Infrastructure aging risk assessment for water distribution systems

Wilmer P. Cantos and Ilan Juran

Wilmer P. Cantos
Ilan Juran (corresponding author)
New York University Tandon School of Engineering,
Urban Infrastructure Institute,
6 MetroTech Center, Brooklyn, NY 11201,
USA
E-mail: *ijuran@nyu.edu*

ABSTRACT

Metropolitan governments and water operators are continuously facing the ever-growing challenges of evaluating the risks and optimizing investment in the rehabilitation of the buried aging infrastructure of water distribution systems (WDS). Proper asset management and efficient rehabilitation planning require monitoring, condition assessment, degradation risk analysis and a data-based model for degradation forecasting to support investment decision-making and significantly reduce the infrastructure rehabilitation cost. This paper presents a statistical and stochastic spatial data analysis of failure records of the WDS of the City of Wattrelos, France. The research objective is to develop and demo-illustrate the application of an operator's experience-based Risk Assessment Method (RAM) for network micro-zone prioritization of rehabilitation/replacement works to optimize preemptive asset management. The data used is a 74-year historical dataset from Wattrelos, France. The database includes approximately 424 observed failures for the period of 1991–2004. The data analysis demonstrates that understanding and using stochastic modeling to characterize the pattern of relationship between Failure Rate (F_R), Age (T) and the Probability (or Risk) of exceeding a specific Failure Rate ($P_r(F_R)$) of a micro-zone can effectively support the operator's assessment, risk management and prioritization in the maintenance and rehabilitation of the WDS.

Key words | aging infrastructure, assets management, case study, life-cycle modeling, risk assessment, water distribution system

INTRODUCTION

Beginning in the 1980s, a number of attempts have been made to statistically model the failure risk of water pipelines (Vanrenterghem-Raven *et al.* 2004; Jafar 2006; Jafar *et al.* 2007, 2010; Alegre *et al.* 2011; Alegre & Coelho 2012; Mamo 2013; Mamo *et al.* 2013, 2014). The development of models for the prediction of failure in the water distribution system (WDS) encounters major difficulties, mainly due to the lack of data on both the WDS and pipe breakage history (Jafar *et al.* 2010; Alegre & Coelho 2012). The assessment of pipe system degradation requires information on the WDS including physical, environmental, and operational parameters that have impacted the pipe failure rate (FR) during its life-cycle (Le Gat & Eisenbeis 2000; Alegre &

Coelho 2012). Pipe system aging modeling has included statistical models for predicting water pipe failure using historical data, survival data analysis and stochastic models (Jafar 2006; Jafar *et al.* 2007, 2010; Mamo 2013; Mamo *et al.* 2013, 2014) such as artificial neural networks (ANN) for forecasting failure occurrence in the network.

The purpose of this research is to establish a Decision Support System (DSS) for optimizing infrastructure asset management of WDS through a statistical analysis of the WDS FR using historical data for degradation risk assessment. To support the operator's decision-making in optimizing the network rehabilitation the proposed statistical analysis method yields

- For a pre-established Failure Rate (F_R) of a cluster of pipes, defined by the pipe characteristics (i.e. year of installation, materials, diameter), the Probability ($P_r(F_R)$) of exceeding the selected F_R value as a function of the Age (T) of the pipes
- For an Acceptable Probability (or Risk) of exceeding a pre-selected F_R value the variation of the F_R value as a function of the Age (T) of the pipes
- An operator's experiencebased Risk Matrix to identify the risk of exceeding a pre-selected F_R value at a specific Age (T).

METHODOLOGY

The proposed Risk Assessment Method (RAM) is based upon the use of both statistical and stochastic models for the network micro-zone degradation rate assessment to enable optimization and prioritization of asset management and rehabilitation works. The use of Geographical Information Systems (GIS) for the construction of the database presents a specific interest in facilitating access to information on the water network (Jafar 2006; Jafar *et al.* 2007, 2010; Mamo 2013; Mamo *et al.* 2013, 2014). For this demoillustration the RAM application used data on pipe failures collected in the City of Wattrelos, France.

The Failure Rate (F_R) of the selected micro-zone indicates the number of failures per unit length of pipes (km) per year, as given by Equation (1):

$$\text{Failure Rate}\,(F_R) = \frac{\text{Number of failures}}{\text{Total length of pipes(km or miles)} * \text{year}} \tag{1}$$

For any age (T) the likelihood of failure is defined by the frequency of the F_R values, which is equivalent to the return period of the selected F_R value. For a sample of pipes with an average sample age (T) the Probability P_i of a selected F_R value over the selected time period is defined by the frequency distribution of the F_R values:

$$P_i = \frac{N_i}{N_F} \tag{2}$$

where:

N_i/N_F is the frequency of a selected F_{Ri} value (or range of values)
N_F is the total number of years for the selected time period
i is the serial number of a selected F_{Ri} value in a time series of N_F values.

