Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 186 (2017) 117 – 126

Procedia Engineering

www.elsevier.com/locate/procedia

XVIII International Conference on Water Distribution Systems Analysis, WDSA2016

Pipe Failure Prediction in Water Distribution Systems Considering Static and Dynamic Factors

Raziyeh Farmani[a]*, Konstantinos Kakoudakis[b], Kourosh Behzadian[c] and David Butler[d]

[a]Associate Professor, Centre for Water Systems, University of Exeter, Exeter, UK
[b]PhD student, Centre for Water Systems, University of Exeter, Exeter, UK
[c]Senior Lecturer, University of West London, London, UK
[d]Professor, Centre for Water Systems, University of Exeter, Exeter, UK

Abstract

Due to high economic, environmental and social costs resulting from pipe bursts in water distribution systems, development of a reliable and accurate prediction model to assess susceptibility of a pipe to failure is of paramount importance. This paper aims to consider the impact of both static and dynamic factors on pipe failure for long and mid-term predications. Length, diameter and age of pipes are the static and weather is the dynamic factors for the prediction model. To improve the performance of the pipe failure prediction models, the *K*-means clustering approach is considered. Evolutionary Polynomial Regression (EPR) is used as the pipe failure prediction model. To prepare the database for the prediction model, homogenous groups of pipes are created by *aggregating* individual pipes using their attributes of age, diameter and soil type. The created groups were divided into training and test datasets using the cross-validation technique. The *K*-means clustering approach is employed to partition the training data into a number of clusters with similar features based on diameter and age of the pipe groups. An EPR model is developed and calibrated for each data cluster. To predict pipe failures for new (unseen) data, the most suitable cluster is identified and the relevant EPR model is used to obtain the most accurate prediction. The proposed approach is demonstrated by application to a water distribution system in the UK. Comparison of the results shows that the cluster-based prediction model is able to significantly reduce the prediction error of pipe failures. Temperature-related factor is identified as the main dynamic factor influencing the t mid-term prediction of pipe failures. An EPR model is employed to predict the annual variation in the number of failures. Mid-term and long-term prediction models are developed to present the relationship between number of pipe failures and temperature-related factors for better operation and long term for capital investment respectively.

* Corresponding author. Tel.: +44 1392 723630
E-mail address: r.farmani@exeter.ac.uk

doi:10.1016/j.proeng.2017.03.217

Keywords: water distribution system; pipe failure; prediction models; Evolutionary polynomial regression; *K*-means clustering

1. Introduction

The failure of water distribution networks' pipes is a global concern due to the potential consequences. The water authorities, in order to cope with pipe failure, can follow either a reactive or a proactive approach. In a proactive strategy pipe rehabilitation is scheduled in advance after assessing and forecasting pipe propensity to fail [1]. The proactive strategy includes scheduling the maintenance, improvement and extension of water mains in order to maintain/improve the current level of service. The use of predictive models is an important step in the implementation of the proactive strategy. They can help water utilities to make more informed and accurate decisions for the future planning of pipe rehabilitation and/or replacement.

The pipe failure is the cumulative effect of various pipe-intrinsic (such as material, diameter, and age), operational (such as corrosion, pressure, external stresses) and environmental factors (such as temperature, rainfall, soil conditions) acting on them. Environmental and pipe-intrinsic factors can be divided into static and dynamic (time-dependent), while operational factors are inherently dynamic. This paper proposes the use of two different approaches for the long-term and the mid-term prediction of pipe failure. The prediction models provide insights into the relationships between pipe failure and all kinds of factors influencing pipe failure. The long-term approach employs the pipe-intrinsic factors as explanatory variables while the mid-term approach employs the environmental factors.

The paper presents the combined use of Evolutionary Polynomial Regression [2] and *K*-means clustering method [3] to achieve accurate long-term predictions of the expected number of pipe failures. The mid-term approach uses EPR to predict the number of failures yearly considering weather-related factors as explanatory variables.

2. Methodology

Figures 1&2 show the framework for the long-term and mid-term prediction of pipe failures. The long-term prediction models consider length, diameter and age and the mid-term prediction models g temperature-related factors as explanatory variables respectively. The prediction models were developed using EPR-MOGA-XL vr.1 [4, 5].

The proposed methodology for the long-term predictions consists of the following steps (figure 1):

Initially, the individual pipes are *aggregated* into homogenous groups assuming that pipes with similar specific intrinsic properties such as material, diameter and age are expected to have the same breakage pattern [6]. In addition, soil type is used as an aggregation criterion because soil properties have been associated with the corrosion of the metallic pipes [7, 8] and are considered as a major factor contributing to their failure [9, 10]. The total length and the total number of failures of each homogenous group (specific age, diameter and soil type) are calculated as a sum. The original database which includes a huge number of individual pipes is converted into a number of groups.

The created homogenous groups were split into training and test datasets using the cross-validation technique [11] for calibration and validation purposes respectively. All the groups are used both for training and test and each group is used for test exactly once [12]. The training dataset is partitioned into *k* clusters based on diameter and age of groups using the *K*-means algorithm. One EPR model is developed for each cluster of the training dataset.

Finally, the performance of the developed models is evaluated by using the test data. The Euclidian distance between the test data sample (i.e. age and diameter of the groups that constitute the test dataset) and the counterpart cluster centre values is calculated to identify the appropriate cluster for each test data. The corresponding model is employed to calculate the number of failures and the performance indicators are evaluated using the predicted values and the observations.

Various number of clusters are analysed here to identify the optimal number until no further improvement was achieved with the test data. For any specific number of *k*, the number of EPR models developed is equal to this number of clusters.