Journal of the Korean Society of Agricultural Engineers
DOI : https://doi.org/10.5389/KSAE.2021.63.2.051

Vol. 63, No. 2, pp.51~60, March 2021
ISSN 1738-3692 | eISSN 2093-7709

Evaluation of Surrogate Monitoring Parameters for SS and T-P Using Multiple Linear Regression and Random Forest

다중 선형 회귀 분석과 랜덤 포레스트를 이용한 SS, T-P 대리모니터링 기법 평가

Jeung, Minhyuk[a] · Beom, Jina[b] · Choi, Dongho[c] · Kim, Young-joo[d] · Her, Younggu[e] · Yoon, Kwangsik[f, †]
정민혁 · 범진아 · 최동호 · 김영주 · 허용구 · 윤광식

ABSTRACT

Effective nonpoint source (NPS) pollution management requires frequent water quality monitoring, which is, however, often costly to be implemented in practice. Statistical techniques and machine learning methods allow us to identify and focus on fundamental environmental variables that have close relationships with NPS pollutants of interest. This study developed surrogate models to predict the concentrations of suspended sediment (SS) and total phosphorus (T-P) from turbidity and runoff discharge rates using multiple linear regression (MLR) and random forest (RF) methods. The RF models provided acceptable performance in predicting SS and T-P, especially when runoff discharge rates were high. The RF models outperformed the MLR models in all the cases. Such finding highlights the potential of RF techniques and models as a tool to identify fundamental environmental variables that are measured in relatively inexpensive ways or freely available but still able to provide information required to quantify the concentrations of NPS pollutants. The analysis of relative importance rates showed that the temporal variations of SS and T-P concentrations could be more effectively explained by that of turbidity than runoff discharge rate. This study demonstrated that the advanced statistical techniques such as machine learning could help to improve the efficiency of NPS pollutants monitoring.

Keywords: Surrogate monitoring; non-point source; machine learning; influence factor

Ⅰ. Introduction

The temporal variability of water quality parameters is information critical for efficient watershed management (Montgomery et al., 2007; Horsburgh et al., 2009). NPS pollutants are loaded to waterbodies with rainfall events that may last several days. In addition, the transport of NPS pollutants is complicated with hydrological processes that greatly vary over time, even in an event. Thus, a frequent monitoring is often required when assessing NPS pollution and developing water quality management plans (Scholefield et al., 2005; Houser et al., 2006; Jordan et al., 2007). However, frequent NPS monitoring is often prohibitive in terms of time and effort.

The use of high-frequency water quality sensors can be one of the solutions for the cost issue of NPS monitoring, but many of water quality parameters such as suspended solids (SS) and total phosphorus (T-P) still require extensive equipment and lab tests to be quantified (Jones et al., 2011, Villa et al., 2019). In this context, surrogate monitoring can be an efficient alternative. Previous studies have tried to predict the water quality parameters with accessible parameters such as pH, dissolved oxygen (DO), temperature using traditional statistical techniques such as multiple linear regression (MLR) (Çamdevýren et al., 2005; Chenini and Khemiri, 2009; DeForest et al., 2018). Some of them showed that SS and T-P is closely associated with turbidity and discharge rates of a river (Jones et al., 2011; Ziegler et al., 2011). It is a relative measure of

[a] Ph.D. Student, Department of Rural and Bio-Systems Engineering, Chonnam National University
[b] Ph.D. Student, Department of Rural and Bio-Systems Engineering, Chonnam National University
[c] Researcher, Presidential Water Commission Support Department Planning and Operation, Republic of Korea Presidential Water Commission
[d] Professor, Department of Cadastre and Civil Engineering, VISION College of Jeonju
[e] Assistant Professor, Tropical Research and Education, Department of Agricultural and Biological Engineering, University of Florida
[f] Professor, Department of Rural and Bio-Systems Engineering, Chonnam National University
† **Corresponding author**
Tel.: +82-10-3081-5320, Fax: +82-62-530-2158
E-mail: ksyoon@chonnam.ac.kr
Received: November 27, 2020
Revised: February 3, 2021
Accepted: February 3, 2021

light diffraction in water caused by particulate form, and the use of turbidity as a surrogate parameter to estimate T-P is based on that T-P transported in streams predominantly particulate form (Settle et al., 2007; Ruegner et al., 2013). Moreover, Johanna et al. (2014) predicted SS, particulate organic carbon, and particulate nitrogen from turbidity, flow discharge rate, and rainfall depth using MLR. Although, MLR is a simple statistical method to use (Jones et al., 2011; Ziegler et al., 2011; Johanna et al., 2014), and its application is limited to linear relationships, and it is known to be very sensitive to outliers (Chatterjee and Hadi, 1988).

