Supplement of Biogeosciences, 23, 683–708, 2026
https://doi.org/10.5194/bg-23-683-2026-supplement


Biogeosciences
Open Access

*Supplement of*

# Upscaling of soil methane fluxes from topographic attributes derived from a digital elevation model in a cold temperate mountain forest

**Sumonta Kumar Paul et al.**

*Correspondence to:* Sumonta Kumar Paul (paulsumonta@gmail.com) and Daniel Epron (daniel.epron.3a@kyoto-u.ac.jp)

**Sect. S1. Non-waterlogged and wetland classification**

To distinguish non-waterlogged and wetland areas, we collected additional GPS positions at the edges and within the three wetland patches, in addition to the positions of the 55 sampling points. These observations were used to train a random forest classification model to predict the non-waterlogged and wetland across the study area. Four topographic attributes—SAGA wetness index (SWI), Profile curvature (PrC), slope, and vertical distance to channel network (VDCN) were used as predictors, as they represent key hydrological controls on moisture accumulation. A Random Forest classifier was used for categorical prediction. The RF classifier was implemented using the R-packages "caret" (Kuhn and Johnson, 2013) and "randomForest" (Liaw and Wiener, 2002). Accuracy was evaluated using a confusion matrix.

The model was trained on 70% of the observations, with the remaining 30% reserved for independent testing. The mtry parameter, which determines the number of randomly selected predictor variables at each node, was tested from 2 to n-1 (n being the total number of predictors), and the ntree parameter was set to 500, ensuring the model constructed an ensemble of 500 decision trees.

The model showed an overall classification accuracy of 84%. When applied across the entire study area, visual inspection of the classification results revealed that 303 of the 455 pixels labeled as wetland were in fact non-waterlogged areas, indicating a tendency of the model to overestimate wetland extent. However, predictions for non-waterlogged areas were highly accurate: no pixels classified as non-waterlogged were found within the wetland patches, demonstrating strong model specificity for the non-waterlogged class. A posteriori, pixels classified as wetland had SWI values above 8.1, profile curvature between -0.003 and 0.001, slope values below 6.8, and VDCN values below 2.2 (Fig. S1).

