Desalination and Water Treatment
www.deswater.com

33 (2011) 97–106
September

doi: 10.5004/dwt.2011.2631

Water pipe network reliability assessment using the DAC method

V. Kanakoudis[a], S. Tsitsifli[b]*

Department of Civil Engineering, Pedion Areos, University of Thessaly, 38334 Volos, Greece
[a]*Tel. +30 24210 74156; Fax +30 24210 74169; email: bkanakoud@civ.uth.gr*
[b]*Tel. +30 2410 531397; email: tsitsifli@civ.uth.gr*

Received 20 January 2010; Accepted in revised form 5 November 2010

ABSTRACT

The Discriminant Analysis and Classification (DAC) method has offered remarkable results regarding the prediction of failures in an oil or a gas pipe network, based on the network characteristics. The DAC method also proved its ability to identify the most crucial network parameters affecting its behavior. The present study attempts to check whether the DAC method can provide safe results regarding the reliability assessment of urban water networks too. The DAC method aims at classifying the network pipes in two groups (failures/successes), based on simple or/and dimensionless joint variables. Serious problems related to the quality, reliability and compatibility of the data provided by the Water Utilities were tackled using dummy variables based on field data. The distinction between the meanings of 'failure' and 'success', for a water pipe network, was also crucial. For the case study water pipe network of Larisa city, in Greece, the criterion used to define the meanings of 'failure' and 'success' was "the total water volume being lost" through a leak or a break in a pipe. The available pipe failure data records for Larisa city were poor and not fully compatible to the DAC method demands. The results showed that discrimination is good enough and would be even better if additional data (in line with the DAC standards) was available. Thus, overall, the DAC method proved to be a useful tool for pipe reliability prediction in urban water pipe networks.

Keywords: Water pipe networks; Reliability assessment; Discriminant analysis and classification

1. Introduction

The Discriminant Analysis and Classification method has been successfully used by the authors in the past to predict whether an oil or a gas pipe will fail or not, based on its characteristics, while at the same time identifying the most crucial factors affecting its behavior [1]. The DAC method classifies the pipes in two groups (failed pipes called 'failures'/not failed pipes called 'successes'), based on simple or/and dimensionless joint variables related to pipe characteristics. It proved to be very effective in developing pipe failure prediction models, as it can analyze the differences of the two groups using the Z-score index, resulting from the thorough study of a large number of variables [1;2]. Pipes are classified as 'failures' or 'successes' utilizing pipe failure data records. Several pipe characteristics (e.g., operating pressure; length; diameter; material; age; fluid supplied by the pipes called 'product') were used as variables. The DAC method revealed the correlations amongst pipe characteristics affecting pipes failure rates. Joint variables resulting from simple ones were introduced in order for the analysis to be based on

* Corresponding author.

Presented at the 2nd International Conference on Environmental Management, Engineering, Planning and Economics (CEMEPE), Mykonos, Greece, June 21–26, 2009.

dimensionless variables. Those variables were also used as they, compared to simple ones, tend to carry more information regarding the behaviour of the pipe. Results showed that the successful implementation of DAC method highly depends on how accurate and extended the available pipe failure data records are [1]. Such records are usually available for oil and gas pipe networks (due to the high revenue value of the fluid supplied).

Therefore, the authors tested the DAC method for two oil/gas pipe networks [1].The main aim in each case was to develop a model that could correctly classify the network pipes to 'successes' or 'failures'; to define the crucial pipe characteristics affecting its behaviour; and to predict whether a pipe will fail or not. The results for both networks were very satisfying, as the correct prediction rate for the failing pipes reached 96.6% [1]. Thus the authors decided to check whether the DAC method could provide the same promising results for urban water networks too. The present paper deals with the implementation of the method in the water pipe network of Larisa city in Greece. The main problem that had to be faced was to come up with a proper way to "manipulate" the available pipe failure data records in order to fit the needs/standards of the DAC method. The distinction between the meanings of "failure" and "success" was also a very intriguing task to handle.

2. Literature review

There have been several attempts to identify statistical relationships between water main break rates and pipe and network characteristics such as age/diameter/material of the pipe; soil aggressiveness; operating pressure; external temperature; possible external loads (including highway traffic); and recorded history of pipe breaks. Shamir and Howard [3] reported an exponential relationship between failure rates and pipe age, and developed a methodology to estimate the optimal pipe replacement time. O'Day et al. [4] also studied pipe failure rates related to their age. Clark et al. [5] developed a linear multivariate equation to determine the time between the pipe installation and the occurrence of its first break. Clark et al. [5] developed also an exponential multivariate equation to determine the pipe failure rate following the first break incident. Kettler and Goulter [6] reported a strong linear correlation between the break rate of a water main and its diameter; and a moderate linear correlation between the break rate of a water main and its age. They also examined the break rate variation regarding the material of the pipe, and they analyzed the types of breaks for different pipe materials. The results of several studies revealed a correlation between the diameter of a pipe and the type and rate of its breaks [7,8]. Other studies [6,7] revealed that small diameter pipes tend to break more often compared to large diameter ones under certain environmental conditions. Marks et al. [9] developed a failure model that can calculate the probability of a pipe break in a small time range dt based on the pipe age; the number of previous breaks; and the time since its last break. Andreou et al. [10] suggested a probabilistic approach, consisting of a proportional hazards model to predict failure at early stages of deterioration. For the later stages of deterioration they developed a Poisson-type model. The base function relating the probability of breakage to pipe age could vary in the same distribution system. Therefore, the layering of the data set into groups (based on specific parameters) would increase the accuracy of the model.

