EGU25-11162 Session HS4.10
Recent advances in (hybrid) hydrological forecasting using physically-based and machine learning models
Fri, 2 May, 14:00–15:45, Hall A, EGU 2025 General Assembly

A multi-stage machine learning application for predicting vegetation distribution and its factors in river channels

Hitoshi Miyamoto and Naoya Maeda

Shibaura Institute of Technology (SIT), Tokyo, JAPAN, e-mail: miyamo@shibaura-it.ac.jp

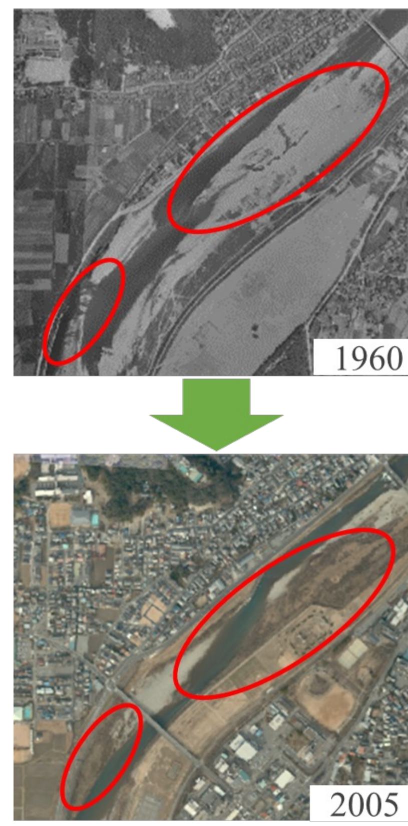
Research Background & Purposes:

Vegetation overgrowth and forest expansion in fluvial floodplains and sand bars have become a serious engineering problem for river management in Japan and around the world.

- decreasing ***Water Flow Capacity at Flooding***
- changing ***Groundwater*** **Levels,**
  Hyporheic **Processes,**
  Sediment and Nutrient Cycles,
  Riparian Ecosystems and
  Original Riverine Landscapes

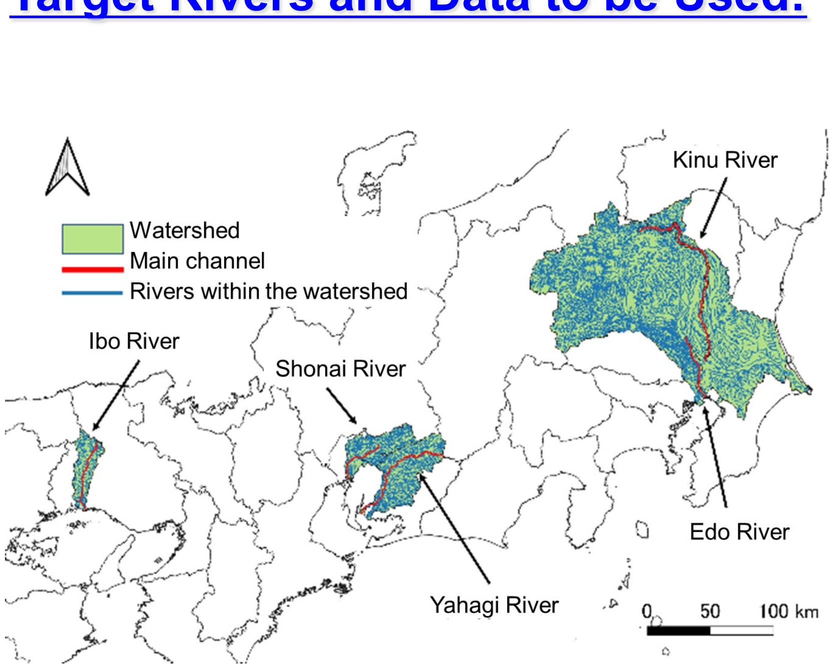
From both viewpoints of flood control and ecological conservation, it would be necessary to continuously monitor the vegetation dynamics for a long period of time.

