

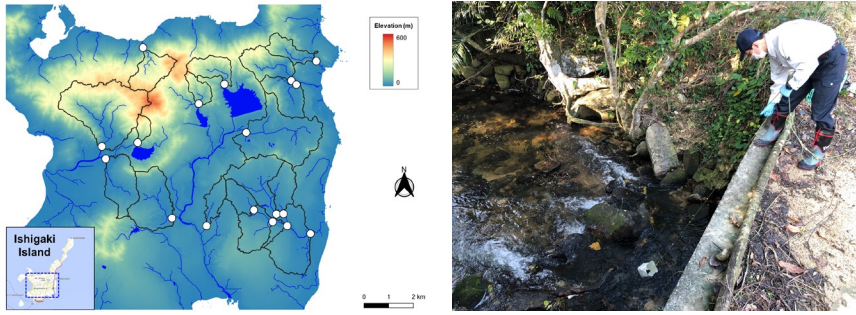
## Predicting riverine nutrient concentrations from catchment characteristics using a machine learning method

Excessive loadings of terrestrial nitrogen and phosphorus, and their imbalances with silicon have been recognized as one of the major causes of water quality and ecosystem deterioration in receiving coastal waters around tropical islands. In this study, water quality was monitored every two months for one year in the rivers and streams of Ishigaki Island, a tropical island in Japan (Fig. 1), to analyze how the catchment characteristics (e.g., land use, geology) influence the concentrations of dissolved inorganic nitrogen (DIN), total phosphorus (TP) and dissolved silicon (DSi).

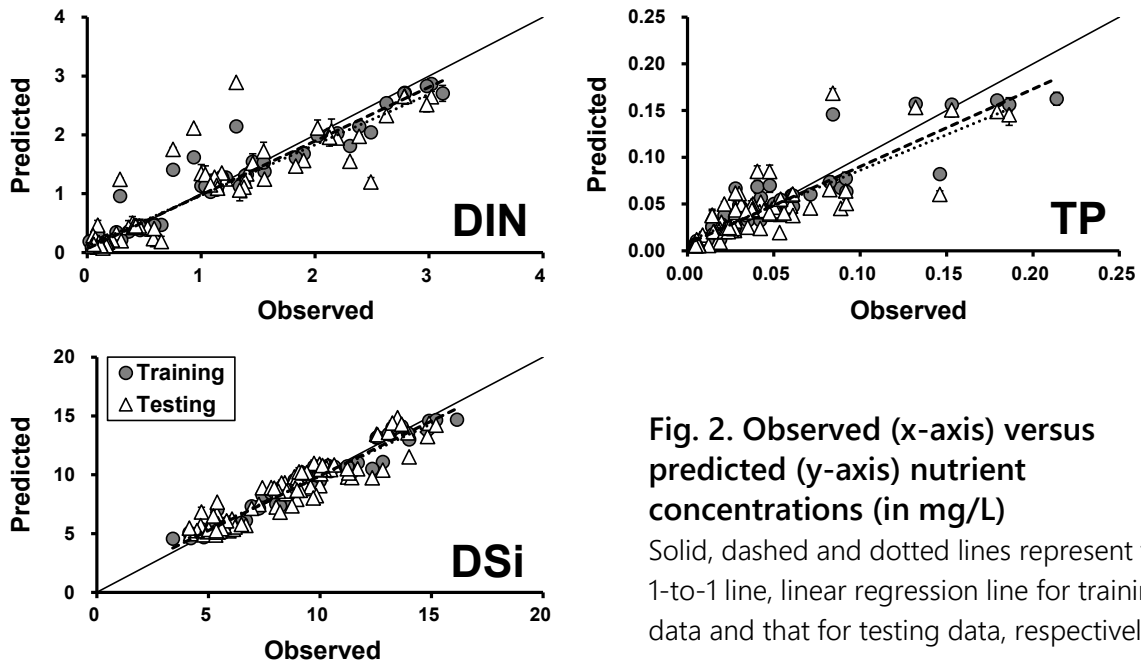
Based on the monitoring results, predictive models for nutrient concentrations from the catchment properties were created using Random Forest (RF) machine learning algorithm. Coefficient of determination ( $R^2$ ) between the observed and predicted values were 0.87 ~ 0.96 and 0.69 ~ 0.88 for training and testing data sets, respectively, and the percent bias (PBIAS) were -0.33 ~ 0.36% and -1.30 ~ 0.06%, demonstrating that the created models can predict nutrient concentrations with sufficient accuracy (Fig. 2). Contributions of the catchment properties to nutrient concentrations can be also evaluated from their importance in the RF models in combination with the correlation coefficients with nutrient concentrations. Agricultural land uses (e.g., livestock barn, sugarcane field) were ranked as the most important parameters in the RF models for DIN and TP, and they also had strong positive correlations (Spearman's rank correlation coefficients being more than 0.7) with these nutrients (Fig. 3). Considering that fact that sugarcane cultivation and beef calf raising are active in Ishigaki Island, these results suggest that these agricultural activities have strong influences on riverine nitrogen and phosphorus concentrations. On the other hand, forest was not only the most important parameter in the RF model for DSi but also strongly positively correlated with DSi concentration (Fig. 3), suggesting that forest serves as a dominant source of riverine Si.