Goulter and Kazemi [11] proposed a break clustering model (non-homogeneous Poisson distribution model) to predict the probability of subsequent breaks, given that at least one break had already occurred. Kleiner et al. [12] and Kleiner and Rajani [13] developed a methodology to assess future rehabilitation needs on the basis of historical water main breakage records available. These records included limited and incomplete data of variables causing pipe breaks. Prasad et al. [14] used genetic algorithm methods introducing network resilience. This is a new reliability measure trying to provide surplus head above the minimum allowable head at nodes and reliable loops with practicable pipe diameters. Vanrenterghem et al. [15] used a proportional hazards model to analyze replacement strategies using failure data records from the New York City water mains. Aslani [16] and Christodoulou et al. [17] reported additional work using the same case study. Christodoulou et al. [18] used the knowledge gained by the New York City case study and reported a developed framework for integrated GIS based management, risk assessment and prioritization of water leakage actions. Kanakoudis and Tolikas [19] developed indices to assess the performance level of the system and hierarchically analyzed the possible preventive maintenance actions in a water system. They also developed a model to calculate the pipe optimal replacement time based on a technical-economical analysis taking into account the costs associated to the repair or replacement of failing system components. Park et al. [20] presented a method focused on modeling the failure rate and estimating economically optimal replacement time of an individual water main by using two widely used indices related to the rate of failure occurrence. This methodology has the limitation of requiring large number of recorded breaks. Also the data of the failure-causing parameters must be collected and recorded in a standardized framework in order to maximize the efficiency of water distribution system maintenance. Finally, Christodoulou [21] sets the "repair or replace" dilemma investigating the failure causing parameters and outlines a multi-criteria decision support system for modeling pipe behavior based on non-parametric survival analysis techniques.

3. Methodology

The DAC method aims at forming two groups of pipes, failed and survived ones. Then, by using the linear discriminant function (1) the Z-scores of each group are identified.

$$Z_m = U_0 + U_1 X_{1m} + U_2 X_{2m} + \ldots + U_i X_{im} \quad (1)$$

where Z_m is the value (score) of the discriminant function for case m; X_{im} is the value of the variable *i* (from now on X_i will indicate variable *i*); U_i is the best discriminant coefficient of the *i* variable. The constant term U_o is the adjustment for the mean values, so that the mean discriminant score equals to zero over all cases [1]. The discrimination and the classification process drove to several results [1]:

- Stability of a discriminant variable. If a discriminant variable is unstable it will be excluded from the further steps of the classification process.
- Importance of a discriminant variable. High importance means greater contribution to the discrimination.
- A statistical technique used to evaluate the discrimination function is Wilk's Lambda (Λ) [2]. As the Λ-value gets closer to zero, the discrimination gets better.
- Another statistical technique used for the same reason is the canonical correlation coefficient (R). As R-value gets closer to 1, the discrimination gets better.
- The classification percentages derive from the classification matrix [1]. The classification percentages for the failed pipes EF_f and for the survived pipes EF_s indicate the discrimination ability related to failures or successes respectively (percentage of population correctly classified), while the total classification percentage EF_t indicates this ability over both populations. In pipe networks to classify a pipe as a 'failure' when it will prove to be a 'success' has less economic and safety implications than vice versa, as this would mean that the pipe is expected to survive but it will actually fail.
- The critical Z-score is the criterion used to classify a new pipe as 'failure' or 'success' [1,22].

For the application of the method to water pipe networks, it is necessary first to define what both meanings (failed and the survived pipes) stand for. This can be done using one of the following two criteria: a) the water loss rate during a failure (in this case the surviving pipes are those experiencing leaks while the failing ones are those facing breaks); b) the total water volume being lost during a failure (in this case the surviving pipes are those experiencing breaks while the failing ones are those facing leaks). The latter criterion is used here as the total amount of water being lost is what any water utility wants to decrease. Field studies proved that the total water volume lost due to leaks is even 5 times bigger that the respective water volume lost due to breaks [23].

From the implementation of the DAC method to oil and gas pipe networks it was found that the quality and the compatibility of the failure data records is crucial for DAC method successful implementation [1]. This also stands for water pipe networks. Therefore, the data related to the pipe characteristics must refer to each pipe according to its actual location. It is not possible to group all pipes having for example the same diameter and material, as each failure is unique and occurs in a specific place. The local conditions met (e.g., soil aggressiveness, external stresses) usually have a significant impact regarding the failure occurrence rate. The same conditions do not apply in all pipes of same material and diameter. Therefore, it is necessary to discriminate the pipes according to their location. Field studies proved that failure clustering in time and space is very common in water pipe networks [11,24].

The DAC method cannot give reliable results when the characteristics of the pipes per material and diameter refer to the total length of the network. Only rough conclusions can be drawn by such results that do not refer to all pipes included in the "sample," as a pipe may fail just due to specific local conditions. It does not mean that all pipes of the same material and same diameter will fail in the future. Furthermore, usually the data records provided by the water utilities containing details regarding the characteristics of the pipes (e.g., pipe diameter, material, length, previous failures, age) are very poor and not accurate enough. Finally, failure causing characteristics, such as soil conditions, pipe exact location, exact pressure, soil temperature, pipe previous failures, etc., are not usually recorded by the water utility. Thus, dummy variables are introduced.