The Cumulative Probability P_{Ci} of a selected F_{Ri} value refers to the probability that the random F_R value is smaller or equal to the specified F_{Ri} value. The probability (or risk) of exceeding the selected F_{Ri} value is given by:

$$P_{Ri} = 1 - P_{Ci} \tag{3}$$

For a selected age (T), a risk matrix, indicating a risk level on a color-coded scale (i.e. 1 to 4, corresponding to increasing deterioration severity level), is established based upon the Failure Rate ($F_{Ri}(T)$) for the cluster of pipes under consideration and the acceptable risk P_{rA} of exceeding the selected F_{Ri} value.

The Weibull Distribution Function is used for the statistical data analysis to model the network life-cycle and the relationship between the Failure Rate (F_R), Age (T) and the Probability (or Risk) of exceeding a specific Failure Rate ($P_{Ri}(F_R)$) of a selected micro-zone. The three-parameter Weibull Cumulative Density Function (CDF) is given by:

$$F(t) = 1 - e^{-\left(\frac{t-\gamma}{n}\right)^{\beta}} \tag{4}$$

where:

$\eta =$ Scale parameter (or characteristic life-cycle). For a selected Age (T) the Weibull CDF of the FR is obtained by substituting $t = F_{Ri}$ and using for a scale parameter $F_{R0} = 1$ Failure per 1 km per 1 year.
$\beta =$ Shape parameter (or degradation rate parameter)
$\gamma =$ Location parameter (or failure-free life-cycle), representing a ($F_{Ri} \leq F_{R\gamma}$) threshold of zero frequency (i.e. $P_i = 0$).

The reliability function is then given by:

$$R(t) = e^{-\left(\frac{t-\gamma}{n}\right)^{\beta}} \tag{5}$$

The three-parameter Weibull Probability Density Function (PDF) is given by:

$$f(t) = \beta \left(\frac{t - \gamma}{\eta} \right)^{\beta - 1} e^{-\left(\frac{t-\gamma}{n} \right)^{\beta}} \tag{6}$$

where:

$f(t) \geq 0, \quad t \geq \gamma$

$\beta > 0, \ \eta > 0$

$-\infty < \gamma < +\infty$

$t =$ the random variable under consideration.

For $\beta > 1$, the F_{Rmax} value for which the PDF reaches its maximum is obtained from:

$$\frac{1}{F_{R0}} (F_{Rmax} - F_{R\gamma}) = \left(\frac{\beta - 1}{\beta} \right)^{\frac{1}{\beta}} \tag{7}$$

The Maximum Probability Density P_{dmax} is therefore given by:

$$P_{dmax} = \left(\frac{\beta}{F_{R0}} \right)(k)^{\frac{1}{\beta}}(e)^{-k} \tag{8}$$

where:

$$k = \left(\frac{\beta - 1}{\beta} \right)$$

Data collection description of the study area

The study area is the City of Wattrelos, France, covering 1,362 hectares with a population of approximately 43,000 inhabitants. The WDS is approximately 162 km in length with the initial layout in 1891. The data used in this research is a 74-year historical dataset.

The pipe network database, including (i) water network characteristics and operational parameters (e.g. variance of pressure) and (ii) failure information with about 424 observations for the period 1991–2004, was compiled by Jafar (2006) and Jafar *et al.* (2007, 2010) using the GIS.

The pilot WDS consists of 200 pipes with a total length of 17.18 km. The oldest pipe was installed in 1927 while the youngest one was installed in 2002. The first recorded breakage dates to 1937 and the last one to 2005. The dataset selected for the pilot study included 51 cast iron pipes, total length of 4 km, installation year – 1927–1936 (average group installation year – 1931) and diameter (40 mm $< D <$ 100 mm). To analyze the influence of the age (T) on the Failure Rate (F_R), the breakage-recorded data, from 1937 to 2004, were set into nine age-groups of 20-year time-period.

RESULTS AND DISCUSSION

Statistical data analysis

Figure 1(a) illustrates the Probability Distribution (P_i) of the F_{Ri} values and Figure 1(b) the cumulative probability function for the selected cluster of cast iron pipes and a 20-year group age (1937 to 1956, average group age 15 years). Figure 1(c) shows the reliability function for this group age with increasing F_R values. Exponential regression is used to establish the Weibull shape factor β for the statistical data analysis, yielding for $\beta = 2$ the best-fit curve. Figure 1(d) shows that the Maximum Probability Density (P_{dmax}) remains practically constant with the group age and its corresponding Maximum Probability (P_{max}) value is quite consistent with the value calculated from Equation (8) for $\beta = 2$. Figure 1(a) shows that the Weibull function with $\beta = 2$ yields F_{Rmax} (Equation (7)) and P_{max} (Equation (8)) values which correspond fairly well to the F_R probability distribution characteristics values.