As the RF models can be operated in a relatively easy way, they can be utilized for formulating action plans toward conservation of coastal water quality and ecosystems (e.g., identification of rivers that need to be monitored with respect to nutrient loading to coastal water). These models would also be applicable for predicting the variations in riverine nutrient concentrations due to the change in catchment land use and management, while the predictions need to be validated through the comparison with those of process-based hydrologic models.

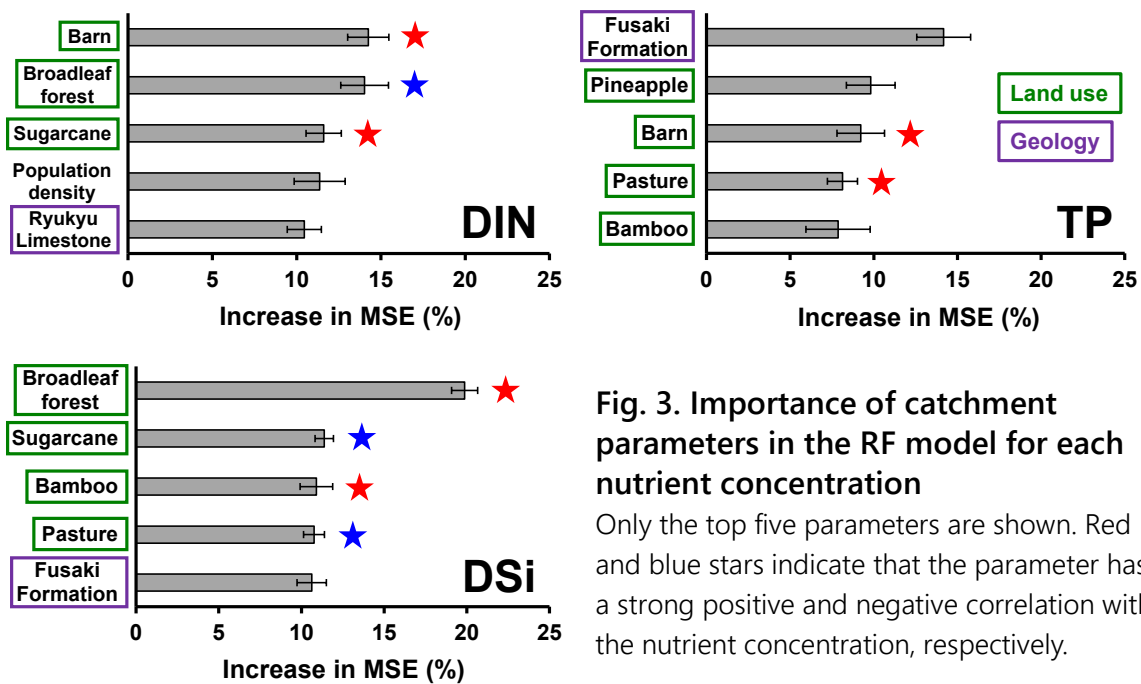
*(T. Kikuchi, T. Anzai)*



**Fig. 1. Study site (left); river water sampling (right)**  
White circles and black lines in the left panel represent the sampling stations and catchment boundaries, respectively.



**Fig. 2. Observed (x-axis) versus predicted (y-axis) nutrient concentrations (in mg/L)**  
Solid, dashed and dotted lines represent the 1-to-1 line, linear regression line for training data and that for testing data, respectively.



**Fig. 3. Importance of catchment parameters in the RF model for each nutrient concentration**  
Only the top five parameters are shown. Red and blue stars indicate that the parameter has a strong positive and negative correlation with the nutrient concentration, respectively.

Reference: Kikuchi et al. (2023) *Environmental Pollution* 316: 120599.  
Figures reprinted/modified with permission.