A Spatial probabilistic model of pipeline failures

István Bogárdi / *Roland* Fülöp

Abstract

A spatial probabilistic model is developed and applied to predict failure patterns over a water supply pipe network. The Zalaegerszeg waterworks is used to illustrate the methodology. The ageing of water supply pipes is a worldwide concern; therefore it is of high interest to identify the best interventions in time and space. Two-part models can be used; one is an economic by formulating time horizon total costs. The other part forecasts the break patterns in time and space. Spatial modeling is necessary since elements of the cost function depend on the location of failures. A network based model is developed for the internal part of the Zalaegerszeg waterworks since the system is homogeneous in a sense that the environmental and network features do not exhibit significant correlation with the occurrence of failures. Network based failure probabilities are described by a space-time Poisson process where non-homogeneous Poisson process (NHPP) refers to time and a stochastic point process refers to space. To estimate the intensity the whole area is covered by a grid system of 250 × 250 meters with an average pipe length of 703 meter in a cell. Both the number of and the distance between failures prove the applicability of the Poisson process. It is shown how a simulation procedure can be applied to generate possible pipe failure patterns which - by combining with the spatial cost functions – leads to estimate total costs per selected time periods.

Keywords

Spatial probabilistic model · spatial-time Poisson · ROCOF function · failure prediction · water network · generating failure patterns

István Bogárdi

Department of Civil Engineering, University of Nebraska–Lincoln, Lincoln, NE, USA; Department of Meteorology Eötvös Loránd University, Budapest, Hungary
e-mail: ibogardi1@unlnotes.unl.edu

Roland Fülöp

Department of Sanitary and Environmental Engineering, Budapest University of Technology and Economics, 1111, Budapest, Műegyetem rkp. 3, Hungary
e-mail: froland@vkkt.bme.hu

1 Introduction

The purpose of the paper is to develop and apply a spatial probabilistic model in order to predict failure patterns over a water supply pipe network. The ageing of water supply pipes is a worldwide concern. In Hungary the average annual failure rate is 38 breaks per 100 km [1] which is similar to 36/100 in the USA [2]. The specific failure rate varies considerably in European cities (Fig. 1) [3].

The break history generally shows an increasing trend, e.g. in the town of Zalaegerszeg used to illustrate the methodology in this paper (Fig. 2).

In the U.S. the costs for upgrading the nation's aging pipe infrastructure is estimated $77 billion over the next 20 years, that is $55000/km [4]. This figure is $117000/km in Hungary [5].

Water utility companies generally have limited amount of available resources for the maintenance, repair and gradual replacement of the ageing water mains. Thus, it is of high interest to identify in time and space the best intervention actions over their system. The best intervention actions commonly aims at minimizing the total discounted costs over the planning horizon. Often, are these intervention actions realized using past experiments, limited available amount of measured information and expert judgment. As a consequence, the actual interventions may considerably deviate from the unknown best interventions. To this end, much effort has been done to develop models to facilitate the identification of the best interventions. However, the lack of information required by many of the models hinders their wide application.