4. Water pipe reliability assessment in Larisa case study

4.1. The case study network

The water supply network of Larisa city (Greece) was used as a case study network in order to check whether the DAC method can provide satisfying results. The local water utility keeps failure data in the form of total failures per pipe material and diameter per year. Pipe variables (mean values, standard deviation, min and max values) considered in the DAC method analysis are presented in Table 1. The network pipes are divided in failures and successes according to the criterion of total water volume losses (leaks/ repairs=failures; breaks/ replacements = successes). For the 18-year study period (1989–2006), there were 319 'failed' pipes recorded and 141 'survived' ones. The variables "operating pressure", "soil" and "external loads" are considered dummy variables and their values are not recorded but estimated based on experience and local conditions. As the local water utility does not keep analytical data records regarding the "operating pressure" of the network pipes, the present study considers its value to range from 3 to 6 atm according to the pipe

Table 1
Mean, standard deviation, min and max values of the variables used for Larisa

Variable	Leaks – Repairs F				Breaks – Replacements S			
	Mean	Standard deviation	MIN	MAX	Mean	Standard deviation	MIN	MAX
Material-MAT, %	22.297	12.073	9.85	39.16	22.946	13.507	9.85	39.16
Diameter-D, mm	140.62	103.42	19	500	105.6	59.55	19	300
Length-L, m	26,461.58	33,961.95	76.32	113,095.8	36,797.81	36,851	363.63	113,095.8
Previous failures-BR, m	11.893	21.944	1	135	329.05	640.73	1	4,281.2
Age-AG, y	24	19	0	61	13	12	0	59
DIM1	73	171	0.69	1,147	29.4	48.5	0.69	392
DIM2	2	7.1	0	79	20.49	62.1	0.009	661.39
Pressure-PRES, atm	4.85	0.84	3	6	5.09	0.63	3	6
Soil-S	1	0.82	0	2	0.99	0.83	0	2
External loads-LO	0.73	0.72	0	2	0.59	0.57	0	2

diameter. The variable "soil" refers to soil conditions (where the pipe is embedded) regarding its "aggressiveness" promoting either the pipe corrosion or damage due to soil water frost stresses developed. This variable was handled as a qualitative size rather than a quantitative one, taking values from 0 to 2, where 0 corresponds to a 'not at all aggressive' soil, while 2 corresponds to a 'very aggressive' soil. The same concept (qualitative vs. quantitative) also stands for the variable "external loads" (frost stresses excluded). Values for this variable range from 0 to 2, where 0 stands for 'minimum loads' from road traffic or for pipes under the pavement, to 2 which stands for 'maximum loads' from very heavy traffic with heavy vehicles. The variable "material" is being expressed as a percentage of each material length over the total pipe length. Two dimensionless joint variables are being used, namely DIM1 and DIM2, which are explained in Table 2. Several variable combinations resulted finally in the analysis of 26 scenarios (Table 3).

Table 2
Joint variables used in the analysis

Joint variable	Stands for	Joint variable	Stands for
DIM1	$[D/L]\ 10^3$	DIM2	$[BR\ /L]\ 10^3$

4.2. Results and discussion

The results of the first part of the analysis (Table 4) show that:

- The most important variable contributing the most to the discrimination of all scenarios is the variable "previous failures". This is justified, since previous studies proved a space and time clustering of failures near an initial failure site [10].
- Although the variable "soil" proved to be not that important (11/26 scenarios), it is considered stable, and therefore, it cannot be ignored and thus excluded from further analysis.
- The variables contributing the most to the discrimination are found to be in a descending order: previ-

Table 3
The 26 scenarios analyzed

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
D	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
MAT	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
AG	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
BR	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
L		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
DIM1			x	x								x	x	x	x	x	x	x	x	x	x	x	x	x	x	
DIM2				x															x	x	x	x	x	x	x	
S					x			x	x		x	x			x	x		x	x			x	x		x	
LO						x		x		x	x		x		x		x	x		x		x		x	x	
PRES							x		x	x	x			x		x	x	x			x		x	x	x	x

Table 4
Results for the Larisa water distribution network (characterizing the model variables)

Scen.	Important variable	Less important variable	Unstable variables	Scen.	Important variable	Less important variable	Unstable variables
1	Previous failures	Diameter	D, BR, AG	14	Previous failures	Length	D, BR, AG, DIM1
2	Previous failures	Diameter	D, BR, AG	15	Previous failures	Soil	D, BR, AG, DIM1
3	Previous failures	Diameter	D, BR, AG, DIM1	16	Previous failures	Soil	D, BR, AG, DIM1
4	Previous failures	Diameter	BR, AG, DIM1, DIM2	17	Previous failures	DIM1	D, BR, AG, DIM1
5	Previous failures	Soil	D, BR, AG, S	18	Previous failures	Soil	D, BR, AG, DIM1
6	Previous failures	Material	D, BR, AG	19	Previous failures	Soil	BR, AG, DIM1, DIM2
7	Previous failures	Material	D, BR, AG	20	Previous failures	Age	D, BR, AG, DIM1, DIM2
8	Previous failures	Soil	D, BR, AG	21	Previous failures	Length	D, BR, AG, DIM1, DIM2
9	Previous failures	Soil	D, BR, AG, S, PRES	22	Previous failures	Soil	D, BR, AG, DIM1, DIM2
10	Previous failures	Material	D, BR, AG	23	Previous failures	Soil	D, BR, AG, DIM1, DIM2
11	Previous failures	Soil	D, BR, AG	24	Previous failures	Age	D, BR, AG, DIM1, DIM2
12	Previous failures	Soil	D, BR, AG, DIM1	25	Previous failures	Soil	D, BR, AG, DIM1, DIM2
13	Previous failures	DIM1	D, BR, AG, DIM1	26	Previous failures	Age	D, BR, AG

