2010) reinforced the conclusions derived from the case of normal operating conditions. The analysis showed a dramatic increase in pipe-burst incidents which can be attributed directly to the intermittent supply policy. This increase is especially evident in deteriorating pipes (with NOPB ≥ 4) and small-diameter pipes (primarily house connections). However, data stratifications of the type, the material and diameter of each pipe, the seasonal weather, the soil conditions and the water pressure levels at the time of each water-leak incident were expected to shed more light on to the behavior of the UWDN under the aforementioned adverse operating conditions. Clustering and spatio-temporal analysis were the missing link in the analysis of the pipe-failure data.

4 The developed spatio-temporal analytical and visualization tools

One of the primary objectives of the NIREAS-IWRC and UCyAMR projects currently in progress is to provide managers of UWDN with automated decision support tools that consider the spatio-temporal aspect of pipe-failure incidents. As a result, an application is being developed that will allow the user to geographically visualize selected data on pipe-failure incidents. The goal is to provide UWDN managers with visual tools that can analyze pipe-failure datasets in both space and time, identify spatio-temporal patterns and derive rules based on data clusters.

Several database software have spatial-extensions built that allow the user to store geometric and geographic data (e.g. Postgres through the PostGIS extension, SQL server 2007 etc). Additionally, these applications provide the user with several spatial functions that can be embedded in SQL code. In the NIREAS-IWRC and UCyAMR projects, numbered street networks of Nicosia and Limassol have been acquired and are used to geocode pipe-failure incidents using street-name, street number and postcode. The user will be able to apply filters through the application on attributes such as (1) from-to dates when the incidents occurred (2) the type of incidents (3) the district metered areas (DMA) where the incidents occurred (if required). The user will also have the option to filter out all incidents which have occurred in pipes that have since been replaced. These filters are parsed as parameters in a SQL query and visualized in a GIS viewer. In addition to viewing the selected data, the user has the choice to calculate the incident occurrence in District Metered Areas (DMAs) (figure 2). A point-in-polygon spatial-database function is used to count the incidents occurring within each DMA polygon stored in the database and then calculate indicators of degradation (e.g. water main breaks per water main km). The user can also execute this procedure at a street level (figure 3). Finally, the user is able to select the option of calculating spatial (figure 4) or spatio-temporal (figure 5) clusters occurring within the selected data.

Past studies have shown the importance of the "number of previous breaks" (NOPB) factor, on a pipe or on a street level, as a predictor in modeling pipe breaks (Andreou et al, 1987; Christodoulou et al, 2009). These findings showed that a pipe-break is more likely to occur in an area where several previous breaks have occurred compared to an area where few or no pipe-breaks have occurred. Consequently, the significance of the NOPB factor implies that pipe-failure incidents tend to be clustered in space. Few studies have studied the spatio-temporal aspect of pipe-failure incidents in UWDNs. Goulter and Kazemi (1988) first studied the spatio-temporal aspect in pipe-breaks occurring in Winnipeg using an ad-hoc methodology. They found that pipe-breaks were significantly clustered in space and furthermore that there is a strong association between the spatial and temporal interval of two incidents meaning that incidents that occurred “close by” tended to have a shorter time occurrence interval between them. Gagatsis (2011) calculated the nearest neighbor distance and Ripley’s K statistic using pipe-failure data from the UWDN of Limassol between 2008 and 2010 and found that the data was significantly clustered compared to a random spatial distribution (NNIndex = 0.382298, p-value < 0.005; NNI=1 indicates random distribution, NNI>1 indicates dispersion and NNI<1 indicates clustering), indicating that pipe failures are not spatially random and that there is a pattern in such failures.

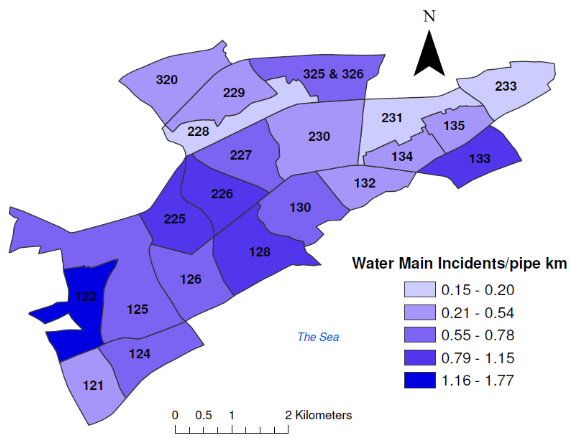

Figure 2. Chloropleth map using a natural jenks categorization, showing the number of water main pipe failures per water-main kilometer in selected DMAs of Limassol (July 2007 - February 2011). Adapted from Gagatsis (2011).