Figure 2 presents the cumulative probability functions for the F_{Ri} values for the selected cluster of pipes and the available group ages. It illustrates the relationships among three parameters, including F_{Ri}, Age (T) and Cumulative Probability (P_{Ci}). For a constant F_{Ri} value (e.g. $F_{Ri} = 1$) the Cumulative Probability (P_{Ci}) as well as the Probability (or Risk) of exceeding the selected F_{Ri} value (i.e. $P_{Ri} = 1 - P_{Ci}$) are therefore a function of the Group Age (T).

For example, the data analysis shows that for a selected value of $F_{Ri} = 1$, a 10-year delay in the network rehabilitation from an average group age of 55 years to 64 years will result in a significant risk increase of the probability of
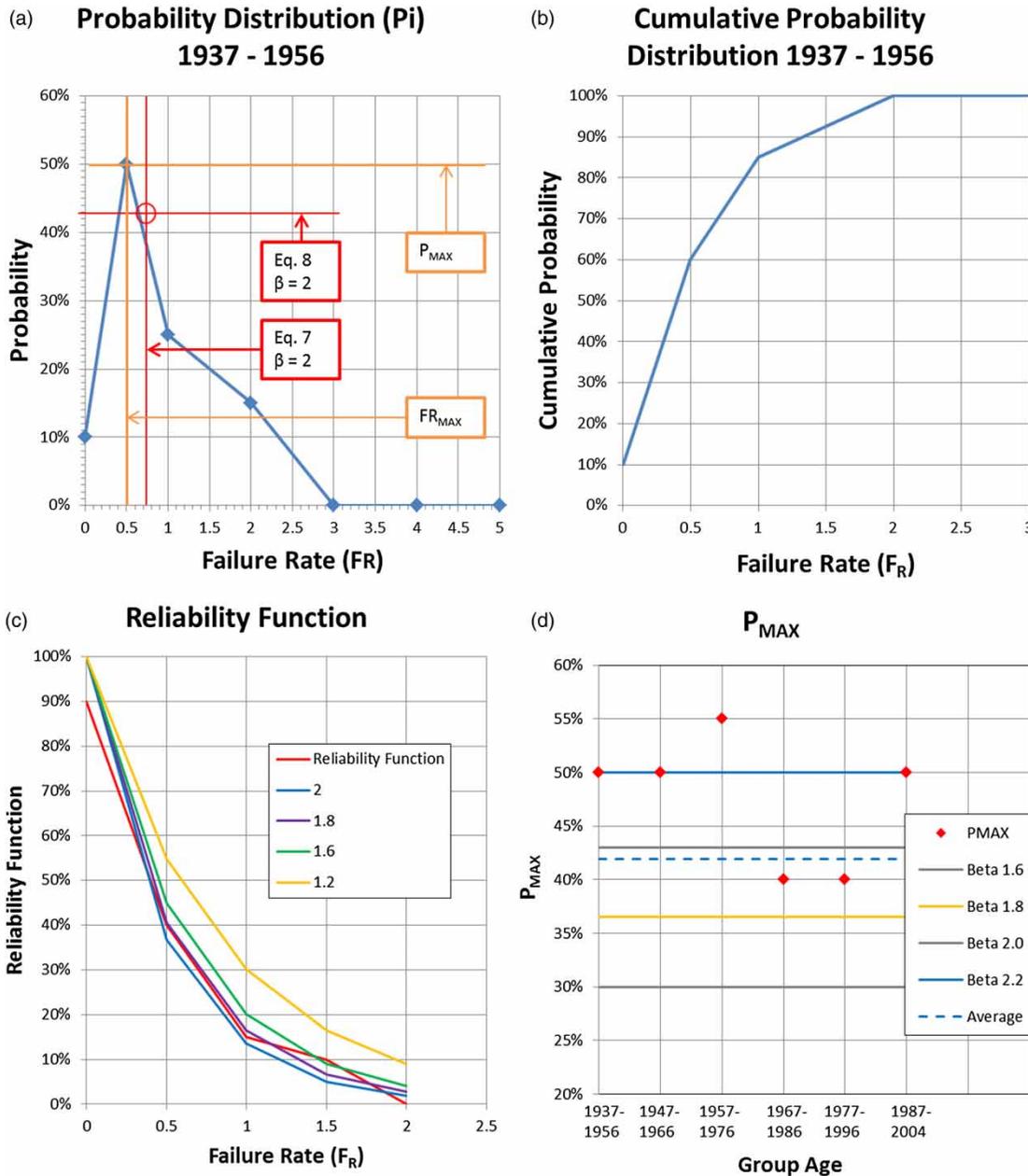
Figure 1 │ (a) Probability distribution group age: 1937–1956; (b) cumulative probability distribution group age: 1937–1956; (c) exponential regression to establish the shape factor – β; (d) maximun probability (P_{MAX}) for specific group age.