ous failures, diameter, DIM2, pressure, length, age, product, soil, load and DIM1. From the international literature, it is found that all these variables affect pipe failures at a different rate for each case (Table 5). Previous studies [25,26] showed that material (product), diameter, length and traffic are the most important risk factors for the New York City case, while previous breaks, diameter, material and traffic are the most important ones for the Limassol case.

The implementation of the DAC method in the water distribution network of Larisa (Table 6, Fig. 1a) resulted in satisfying classification percentages, ranging from: a) 73.7% to 86.5% for the 'failed' pipes (EFf); b) from 63.8% to 74.5% for the 'survived' pipes (EFs); and c) from 72.2% to 82.0% in total (EFt). Thus, the majority of the pipes that actually survived or failed were correctly classified as "survivals" or "failures," respectively. Regarding the discrimination level, Wilk's Λ values ranged from 0.72 to 0.79, revealing that the discrimination level achieved was quite good. Finally, the *R* values agree with the Λ values ranging from 0.46 to 0.53 (Table 6, Fig. 1b). Scenarios no. 24 and 25 (the latter includes all variables) result in the best Wilk's Λ, CCC and EF_s values, while scenarios no. 21 and 23 result in the best EF_f and EF_t values (quite close to no. 24 and 25 respective ones).

Non-standardized coefficients (U_i) are those used for the calculation of Z-score for both groups of pipes (Table 7, Fig. 2). In a negative non-standardized coefficient (U_i), an increase of the related variable value results in decreasing its Z-score value and its probability to belong in the group of "successes-breaks". Such variables are the DIM1 and the pipe's material; age; and pressure. As the "age" and "pressure" values increase the pipes will tend to fail. The variable "pipe material" gets its greatest values for PVC pipes that are most commonly met in the network (39% of the total pipe length compared to 29% of asbestos-cement pipes; 12% of cast iron pipes; 10% of steel pipes; and 10% of PE pipes). Exactly the opposite takes place regarding the variables related to positive non-standardized coefficients, like the number of previous failure incidents; the external loads; and the DIM2. It is

Table 5
Parameters affecting the pipes failures rate [27]

Pipe section factors	Operational / maintenance factors	Environmental / climate factors
Pipe material	Operating pressure	Soil type
Pipe diameter	Nature/date of last failure (e.g. type, cause, severity)	Soil temperature or frost depth
Joint type	Nature of maintenance operations (e.g. TV inspections, pipe cleaning, cathodic protection)	Rainfall
Pipe age	Nature and date of last repair (e.g. type, length)	Soil moisture content
Pipe depth below surface	Water quality	Temperature
Pipe condition (e.g. wall thickness, defects)	Construction method	Traffic and loading

Table 6
Results for the Larisa water distribution network (classification achieved)

Scenario	EF_f	EF_s	EF_t	Wilk's Λ	*R*	Scenario	EF_f	EF_s	EF_t	Wilk's Λ	*R*
1	75.5%	68.1%	73.3%	0.787	0.461	14	80.9%	70.2%	77.6%	0.757	0.493
2	74.6%	68.8%	72.8%	0.787	0.462	15	77.4%	70.9%	75.4%	0.749	0.501
3	73.7%	69.5%	72.4%	0.785	0.463	16	80.3%	70.2%	77.2%	0.757	0.493
4	82.4%	63.8%	76.7%	0.759	0.491	17	79.6%	70.2%	76.7%	0.746	0.504
5	74.6%	68.8%	72.8%	0.787	0.462	18	78.4%	71.6%	76.3%	0.746	0.504
6	77.4%	72.3%	75.9%	0.749	0.501	19	82.4%	63.8%	76.7%	0.759	0.491
7	80.6%	70.2%	77.4%	0.757	0.493	20	79.9%	70.9%	77.2%	0.722	0.527
8	77.4%	70.9%	75.4%	0.749	0.501	21	86.5%	71.6%	82.0%	0.725	0.525
9	80.3%	70.9%	77.4%	0.757	0.493	22	79.9%	70.9%	77.2%	0.722	0.527
10	79.3%	70.2%	76.5%	0.746	0.504	23	86.5%	70.9%	81.7%	0.725	0.525
11	78.4%	71.6%	76.3%	0.746	0.504	24	83.7%	74.5%	80.9%	0.717	0.532
12	73.7%	68.8%	72.2%	0.785	0.463	25	83.7%	72.3%	80.2%	0.716	0.533
13	77.4%	72.3%	75.9%	0.749	0.501	26	79.6%	70.2%	76.7%	0.758	0492