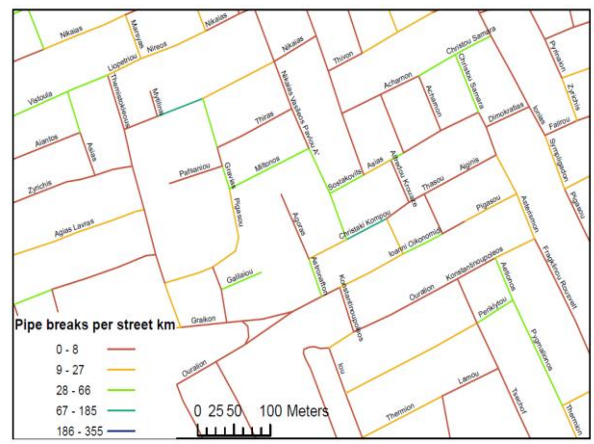

Figure 3. A cloropleth map illustrating the number of pipe breaks per street segment kilometer, in streets of Limassol during the year 2010. A natural jenks classification with 5 categories is used for the illustration.

It should be noted that previous data-mining analysis (Christodoulou and Deligianni, 2009) had only focused on pipe attributes. Spatial or geographical clustering is aimed at identifying regions in space where there is a higher concentration of points (data objects) compared to the rest of the study site. This is accomplished by grouping together points which are in relative proximity from one another. Spatio-temporal clustering is aimed at identifying regions in space with a relatively high concentration of points over a certain period of time. These techniques have been used extensively in crime pattern analysis (Rey et al., 2011) and epidemiological data analysis (Mcnally et al., 2006). In the context of clustering pipe-failure incidents, clusters represent parts of the UWDN that had an abnormally high number of failure incidents over a certain period of time. Identifying these clusters is useful because they can be used to: (a) Provide an understanding on why they occur in certain regions over certain times (e.g. by linking the clusters to soil maps using GIS), (b) help decide whether to repair or replace pipes over certain areas of the network when they break again.

Oliveira et al. (2011) suggested a spatial-clustering approach based on a network implementation of the DBSCAN clustering algorithm created by Ester et al. (1996). DBSCAN is a density based algorithm that groups points together based on how densely they are located from one another, without requiring a-priori knowledge of the number of clusters. The algorithm can identify clusters of arbitrary shapes (such as a cross road) in contrast to popular partitioning algorithms that produce ellipses and convex hull clusters. Additionally, DBSCAN can identify points that do not belong to any cluster and classify them as noise. These characteristics are essential for clustering pipe-failure incidents which a restricted to a network. Oliveira et al. (2011) suggested that using the network shortest distance instead of the Euclidean shortest distance between two pipes is more appropriate since this represents more accurately their proximity within the network. Oliveira et al. (2011) also discussed the disadvantage of DBSCAN in discovering clusters of different densities and suggested an approach of extracting clusters of different densities based on the OPTICS clustering algorithm suggested by Ankerst et al. (1999).

The application designed in the NIREAS-IWRC and UCyAMR projects uses an extended DBSCAN approach which uses an additional user input parameter in order to cluster incidents in both space and time. The user can still opt to ignore the temporal aspect and only cluster incidents spatially, if desired. The clustering algorithm requires the following input parameters: (1) The MinPts parameter that states the minimum number of breaks required to form a cluster (2) The distance threshold (the radius if euclidean distance is used or the network distance threshold around an incident). (3) The time window before and after an incident. The DBSCAN algorithm is based on the following notions: (1) A Core Point is a point which has at least MinPts of points within its threshold distance. A Core Point forms a cluster. (2) A directly density reachable point is a point that is within the distance threshold of a Core Point. (3) A point **p** is density reachable from another point **q** if there is a chain of points $\mathbf{p_1, \ldots, p_n, p_1 = q, p_n = p}$ where $\mathbf{p_{i+1}}$ is directly density reachable from $\mathbf{p_i}$. (4) A point belongs to a cluster if it is directly density reachable from a core point or density reachable from a point belonging to a cluster. (5) All points that are not part of a cluster are considered as noise. The algorithm used in the NIREAS-IWRC/UCyAMR application extends these notions by adding the time window parameter. Now a Core Point is a point which has at least MinPts of points within its threshold distance that also occur within the time window defined. Similarly, a directly density reachable point is a point that is within the distance threshold of a point and also within the time window of a core point. A brief pseudo-algorithmic representation of the algorithm is presented in figure 6. For a more detailed explanation of how DBSCAN workers refer to Ester et al. (1996).