Models helping water companies should have two integrated parts. One is an economic by formulating a cost function that summarizes time horizon costs of break repairs, losses due to breaks, pipe replacements and lack of service. The other part forecasts the break patterns in time and space. Spatial modeling is necessary since elements of the cost function depend on the location of failures. The model presented in this paper will forecast the break patterns in space using commonly available company-based information.
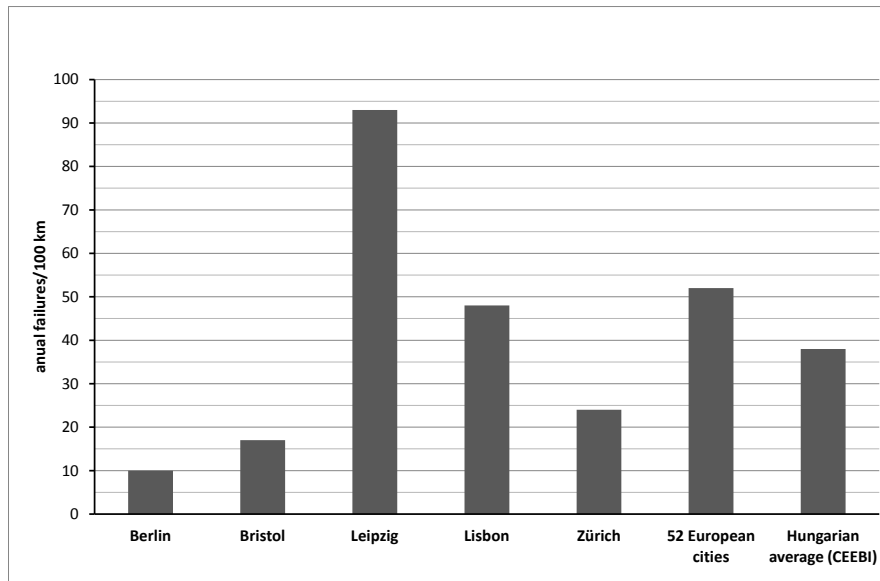
Fig. 1. Specific failure rates of water supply pipes in European cities
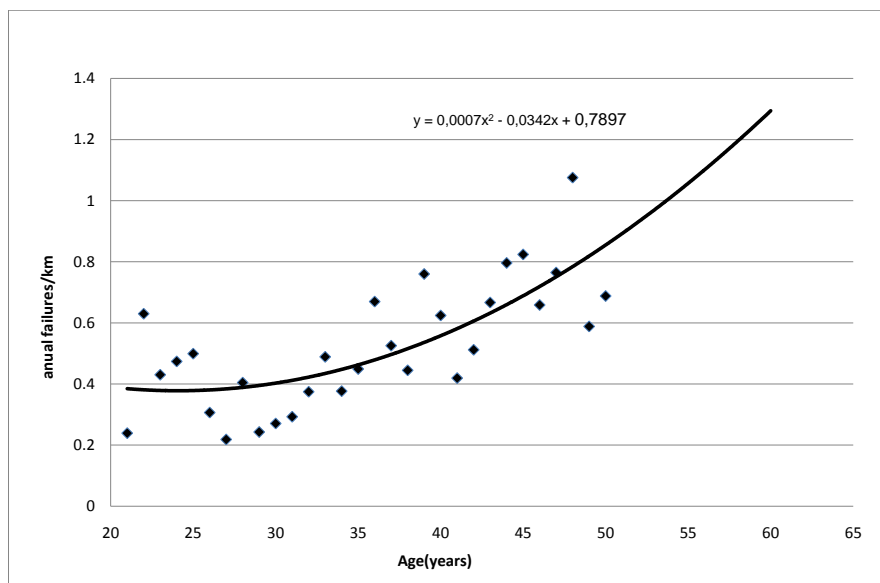

Fig. 2. Relationship with age and specific pipe failure rate (ROCOF) in Zalaegerszeg

2 Pipeline Failure Models

Four main groups of pipeline failure models can be distinguished:

- Expert based point assignment methods

- Mechanistic models

- Regression/neural network models

- Failure probability models

Expert based point assignment methods are popular in asset management [6], and can be used with data commonly available at the waterworks. Earlier the waterworks of Budapest developed a scoring system for reconstruction planning [7]. This company nowadays also applies a fuzzy logic method for reconstruction planning [8]. This software operates with a large data demand because it works with 58 input data field, and suitable for ranking the investigated pipe sections based on their risk. The advantages of the expert based models are the relatively low data demand, and their easy applicability. However, subjective

factors may limit the accuracy of the results.

The base of mechanistic models is a comparison between the overall load (e.g. traffic load, pressure, etc.) and the strength/capacity of the pipe, either or both formulated as random variables leading to the probability of failure [9]. The data demand of these models is quite high, since statistical information on the parameters influencing load/strength are rarely available (e.g. the bedding factor, the exact traffic and soil load, the decrease of wall thickness, etc. [10]). To this end, pipe material inspections may be necessary for verification. Pipe inspection may be in-situ (e.g. acoustic investigation) or laboratory inspection. These inspections are quite expensive in water supply systems. Consequently, pipe inspections alone are inappropriate for reconstruction planning for the whole system. Thus, in most cases mechanistic models can be used for pipe sections or for major transmission water mains, and not for the whole water supply system.

Regression models describe the relationship between failures or failures rates and time [11] or several influencing factors. Commonly used influencing factors include network properties (material, diameter, age, pressure, water quality), environmental (soil, groundwater) and external loads (traffic, bedding factor, depth of the laying) [12–14]. These models are relatively easy to connect with an economic model but they need much data and the relationship may be quite weak [15]. Artificial neural networks may be more appropriate since they detect the best possible relationship obtained from a learning data set [16].

Failure probability models do not intend to investigate the physical processes causing failures but use statistical data to estimate the probability of the time between failures, the location of failures or the number of failures per time (e.g. years). In this respect, a water supply networks is considered as repairable system which, after failure can be restored to performing all of its functions by any method (say repair), other than replacement of the entire system. The model, which is mostly used with repairable systems, is the NHHP to take changing failure rates in time into consideration. The basic parameter of NHPP is the rate of occurrence of failures (ROCOF) as shown in Fig. 2. Then, in a given time period the probability of the number of failures can be represented by a Poisson distribution with a mean of the average ROCOF for that period.