This study developed an ML (machine learning) model that predicts the vegetation distribution in the following year by the ML model in multiple stages with current year's conditions



River channel in Ibo River (Japan Map, GSI Map, Geospatial Information Authority of Japan)

Target Rivers and Data to be Used:



Background Japan Map: National Land Information Division, National Spatial Planning and Regional Policy Bureau, MLIT of Japan

The target rivers were five Japanese large rivers, i.e., *Kinu River*, *Edo River*, *Yahagi River*, *Shonai River*, and *Ibo River*. The multi-stage ML model's explanatory and target variables were created for each river segment using *DEMs (Digital Elevation Models)* and *river environment base maps*.

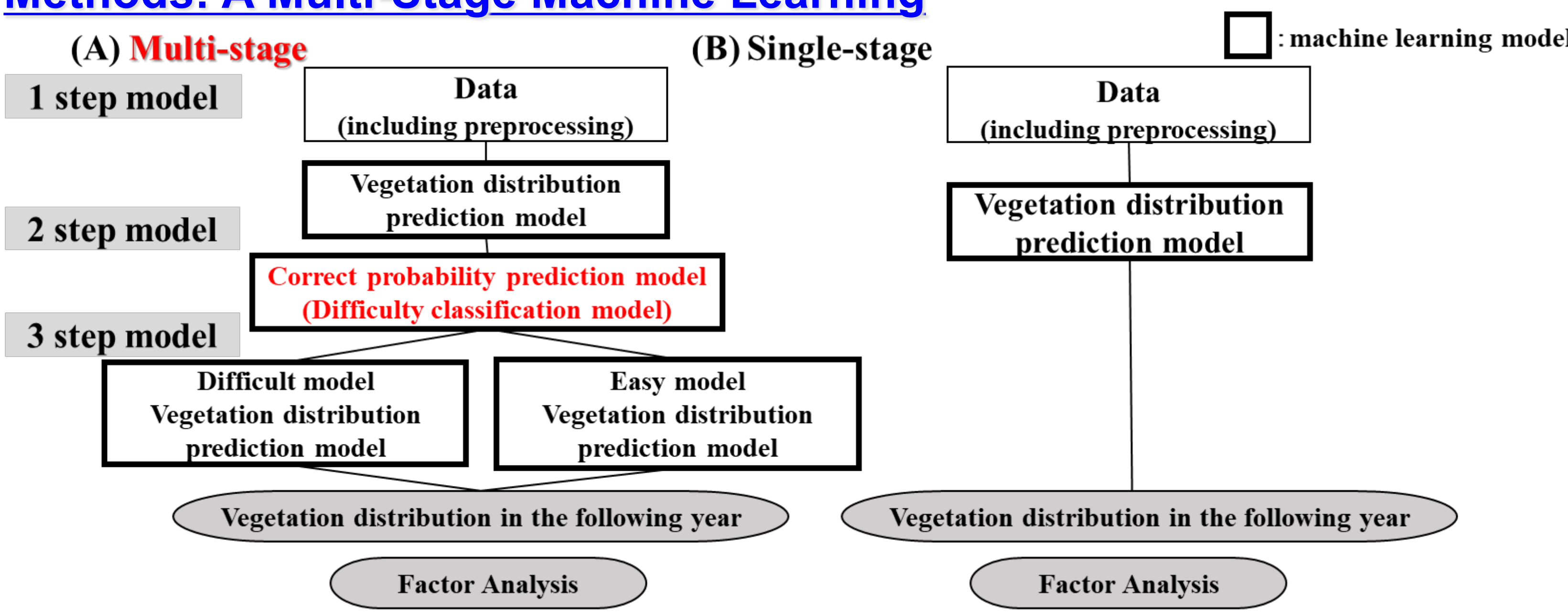
Explanatory variables:

- **Distance from Water (*WD*),**
- **Relative Height (*RH*),**
- **Distance from Vegetation History (*VD*),**
- **Distance from Grass/Tree History(*GTD*)**

Objective variable:

