Theme 2 | Advances in soil mapping and monitoring

Digital Soil Mapping of Surface Texture using Random Forest Classifier

Kalaiselvi. B*, S. Dharumarajan, M. Lalitha, R. Vasundhara, Rajendra Hegde and V. Ramamurthy

ICAR-National Bureau of Soil Survey and Land Use Planning, Regional Centre, Bangalore-560024

Introduction

- Knowledge of spatial soil textural information at higher resolution is vital for optimal agricultural management, modeling, and monitoring.
- Soil static properties ie., soil texture plays imperative role in influencing soil overall productivity and also in land suitability classes and soil taxonomic classifications.
- Digital soil mapping (DSM) aids in establishing dependent spatial soil information using SCORPAN model by interrelating the environmental co-variables with legacy soil information (McBratney et al., 2003).
- Many researchers focused on particle size fractions (PSF) prediction instead textural classification for mapping soil textural classes.
- The present study aimed to predict the surface soil textural classes over 13 Lakh sq. km area of Tamil Nadu in the southern end of India using RF classifier algorithms through DSM approaches.

Methodology

A total of 2109 soil profile information of soil resource mapping (SRM) of Tamil Nadu at 1:250,000 scale was compiled as legacy soil information after removing the outliers for predicting soil textural classes as categorical variables using the Random Forest (RF) classifier. About 34 environmental covariates were taken into account for prediction such as derivatives of digital elevation models, IRS LISS-III data (4 bands), vegetation indices, rainfall and temperature data, and bioclimatic variables.

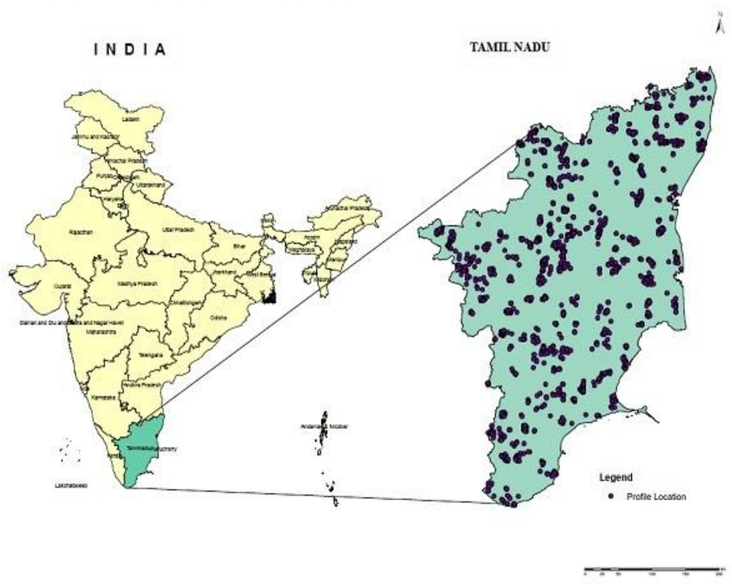

Fig.1. Study area with profile locations

Table 1. Different covariates used for modeling

Predictor	Source	Resolution
Elevation (m)	SRTM DEM	30 m
Slope (%)	SRTM DEM	30 m
Aspect	SRTM DEM	30 m
TPI	SRTM DEM	30 m
TWI	SRTM DEM	30 m
Plan curvature	SRTM DEM	30 m
LS-factor	SRTM DEM	30 m
MrVBF	SRTM DEM	30 m
MrRTF	SRTM DEM	30 m
NDVI	MOD13Q1(2011-2015)	250m_16 days
EVI	MOD13Q1(2011-2015)	
LISS-III	4 bands	23.5 m
Precipitation (mm)	WorldClim2	10 min
Minimum temperature (Tmin) (°C)	WorldClim2	10 min
Maximum temperature (Tmax) (°C)	WorldClim2	10 min
Isothermality	WorldClim2	10 min
Precipitation of driest month (mm)	WorldClim2	10 min
Precipitation of wettest month (mm)	WorldClim2	10 min
Precipitation seasonality	WorldClim2	10 min
Mean diurnal range (°C)	WorldClim2	10 min

Random forest classifier algorithm was used to classify the surface textural classes. *randomForest package* used to establish the relationship between different environmental variables. The performance of the RF classifier was determined based on overall accuracy and kappa index. OA is used to measure the effectiveness of the model and estimated by the formula

$$OA = \sum_{i=1}^{n} \frac{X_{ii}}{N}$$

where n is the number of columns or rows of error matrix, Xii is the classes predicted properly, and N is the total number of observations. Kappa indices >0.80, 0.4-0.8 and <0.4 indicates the strong, moderate and poor agreements of classification, respectively (Congalton & Green, 1998).