The use of machine learning techniques such as random forest (RF) became affordable with recent advances made in computing resources and techniques, and they have been employed as an analysis tool for hydrologic research (Rodriguez-Galiano et al., 2014; Singh et al., 2017). The RF is a computational algorithm that can predict the values of continuous variables from continuous and/or categorical predictor variables by constructing a decision tree or classifying data (Razi et al., 2005). The algorithm is known not to overfit and efficient in predicting irregular variables that exhibit little periodicity (Diaz-Uriate and de Andres, 2006).

This study explored the potential of the RF algorithm as a tool to efficiently identify primary variables that can serve as surrogate for NPS pollutants. We developed prediction models using MLR (traditional statistical model) and RF (the latest statistical data mining technique) and compared their performance to demonstrate their capacities and potentials. This study focused on SS and T-P that are common NPS pollutants in Korea.

II. Material and methods

1. Study areas and NPS monitoring

This study has monitored SS and T-P at the outlet of a catchment, Wol-jeong (WJ), located within the Pungyeongjeongcheon watershed in Korea for two years, from 01/01/2017 and 12/31/2018 (Fig. 1). WJ is mostly covered by agricultural land uses (86%) including rice paddy fields and upland. The mean annual temperature and precipitation of the study catchment are 13.8°C and 1,391 mm, respectively.

A set of water pressure sensor and logger (OTT Orpheus mini, Germany) was installed to monitor the level of streamflow, and the velocity (VALEPORT model 002, UK) and cross section of flow were measured six times during the monitoring period. The flow level measurements were converted to discharge data using a flow rating curve developed for the outlet of the study catchment. A digital turbidity sensor (FTS DTS-12, Canada) was installed to monitor the turbidity of flow at the monitoring site (or the outlet). The flow duration curve was then developed from flow observations and the flow conditions was seperated using Table 1. Then, we separated the flow conditions into high-flow periods (the exceedance

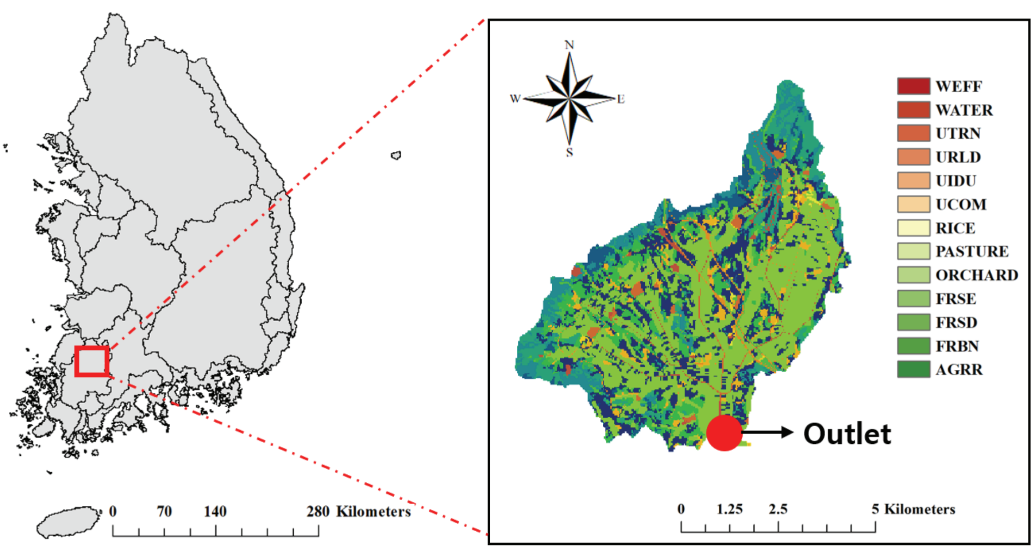


Fig. 1 Locations of the study watersheds and NPS pollutant monitoring stations

Table 1 Classifications of hydrologic conditions

Hydrologic condition class	Flow duration interval (%)
High flows	0～10
Moist conditions	10～40
Mid-range conditions	40～60
Dry conditions	60～90
Low flows	90～100

* Source: USEPA, 2007

probability equal to or less than 10%) and low-flow ones (the exceedance probability greater than 10%) so that the flow-dependent behavior of NPS pollutant transport can be examined in the study. Stormwater samples were taken every hour in the first 24 hours of a storm event using an automatic sampler (ISCO portable sampler 6712, USA), and the sampling interval was increased to 6 hours for the rest of the event. The concentrations of SS and T-P were analyzed in a laboratory according to the standard water pollution test method (APHA, 2001). The monitored water quality variables and predictors such as turbidity and runoff discharge rates are summarized in Table 2.

