Knowledge-Based Systems 110 (2016) 144–156

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Predicting water main failures: A Bayesian model updating approach

Golam Kabir[a,*], Solomon Tesfamariam[a], Jason Loeppky[b], Rehan Sadiq[a]

[a] School of Engineering, University of British Columbia (UBC), Kelowna, BC, Canada, V1V1V7
[b] Department of Statistics, University of British Columbia (UBC), Kelowna, BC, Canada, V1V1V7

ARTICLE INFO

Article history:
Received 13 December 2015
Revised 12 July 2016
Accepted 16 July 2016
Available online 21 July 2016

Keywords:
Water main failure
Bayesian updating
Bayesian model averaging (BMA)
Survival analysis
Weibull proportional hazard model (PHM)
Reliability
Uncertainty

ABSTRACT

Water utilities often rely on water main failure prediction models to develop an effective maintenance, rehabilitation and replacement (M/R/R) action plan. However, the understanding of water main failure becomes difficult due to various uncertainties. In this study, a Bayesian updating based water main failure prediction framework is developed to update the performance of the Bayesian Weibull proportional hazard (BWPHM) model. Applicability of the proposed framework is illustrated with modeling failure prediction of cast iron and ductile iron pipes of the water distribution network of the City of Calgary, Alberta, Canada. The Bayesian updating models have effectively improved the water main failure prediction whenever new data or information becomes available. The proposed framework can assess the model performance in the light of uncertain and evolving information and will help the water utility authorities to attain an acceptable level of service considering financial constraints.

1. Introduction

To develop an effective maintenance, rehabilitation, and replacement (M/R/R) action plan of water utilities, an appropriate water main failure prediction model is required [10]. However, the full understanding of the buried water mains failure processes is a complex undertaking [26]. Substantial efforts have been made to develop statistical water main failure prediction models. Statistical models attempting to predict the failure behavior of water mains are not only affected both by the quality and quantity of available data, but also by the adopted statistical techniques [22,23]. It is often challenging to develop statistical models for water main failures due to multiple factors affecting this failure [2,36].

Survival analysis is the most widely used statistical models for water main failures which are a branch of statistics dealing with deterioration and failure over time and involves the modeling of the elapsed time between an initiating event and a terminal event [10,23]. The analysis incorporates the fact that while some pipes break, others do not and this information has a strong impact on pipe failure analysis [36]. The models use covariates (i.e., diameter, length, soil resistivity) to differentiate the pipe failure distributions without splitting the failure data, thereby giving a better understanding of how covariates influence the failure of the pipe [26]. Different researchers applied different survival analysis methods like Kaplan–Meier estimator [6], homogeneous Poisson process or Poisson regression [3,4], Nonhomogeneous Poisson Process (NHPP) [21,35], zero-inflated NHPP [12,34], exponential/Weibull model [10,11,30], multivariate exponential model [22,28], Cox proportional hazard model (Cox-PHM) [7,24,30,31,40], and Weibull proportional hazard model (WPHM) [24,26,40].

Uncertainties become an integral part of the water main failure prediction models because of the integration or fusion of data/information from different sources, involvement of human (expert) judgment for the interpretation of data and observations, partial and incomplete information [10,14,20]. Due to inaccurate measurements of the different factors, unreliable recording of failure times and lack of the actual failure times, data quality also becomes a serious issue as many datasets contain such uncertainties [12]. Furthermore, because of the involvement of multiple experts who have different levels of knowledge and credibility related to the problem, the decision-making problem becomes more complex and uncertain [19,37].

To deal with these uncertainties in the water main failure prediction model, Bayesian inference or analysis where all the model parameters are expressed as random have been previously been considered (e.g. [12,38]). The Bayesian models can incorporate external information like elicited expert opinions, judgment, beliefs, relevant historical information into the model by constructing a probability distribution that describes the uncertainty in the model parameters (prior to the observing data from the experiment) [10,12,20,38].

* Corresponding author. Fax: +12508079850.
E-mail addresses: golam.kabir@ubc.ca (G. Kabir), solomon.tesfamariam@ubc.ca (S. Tesfamariam), jason.loeppky@ubc.ca (J. Loeppky), rehan.sadiq@ubc.ca (R. Sadiq).