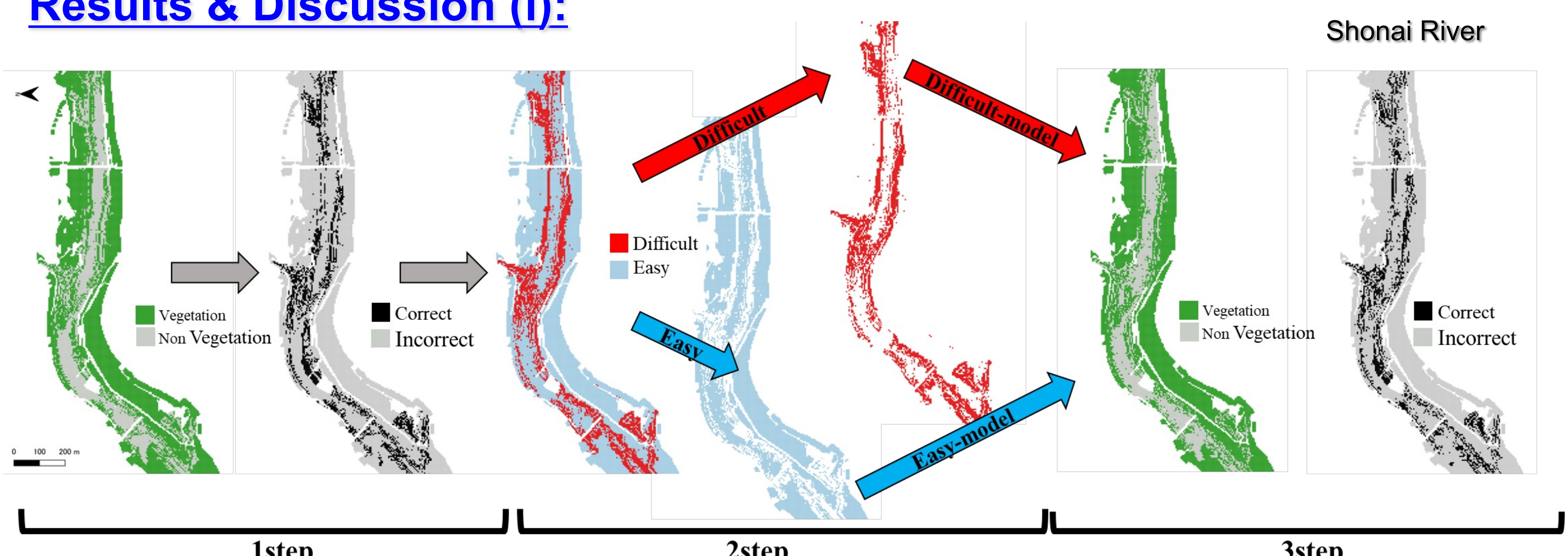
- **Vegetation Distribution (*VG*)**

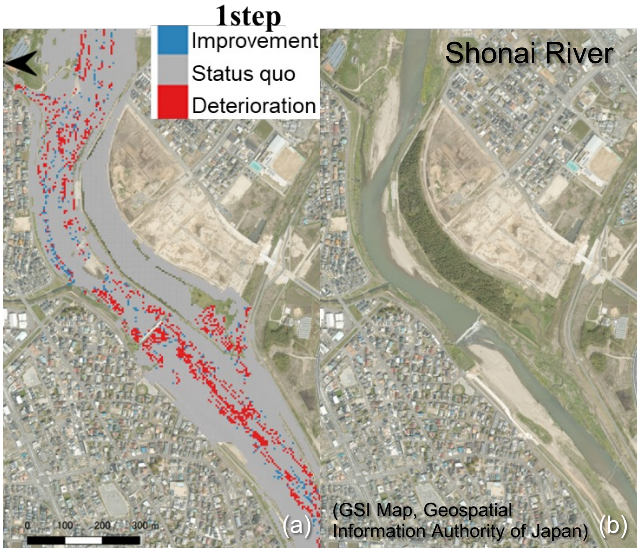
Methods: A Multi-Stage Machine Learning



- **The multi-stage ML model consisted of three ML stages to predict the vegetation distribution of the following year from the current river vegetation distribution and topographical information.**
- **The advantage of the multi-stage ML model was that a third-stage vegetation distribution prediction model could be constructed according to the difficulty of prediction using the second-stage classification result.**
- **XGB (eXtreme Gradient Boosting) was used as the machine learning model. SHAP (SHapley Additive exPlanations) was used for factor analysis.**

Results & Discussion (i):