2. Surrogate model development

The potential relationships between the pollutants of interest (SS and T-P) and the primary environmental variables (discharge rates and turbidity) were explored using the MLR and RF methods. RF is the combination of tree-based classification methods such as a regression tree developed by Breiman, (2001). The processes of growing regression trees can be summarized into two steps. First, data of interest are split into multiple tree branches in the way toward minimizing least-square deviations between observed and predicted variables, which is called the optimal split (OP_s) (Hasanipanah et al., 2017). The split is calculated as the difference of variance between the mother node and the left child node, and the right child node is maximized by using Eqs. 1 and 2 (Ließ et al., 2012; Granata et al., 2017).

$$R(t) = 1/n(t) \times \sum_{i=1}^{n} [y_i(t) - \bar{y}(t)]^2 \quad (1)$$

$$OP_s = R(t) - [R(t_l) + R(t_r)] \quad (2)$$

where *R(t)* is the variance in a node *t*, which is divided into the left and right child nodes, t_l and t_r. *n* is the number of observations, and $\bar{y}$ is the mean value of the predicted variable.

Once tree branches are constructed in the first step, RF prunes the trees by removing nodes that have low explanatory power. Here, the existence of nodes that may be removed imply overfitting. An overfitted model does not sensitively respond to newly added data, and it creates additional uncertainty to the prediction (Breiman et al., 1984). Thus, the pruning process is necessary to reduce prediction uncertainty. The overall procedures of predicting SS and T-P using the MLR and RF models are presented in Fig. 2.

The RF uses a bagging technique with a randomized subset of variables, which is mathematically described using Eq. 3 (Prasad et al., 2006).

$$\hat{f}(x) = \frac{1}{K}\sum_{k=1}^{K} T(x) \quad (3)$$

where $\hat{f}(x)$ is a RF prediction, *K* is the number of trees (or branches), and *T(x)* is the result of each regression tree.

The node size (five) was adapted from literature (Wang et al., 2015) and the'mtry' is the number of predictors to select at random for each split in the tree model. However, we used only two variables in this study, so we used'mtry' of two. The RF model calculates the relative importance of each parameter using the Gini impurity (GI) and mean square error (MSE) statistics. GI quantifies the quality of each split, and MSE calculates the mean decrease of prediction accuracy (Breiman, 2001; Ouedraogo et al., 2018). The GI method is used to

Table 2 Summary statistics of predictors and variables to predict SS and T-P

Variables	Unit	Min.	Median	Mean	Max.	Standard deviation	Number of samples
SS	mg/L	1.1	26.5	35.3	376.4	41.7	134
T-P	mg/L	0.067	0.128	0.143	0.770	0.083	134
Turbidity	NTU	6.6	29.1	42.9	383.0	48.5	134
Runoff	m^3/s	0.4	0.8	0.9	10.8	0.7	134

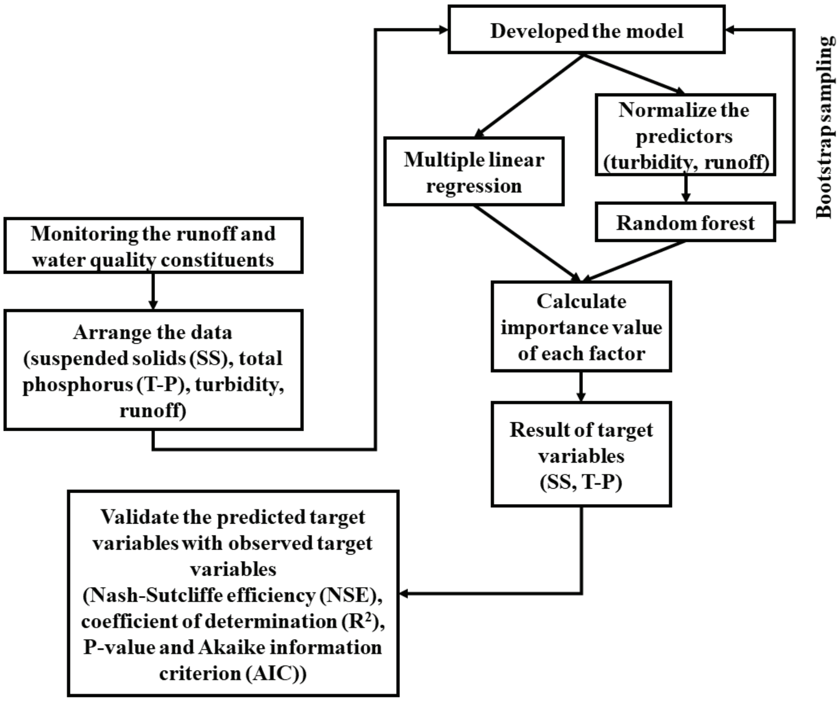


Fig. 2 Overall procedures of predicting SS and TP using random forest and multiple linear regression