exceeding the selected value of $F_{Ri} = 1$ from $P_{Ri} = 25\%$ to 60%.

Similarly, for a constant Cumulative Probability (P_{Ci}) value the F_{Ri} value is a function of the average group age (T) as illustrated in Figure 2 for $P_{Ci} = 0.6$. The relationships between these three fundamental aging parameters could be efficiently used for risk assessment to enable the operator optimizing the WDS asset management.

Stochastic data modeling for forecasting Aging (T) effect on Failure Rate (F_R)

The 'Casses' software, developed by CEMAGREF (Casses Software; Le Gat & Eisenbeis 2000) for analyzing failure data and forecasting breakage risks in water distribution pipelines was used for the data analysis following its three steps of calibration, validation and forecasting. The 'Casses' algorithm
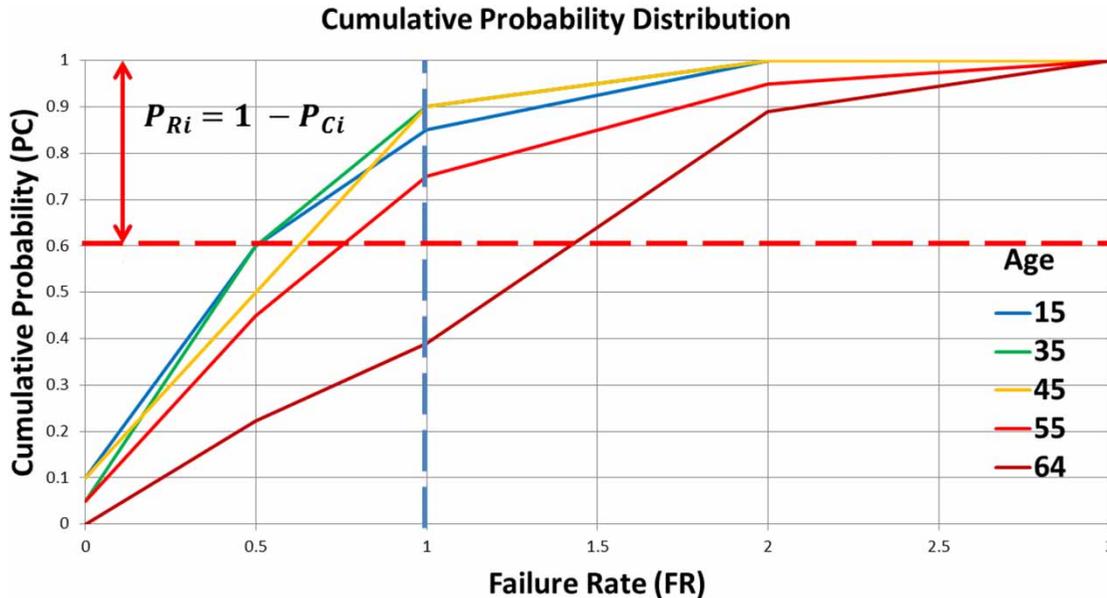
Figure 2 │ Cumulative probability functions for all the group ages.

uses the Poisson model for stochastic data analysis. The model calibration used record data up to 2004, including the validation period of 2000 to 2004, for forecasting the behavior trend over a period extending up to 2055.

Figure 3(a) and 3(b) show respectively the probability distribution of the F_R values and the reliability function forecast for the 20-year time-period group age of 2035 to 2055 (average group age of $T = 114$ years) for the cast iron pipes.

Figure 3(b) illustrates that the reliability function for this group age corresponds fairly well to the Weibull reliability function with $\beta = 2$. Figure 3(a) shows that the Weibull function with $\beta = 2$ and $F_{R\gamma} = 2.8$ yields consistent F_{Rmax} (Equation (7)) and P_{max} (Equation (8)) values of the F_R probability distribution.

Figure 3(c) shows the forecast cumulative probability functions obtained for all the average group ages analyzed, indicating that for average group ages of 73 years and 113 years the probability of exceeding a selected F_{Ri} value of 3 increases, respectively, from $P_{Ri} = 0\%$ to 90%.