http://dx.doi.org/10.1016/j.knosys.2016.07.024
0950-7051/

Notation

p	precipitation
e	evapotranspiration
I	heat index constant
a	constant, function of heat index constant
Tm	mean monthly temperature
FI_p	freezing index
D_i	daily temperature
DD_p	all the days below the threshold temperature
ϕ	threshold temperature
M	vector of model
ω_k	vector of parameters of model M_k
Δ	quantity of interest
$Pr(M_k)$	prior probability of model M_k
$Pr(\Delta\|D)$	posterior distribution of Δ given the data D
$Pr(M_k\|D)$	posterior probability of model M_k
$Pr(D\|M_k)$	marginal likelihood of model M_k
$Pr(\omega_k\|M_k)$	prior density of ω_k under model M_k
$\hat{\omega}_k$	posterior mean of ω_k
d_k	dimension of ω_k,
N	total number of uncensored cases
χ	threshold window
E $(\omega\|D)$	mean of the coefficients of models
SD $(\omega\|D)$	standard deviation of the coefficients of models
$Pr\ (\omega_i{\neq}0\|D)$	posterior probabilities of the coefficients of models
T	interarrival time
$h\ (t\|X)$	hazard function
$h_0(t)$	baseline hazard function
θ	vector of covariate coefficients
ln	natural logarithm
α	scale parameter
β	vector of unknown parameters
ϵ	random component
F	Survival function
L	maximum likelihood function
rc	right censored
nrc	not right censored
O	observed values
P	predicted values
$\bar{P}$	mean of predicted values
n	number of observation

Watson et al. [38] proposed an NHPP based Bayesian hierarchical model for pipe failure. Dridi et al. [11] developed a Bayesian exponential/Weibull model and integrated structural and hydraulic indicators for developing optimal replacement strategies of water pipes. Watson [38] used expert elicitation to produce informative priors and thus account for the lack of covariate information. Dridi et al. [10] proposed a Bayesian exponential/Weibull-based pipe failure model and developed a pipe replacement strategy considering the replacement cost, expected cost of pipe break repairs, and hydraulic performance. In their analysis, they used the Weibull probability distribution function for the first breaks and the exponential probability distribution function for successive breaks. Economou et al. [12] compared Bayesian NHPP model with Bayesian zero-inflated NHPP model to handle the excess amount of zeros in the number of failures (known as zero-inflation). The authors found that zero-inflated NHPP model fitted the data better than the NHPP model for the calibration. However, pipe age was the only governing factor investigated in most of these studies ignoring other influential physical (i.e., diameter, length, and manufacturing period) and environmental (i.e., soil condition, temperature) factors. Kabir et al. [19] developed a Bayesian linear regression based water main failure model and integrated it with the ordered weighted averaging operator to handle the optimism degree of the decision makers. The authors considered age, length, diameter, vintage, temperature, freezing index, rain deficit, soil resistivity and land use for their analysis.

However, very few water main failure prediction models or studies presented any preliminary covariate or model selection method that considers uncertainties. Kabir et al. [20] proposed Bayesian model averaging (BMA) to select the influential covariate taking account of model uncertainties. Nevertheless, most of the researchers develop their model using the entire water main failure datasets. To find out the performance of the water main failure prediction models, some researcher divided their dataset for training and testing or validation purposes [12,21,22,26,40]. Most of the studies concluded that Bayesian analysis can update the model whenever new data or information is available. However, very few water main failure prediction model or study mentioned any formal guideline or framework how to deal with the new data or information or tried to improve or update the performance of the model whenever new data or information is available. To handle the new failure data or pipe information, the full analyses have to be performed again for most of the models.

The objective of this study is to develop an effective Bayesian updating based water main failure prediction framework that not only incorporates the uncertainties but also provide a rational framework on how to update performance of the model. The proposed framework provides an efficient framework for probabilistic updating and the assessment of model performance in light of uncertain and evolving information, particularly for post-event failures or pipe replacement. The proposed framework will aid the water utility authorities to proactively address the failures of water mains.

In the next section, the proposed Bayesian updating based water main failure prediction framework is described. The applicability of the proposed framework is illustrated with the City of Calgary pipe failure data. The performance comparison of the proposed model with the existing models is also presented. The last section presents the conclusion and discusses the limitations and scope for future research.

2. Proposed framework

The Bayesian updating framework (Fig. 1) for the water main failure prediction is developed using the following steps.

1. Data Collection: The pipe characteristics data, soil information and pipe breakage data will be collected from the water utility's Geographic Information System (GIS).
2. Covariate Selection: The influential and significant covariates will be selected using the BMA approach.
3. Model Development & Updating: The entire water main failure dataset will be divided into multiple periods and the BWPHM based water main failure prediction models will be developed using the failure data of the first period. After that, the model parameters will be updated using the Bayesian updating approach by the water main failure data of the second period. Similarly, Bayesian updating approach will be followed for the water main failure data of the remaining periods.
4. Model Performance Evaluation: The performance of the models after each period will be assessed using different error measures.

Each step will be discussed elaborately in the next sections with appropriate example.