Spatial Mapping of Soil Depth, Coarse Fragments and Textural Classes of Tamil Nadu Using Digital Soil Mapping Approach

B Kalaiselvi (kalaimitra15@gmail.com)
ICAR-National Bureau of Soil Survey and Land Use Planning https://orcid.org/0000-0002-1452-7255
S. Dharumarajan
ICAR-National Bureau of Soil Survey and Land Use Planning
M. Lalitha
ICAR-National Bureau of Soil Survey and Land Use Planning
R. Srinivasan
ICAR-National Bureau of Soil Survey and Land Use Planning
R. Vasundhara
ICAR-National Bureau of Soil Survey and Land Use Planning
Amar Suputhra
ICAR-National Bureau of Soil Survey and Land Use Planning
Rajendra Hegde
ICAR-National Bureau of Soil Survey and Land Use Planning

Research Article

Keywords: Spatial distribution, digital soil mapping, environmental covariates, soil properties, random forest model.

Posted Date: June 16th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-596172/v1

Spatial mapping of soil depth, coarse fragments and textural classes of Tamil Nadu using Digital soil mapping approach

Kalaiselvi, B*., S. Dharumarajan, M. Lalitha, R. Srinivasan, R. Vasundhara, Amar Suputhra and Rajendra Hegde

Kalaiselvi B*, Scientist

S. Dharumarajan, Senior scientist

M. Lalitha, Scientist

R. Srinivasan, Scientist

R. Vasundhara, Scientist

Amar Suputhra, Young Profession II

Rajendra Hegde, Principal Scientist

ICAR-National Bureau of Soil Survey and Land Use Planning,

Regional Centre, Hebbal, Bangalore, Karnataka, India

***Corresponding author:** kalaimitra15@gmail.com

Orcid ID: 0000-0002-1452-7255

Abstract

Knowledge on spatial distribution of soil depth, coarse fragments and texture are crucial for land resource management and environmental soil modeling. Digital soil mapping approach helps in prediction of spatial soil information by establishing the relationship between soil and environmental covariates. In the present study, we assessed spatial

distribution of soil depth, coarse fragments (CF) and soil textural classes over 0.13 M sq.km area of Tamil Nadu state. About 2100 samples were used for the prediction of soil properties using random forest model (RF). Out of which, 80 per cent samples were used for training and 20 percent samples were used for testing. Different environmental covariates such as digital elevation model outputs, landsat data and bioclimatic variables were related to predict the soil properties. The predicted soil depth and CF ranged from 46-200 cm and 1-42 per cent respectively. The RF model performed well by explaining the variability (R^2) of 43% for soil depth and 21% for coarse fragments with RMSE of 38 cm and 13%, respectively. The RF classifier classified the soil textural classes with 64% overall accuracy and 43% kappa index. Variable importance ranking of Random forest model showed that elevation, MrVBF are the important predictors used for prediction of soil depth and CF, whereas remote sensing vegetation indices such as NDVI, EVI were acted as primary variable for prediction of soil textural classes. In this study, 250 m resolution detailed soil depth, CF and textural class maps were prepared which will be useful for different environmental modeling and proper agricultural management purposes.

Keywords: Spatial distribution, digital soil mapping, environmental covariates, soil properties, random forest model.

Declarations

Conflicts of interest/Competing interests: The authors declare that they have no potential competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of data and material: The data that utilized for the study available from the corresponding-author upon reasonable request.

Spatial mapping of soil depth, coarse fragments and textural classes of Tamil Nadu using Digital soil mapping approach

Kalaiselvi, B*., S. Dharumarajan, M. Lalitha, R. Srinivasan, R. Vasundhara, Amar Suputhra and Rajendra Hegde

***Corresponding author: kalaimitra15@gmail.com**

ICAR-National Bureau of Soil Survey and Land Use Planning, Regional Centre, Hebbal, Bangalore, Karnataka, India