Failure probability models have two main groups: the first considers individual pipe sections (commonly called pipes) while the other endeavors to cover the whole network. The pipe based models work with homogenous pipes (same environmental and network features) and applies ROCOF function to each of them. However, at least five historical failures per pipes are necessary to estimate the ROCOF functions [17]. So many known failures rarely happen in practice. As a consequence pipe based models can applied mostly to a small part of the whole network [18]. Network based models can be used if the investigated system is homogeneous in a sense that the environmental and network features do not exhibit significant correlation with the occurrence of failures. The case of Zalaegerszeg has this characteristic. Network based failure probabilities then may be described by a space-time Poisson process where NHPP refers to time and a stochastic point process refers to space.

3 Case of Zalaegerszeg Waterworks

The city of Zalaegerszeg in Hungary has 60 000 inhabitants. The development of Zalaegerszeg was significant during the last century. It was influenced by the inhabitants, location, industry of the town and also by its role as a cultural center similar to other European cities [19]. The predecessor of ZALAVIZ Waterworks Company founded in 1959 began water supply that year. Most conduits were built of asbestos cement until 1980's. The length of the water distribution system is about 292 km with diameters of 80-400 mm. The water supply in the town is close to 100%. The average supplied water is $11000\,\mathrm{m}^3$/day. This waterworks is interested in learning how the reconstruction budget in Zalaegerszeg can be spent effectively in time.

The internal part of the city without the suburbs (Fig. 3) was selected for the investigation because the pipes (70 km.) are quite homogenous (material, age, diameter, etc.), and the spatial distribution of the conduits on the assigned area is uniform. The ratio of the asbestos cement pipes is above 70%. The size of investigated area is $5,875\,\mathrm{km}^2$.
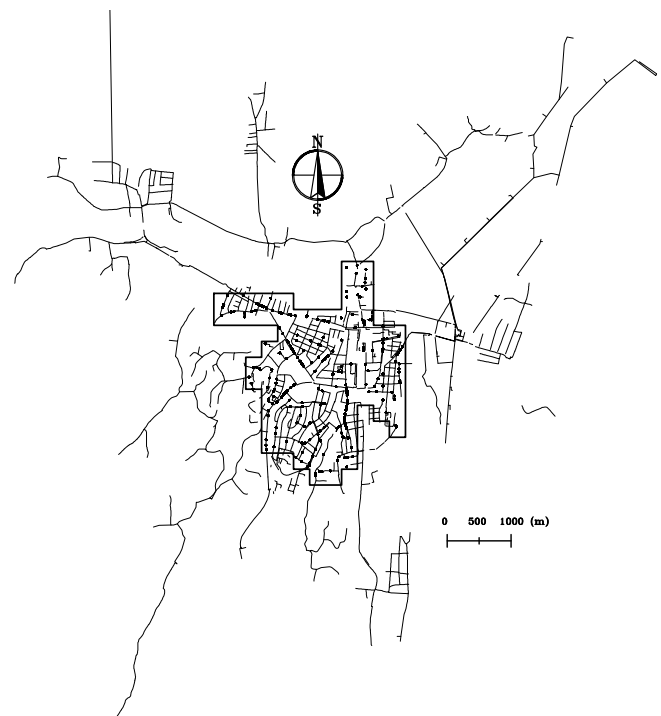

Fig. 3. The boundary of the investigated area

The following data have been available from the technical information system of ZALAVIZ:

- pipes (coordinates, material, diameter, year of construction, length)

- failures (year, coordinates)

- hydraulic data from HCWP 6.1 [20]

219 failure data are available for six complete years (2004-2009). This short data time series is not suitable for classical statistical analysis. The average annual value of failure rate is 54,6 /100 km. ZALAVIZ has not collected technically relevant information about the environment (soil, traffic, corrosivity, etc.) The environmental and traffic data was obtained from other sources. Soil type information was provided by the Geological Institute of Hungary [21]. The typical soil in Zalaegerszeg is clay.

The regression analysis between the actual environmental and network factors and the occurrences of failures did not result in any significant correlations (r < 0.2). Similarly, the neural network could not show any casual relationship.

