Simple Optimal Sampling Algorithm to Strengthen Digital Soil Mapping Using the Spatial Distribution of Machine Learning Predictive Uncertainty: A Case Study for Field Capacity Prediction

Hyunje Yang, Honggeun Lim, Haewon Moon, Qiwen Li, Sooyoun Nam, Jaehoon Kim and Hyung Tae Choi *

Forest Environment and Conservation Department, National Institute of Forest Science, Seoul 02455, Republic of Korea
* Correspondence: choiht@korea.kr; Tel.: +82-2-961-2643

Supplementary Materials

Table S1. Forest environmental covariates from geographically referenced database used in this study.

Environmental covariates	Unit	Data type	Source	Definition	Description
Soil texture class (STC)	-	12 categories	FSSM	12 Soil texture classes by USDA textural triangle	Clay (c), Silty clay (sic), Silty clay loam (sicl), Clay loam (cl), Silt loam (sil), Silt (si), Loam (l), Sandy clay (sc), Sandy clay loam (scl), Sandy loam (sl), Loamy sand (ls), Sand (s)
Total soil depth (TSD)	cm	Continuous	FSSM	Total soil layer depth from surface to bedrock	Considering the relatively shallow soil layer in South Korea, the maximum soil depth is 100 cm.
Stone contents (SC)	%	5 categories	FSSM	Gravel fraction or gravel contents	Gravel is >2 mm in diameter; class 1 (<5 %), class 2 (5~15 %), class 3 (15~30 %), class 4 (30~50 %), and class 5 (≥50 %)
Hardness	kg cm^2	Continuous	FSSM	Resistance of soil to deformation	
Elevation	m	Continuous	DEM	Height above sea level	
Slope	%	Continuous	DEM	Steepness of a surface	
Aspect	°	Continuous	DEM	Orientation of slope in degrees	
Catchment area (CA)	m^2	Continuous	DEM	Upper catchment area related to the corresponding point	

Topographic wetness index (TWI)	-	Continuous	DEM	Index of quantifying topographic effects on hydrological processes	$\ln(\alpha/\tan(\beta))$, where α is the total catchment area of unit contour length and β is slope gradient (Sorensen et al., 2006)
Topographic position index (TPI)	-	Continuous	DEM	Index of classifying slope position and landform types	Index of topographical information using the vicinity of the corresponding cell in DEM. Positive value means ridge, negative value means valleys, and zero means flat area or constant slope (De Reu et al., 2013)
Profile curvature (ProC)	-	Continuous	DEM	Slope curvature parallel to the slope	Curvature parallel to the direction of the maximum slope. A positive profile curvature value means convex up, and zero value means flat surface. Profile curvature is analyzed with the ArcGIS spatial analysis tools.
Plan curvature (PlanC)	-	Continuous	DEM	Slope curvature perpendicular to the slope	Curvature perpendicular to the direction of the maximum slope. A positive plan curvature value means laterally convex, and zero value means linear surface. Plan curvature is analyzed with the ArcGIS spatial analysis tools.
Bedrock	-	3 categories	GM	Types of bedrock	Igneous, Sedimentary, Metamorphic
Forest type (FT)	-	3 categories	FTM	Dominant tree type	Coniferous, Broadleaf, Mixed