Abstract

Knowledge on spatial distribution of soil depth, coarse fragments and texture are crucial for land resource management and environmental soil modeling. Digital soil mapping approach helps in prediction of spatial soil information by establishing the relationship between soil and environmental covariates. In the present study, we assessed spatial distribution of soil depth, coarse fragments (CF) and soil textural classes over 0.13 M sq.km area of Tamil Nadu state. About 2100 samples were used for the prediction of soil properties using random forest model (RF). Out of which, 80 per cent samples were used for training and 20 percent samples were used for testing. Different environmental covariates such as digital elevation model outputs, landsat data and bioclimatic variables were related to predict the soil properties. The predicted soil depth and CF ranged from 46-200 cm and 1-42 per cent respectively. The RF model performed well by explaining the variability (R^2) of 43% for soil depth and 21% for coarse fragments with RMSE of 38 cm and 13%, respectively. The RF classifier classified the soil textural classes with 64% overall accuracy and 43% kappa index. Variable importance ranking of Random Forest model showed that elevation, MrVBF are the important predictors used for prediction of soil

depth and CF, whereas remote sensing vegetation indices such as NDVI, EVI were acted as primary variable for prediction of soil textural classes. In this study, 250 m resolution detailed soil depth, CF and textural class maps were prepared which will be useful for different environmental modeling and proper agricultural management purposes.

Keywords: Spatial distribution, digital soil mapping, environmental covariates, soil properties, random forest model.

Introduction

With increasing need for site-specific management in agriculture, the planners and researchers require spatially continuous soil information at higher resolution for better representation, resource assessment, spatial crop planning, land management and effective policy making at regional and national level (McBratney et al. 2003). Spatial pattern of soil depth influences soil moisture, plant available water content, nutrient capacity, runoff generation (Gribb et al. 2009) and biological productivity (Gessler et al. 1995). Similarly, soil texture plays a key role in nutrient retention (Kettler et al. 2001), soil aeration and porosity, soil aggregation and structure development (Carrizo 2015), soil organic carbon and soil quality (Kong et al. 2009), soil water transportation, soil degradation (Hillel 1980) and soil productivity (Katerji & Mastrorilli 2009). These soil properties are also act as important criteria in grouping the soil series for similar management purpose, land suitability classes (Thompson et al. 2012) and involve in soil taxonomic classifications (Soil Survey Staff 2014). Assessing spatial distribution of soil depth and coarse fragments is laborious, difficult to measure practically as they are highly variable spatially (Wilding 1985).

Need for spatially seamless soil information dealt by prediction of soil properties by the concept of digital soil mapping which predicts the dependent soil properties (Florinsky et al. 2002) using the interrelationship between environmental covariates (DEM, climatic parameters, remote sensing imageries) (McBratney et al. 2003) and legacy soil information (Akpa et al. 2014) through numerical, statistical and geostatistical models (Minasny and McBratney 2016). It takes the advantage of ability to produce uncertainty and able to update when new data collected (Lagacherie and McBratney 2007). Spatial patterns in soil

properties (soil depth, CF and texture) are resultant of complex interactions of different factors such as geomorphic positions, topography, parent material, climate, weathering rate, vegetation cover, lithology (Kuriakose et al. 2009), biological, chemical and physical processes (Jenny 1941; Odeh et al. 1991). Thus, the environmental covariates related with soil development selected for establishing the DSM models such as legacy soil data (S), climatic information (C), Vegetation/landuse (O), derivatives of digital elevation models (R), lithology (P) and geographic co-ordinates (N) (Minasny and McBratney 2016).

Tsai et al. (2001) used linear soil landscape regression model to map soil depth in Taiwan and estimated the spatial solumn thickness with RMSE of 16.13 cm. Tesfa et al. (2009) developed statistical models for prediction of soil depth over complex terrain in DCE watershed, USA (28 sq.km of 819 sub-watersheds) and found that with increasing complexity of variabilities, RF outperforms as it is robust against overfitting. Kuriakose et al. (2009) predicted soil depth with maximum capturable variability of 52% (R^2) in part of Western ghats, Kerala. Mehnatkesh et al. (2013) identified the relationship between soil depth and topographic variables (slope, WI, catchment area and sediment transport index) using MLR and explained 76% total variability of soil depth. Bonfatti et al. (2016) used multilinear regression model in Southern Brazil (8118 ha) and predicted soil depth with R^2, RMSE and CCC of 0.43, 34.8 cm and 0.59 respectively. Lacoste et al. (2014) investigated soil depth for a region in France of approximately 5,40,000 km^2 and predicted the soil depth with R^2 and RMSE values between 50–58% and 35–39 cm. Vaysse and Lagacherie (2017) attempted to predict soil depth of Languedoc-Roussillon region, France and could capture the maximum variability of R^2= 23% with RMSE of 34 cm. In India, Dharumarajan et al.

(2020a) prepared 250 m resolution soil depth map of Karnataka using Regression kriging and QRF model which could explain maximum variability of 30% (R^2) and 34 cm (RMSE).