4 The Spatial Probabilistic Model

The purpose of this modeling effort is to describe the location of failures stemming from a spatial point process over the investigated pipe network. A spatial point process is a random pattern of points in d-dimensional space. Spatial point processes are useful as statistical models in the analysis of observed patterns of points, where the points represent the locations of some object of study, here pipe failures [22]. One of most practically applied spatial point process is the homogeneous spatial Poisson process [23, 24]. Let the parameter S denote a bounded region of the plane (or higher- dimensional space, for that matter). Let X(S) be the number of entities, here failures contained in S. Then X(S) is a homogeneous spatial Poisson process if it obeys the Poisson postulates, yielding a probability distribution:

$$P\{X(S) = k\} = \frac{[\lambda A(S)]^k e^{-\lambda A(S)}}{k!}$$
$$\text{for} \quad A(S) \geq 0, k = 0, 1, 2, \cdots \tag{1}$$

In this case λ is a positive constant called the *intensity parameter* of the process and A(S) represents the area or volume of S, depending on whether S is a region in the plane or higher-dimensional space. In the present case, d = 2 and so A is the area of a regular grid of the pipe network.

The underlying mathematical preconditions of the homogeneous spatial Poisson process are satisfied:

1 Only nonnegative integer values are assumed by $X(S)$ and $0 < P\{X(S) > 0\} < 1$ if $A(S) \geq 0$

2 The probability distribution of $X(S)$ depends on S only through the value of $A(S)$ with the further property that if $A(S) \rightarrow 0$, then $P\{X(S) \geq 1\} \rightarrow 0$

3 If $S_1, S_2, ...S_n(n \geq 1)$ are disjoint regions, then $X(S_1), ..., X(S_n)$ are mutually independent random variables and $X(S_1 \cup ... \cup S_n) = X(S_1) + ... + X(S_n)$

4 There is a probability zero of points overlapping:

$$\lim_{A(S) \rightarrow 0} \frac{P\{X(S) \geq 1\}}{P\{X(S) = 1\}} = 1 \tag{2}$$

To estimate the intensity the whole area is covered by a grid system of 250×250 meters cell sizes with an average pipe length of 703 meter in the 101 cells. The average ROCOF of 2004-2009 (age 31.7-37.7) is 0.5 failures/km/year (Fig. 2). Thus, the intensity: $\lambda = 204/101 = 2.01$ failures/cell. Fig. 4 shows the relative frequency distribution and the calculated Poisson probabilities. The expected distance between failures can be calculated as

$$\frac{1}{\lambda} = \frac{\sqrt{\frac{A_{inv}}{N}}}{2} = 97\,\text{m} \tag{3}$$

where:

- A_{inv} - investigated area (m^2)

- N - total number of failures

Fig. 5 showing the cumulative frequency distribution and the exponential distribution with the calculated expected distance proves that the Poisson assumption works.

Next, the homogeneous spatial Poisson process is validated by doubling the areas of the cells. To this end, Eq (1) is used with parameter $2\lambda = \textbf{4.02}$ failures/two cells to calculate a Poisson distribution and compare it with the corresponding relative frequencies. Fig. 6 suggests the validity of the assumption that the location of failures can be modeled as a homogeneous spatial Poisson process.

Using the available 6 year spatial failure data the spatial stationarity of the Poisson process can be checked by comparing the relative frequencies of failures stemming from 2004 to 2006 and those from 2007-2009. Fig. 7 shows that the two distributions are very close to each other. The present random spatial failure occurrence may be used for network rehabilitation planning until there are no significant changes in the environmental factors discussed earlier.

5 Generating Failure Patterns

Generation of possible future failure patterns based on the above spatial probabilistic model can assist waterworks with network restoration planning and failure scenario analysis. Consider the total cost of break repairs, losses due to breaks, pipe replacements and lack of service in a selected future time period t as $C(t)$. This cost depends on the number X and the location of failures called here as failure pattern. The number of failures follows a NHPP in time characterized by the ROCOF function (Fig. 2), while failure pattern is described by the above homogeneous spatial Poisson process. Thus $C(t)$ can be considered as an expected value obtained by double expectation:

- first, over the possible failure patterns for a generated number of failures

- next, over the NHPP based Poisson number of failures.

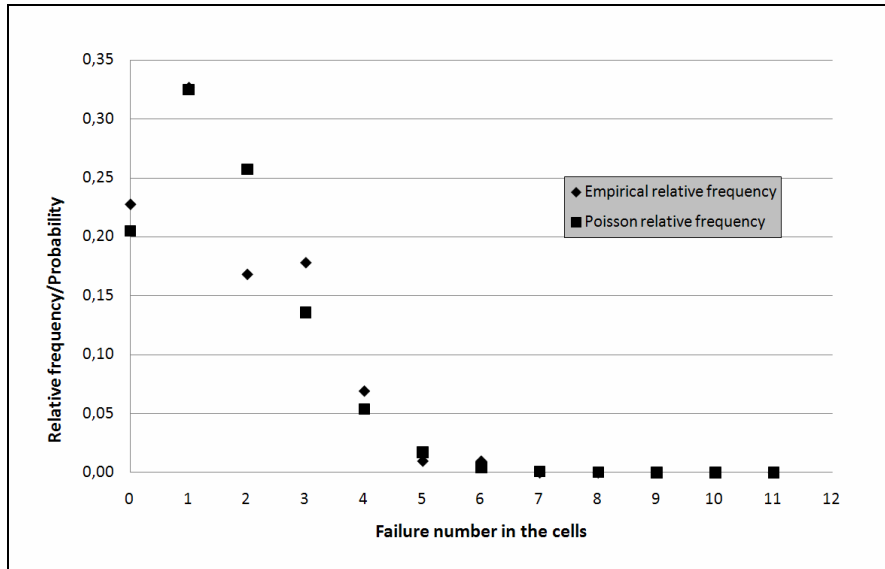The expected $C(t)$ can be obtained by the following procedure.