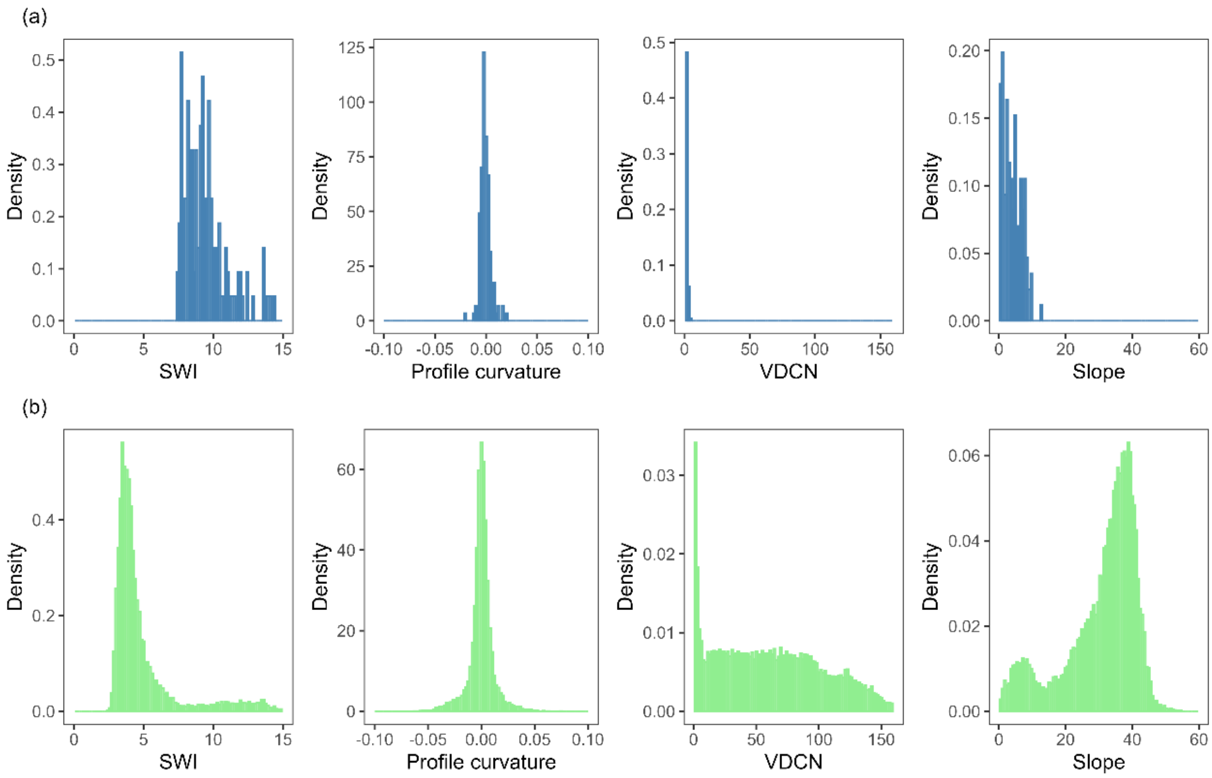


**Fig. S1: Plot comparing the distribution of the topographic variables between (a) wetland pixels (in blue) and (b) non-waterlogged pixels (in green).**

**Sect. S2. Vegetation classification**

Tree inventory was conducted to categorized the vegetation surrounding the flux measurement points. A circular plot with a 10-meter radius was established, centered at each flux measurement point. We calculated the plot basal area (BA) as the sum of the cross-sectional areas (CSA) at breast height of all tree trunks in each plot, and subsequently determined the proportional contribution of coniferous trees ($RBA_{CON}$) for each plot.

A random forest regression (RF) model was employed to create spatially explicit maps of plot basal area (BA) and the relative basal area of coniferous species ($RBA_{CON}$) across the study area. Topographic predictors used as predictors in the RF models to estimate BA and $RBA_{CON}$ were derived from a digital elevation model (DEM), including the SAGA wetness index (SWI), topographic position index (TPI), and vertical distance to channel network (VDCN). A NDVI layer produced from orthoimages acquired during the flux measurement period to represent spatial variation in canopy greenness was added. All predictor variables were first harmonized to a 10-m resolution to calibrate and validate the RF models against the field observations. RF models were trained to predict BA and $RBA_{CON}$ using the R-packages "caret" (Kuhn and Johnson, 2013) and "randomForest" (Liaw and Wiener, 2002). The mtry parameter, which determines the number of randomly selected predictor variables at each node, was tested from 2 to n-1 (n being the total number of predictors) using leave-one-out cross-validation to minimize prediction error and maximize the variance explained by the model. The ntree parameter was set to 500, ensuring the model constructed an ensemble of 500 decision trees.

Model accuracy was assessed by the coefficient of determination ($R^2$) and root mean square error (RMSE), with $R^2$ = 0.45, RMSE = 0.87, and $R^2$ = 0.51, RMSE = 0.26 for BA and $RBA_{CON}$, respectively. Model performance was evaluated by comparing predicted and observed values. For both BA and $RBA_{CON}$, the slopes did not differ significantly from 1 (Fig. S2).