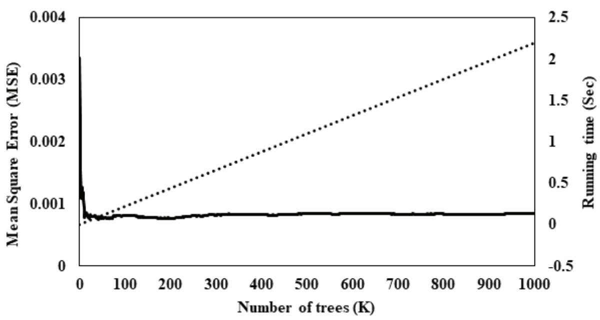


Fig. 3 Responses of MSE to the number of trees (K)

calculate the quality of each split for the variables in a tree model and calculate the mean decrease using MSE due to splits on every predictor (Breiman, 2001). In this study, we found that *K* of 100 is enough to stabilize the MSE. However, the running time of RF slightly increased until *K* value hit 1,000. We selected K value of 1,000 while still providing reasonably short computing time (Fig. 3). The relative importance was calculated by running the RF algorithm 100 times and then averaging the importance rates calculated from the iterations.

It is known that one of the considerations in the machine learning technique is that training variables with different range affects the learning rates and it can be crucial problems with prediction performance (Ioffe and Szegedy, 2015). For example, turbidity ranges from 6.6 to 383 NTU, and runoff ranges from 0.6 to 3.2 m^3/s in this study. Then, the turbidity and runoff discharge rate data were normalized into the range from 0 to 1 using Eq. 4 to assure that each variable can have the similar level of explanatory power in the analysis.

$$\grave{x} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{4}$$

where x' is the normalized value of the data set (range of 0 to 1) and x is the original value.

The flow observations made from 2017 to 2018 showed that the flow discharge rate of 1.08 m^3/s correspond to the 10% discharge of the study catchment (Fig. 4). The threshold rate of 1.08 m^3/s was used to separate discharge into high and low flow, and then the models were separately applied to the two flow regimes.

The performance of the MLR and RF models was evaluated with the split-sample scheme of 80/20 (80% for training and 20% for validation; Afendras and Markatou, 2019). To validate the models, we used four different statistics and criteria: the Nash-Sutcliffe efficiency (NSE) coefficient, the coefficients of

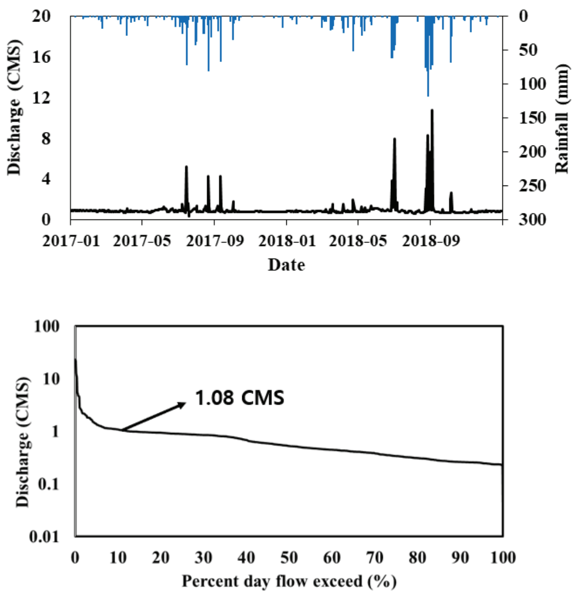


Fig. 4 Daily runoff and flow duration curve of the study site

determination (R^2), p-value (P), and the Akaike information criterion (AIC) (Nash and Sutcliffe, 1970; Akaike, 1973; Snipes and Taylor, 2014).

$$NSE = 1 - \frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \tag{5}$$

$$R^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \tag{6}$$

$$AIC = 2k - 2\ln(L) \tag{7}$$

$$P = \Pr(T \le -|t|H) + \Pr(T \ge +|t|H) \tag{8}$$

where X_i and Y_i is the predicted value and observed values from site i and n is the number of samples. In the AIC method, k represents the number of parameters, and L is the likelihood function, which represent the model fit and it should represented by log-value to limit the maximum value of model fit.

A lower AIC score is interpreted as a more accurate model. The p-value is defined as the probability (*Pr*), under the null htpothesis H about the unknown distribution of the T. If the p-value shows very small value, then the statistical significance is thought to be very large. On the other hand, the criteria commonly set to 0.05, 0.01, 0.005, or 0.001 and we choose 0.001 as criteria in this study.