Fig. 4. Relative frequency and probability of failures in the investigated area
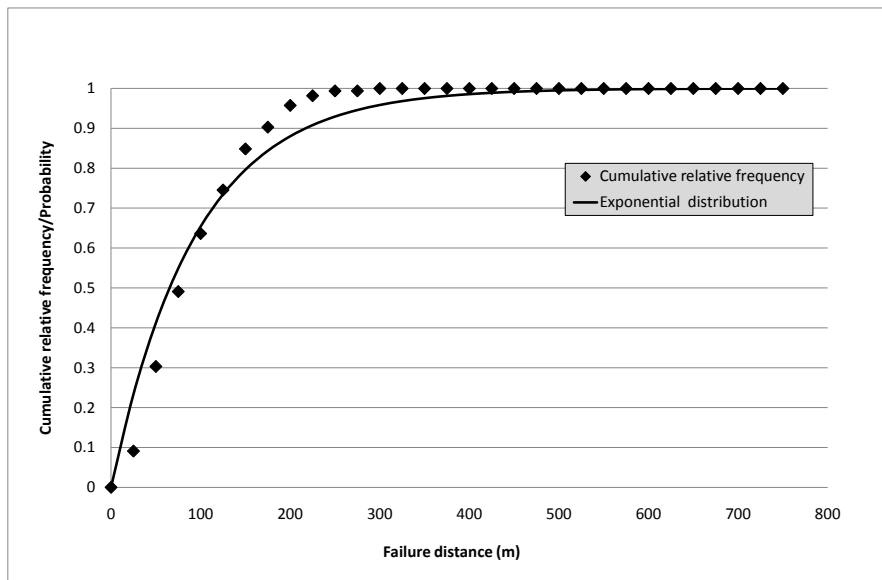

Fig. 5. Cumulative relative frequency and distribution of failures distance in the investigated area

1 Simulate a Poisson number of failures covering the whole region for a selected period (annual, five-year) As an example, consider the period 2016-2020 (age 43.7-48.7) represented by an average ROCOF value of 0.73 failure/km/year (Fig. 2). Using this value as the expected value a possible failure number of $i = 301$ is generated with the Poisson distribution. The next steps involve the simulation of spatial patterns with a given intensity over a region [22].

2 Scatter that number of failures uniformly over the analyzed region.

3 Generate a random radius D between a uniformly scattered point and its nearest neighbor (a possible failure location)

from the distribution function:

$$F_D(x) = P(D \le x) = 1 - e^{-\lambda \pi x^2} \qquad (4)$$

In the present case $\lambda = 0.015$ 1/m. Next, find a possible failure location where D reaches a pipe. If D crosses several pipes select one of them randomly (Fig. 8).