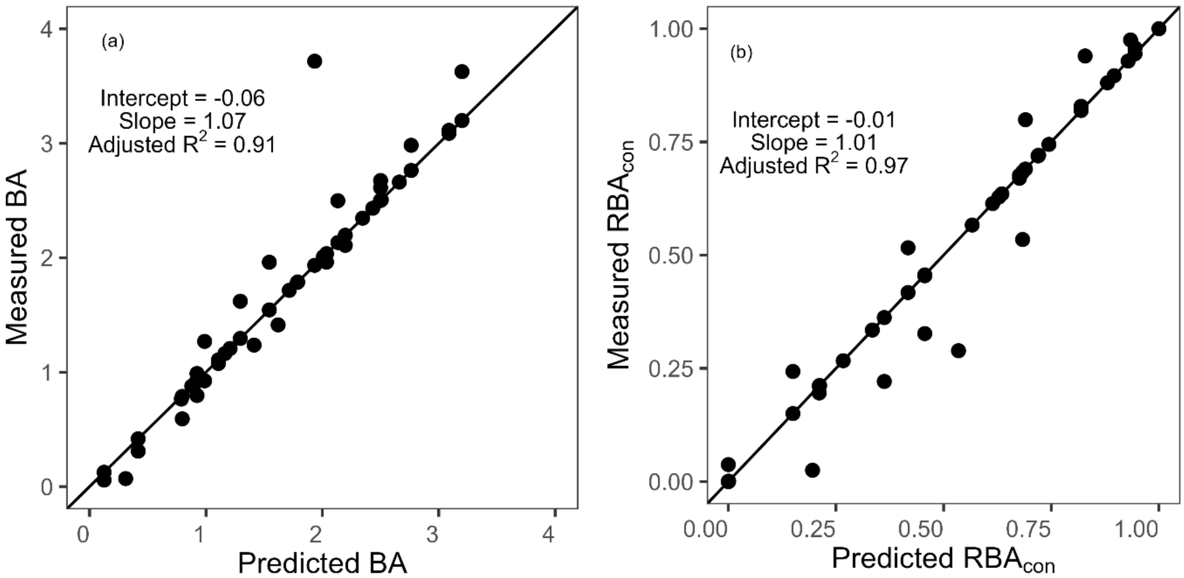


**Fig. S2: Comparison between predicted and measured (a) basal area (BA) and (b) proportional contribution to the coniferous ($RBA_{CON}$). The diagonals are the identity (1:1) lines.**

Following model validation, the trained RF models were applied to 5-m resolution predictors to produce high-resolution maps of basal area (BA) and relative basal area of coniferous trees ($RBA_{CON}$) across the entire study area (Fig S2). In the plain area, vegetation density was generally low, while the ridges were characterized by high-density. $RBA_{CON}$ varied according to the topographic gradient: broadleaf species dominate the plains, while foot slopes, slopes and ridges were dominated by a mixture of broadleaved and coniferous trees.

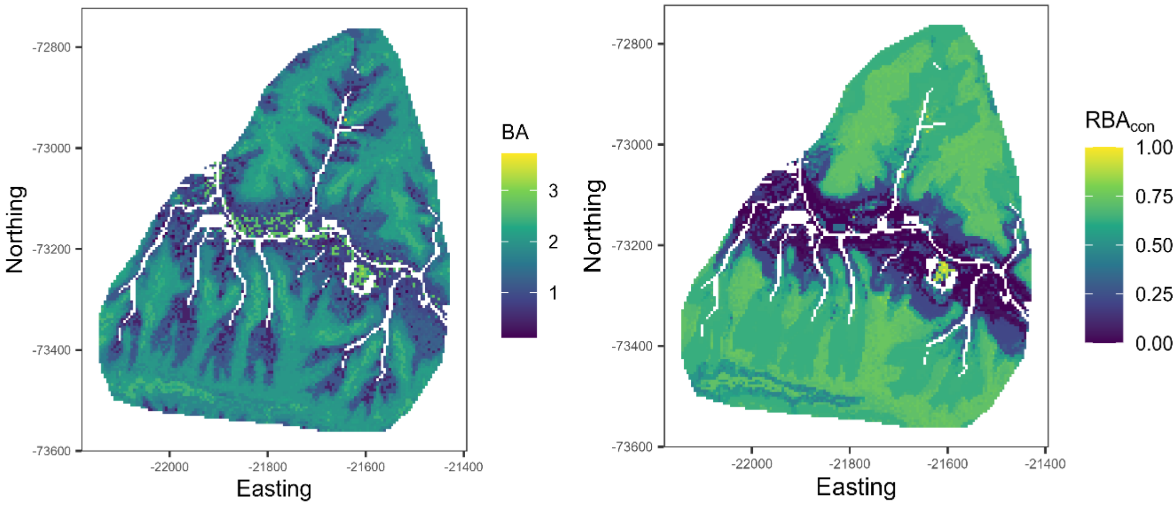


**Fig. S3: Maps of predicted BA and RBA$_{CON}$ at each pixel of the study area.**

**Reference:**


Kuhn, M. and Johnson, K.: Applied Predictive Modeling, Springer New York, New York, NY, https://doi.org/10.1007/978-1-4614-6849-3, 2013.

Liaw, A. and Wiener, M.: Classification and Regression by randomForest, R news, 2, 18–22, 2002.