For the clustering procedure, the application retrieves the filtered dataset from the database and sends the list of incidents to the algorithm code. The application adds 2 columns to the dataset: 'Column A' stating whether an incident is a core incident or noise. 'Column B' assigns a cluster id, a number that states the cluster to which an incident belongs to. This number can then be used in a GIS viewer to display the different clusters in different colors (figure 4). Additionally, the application creates a table which labels each cluster using the dates of the first and last incidents occurring within the cluster and calculates spatio-temporal (figure 5) clusters occurring within the selected data.

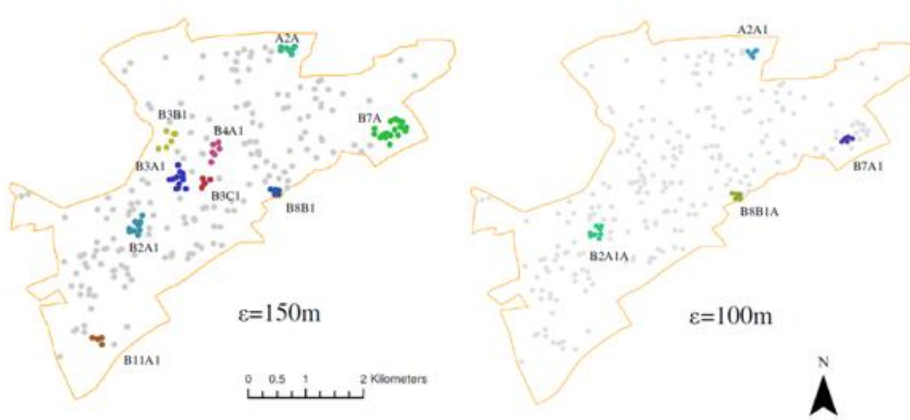

Figure 4. Spatial Clusters of Water Main pipe failure incidents discovered using the DBSCAN algorithm in an area of Limassol between 2008 and 2010. The parameter "ε" indicates the distance threshold. A smaller ε results in the discovery of denser clusters with more incidents assigned as noise. The different colors indicate the different clusters. Adapted from Gagatsis (2011).

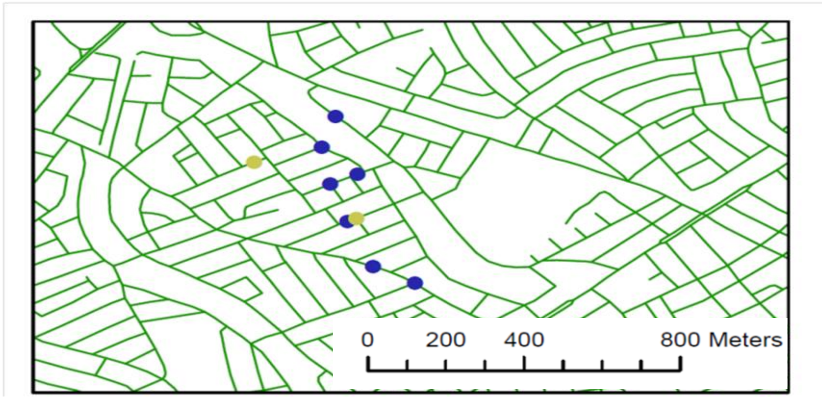

Figure 5. An example of a spatio-temporal cluster of pipe breaks detected in the Limassol municipality in Cyprus between late June (green colored points) and the end of July (blue) in the year 2010.

```
- While list is not empty
- Start processing first incident in the list
- If number of incidents within distance threshold d and time window t of incident >= MinPts

THEN

- The incident is a core incident and a spatiotemporal cluster is formed
- The algorithm iteratively collects all density reachable incidents from the incident w.r.t. d,t,MinPts
- The incidents are assigned a cluster id and removed from the list
- The algorithm proceeds with the first incident on the modified list

ELSE

- The incident is not a core incident
- The algorithm continues with the following incident on the list

- All incidents not belonging to a cluster as assigned as noise
```

Figure 6. A simple representation of the spatio-temporal clustering algorithm. The algorithm is essentially the same as the one presented by Ester et al. (1996) simply adding time as a condition. The algorithm is run until all the incidents in the list have been processed.