4 Perform step 3. for every failure point (obtained in step 2) to obtain one possible failure pattern, j. A possible generated failure pattern is on Fig. 9. Fig. 10 shows that failure distances are fairly well represented by the theoretical exponential distribution even for this single generated failure pattern. Given the cost consequences as functions of the failure locations the total cost C(i,j,t) can be calculated.
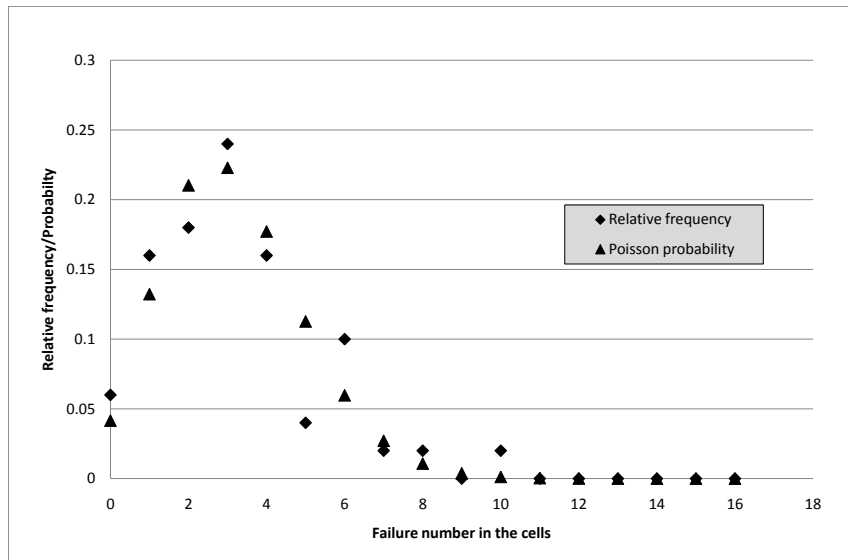
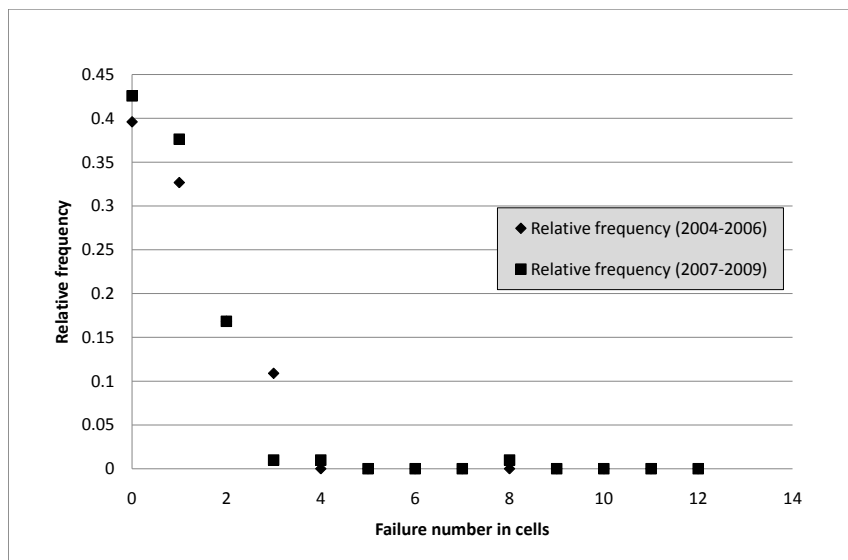**Fig. 6.** Relative frequency and probability of failure numbers in double size cells

Fig. 7. Relative frequency of failure numbers observed in two periods

5 Repeat steps 3-4 n-times to obtain other possible failure patterns and estimate the expected costs over all generated failure patterns as $C(i,t) = \sum C(i,j,t) /n$

6 Repeat steps 1-5 m-times to obtain possible failure patterns for all simulated number of failures and estimate the double expected costs as $C(t) = \sum C(i,t) /m$

6 Discussion and Conclusions

A network based spatial probabilistic model is presented and applied to a medium size city to forecast failure locations of water supply mains. This homogeneous Poisson process is suitable for failure forecasting if short term historical failure data are available for reconstruction planning. Network based probabilistic modeling can be used for cases when the following conditions are satisfied:

- uniform pipe distribution over area,

- similar features of the pipes (size, material, age, pressure)

