Desalination and Water Treatment
www.deswater.com

doi:10.5004/dwt.2018.22714

120 (2018) 190–197
July

Statistical approach to developing screening models for pipe failure events in water network systems

Bumjo Kim[a], Seo Jin Ki[b], Dong Jin Jeon[a], Joon Ha Kim[a,*]

[a]*School of Environmental Science and Engineering, Gwangju Institute of Science and Technology (GIST), 123 Cheomdangwagi-ro, Buk-gu, Gwangju 61005, Korea, Tel. +82-62-715-3391, email: joonkim@gist.ac.kr (J.H. Kim)*
[b]*Department of Environmental Engineering, Gyeongnam National University of Science and Technology, 33 Dongjin-ro, Jinju-si, Gyeongsangnam-do 52725, Korea*

Received 9 April 2018; Accepted 20 June 2018

ABSTRACT

Accurate assessment of piping systems' risk to damage reduces annual operation and maintenance costs. Recently, extreme climate events (e.g. cold snaps or heavy snow) due to global climate change have increased pipe system failure. The objective of this study is to establish a framework for developing screening models of pipe failure events, due to water network systems freezing using two statistical approaches. More specifically, logistic regression was used to estimate the probability of failure at a household level, whereas the customized model developed to predict the frequency of community-wide failure events. The data recorded at least one failure event in Korea from 2008 to 2015, which was provided to the logistic regression model. The customized model, however, only used the data set compiled from three areas of concern with the highest frequency of the failures. Results showed that the logistic model showed the best performance out of the 11 constructed models, in terms of *R* and the variance inflation factor (of lower than two). The logistic model incorporated three variables: the minimum temperature on the day of failure, the natural logarithm of the total water usage in the previous month and the mean minimum temperature over the previous 10 days. The selected model had an overall prediction accuracy of 66.4%. When the customized model at the community level was examined the three models not only yielded moderate R^2 values ranging from 0.53 to 0.66, but also helped identify water network systems at risk of failures. Overall, this study demonstrated that the proposed methodology can be used to highlight areas of concern at different geographic scales, along with refining existing statistical models with new variables updated in real time.

Keywords: Pipe failure event; Logistic regression; Customized model; Minimum temperature; Total water usage; Variance inflation factor

*Corresponding author.