Risk Assessment Methodology for a Decision Support System (RAM-DSS)

The integration of the statistical and stochastic data analyses of the recorded and forecast breakage rates enables the operator to establish a Risk Assessment Methodology (RAM) for optimizing WDS asset management. Figure 4(a) illustrates for constant Cumulative Probability (P_{Ci}) levels of 20%, 40%, 60% and 80%, the variation of the F_{Ri} value with the Average Group Age (T) over a time period of 123 years. It shows that for the data of the cast iron pipes, the aging effect starts to significantly increase after 55 years.

Probability (P_{Ri}) of exceeding a selected F_{Ri} value

Figure 4(b) illustrates for selected F_{Ri} values of 0, 1, 2, 3, and 4, the risk increase with the network aging through variation of the Probability (P_{Ri}) of exceeding the selected F_{Ri} value vs the Average Group Age (T). For example, for a selected value of $F_{Ri} = 1$ a delayed rehabilitation of 9 years (from 55 to 64 years) will increase the risk of exceeding the selected value of $F_{Ri} = 1$ from 25% to 60%.

The purpose of the proposed Decision Support System (RAM-DSS) is to enable the operator to assess the delayed rehabilitation investment impacts on the risk of exceeding a selected F_{Ri} value and its cost consequences. For this purpose a risk matrix is defined as the product of the likelihood of exceeding the selected F_R value and the consequences of its occurrence.
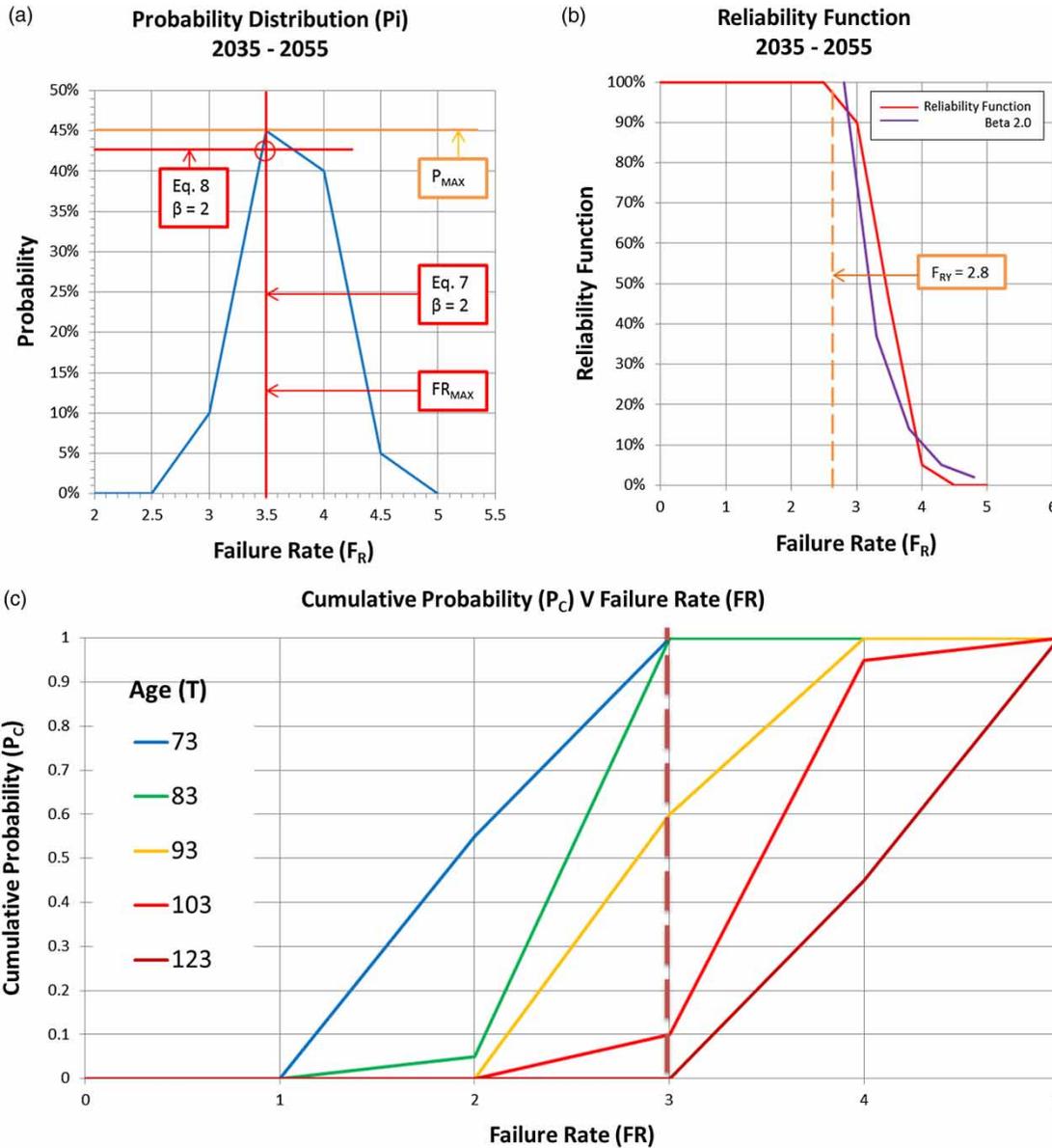
Figure 3 │ (a) Probability distribution group age: 2035–2055; (b) reliability function for group age: 2035–2055; (c) Cumulative Probability (P_{Ci}) vs Failure Rate (F_R) for data set 1.