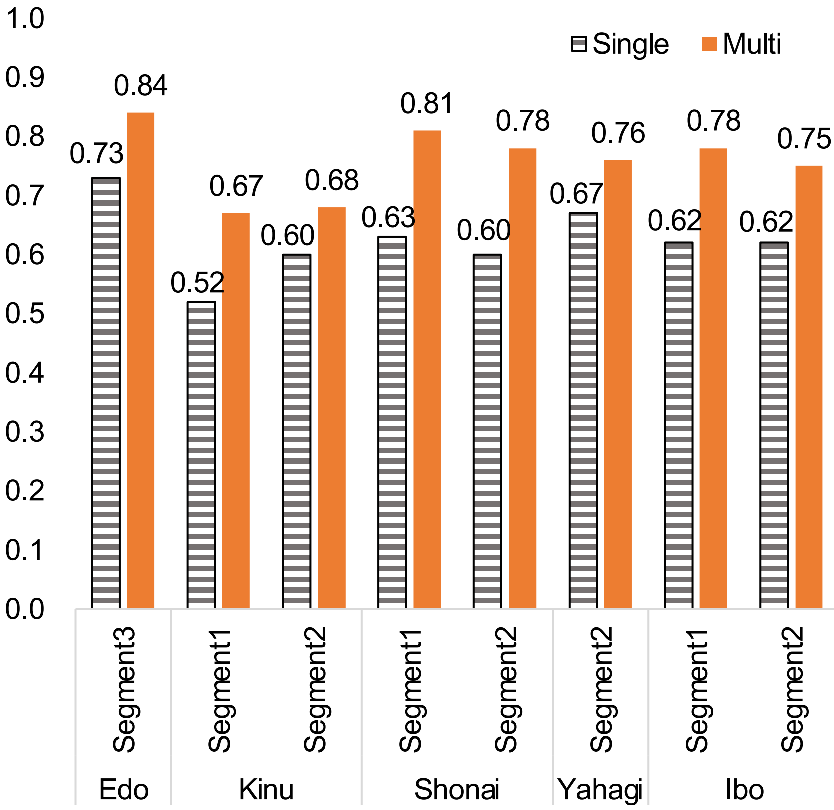




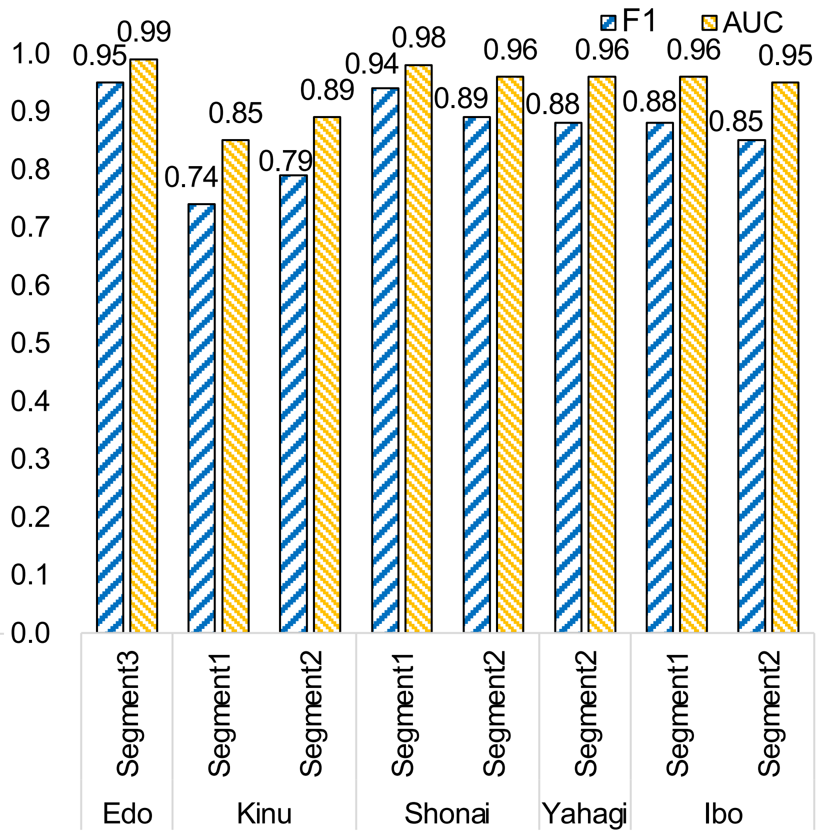
- **The 2step in the ML model** showed that the data **were well divided into the two classes with different difficulty levels**. The difficult ML model selected the locations in **the boundaries between vegetation and water surface, sand and gravel**.
- The final results of the multi-stage ML model indicated that the **3step ML model detected more accurate locations** than the 1step model **at the boundaries**.