***Abbreviations*:** FSSM, forest site and soil map; DEM; digital elevation map; GM, geologic map; FTM, forest type map.

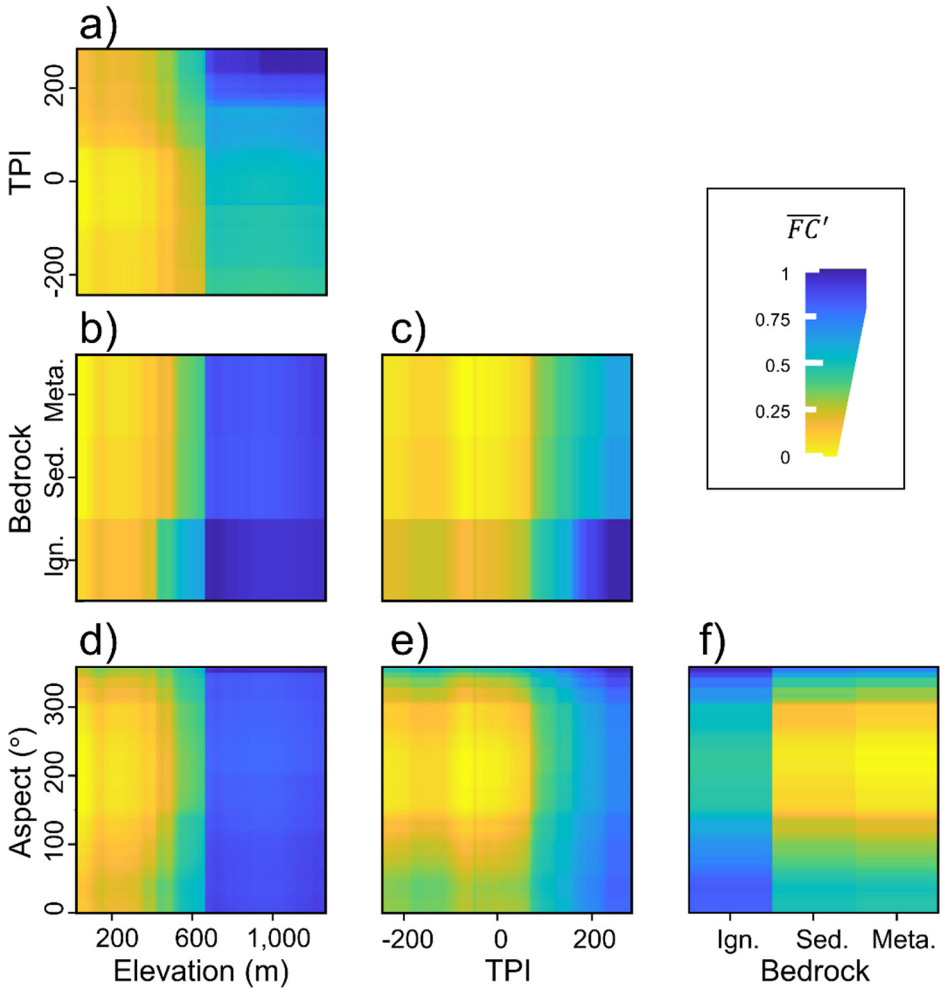

Figure S1. 2D heatmap matrix of predicted field capacity (FC) for four environmental covariates. A randomly selected soil sample was selected and two variables of interest were changed while the other variables were fixed. This process was iterated 500 times and all predicted FCs were averaged and standardized to show the effect of only two variables of interest on FC. Note that Ign. is igneous rock, Sed. is sedimentary rock, Meta. is metamorphic rock. When it comes to aspect, 0° is the northern aspect and 180° is the southern aspect.

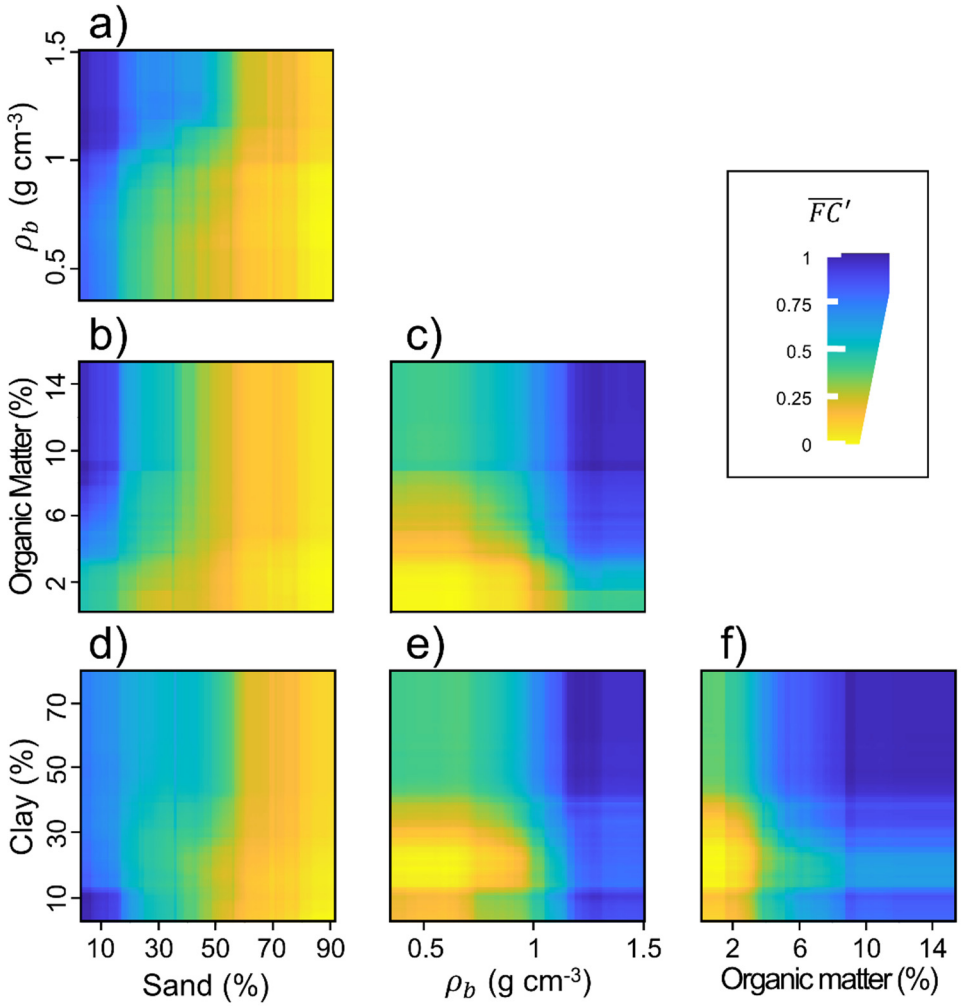

Figure S2. 2D heatmap matrix of predicted FC for four soil properties. A randomly selected soil sample was selected and two variables of interest were changed while the other variables were fixed. This process was iterated 500 times and all predicted FCs were averaged and standardized to show the effect of only two variables of interest on FC.