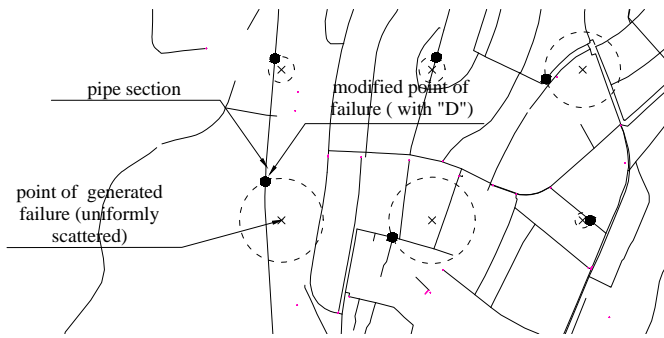- similar environmental features,

- known failure locations

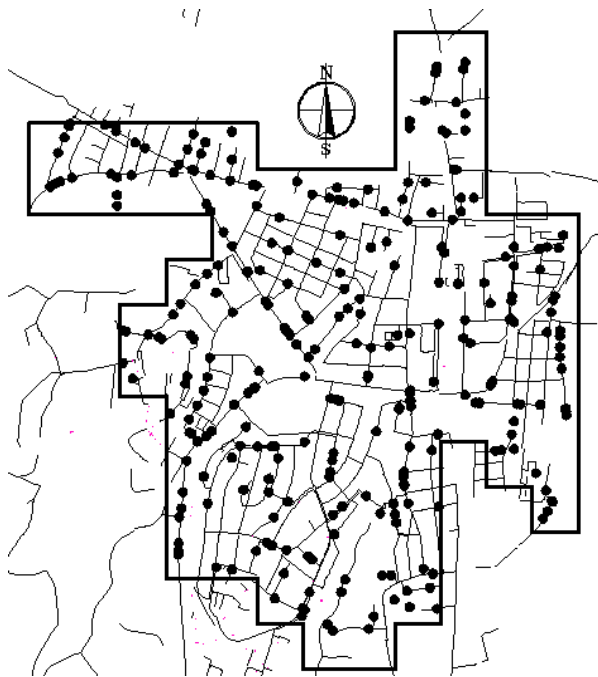**Fig. 8.** Method of generating failure points

Fig. 9. A generated failure pattern for the period 2016-2020

- minimum 5 year long failure data.

If these criteria are not met the network should be divided into and analyzed according to individual pipes without a formal spatial analysis. However, detailed pipe based failure data are necessary in this case. Often a non-homogeneous Poisson process based on a ROCOF function describes the statistical properties of failures in time. Pipe-based models require specific ROCOF function /pipe. On the other hand, network based models can be used with one ROCOF function but a joint space-time modeling is necessary as shown in this paper. Total life time cost can be obtained by summarizing time period costs calculated according to the six step procedure over the planning horizon.

Probabilistic models, in general, assume that the future behavior of the system will correspond statistically to the observational data upon which parameters of probabilistic models are estimated. Here, specifically, it is assumed that the ROCOF function will stay the same in the future and the spatial Poisson process will be maintained.

The following conclusions can be drawn:

1 The spatial probabilistic model presented in this paper can be used to forecast future failure patterns over water supply mains.

2 A main precondition of using the model is homogeneous environmental and pipe network features.

3 The spatial model is a homogeneous Poisson process that can be connected to a non-homogeneous Poisson process describing pipe failure changes in time.

4 Water main reconstruction planning requires cost estimation of break repairs, losses due to breaks, pipe replacements and lack of service. Spatial modeling is necessary since elements of the cost function depend on the location of failures.

5 The space-time model enables us to calculate to any future time period the total costs expressed as an expected value over space and time.

6 In the investigated Zalaegerszeg system the environmental and network features do not exhibit significant correlation with the occurrence of failures.

7 The homogeneous spatial Poisson process could be calibrated and validated with the available 219 failure data stemming from six years.

8 It is shown how a simulation procedure can be applied to generate possible pipe failure patterns which - by combining with the spatial cost functions – leads to estimate total costs per selected time periods.

References

1 *Data Collection in Hungary*, Central and Eastern European Benchmarking Initiative for Water Supply Utilities, 2006.

2 **American Water Works Service Co. Inc. Engineering Department**, *Deteriorating buried infrastructure management challenges and strategies*, American Water Works Association (AWWA), 2002, http://www.epa.gov/ogwdw000/disinfection/tcr/pdfs/whitepaper_tcr_infrastructure.pdf.

3 **Sægrov S.**, *Ageing of Water and wastewater networks International status and responding strategies*, Sintef, 2007.

4 **Selvakumar A., Clark R. M., Sivaganesan M.**, *Costs for water supply distribution system rehabilitation*, Journal of Water Resources Planning and Management **128** (2002), no. 4, 303-307.

5 **Hungarian National Water Technology Platform**, *Strategic Researches Plan second Edition*, Hungarian National Water Technology Platform, 2009, http://www.nvp.hu/sites/default/files/NVP_skt_2_kiadas.pdf.
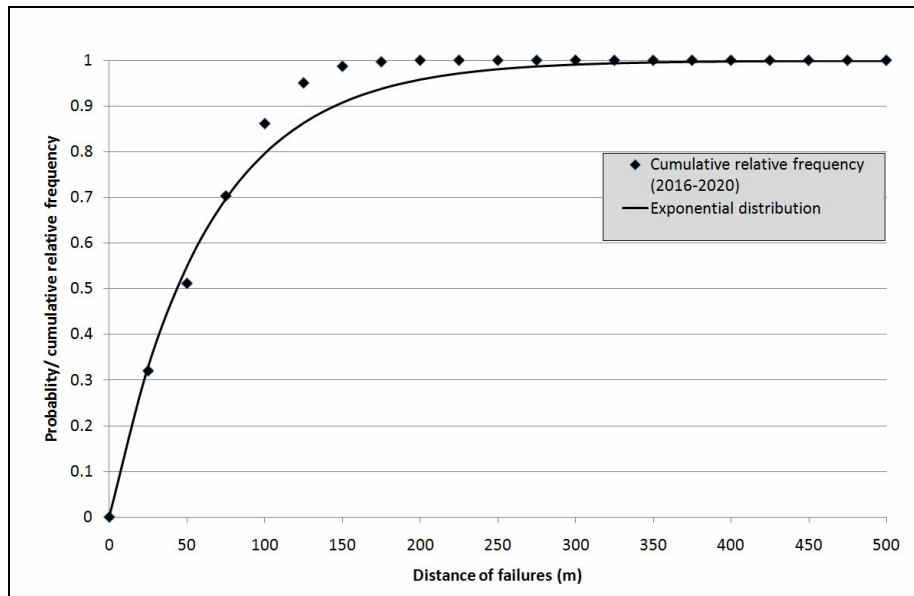
Fig. 10. Cumulative relative frequency and distribution of failures distances for a generated failure pattern