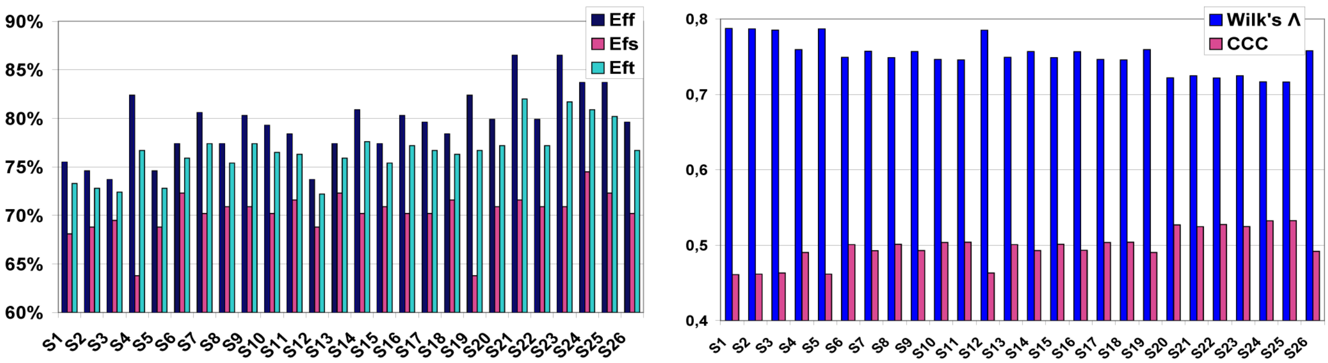

Fig. 1. (a) Ef_f, EF_s, EF_t; (b) Wilk's Λ and CCC for the scenarios analyzed.

quite expected that as the number of previous failure incidents will increase, the pipes will break (breaks = successes). The same happens for the "external loads". As the road will suffer from heavier traffic, the pipes will break (breaks = successes).

4.3. Time step analysis

In order to validate the DAC method, a time step analysis is used. The available data for the water supply network of Larisa city refer to a total time period of 18

Table 7
Non-standardized coefficients U_i for each variable in all scenarios

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
U_i (diameter)	–0.0003	–0.0015	–0.0003	0.0018	–0.0015	–0.0130	–0.0312	–0.0134	–0.0313	–0.0286	–0.0287	–0.0003	–0.0112
U_i (failures)	0.0281	0.0456	0.0471	0.0371	0.0456	0.0478	0.0479	0.0480	0.0479	0.0487	0.0488	0.0471	0.0485
U_i (material)	–0.0480	–0.0187	–0.0219	–0.0227	–0.0186	–0.0256	–0.0222	–0.0264	–0.0223	–0.0245	–0.0251	–0.0220	–0.0274
U_i (age)	–0.0287	–0.0221	–0.0206	–0.0172	–0.0222	–0.0265	–0.0275	–0.0263	–0.0274	–0.0281	–0.0279	–0.0205	–0.0254
U_i (length)		–0.0343	–0.0382	–0.0279	–0.0343	–0.0309	–0.0330	–0.0305	–0.0329	–0.0318	–0.0314	–0.0382	–0.0331
U_i (DIM1)			–0.0034	–0.0074								–0.0035	–0.0019
U_i (DIM2)				0.0645									
U_i (soil)					–0.0143			0.1138	0.0323		0.0926	0.0212	
U_i (pressure)							–3.2883		–3.2979	–2.2850	–2.2550		
U_i (loads)						1.7970		1.8553		1.0197	1.0774		1.6481
	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26
U_i (diameter)	–0.0273	–0.0115	–0.0274	–0.0262	–0.0263	0.0018	–0.0089	–0.0274	–0.0092	–0.0275	–0.0269	–0.0270	–0.0324
U_i (failures)	0.0488	0.0487	0.0488	0.0492	0.0493	0.0371	0.0370	0.0357	0.0371	0.0357	0.0360	0.0360	0.0305
U_i (material)	–0.0247	–0.0282	–0.0251	–0.0259	–0.0266	–0.0229	–0.0351	–0.0343	–0.0363	–0.0349	–0.0363	–0.0373	–0.0522
U_i (age)	–0.0269	–0.0252	–0.0267	–0.0275	–0.0272	–0.0171	–0.0218	–0.0248	–0.0217	–0.0247	–0.0248	–0.0247	–0.0334
U_i (length)	0.0000	–0.0327	–0.0355	–0.0338	–0.0334	–0.0278	–0.0189	0.0000	–0.0184	–0.0190	–0.0171	–0.0267	
U_i (DIM1)	–0.0024	–0.0019	–0.0025	–0.0016	–0.0017	–0.0074	–0.0062	–0.0078	–0.0062	–0.0079	–0.0069	–0.0070	
U_i (DIM2)						0.0646	0.0711	0.0783	0.0711	0.0784	0.0771	0.0772	
U_i (soil)		0.1203	0.0564		0.1005	0.0369			0.1414	0.0848		0.1208	
Ui(pressure)	–2.9797		–2.9888	–2.1880	–2.1560			–3.5002		–3.5135	–2.8810	–2.8510	–3.6412
Ui(loads)		1.7034		0.8967	0.9531		1.8106		1.8726		0.7583	0.8226	