For optimizing asset management the consequences could be defined by the selected F_R values for different severity levels of Insignificant, Minor, Moderate, Major, etc. For this pilot study, as illustrated in Table 1, considering a WDS, 4 km long, five severity levels of consequences were identified considering F_R ranges and the related average number of breakages per month. The event likelihood of exceeding a selected F_R value is identified by its frequency of occurrence over the selected time period (or its return period) (Table 2).

Figure 5(a) presents the color-coded risk matrix for the operator's experience-based DSS used for the demo-illustration in this pilot study. As illustrated in Figure 5(b), its deployment can be used for graphic color-coded risk visualization of the Probability (P_{Ri}) of exceeding a selected F_{Ri} value. The color-coded risk visualization can be efficiently integrated with multilayer input data on a GIS platform (Jafar 2006; Jafar *et al.* 2007, 2010) to support risk-based optimization and prioritization of investments in
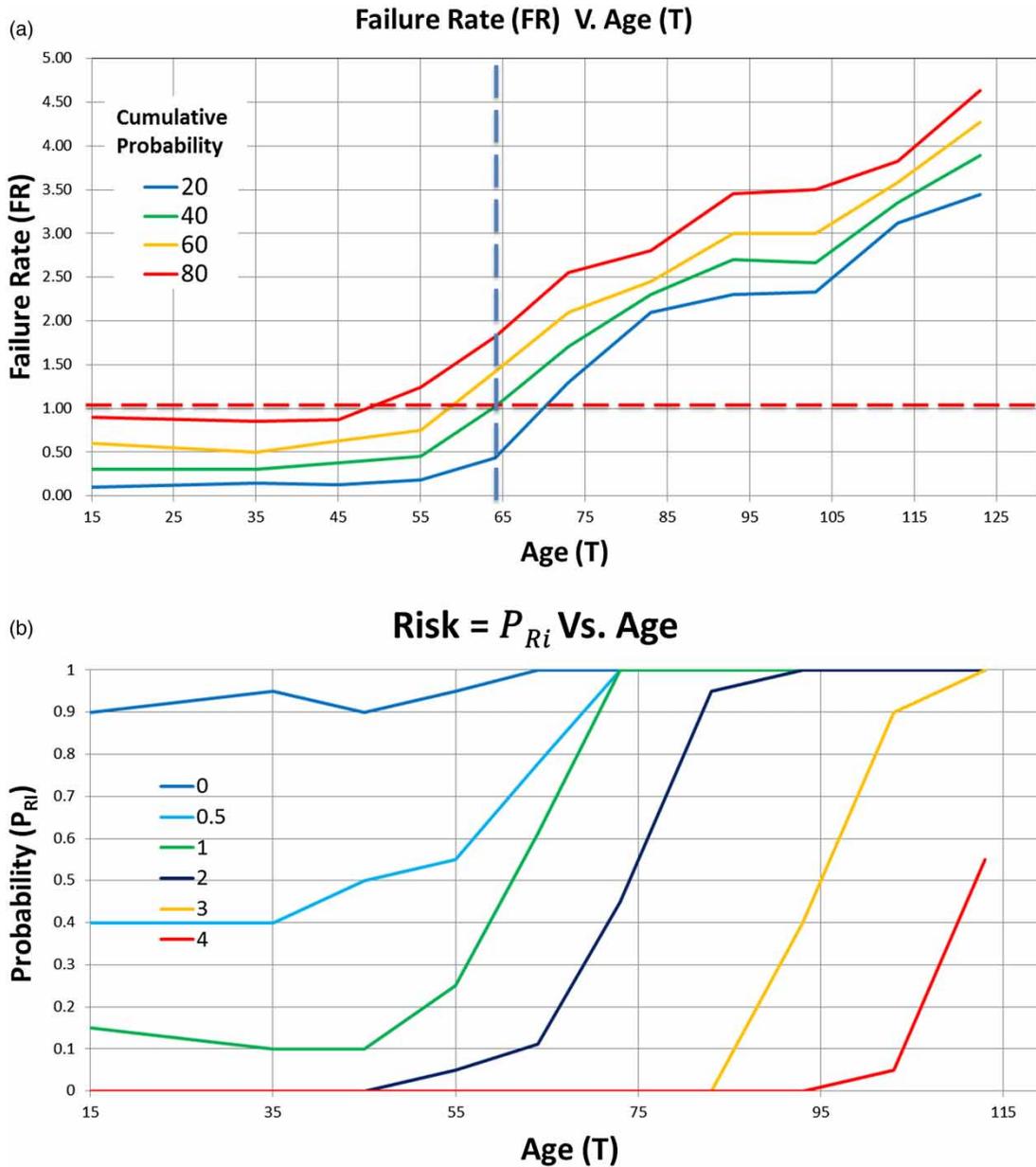
(a)