5 Conclusion

The manuscript presents tools developed for spatio-temporal analysis of pipe-failure data and conclusions drawn from such analysis. The reported analysis of normal and abnormal operating conditions uphold conclusions drawn from statistical and survival analysis of the same datasets previously reported on in literature, complementing the previously obtained numerical results with visual, spatial and temporal data-clustering results. The results reinforce the existence of patters in the behavior of UWDN and help deduce repair-or-replace rules for such networks. Ongoing work on devising decision support systems for the sustainable management of UWDN incorporates such spatio-temporal tools and complements it with live UWDN performance data through the use of automatic meter reading for dynamically detecting pipe-failure clusters.

Acknowledgements

The NIREAS-IWRC and UCyAMR research projects are co-financed by the European Regional Development Fund and the Republic of Cyprus through the Cyprus Research Promotion Foundation (Grants No. AEIFORIA/ASTI/0609(BIE)/07, PENEK/ENISH/0308/34 and NEA

YPODOMI/STRAT/0308/09). Special thanks are also extended to the Water Boards of Limassol and Nicosia for providing operational data on their piping networks.

References

ANDREOU, S.A., MARKS, D.H. AND CLARK, R.M., 1987. A new methodology for modeling break failure patterns in deteriorating water distribution systems: theory. *Advances in Water Resources*, 10(3), 2-10.

ANKEEST, M., BREUNIG, M. KRIEGEL, H.P., SNDER, J., 1999. OPTICS: Ordering Points to Identify the Clustering Structure. Proc., *ACM SIGMOD International Conference on Management of data*, ACM Press, 49-60.

CHRISTODOULOU, S., DELIGIANNI, A., ASLANI, P. AND AGATHKLEOUS, A., 2009. Risk-based asset management of water piping networks using neurofuzzy systems. *Computers, Environment and Urban Systems*, 33(2), 138-149.

CHRISTODOULOU, S. AND DELIGIANNI, A., 2010a. A Neurofuzzy Decision Framework for the Management of Water Distribution Networks, *Water Resources Management*, 24(1), 139-156.

CHRISTODOULOU, S., AGATHOKLEOUS, A., CHARALAMBOUS, B. AND ADAMOU, A., 2010b. Proactive Risk-Based Integrity Assessment of Water Distribution Networks. *Water Resources Management*, 24(13), 3715-3730.

CHRISTODOULOU, S., AGATHOCLEOUS, A. AND CHARALAMBOUS, B., 2011. Urban Water Distribution Network Perforamnce Under Intermittent Water Supply Conditions. *Proc. EWRA2011 International Conference*, Catania, Italy.

CLARK, M.A., STAFFORD, C.L. AND GOODRICH, J.A., 1982. Water distribution systems: a spatial and cost evaluation. *Journal of the American Water Works Association (AWWA)*, Planning and Management Division, 108, 243-256.

ESTER, M., KRIEGEL, H.P., SANDER, J. AMD XIAOWEI, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In E. Simoudis, J. Han & U.M. Fayyad (eds), *Proc., Second International Conference on Knowledge Discovery and Data Mining* (KDD-96). AAAI Press, pp. 226–231.

GAGATSIS, A., 2011. *A Spatio-Temporal exploratory analysis of pipe-failure incidents in the Water Distribution Network of Limassol, Cyprus*. MSc. thesis, Edinburgh University

GOULTER, I.C. AND KAZEMI, A., 1988. Spatial and temporal groupings of water main pipe breakage in Winipeg. *Canadian J. of Civil Eng.*, 15(1), 91-100.

KLEINER, Y., ADAMS, B.J. AND ROGERS, J.S., 1998. Long-term planning methodology for water distribution system rehabilitation. *Water Resources Research*, 34(8), 2039-2051.

MCNALLY, R.J.Q., ALEXANDER, F.E. AND BITHELL, J.F., 2006. Space-time clustering of childhood cancer in Great Britain: A national study, 1969-1993. *International Journal of Cancer*, 118, 2840-2846.

OLIVEIRA, D.P., GARET, J.H., SOIBELMAN, L., 2011. A density-based spatial clustering approach for defining local indicators of drinking water distribution pipe breakage. *Advanced Engineering Informatics*, 25, 380-389.

REY, S.J., MACK, E. KOSCHINSKY, J., 2011. Exploratory Space–Time Analysis of Burglary Patterns. *Journal of Quantitative Criminology*, 1-23, Doi: 10.1007/s10940-011-9151-9.

SHAMIR, U. AND HOWARD, D.D., 1979. An analytic approach to scheduling pipe replacement. *Journal of the American Water Works Association (AWWA)*, 71(5), 248-258.