

## Accepted Manuscript

Prediction of Fecal Coliform using Logistic Regression and Tree-based Classification Models in the North Han River, South Korea

Soo Yeon Choi, Il Won Seo

PII: S1570-6443(17)30238-1  
DOI: <https://doi.org/10.1016/j.jher.2018.09.002>  
Reference: JHER 458

To appear in: *Journal of Hydro-environment Research*

Received Date: 30 June 2017  
Revised Date: 1 June 2018  
Accepted Date: 12 September 2018

Please cite this article as: S.Y. Choi, I.W. Seo, Prediction of Fecal Coliform using Logistic Regression and Tree-based Classification Models in the North Han River, South Korea, *Journal of Hydro-environment Research* (2018), doi: <https://doi.org/10.1016/j.jher.2018.09.002>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



**Prediction of Fecal Coliform using Logistic Regression and Tree-based Classification  
Models in the North Han River, South Korea**

**Soo Yeon Choi<sup>a</sup>, Il Won Seo<sup>a\*</sup>,**

<sup>a</sup>Dept. of Civil and Environmental Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, Korea

\*Corresponding author, E-mail address: seoilwon@snu.ac.kr, Fax number: +82-2-873-2684, Telephone number: +82-2-880-7345

**ABSTRACT**

In this study, data-based classification models were developed for real-time prediction of the exceedance of the safety level on fecal coliform in Daesung-ri site of North Han River. The prediction models were developed using the logistic regression model (LRM) and the tree-based models such as classification and regression model (CART), bagging model (BGM), and random forest model (RFM). For model development, rainfall, water quality, and dam discharge data from 2010 to 2015 were collected from the study site. Clustering methods were applied to reduce the sampling bias of training and test datasets and to improve the prediction accuracy. The developed four models were compared with each other in terms of prediction accuracy and applicability. The test results of developed models showed that the total correct classification rate of the four models ranged from 83.7% to 93.0%. Each classification model showed its own strengths; LRM offered flexibility by tuning cutoff values, while RFM showed

Download English Version:

<https://daneshyari.com/en/article/11033236>

Download Persian Version:

<https://daneshyari.com/article/11033236>

[Daneshyari.com](https://daneshyari.com)