Ⅲ. Results and discussion

The comparison between observation and prediction result was conducted with normalized value, and the normalized SS and T-P prediction results of the MLR and RF models are compared with the normalized observation values in Fig. 5 and Table 3. Overall, the MLR model prediction of SS and T-P in the low and high flow conditions were acceptable with R^2 and NSE greater than 0.5, except T-P in low flow condition. The T-P prediction for the low flow condition shows acceptable R^2 but low NSE. The reason for this result is that T-P in agricultural area shows high concentrations for high flow conditions during farming period (Villa et al., 2019). Thus, high flow conditions shows acceptable prediction performance with high concentrations of T-P which shows high fluctuations and low flow condition shows unacceptable prediction performance due to low concentrations of T-P. In general, SS and T-P were more accurately predicted by the MLR models when runoff discharge is relatively large (or the high flow condition). The RF model provided accuracy and its patterns similar to that of the MLR model.

As both models provided similar accuracy when predicting SS and T-P from turbidity and runoff discharge rates, their performance was further investigated in terms of AIC values (Table 3). The AIC values of the RF model are lower than those of the MLR model, indicating that the RF model is more suitable for predicting SS and T-P that the MLR model in this study.

The relative importance of the primary environmental variables, runoff discharge rates and turbidity, to predict SS and T-P was quantified as part of the RF modeling (Fig. 6). The analysis of relative importance showed that turbidity was more closely associated with the SS and T-P concentrations than discharge rates with importance rates ranging from 52% to 86% depending on the flow conditions. As mentioned above, the runoff in high flow condition shows higher relative importance than low flow condition, which is related to the high concentrations of T-P in farming period.

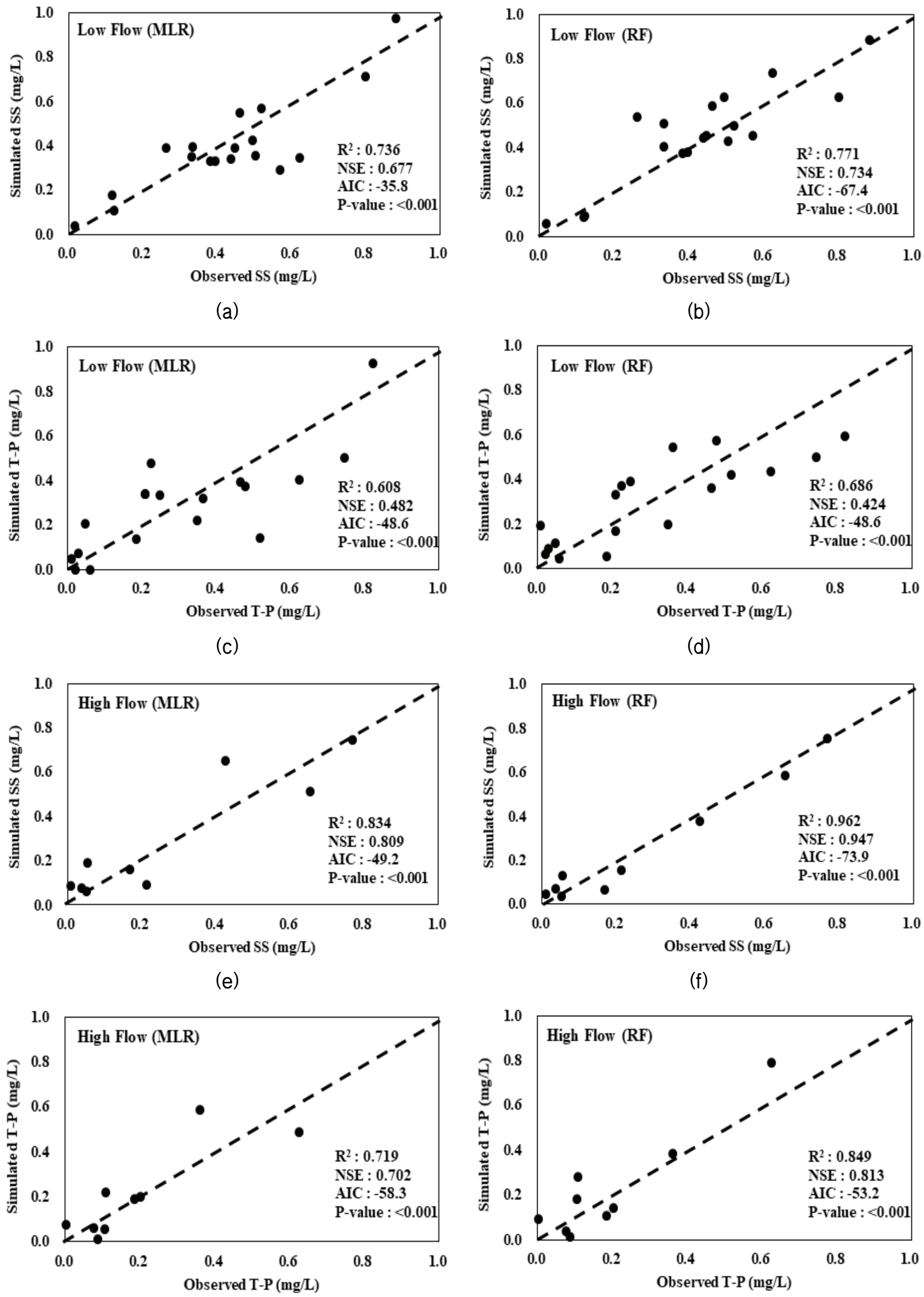
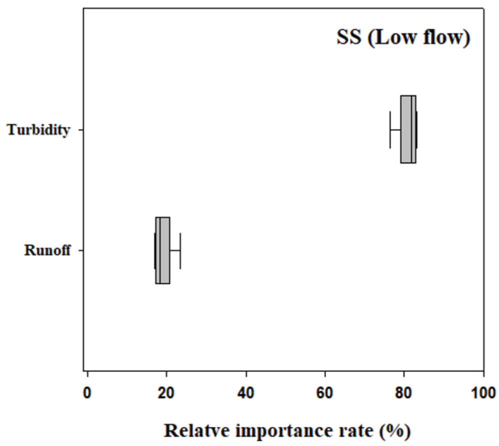