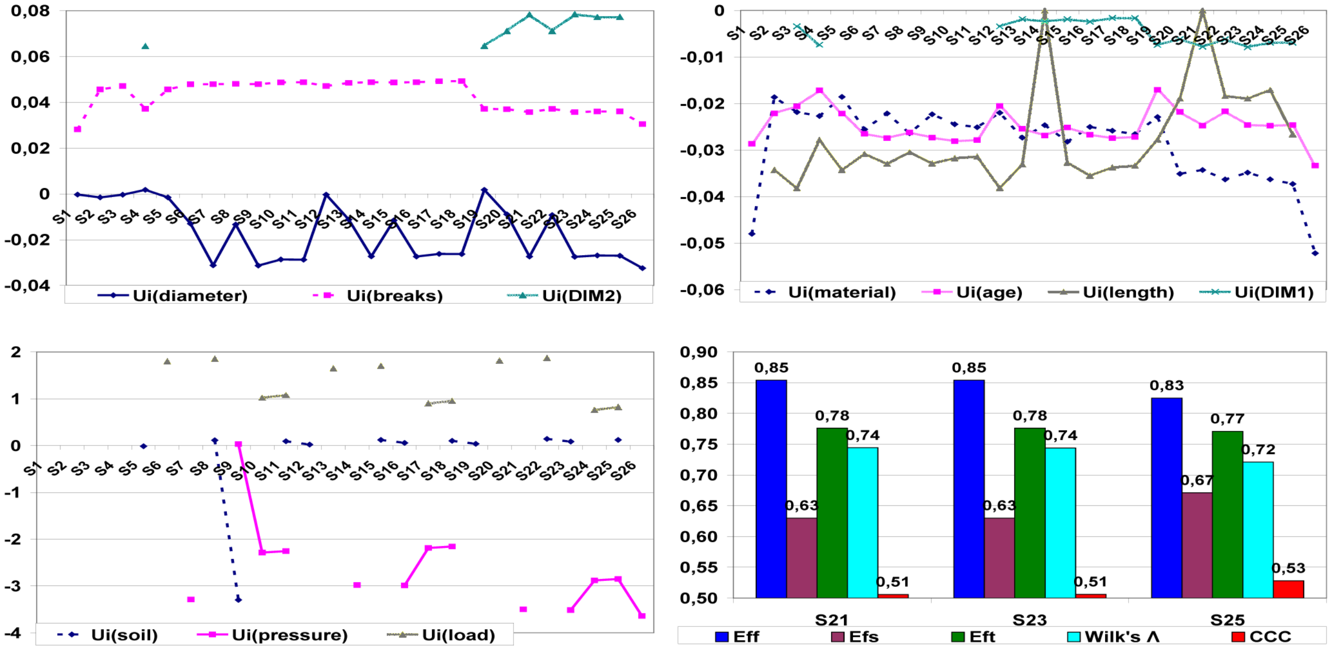

Fig. 2. (a) Non-standardized coefficients (U_i) of D, BR, DIM2; (b) U_i of MAT, AGE, L, DIM1; (c) U_i of S, PRES, LO and (d) EF_f, EF_s, EF_t, Λ and CCC for the time step analysis.

years, divided in two time sub-periods of 9 years each. The time-step analysis is applied only for the scenarios no. 21, 23 and 25 that previously gave the best discrimination and failure prediction results. While the failure prediction model was calibrated using the actual failure data records of both sub-periods, its validation was based on what actually happened during the second sub-period. The results of the analysis for the first 9-year time step (Table 8) revealed that:

- The variable "previous failures" is the most important variable contributing the most in the discrimination for all scenarios. This was expected since this variable will mainly determine if the pipe with the same characteristics will fail again. That was the case when the total period was examined.
- The variable "soil" proved not that important (in 2 out of 3 scenarios). When the total period was examined, the less important variable for the scenario no. 21 was "length". For the scenarios no. 23 and 25 the variable "soil" was added as an unstable variable.
- The time step analysis resulted in slightly decreased correct classification percentages. Wilk's Λ and CCC's values are also worse (Fig. 2d). This was expected as the DAC method gives better discrimination when the number of variables is bigger as proven in the case of oil and gas pipes [1].

Scenario no. 25 remained to be the best regarding the Wilk's Λ, CCC and EF_s levels, while scenarios no. 21 and 23 are the best regarding EF_f and EF_t levels (scenario no. 25 followed). The non-standardized coefficients (U_i) are those used for the calculation of the Z-score for both groups of pipes (Table 9, Fig. 3). Negative non-standardized coefficients (U_i) were related to the variables DIM1, soil and to the pipe characteristics (diameter; material; age; length; and pressure). On the other hand, positive non-standardized coefficients were related to the following variables: previous failures; external loads; and DIM2. The pipe failure probability assessment process using the time step analysis faced a number of difficulties due to the format of the available data records regarding the pipe characteristics. The pipes that actually failed were predicted with satisfying accuracy. If the format of the available data records was similar to the respective one met in oil/gas pipe networks then the correct prediction percentages would have been even better.