(b)

Figure 4 │ (a) Failure Rate (F_R) vs Age (T) for constant Cumulative Probability (P_{Cj}); (b) risk assessment – Probability (P_{Ri}) of exceeding a selected F_{Ri} value vs Average Group Age (T).

Table 1 │ Consequence severity levels

FR	Consequence	Number of breakages per month
<0.5	Insignificant	0.17 (1 per 6 months)
0.5–1	Minor	0.33 (1 per 3 months)
1–2	Moderate	0.67 (1 per 1.5 months)
2–3	Major	1.00 (1 per month)
>3	Severe	> 1.00 (1 per Month)

Table 2 │ F_R likelihood

F_R Likelihood	
>80%	Highly probable
60%–80%	Probable
40%–60%	Occasional
20%–40%	Remote
<20%	Improbable

(a)

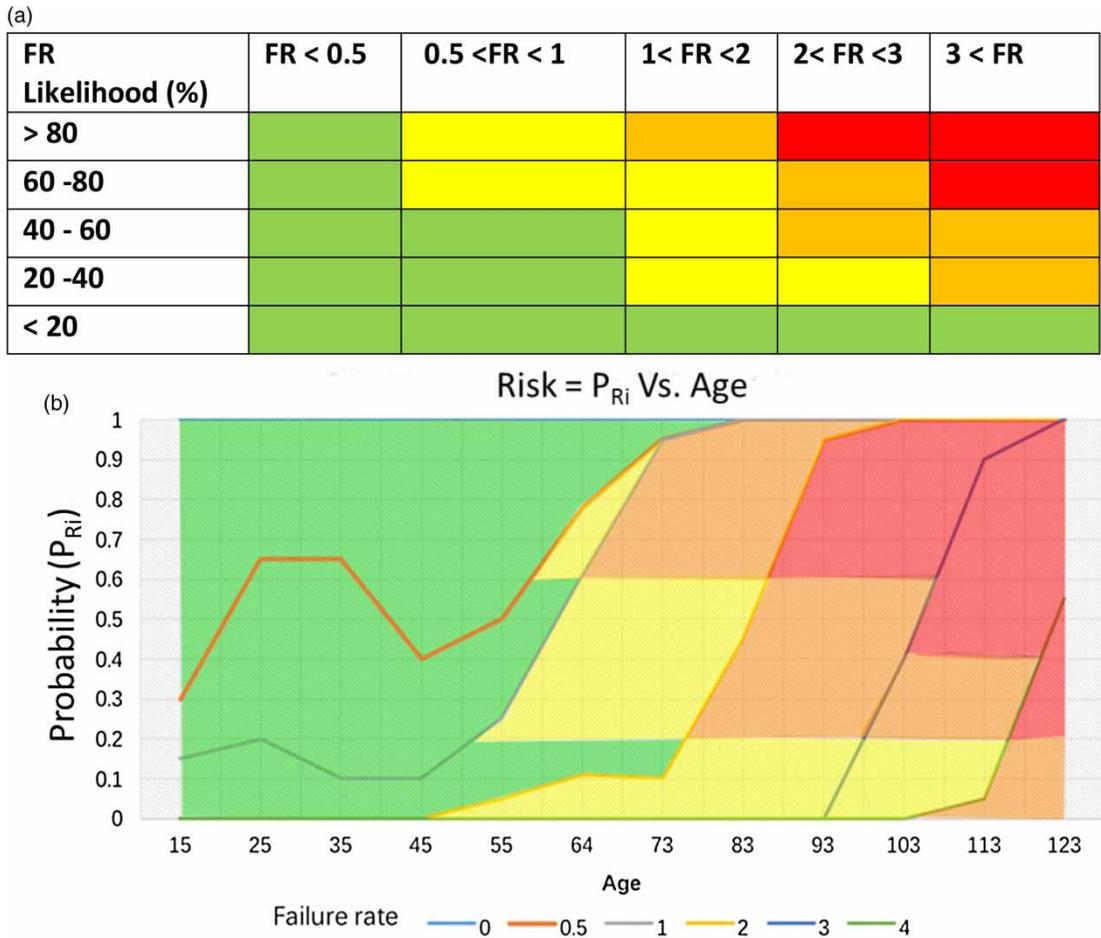| FR Likelihood (%)	FR < 0.5	0.5 <FR < 1	1< FR <2	2< FR <3	3 < FR
> 80					
60 -80					
40 - 60					
20 -40					
< 20					