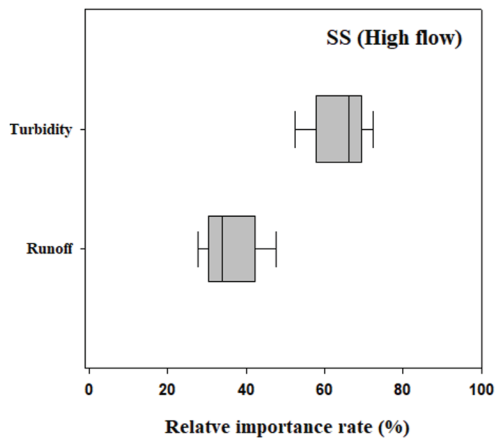
Fig. 5 Comparison between the NPS pollutants observed and simulated using the MLR and RF models (a) SS using MLR in low flow condition; (b) SS using RF in low flow condition; (c) T-P using MLR in low flow condition; (d) T-P using RF in low flow condition. (e) SS using MLR in high flow condition; (f) SS using RF in high flow condition; (g) T-P using MLR in high flow condition; (h) T-P using RF in high flow condition

Table 3 Performance evaluation of the MLR and RF models in predicting SS and T-P concentrations

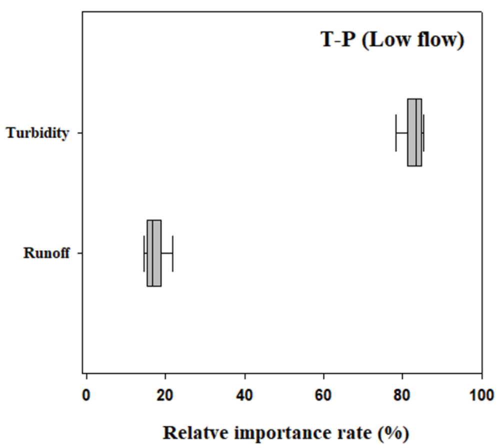
Model	Indicator	Low flow condition		High flow condition	
		SS	T-P	SS	T-P
MLR	R^2	0.74	0.61	0.83	0.72
	NSE	0.68	0.48	0.81	0.70
	AIC	-35.8	-48.6	-49.2	-58.3
	P-value	< 0.001	< 0.001	< 0.001	< 0.001
RF	R^2	0.77	0.69	0.96	0.85
	NSE	0.73	0.42	0.95	0.81
	AIC	-67.4	-48.6	-73.9	-53.2
	P-value	< 0.001	< 0.001	< 0.001	< 0.001



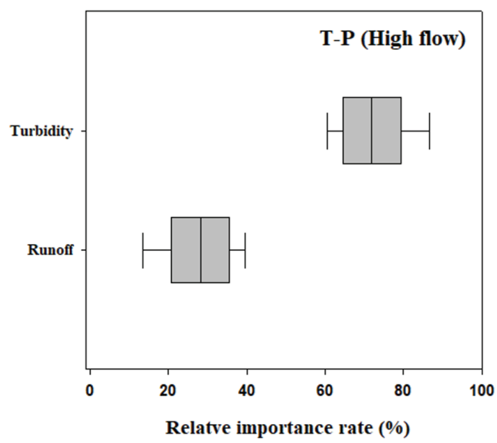
(a) SS in low flow condition



(b) SS in high flow condition



(c) T-P in low flow condition



(d) T-P in high flow condition

Fig. 6 Relative importance rates of runoff discharge rates and turbidity in predicting SS and T-P concentrations