(a) Accuracy Comparison Map (b) Aerial Photo
Improvement and Deterioration of the Multi-Stage ML Models

Results & Discussion (ii):



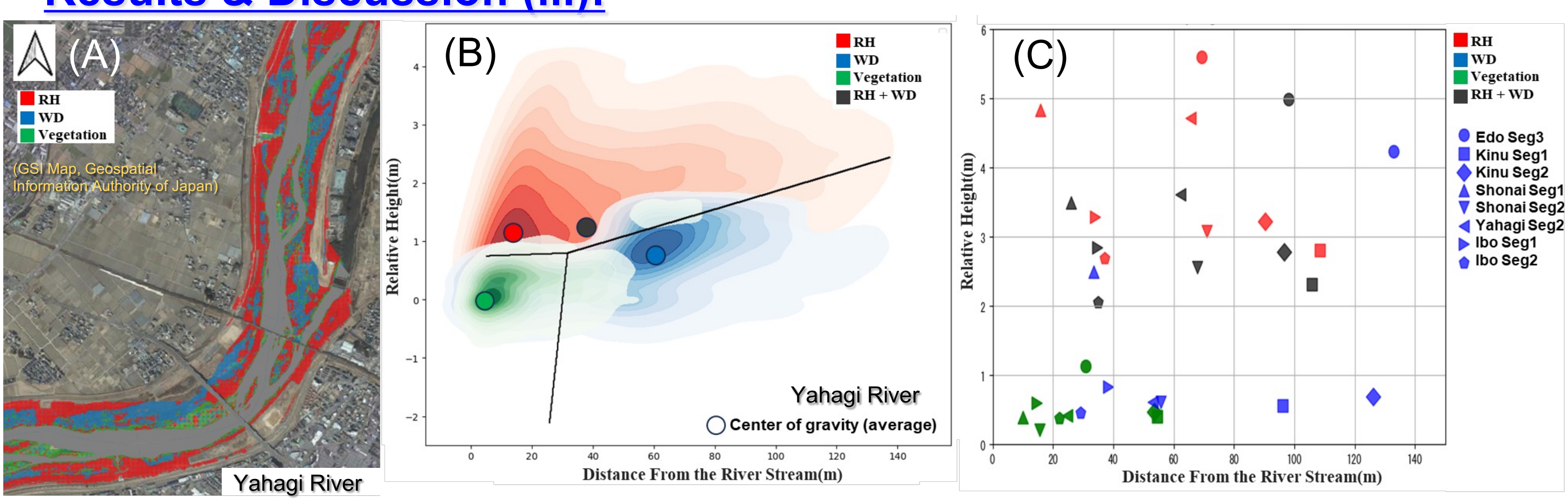
F1 scores at hard-to-predict locations



F1 scores and AUCs in the multi-stage ML model

- **The result of the multi-stage ML model for the five target rivers showed that *the F1 score was 0.8 or higher* for all rivers except the Kinu River.**
- **The multi-stage ML model had an accuracy of *10% higher F1 score* than the conventional single ML model.**

Results & Discussion (iii):



(A) Distribution of the main factor contributing to the prediction, (B) Density map of the main factors (C) Distribution of the spatial locations of the main factors in the five target rivers.

- ***SHAP analysis* revealed the three prominent factors for vegetation existence: *(i) the relative height near the levee and in the center of the floodplain*, *(ii) the distance from the river water's edge near the low water channel*, and *(iii) the vegetation existence history at the boundary between the low water channel and the floodplain.***
- **These results suggest that combining the prediction map with factor analysis could identify the factors influencing where vegetation recruits in a river course.**

References: 1)Caponi, F.,et al. (2019). When Does Vegetation Establish on Gravel Bars? Observations and Modeling in the Alpine Rhine River. *Frontiers in Environmental Science*. 2) Maeda, N., & Miyamoto, H. (2024). A Logistic Regression Model for the Prediction of Vegetation Recruitment in the Kinu River, Japan. *River Research and Applications*.