Mosleh et al. (2016) used different Machine learning techniques (ANN, BRT, GLM and MLR) and found CF with 0.07 R^2 with RMSE of 3.2-3.7 which had high variability about 260%. Similarly, Vaysse and Lagacherie (2015) evaluated simulation models, RFM and kriging approaches to coarse fragments and could capture only 4% variability of coarse fragments. Nussbaum et al. (2018) found that RF model performed better in prediction of soil gravel content with minimal RMSE of 2.64%.

Few studies have been conducted in India to map soil textural classes either through prediction of sand, silt and clay particle size fractions (PSF) (Mitran et al. 2018; Dharumarajan et al. 2020) or categorical classes (Dharumarajan et al. 2020b). Many researches focused on PSF prediction for hydrological modelling, SOC stock estimation and textural class mapping using textural triangles (Akpa et al. 2014; Chagas et al. 2016; Mehrabi-Gohari et al. 2019; Zeraatpisheh et al. 2019; Dharumarajan et al. 2020b). RF classifier algorithms outperformed in classification of soil taxonomic classes (Silva et al. 2019; Boroujeni et al. 2020), parent material (Heung et al. 2014), land suitability (Senagi et al. 2017) etc. Lieβ et al. (2012) found the maximum efficacy of RF model over Regression Tree in PSF prediction (30-40% R^2). Vaysse and Lagacherie (2015) predicted PSF using RF with 33-35% R^2 of sand, 31-35% R^2 of clay and 23-29% R^2 of silt content. Camera et al. (2017) reported that random forest classifier was more remarkable than MLR in soil texture classification. Zhang et al. (2020) compared five different models for the classification of soil texture class and found that random forest classifier performed significantly with overall accuracy of 65 per cent and profound improvement in kappa

index of 21 per cent. Abraham et al. (2020) predicted soil particle size fractions using the Random Forest algorithm to get soil textural classes and got the highest overall accuracy of 84.70 percent compared to other models with kappa index of 55-70 percent.

After inception of GlobalSoilMap project, many researches focused on spatial mapping of soil properties worldwide (Arrouays et al. 2014). In India, very limited attempts only made in digital soil mapping (Sreenivas et al. 2016; Dharumarajan et al. 2017, 2019 & 2020; Reddy et al. 2021). Accurate prediction of soil properties would help in crop productivity and comprehensive land management especially ecological, crop suitability and hydrological modelling. With this context, an exploratory study was conducted to map soil depth, CF and texture of Tamil Nadu experiencing semiarid climate using Random Forest model techniques.

Materials and Methods:

Study Area

Tamil Nadu is the south most state in India with 130058 sq.km area extended between 8°5' and 13°35' N and 76°15' and 80°12' E which accounts for 3.96 percent of the total geographical area (TGA) of the country (Fig.1). Tamil Nadu state has been divided into three physiographic units namely South deccan plateau (uplands), Hill ranges and Coastal Plains (Inland plain and Marine plain) (Natarajan et al. 1997). The geological formations of study area are dominant with Precambrian granitic, pink gneiss and charnockite followed by Pleistocenic alluvium and in someparts Pliocene sandstone. Tamil Nadu comes under semiarid agro-eco region which receives bi-model rainfall ie., 48 % from NEM and 32% from SWM. Mean minimum and maximum temperatures ranged between 21.6°C-31.8°C.

Dominant soils are Inceptisols (50%), Alfisols (30%), followed by Vertisols (7%), Entisols (6%) and Ultisols (1%) and negligible areas of Mollisols (Natarajan et al. 1997). Major crops being cultivated are Paddy, Pulses, Groundnut, Sugarcane, Cotton, Coconut, fruit and vegetable crops.