Ⅳ. Conclusions

This study explored the potential of the MLR and RF model as a tool to identify surrogate environmental variables in two different flow conditions for the concentrations of NPS pollutants such as SS and T-P in agricultural sub-watershed. The performance of the MLR and RF models was evaluated with the 80% for training set and 20% for validation set. Both MLR and RF models provided acceptable performance when predicting SS and T-P in the high-flow condition. The relative importance rate analysis showed that turbidity had relatively high explanatory power (52% to 86%) than runoff discharge rates. Moreover, the runoff in high flow condition shows higher relative importance than low flow condition, which is related to the high concentrations of T-P in farming period. This study demonstrated the machine learning techniques could help to improve the efficiency of NPS pollutant monitoring and prediction by identifying fundamental environmental variables that are relatively easily obtained but still closely related to the NPS pollutants. The results also showed that turbidity could serve as a surrogate water quality parameter for SS and T-P concentrations at acceptable accuracy levels.

감사의 글

본 연구는 BK21 기후지능형간척지농업 교육연구팀 장학금 지원과 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2017R1D1A3B03029713).

REFERENCES

1. Afendras, G., and M. Markatou, 2019. Optimality of training/test size and resampling effectiveness in cross-validation. *Journal of Statistical Planning and Inference* 199: 286-301. doi:10.1016/j.jspi.2018.07.005.
2. Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. *In Second International Symposium on Information Theory*, 267-281.
3. American Public Health Association (APHA), 2001. Standard Methods for the Examination of Water and Waste Water. 21st ed, Washington DC, USA.
4. Breiman, L., 2001. Random forests. *Machine Learning* 45: 5-32.
5. Breiman, L., J. Friedman, R. A. Olshen, and C. J. Stone, 1984. Classification and Regression Trees. Wadsworth: CRC press.
6. Camdevyren, H., N. Demyr, A. Kanik, and S. Keskyn, 2005. Use if principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs. *Ecological Modelling* 181(4): 581-589. doi:10.1016/j.ecolmodel.2004.06.043.
7. Chattergee, S., and A. S. Hadi, 1988. Sensitivity Analysis in Linear Regression. USA, Wiley.
8. Chenini, I., and S. Khemiri, 2009. Evaluation of ground water quality using multiple linear regression and structural equation modeling. *International Journal of Environmental Science and Technology* 6(3): 509-519. doi:10.1007/BF03326090.
9. DeForest, D. K., K. V. Brix, L. M. Tear, and W. J. Adams, 2018. Multiple linear regression models for predicting chronic aluminum toxicity to freshwater aquatic organisms and developing water quality guidelines. *Environmental Toxicology and Chemistry* 37(1): 80-90. doi:10.1002/etc.3922.
10. Diaz-Uriate, R., and S. A. de Andrés, 2006. Gene selection and classification of microarray data using random forest. *BMC Bio*informatics 7: 3. doi:10.1186/1471-2105-7-3.
11. Granata, F., S. Papirio, G. Esposito, R. Gargano, and G. Marinis, 2017. Machine learning algorithms for the forecasting of wastewater quality indicators. *Water* 9(2): 105. doi:10.3390/w9020105.
12. Hasanipanah, M., R. S. Faradonbeh, H. B. Amnieh, D. J. Armaghani, and M. Monjezi, 2017. Forecasting blast-induced ground vibration developing a CART model. *Engineering with Computers* 33: 307-316. doi:10.1007/s00366-016-0475-9.
13. Horsburgh, J. S., A. S. Jones, D. K. Stevens, D. G. Tarboton, and N. O. Mesner, 2009. A sensor network for high frequency estimation of water quality constituent fluxes using surrogates. *Environmental Modelling and Software* 25(9): 1031-1044. doi:10.1016/j.envsoft.2009.10.012.
14. Houser, J. N., P. J. Mulholland, and K. O. Maloney, 2006. Upland disturbance affects headwater stream nutrients and suspended sediments during baseflow and stormflow. *Journal of Environmental Quality* 35: 352-365. doi:10.2134/jeq2005.0102.
15. Johanna, I. F., and S. Petra, 2014. A turbidity-based

method to continuously monitor sediment, carbon and nitrogen flows in mountainous watersheds. *Journal of Hydrology* 513: 45-57. doi:10.1016/j.jhydrol.2014.03.034.