5. Conclusions

As mentioned in previous studies [1,2] the DAC method proved once more to be a useful tool for pipe reliability prediction, since it considers different and complex pipe characteristics. It examines how each characteristic is affecting whether the pipe will fail or not and provides a reliable risk assessment model. The method gave excellent results when used to predict the future behavior of oil/gas pipes, specifically the correct prediction probabilities reached 96.6%. Its application in water pipe networks had to overcome several problems caused mainly due to the poor pipe failure data records available from the water utilities. The definition of the failure criterion was another obstacle. The Larissa water

Table 8
Results for the time step analysis

Scen.	Important variable	Less important variable	Unstable variable	EFf	EFs	EFt	Wilk's Λ	R
21	Previous failures	Age	D, BR, MAT, AG, DIM1, DIM2	85.4%	63.0%	77.6%	0.744	0.506
23	Previous failures	Soil	D, BR, MAT, AG, DIM1, DIM2, S	85.4%	63.0%	77.6%	0.744	0.506
25	Previous failures	Soil	D, BR, MAT, AG, DIM1, DIM2, S	82.5%	67.1%	77.1%	0.721	0.528

Table 9
Non-standardized coefficients for the time step analysis

	Scen. 21	Scen. 23	Scen. 25		Scen. 21	Scen. 23	Scen. 25
U_i (diameter)	–0.02385	–0.02371	–0.02405	U_i (DIM1)	–0.007314	–0.007337	–0.005829
U_i (failures)	0.06605	0.06582	0.06732	U_i (DIM2)	0.052930	0.054740	0.051950
U_i (material)	–0.04761	–0.04587	–0.04796	U_i (soil)		–0.161700	–0.128600
U_i (age)	–0.01180	–0.01363	–0.01309	U_i (pressure)	–3.453000	–3.453000	–2.344000
U_i (length)	–0.05966	–0.05996	–0.05734	U_i (loads)			1.441000

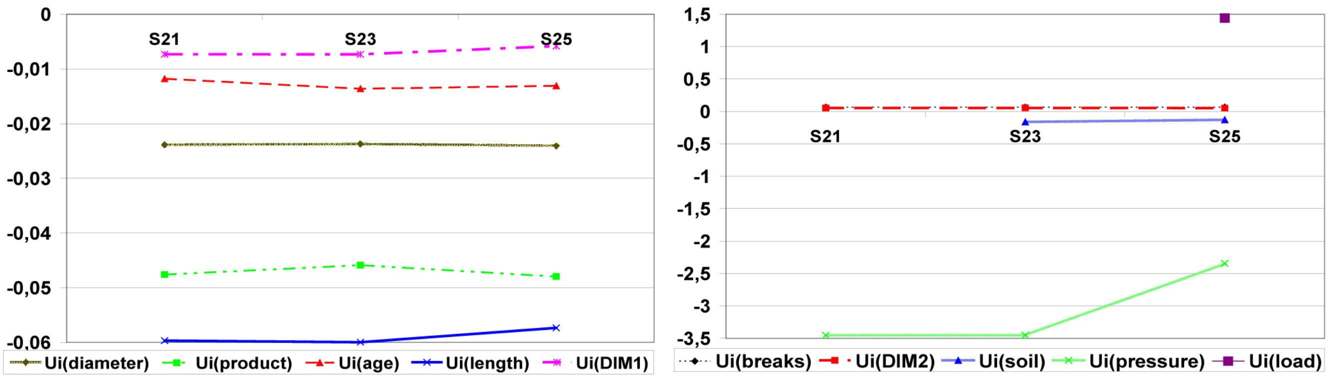

Fig. 3. Non-standardized coefficients for the time step analysis (first 9-year period).

distribution pipe network was used as a case study. The Utility's data failure records were poor and not fully compatible to the form the DAC method demands. In order to overcome this problem and implement the DAC method some assumptions were made and dummy variables were used. In order to classify pipes into 'failures' and 'successes', the total water loss volume criterion was used. The pipes facing leaks were considered as 'failed ones', while those facing breaks were considered as 'successes'. From the analysis and the results obtained it is found that the DAC method can be successfully used for risk assessment in water distribution networks when the data available meets the requested standards. Results are expected to be even better if additional data is available regarding the type of soil, the traffic loads, neighboring constructions and other pipe failure generating causes [28]. Water Utilities should develop systems in order to monitor their networks regarding pipes parameters and also parameters connected to the network in general. Many studies [1,7,8,13,29,30] show that pipe reliability estimation models are getting better and better when full and complete data records exist.

Glossary

AG	—	Pipe age, y
BR	—	Previous breaks, m
D	—	Pipe diameter, mm
DIM1-2	—	Pipe joint variables
EF_f, EF_s	—	Classification indicator regarding the discrimination ability related to each group (failures, successes)
EF_t	—	Classification indicator regarding the discrimination ability over all populations
L	—	Pipe length, m
LO	—	External load

MAT — Pipe material
PRES — Pipe operating pressure, atm
R_i — Canonical correlation coefficient
U_i — Best discriminant coefficient for the "*i*" variable
U_o — Adjustment for the mean values so that the mean discriminant score equals to zero over all cases
S — Soil type
X_i — Value of the sample's characteristic
X_{im} — Value of the "i" variable
Z_m — Value (score) of the discriminant function for case m
Λ — Wilk's Λ