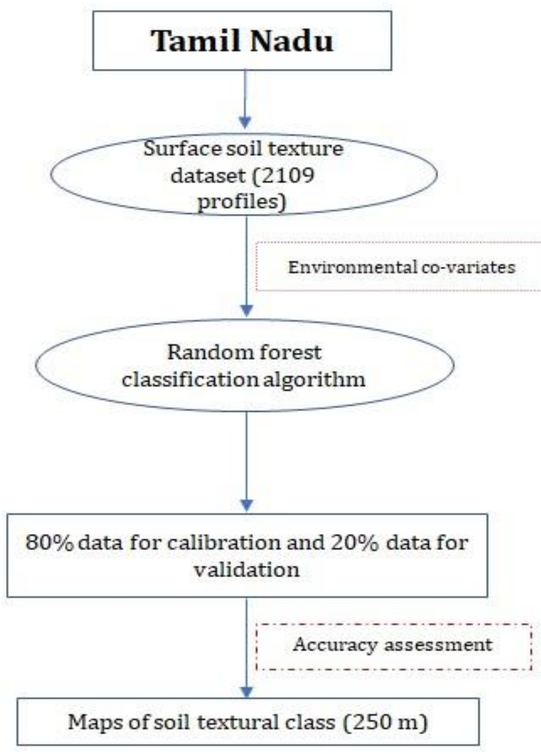

Fig. 2. Methodology

Results

- Results revealed that Tamil Nadu surface soils are dominated with sandy texture (sandy loam and loamy sand), followed by clayey (clay, sandy clay) and loamy soils (clay loam and sandy clay loam).

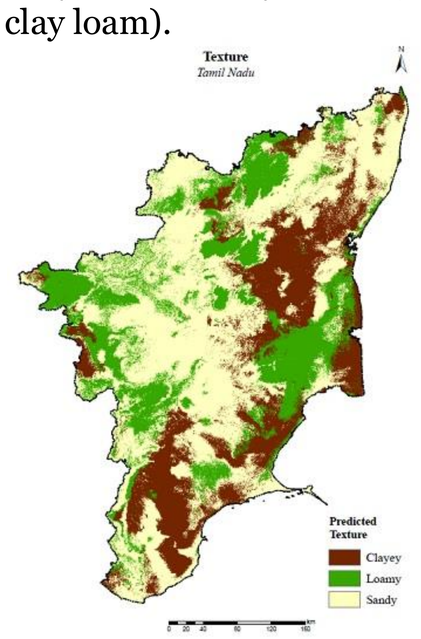

Fig. 3. Surface soil texture class map of Tamil Nadu by Random Forest Classifier

- Mean confusion matrix achieved through 50 iterations in which, 187 soil samples were correctly classified as clay out of 464 soil samples, 275 soil samples were classified as loamy out of 641 samples.
- 60 percent of the soils were classified to the correct class (sandy texture). The error matrix shows that the classification of sandy classes (error= 0.261) is much better than other classes (clay= 0.497; loam= 0.464) concerning error estimates.
- The performance of the RF classifier algorithm was registered with an overall accuracy of 63.8% and kappa index of 0.43.
- Variable importance ranking showed that remote sensing variables (NDVI & EVI), temperature, multiresolution ridge top flatness (MrRTF), and elevation played crucial role in predicting the soil textural classes.

Table 2. Confusion matrix obtained by RF classifier for soil textural classification

		Observed			Classification error	Overall accuracy (%)	Kappa index
Predicted	Soil textural class	Clay soils	Loamy soils	Sandy soils			
	Clay soils	187	79	106	0.50	63.8	0.43
	Loamy soils	63	275	175	0.46		
	Sandy soils	76	133	591	0.26		

Conclusion

- Digital soil mapping of surface soil texture using Random forest classifier was found to be good with overall accuracy and Kappa index
- The prediction performance could be improved by incorporating high resolution co-variates with additional covariates such as lithological data, increasing sampling density by targeted sampling and stratification of dataset for complex landscape units.