16. Jones, A. S., K. S. David, S. H. Jefery, and O. Nancy, 2011. Surrogate measures for providing high frequency estimates of total suspended solids and total phosphorus concentrations. *Journal of the American Water Resources Association* 47(2): 239-253. doi:10.1111/j.1752-1688.2010.00505.x.
17. Jordan, P., A. Arnscheidt, and H. Mcgrogan, 2007. Characterising phosphorus transfers in rural catchments using a continuous bankside analyser. *Hydrology and Earth System Science* 11: 372-381. doi:10.5194/hess-11-372-2007.
18. Ließ, M., B. Glaser, and B. Huwe, 2012. Uncertainty in the spatial prediction of soil texture comparison of regression tree and random forest models. *Geoderma* 170: 70-79. doi:10.1016/j.geoderma.2011.10.010.
19. Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith, 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE.* 50(3): 885-900. doi:10.13031/2013.23153.
20. Montgomery, J. L., T. C. Harmon, C. N. Haas, R. Hooper, N. L. Clesceri, W. Graham, W. J. Kaiser, A. Snaderson, B. Minsker, J. Schnoor, and P. Brezonik, 2007. The waters network: an integrated environmental observatory network for water research. *Environmental Science and Technology* 41(19): 6642-6647. doi:10.1021/es072618f.
21. Nash, J. E., and J. V. Sutcliffe, 1970. River flow forecasting through conceptual models part I – a discussion of principles. *Journal of Hydrology.* 10(3): 282-290. doi:10.1016/0022-1694(70)90255-6.
22. Ouedraogo, I., P. Defourny, and M. Vanclooster, 2018. Application of random forest regression and comparison of its performance to multiple linear regression in modeling groundwater nitrate concentration at the African continent scale. *Hydrogeology* Journal 27(3): 1-18. doi:10.1007/s10040-018-1900-5.
23. Prasad, A. M., L. R. Iverson, and A. Liaw, 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9: 181-199. doi:10.1007/s10021-005-0054-1.
24. Razi, M. A., and K. A. Athappilly, 2005. A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Systems with Applications* 29(1): 65-74. doi:10.1016/j.eswa.2005.01.006.
25. Rodriguez-Galiano, V., M. P. Mendes, M. J. Garcia-soldado, M. Chica-Olmo, and L. Ribeiro, 2014. Predictive modeling of groundwater nitrate pollution using random forest and multisource variables related to intrinsic and specific vulnerability: a case study in an agricultural setting (Southern Spain). *Science of the Total Environment* 476-477: 189-206. doi:10.1016/j.scitotenv.2014.01.001.
26. Ruegner, H., M. Schwientek, B. Beckingham, B. Kuch, and P. Grathwohl, 2013. Turbidity as a proxy for total suspended solids (TSS) and particle facilitated pollutant transport in catchments. *Environmental Earth Sciences* 69: 373-380. doi:10.1007/s12665-013-2307-1.
27. Scholefield, D., T. L. Goff, J. Braven, L. Dbdon, T. Long, and M. Butler, 2005. Concerted diurnal patterns in riverine nutrient concentrations and physical conditions. *Science of the Total Environment* 344: 201-210. doi:10.1016/j.scitotenv.2005.02.014.
28. Settle, S., A. Goonetilleke, and G. Ayoko, 2007. Determination of surrogate indicators for phosphorus and solids in urban stormwater: application of multivariate data analysis techniques. *Water, Air, and Soil Pollution* 182: 149-161. doi:10.1007/s11270-006-9328-2.
29. Singh, B., P. Sihag, and K. Singh, 2017. Modeling of impact of water quality on infiltration rate of soil by random forest regression. *Modeling Earth Systems and Environment* 3: 999-1004. doi:10.1007/s40808-017-0347-3.
30. Snipes, M., and C. D. Taylor, 2014. Model selection and Akaike information criteria: an example from wine ratings and prices. *Wine Economics and Policy* 3(1): 3-9. doi:10.1016/j.wep.2014.03.001.
31. United States Environmental Protection Agency (USEPA), 2007. “An approach for using load duration curves in the development of TMDLs.” 841-B-07-006, United States Environmentl Protection Agency, 1-68.
32. Verzani, J., 2018. Data sets, etc. for the text “Using R for introductory statistics”. 2nd ed. Version 2.0-6.
33. Villa, A., J. Fölster, and K. Kyllmar, 2019. Determining suspended solids and total phosphorus from turbidity: comparison of high-frequency sampling with conventional monitoring methods. *Environmental Monioring and Assessment* 191: 605. doi:10.1007/s10661-019-7775-7.
34. Wang, H., Y. Zhao, R. L. Pu, and Z. Z. Zhang, 2015.

Mapping Robinia pseudoacacia forest health conditions by using combined spectral, spatial, and textural information extracted from IKONOS imagery and random forest classifier. *Remote Sensing* 7(7): 9020-9044. doi:10.3390/rs70709020.
35. Zambrano-Bigiarini, M., 2017. Goodness-of-fit functions for comparison of simulated and observed hydrological time series. Version 0.3-10.
36. Ziegler, A. D., L. X. Xi, and C. Tantasarin, 2011. Sediment load monitoring in the Mae Sa catchment in Northern Thailand, *IAHS-AISH publication* 86-91.