(b)

Figure 5 │ (a) Risk matrix; (b) risk assessment visualization for pre-emptive asset management.

infrastructure repair and rehabilitation for pre-emptive asset management.

CONCLUSION

The analysis of the relationships between the Failure Rate (F_R) of a water distribution network (or sub-network), the Age (T) of its classified pipes and the Risk (or probability – P_R) of exceeding the selected F_R value have been used in this study to develop and demo-illustrate a Risk Assessment Method Decision Support System (RAM-DSS) for optimizing asset management, forecasting the consequences of delayed rehabilitation and prioritizing the required investments.

The Weibull distribution function, often used for relatively small samples, is used for the statistical data analysis to model the network life-cycle and the risk of exceeding a selected Failure Rate (F_R) as a function of the Age (T). For this case study, the shape factor, representing the degradation rate, remains practically constant with the Group Age (T) and consistently yields the F_{Rmax} and P_{max} values of the Probability Density function of the F_R values.

An operator's experience-based risk management matrix is proposed for optimizing WDS asset management. The research demo-illustrates the value of integrating statistical and stochastic data analyses in modeling the vulnerability of the system at a certain age of its life-cycle and forecasting its degradation for preemptive asset management and rehabilitation investment prioritization.

ACKNOWLEDGEMENTS

REFERENCES

Alegre, H. & Coelho, S. T. 2012 *Infrastructure Asset Management of Urban Water Systems*. National Civil Engineering Laboratory, Lisbon, Portugal.

Alegre, H., Almeida, M. C., Covas, D. I. C., Cardoso, M. A. & Coelho, S. T. 2011 Integrated approach for infrastructure asset management of urban water systems. In: *IWA 4th Leading Edge Conference on Strategic Asset Management*, 27–30 September, Mülheim an der Ruhr, Germany.

Casses Software – Version 2.0.0. User Manual, available at https://casses.cemagref.fr/.

Jafar, R. 2006 *Modélisation de la dégradation des réseaux d'eau en vue d'une gestion prévisionnelle*. Doctoral Dissertation (UMR 8107), University of Lille, France.

Jafar, R., Shahrour, I. & Juran, I. 2007 Modelling the structural degradation in water distribution systems using the artificial neural networks (ANN). *Water Asset Management International* 3 (3), 14–18.

Jafar, R., Shahrour, I. & Juran, I. 2010 Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. *Mathematical and Computer Modelling* 51 (9–10), 1170–1180.

Le Gat, Y. & Eisenbeis, P. 2000 Using maintenance records to forecast failures in water networks. *Urban Water* 2 (3), 173–181.

Mamo, T. 2013 *Virtual District Meter Area Municipal Water Supply Pipeline Leak Detection and Classification Using Advance Pattern Recognizer Multi-Class Support Vector Machine for Risk-Based Asset Management*. Doctoral Dissertation, NYU School of Engineering, New York, USA.

Mamo, T., Juran, I. & Shahrour, I. 2013 Urban water demand forecasting using the stochastic nature of short term historical water demand and supply pattern. *Journal of Water Resource and Hydraulic Engineering* 2 (3), 92–103.

Mamo, T., Juran, I. & Shahrour, I. 2014 Virtual DMA municipal water supply pipeline leak detection and classification using advance pattern recognizer multi-class SVM. *J. Pattern Recog. Res.* 1, 25–42.

Vanrenterghem-Raven, A., Eisenbeis, P., Juran, I. & Christodoulou, S. 2004 Statistical modeling of the structural degradation of an urban water distribution system: case study of New York City. In: *World Water & Environmental Resources, Congress 2003* (P. Bizier & P. DeBarry, eds), ASCE, Reston, VA, USA.