References

[1] S. Tsitsifli and V. Kanakoudis, Predicting the behavior of a pipe network using the "critical z-score" as its performance indicator, Desalination, 250 (2010) 258–265.
[2] Y. Bakouros, Offshore Pipeline Prediction. PhD Thesis, University of Bradford, 1988.
[3] U. Shamir and C. Howard, An analytical approach to scheduling pipe replacement, J. AWWA, 71 (1979) 248–258.
[4] D.K. O'Day, C.M. Fox and G.M. Huguet, Aging water distribution systems: A computerized case study, Public Works, 3(8) (1980) 61–111.
[5] R. Clark, C. Stafford and J. Goodrich, Water distribution systems: A spatial and cost evaluation, Wat. Resources Planning Management, ASCE, 108 (1982) 243–256.
[6] A. Kettler and I. Goulter, An analysis of pipe breakage in urban water distribution networks, Canad. J. Civil Eng., 12(2) (1985) 286–293.
[7] D.K. O'Day, Organizing and analyzing leak and break data for making main replacement decisions, J. AWWA, 74(11) (1982) 589–596.
[8] R. Clark and J. Goodrich, Developing a data base on infrastructure needs, J. AWWA, 87(1) (1989) 81–87.
[9] D. Marks, S. Andreou and C. Park, Predicting urban water distribution maintenance strategies: a case study of the New Haven, Connecticut, EPA Draft Report, 1985.
[10] S.A. Andreou, D.H. Marks and R.M.Clark, New methodology for modeling break failure patterns in deteriorating water distribution systems: Applications, Adv. Water Resources, 10 (1987) 2–10.
[11] I. Goulter and A. Kazemi, Spatial and temporal groupings of water main pipe breakage in Winnipeg. Transport Eng., ASCE, 115(2) (1988) 95–111.
[12] Y. Kleiner, B.J. Adams and J.S. Rogers, Long-term planning methodology for water distribution system rehabilitation, Wat. Resources Res., 34(8) (1998) 2039–2051.
[13] Y. Kleiner and B. Rajani, Using limited data to assess future needs, J. AWWA, 91 (1999) 47–61.
[14] T.D. Prasad, S.-H. Hong and N. Park, Reliability based design of water distribution networks using multi-objective genetic algorithms, KSCE Civil Eng., 7(3) (2003) 351–361.
[15] A. Vanrenterghem-Raven, P. Eisenbeis, I. Juran and S. Christodoulou, Statistical modeling of the structural degradation of an urban water distribution system: Case study of New York City, World Water & Environmental Resources Congress and Related Symposia, Pennsylvania, USA, 2003.
[16] P. Aslani, Hazard rate modelling and risk analysis of water mains, MSc. thesis, 2003.
[17] S. Christodoulou, P. Aslani and A. Vanrenterghem, A risk analysis framework for evaluating structural degradation of water mains in urban settings, using neurofuzzy systems and statistical modeling techniques, World Water & Environmental Resources Congress and Related Symposia, Pennsylvania, USA, 2003.
[18] S. Christodoulou, P. Aslani and A. Deligianni, Integrated GIS-based management of water distribution networks, Conference on Computing and Decision-Making in Civil and Building Engineering. Montreal,.Canada, ASCE, 2006.
[19] V.K. Kanakoudis and D.K. Tolikas, Assessing the Performance Level of a Water System. Water, Air, Soil Pollut.: Focus, 4(4) (2004) 307–318.
[20] S. Park, B.J. Kim and G.C. Im, Modeling of water main failure rates using the log-linear ROCOF and the power law process. Wat. Resources Manage., 22 (2008) 1311–1324.
[21] S. Christodoulou, Water network assessment and reliability analysis by use of survival analysis, EWRA 7th Conference, Limasssol, Cyprus, 2009, pp. 503–510.
[22] Y. Bakouros, Predicting pipeline reliability using discriminant analysis, 10th Advances in Reliability Technology Symposium, G.R. Libberton, ed., Elsevier Applied Science, 1988.
[23] V. Kanakoudis and D. Tolikas, The role of leaks and breaks in water networks – Technical and economical solutions, Wat. Supply Res. Techno. - AQUA, 50(5) (2001) 301–311.
[24] I. Goulter, J. Davidson and P. Jacobs, Predicting water-main breakage rates, Water Resources Planning Manage., ASCE, 119(4) (1993) 419–436.
[25] S. Christodoulou and A. Deligianni, A neurofuzzy decision framework for the management of water distribution networks, Wat. Resources Manage., DOI 10.1007/s11269-009-9441-2, 2009.
[26] S. Christodoulou, A. Deligianni, P. Aslani and A. Agathokleous, Risk-based asset management of water piping networks using neurofuzzy systems. Computers, Environ. Urban Systems, 33 (2009) 138–149.
[27] US EPA, Decision-Support Tools for Predicting the Performance of Water Distribution and Wastewater Collection Systems, Vol. EPA/600/R-02/029, Washington, D.C. USA, US EPA, 2006.
[28] V. Kanakoudis, A troubleshooting manual for handling operational problems in water pipe networks, Wat. Supply: Res. Technol.-AQUA, 53(2) (2004) 109–124.
[29] R.A. Luksa, The why, what and how of maintaining system records, J. AWWA, 72(9) (1980) 496–550.
[30] A. Stourm, Optimization of reliability prediction using joint variables, Thesis, Thessaly University, 1997.