6 **Barata P. M.C., Alegre H., Vieira J.M.P.**, *Application of a DPA method for asset management in small water distribution system journal=Water Asset Management International, date=2008, volume=4, pages=2–7, number=4*.

7 **Laky D., Somlyódy L., Darabos P., Honti M.**, *Introduction of the risk analysis model of Water Works of Budapest in Hungarian*, 2004. CD Proceedings.

8 **Hetényi Zs., Zimmer P., Tolnai B.**, *GIS model based on fuzzy logic in Hungarian*, Vízmű Panoráma **14** (2006), no. 5, 9–11.

9 **Pampoukis G., Camarinopoulos L.**, *Reliability of a water supply network*, PSAM 7 - ESREL '03 Conference (1996), Probabilistic safety assessment and management, Spinger, 1996, pp. 1872– 1877.

10 **Davis P., De Silva D., Marlow D., Moglia M., Gould S., Burn S.t**, *Failure prediction and optimal scheduling of replacements in asbestos cement water pipes*, Journal of Water Supply: Research and Technology–AQUA **54** (2008), no. 4, 239–252, DOI doi:10.2166/aqua.2008.035.

11 **Shamir U., Howard C.D.D.**, *An analytic approach to scheduling pipe replacement*, Journal AWWA **71** (1979), no. 5, 248–258.

12 **Rostum J., Dören L., Schilling W.**, *Deterioration of Built Environment: Buildings, Roads and Water Systems*, Norwegian University of Science and Technology, may 1997. IVB-report B2-1997-2, ISBN 82-7598-040-2.

13 **Andreou S.**, *Predictive models for pipe break failures and their implications on maintenance planning strategies for deteriorating water distribution systems*, PhD thesis, Massachusetts Institute of Technology Cambridge, 1986.

14 **Bartos S., Meszaros P., Solti D.**, *Reconstruction of the water and sewer systems*, Műszaki Könyvkiadó, 1989.

15 **Clark R. M., Stafford C. L., Goodrich J. A.**, *Water distribution systems: A spatialand cost evaluation*, Journal of the Water Resources Planning and Management Division **108** (1982), no. 3, 243–256.

16 **Achim D., Ghotb F., McManus K. J.**, *Prediction of water pipe asset life using artificial neural networks*, Journal of infrastructure systems **13** (2007), no. 1, 26–30.

17 **Park S., Jun H., Kim B. J., Im G. C.**, *Modeling of Water Main Failure Rates Using the Log-linear ROCOF and the Power Law Process*, Water Resource Management **22** (2008), no. 9, 1311–1324, DOI DOI: 10.1007/s11269-007-9227-3.

18 **Rostum J.**, *Statistical modeling of pipe failures in water networks*, PhD thesis, University of Science and Technology, 2000.

19 **Ládai A. D.**, *Historical development of Kosice (Kassa) and its region*

– *geoinformatical analysis*, Periodica Polytechnica Civil Engineering **52** (2008), no. 1, 35–38, DOI doi: 10.3311/pp.ci.2008-1.05.

20 **Darabos P.**, *Software package for controlling water supply systems*, Periodica Polytechnica Civil Engineering **41** (1997), no. 2, 135–145.

21 **Geological Institute of Hungary**, available at `http://mafi-loczy.mafi.hu/metadataexplorer`. last visited: October 2009.

22 **Baddeley A, Turner R.**, *spatstat: An R Package for Analyzing Spatial Point Patterns*, Journal of Statistical Software **12** (2005), no. 6, 1–42.

23 **Larson R.C., Odoni A.R.**, *Urban Operations Research*, Prentice- Hall, 1999.

24 **Møller J., Waagepetersen R. P.**, *Statistical inference and simulation for spatial point processes*, Chapman & Hall/CRC, 2004.