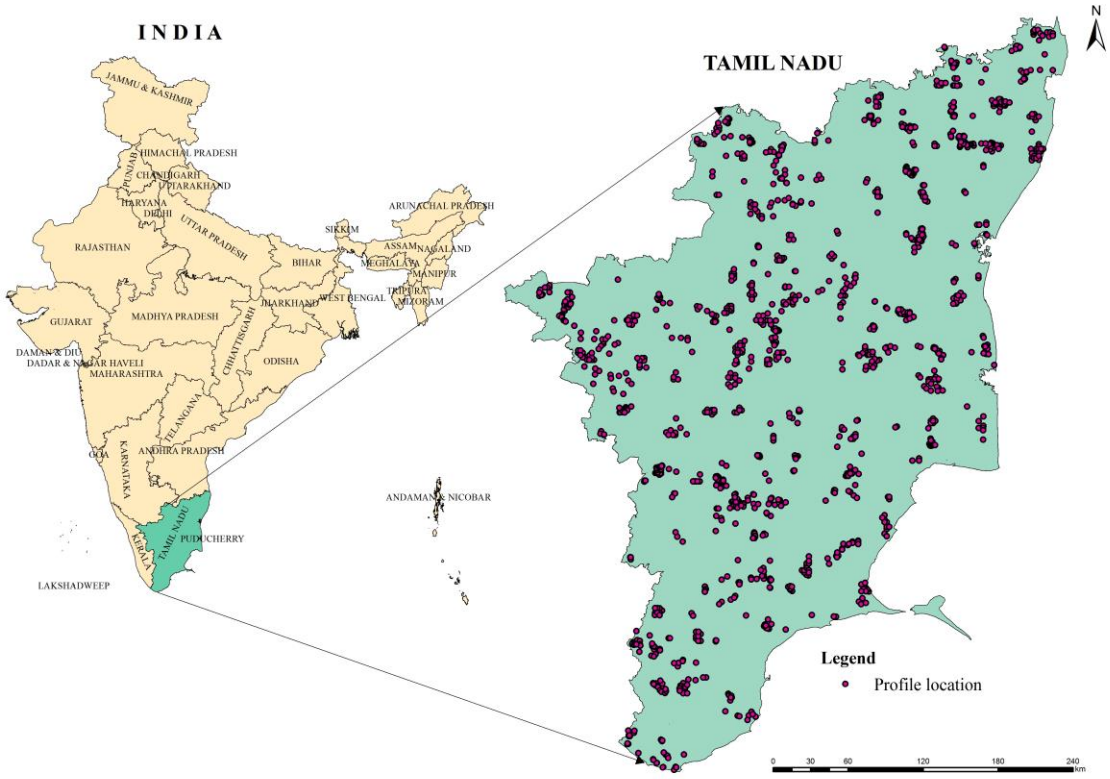


Fig.1. Study area with profile locations

Sampling Methodology

Soils studied during soil resource mapping (SRM) of Tamil Nadu at 1:250,000 scale (1989–1999) and soils studied by Tamil Nadu state soil survey department and other ICAR-NBSS&LUP projects are collected and organized for mapping of soil depth, coarse fragments and texture in Tamil Nadu. The geospatial information of the soils studied during soil resource mapping were limited with degree-minutes which were again assigned with

specific coordinate pertained to the village information. A total of 2119 soil profiles were used for depth prediction study after removing the outliers whereas 2105 surface soil samples were used for coarse fragments and texture class mapping. While using the soil textural classes for spatial mapping, neighbouring soil textural classes are merged into broader classes such as clay and sandy clay into clay (C), clay loam and sandy clay loam in to loam (L) and loamy sand and sandy loam into sandy (S) textural classes based on the earlier studies proven that more number of soil classes with less frequencies may result in significant decrease in prediction accuracy (Lagacharie et al. 2019; Boroujeni et al. 2020).

Environmental co-variates:

Digital elevation model (DEM) of 30 m resolution obtained from shuttle radar topography mission (SRTM) was used to derive different topographic derivatives such as, elevation, slope, aspect, topographic position index (TPI), topographic wetness Index (TWI), LS factor, Multi-resolution Index of Valley Bottom Flatness (MrVBF) and multi-resolution ridge top flatness (MrRTF) as topography is the main soil forming factor in semi-arid regions (Mehnatkesh et al. 2013). Remote sensing variables such as, LISS III band data (1-4), normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI) were also used to relate with the soil properties. Annual mean rainfall, maximum mean temperature (Tmax) and minimum mean temperature (Tmin) are used to represent climate in the models. The bio-climatic variables (19) such as isothermality, precipitation of driest month, precipitation of wettest month, precipitation seasonality and mean diurnal range were also included as environmental covariates for soil properties prediction. Correlation between the soil properties and the environmental co-variates was analyzed

using pearson correlation coefficient analysis with confidence level of 95%. Soil properties prediction methodology presented in Fig. 2.

Table 1. Different covariates used in the model

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	26 m
Precipitation (mm)	WorldClim2	10 min
Minimum temperature (Tmin) (℃)	WorldClim2	10 min
Maximum temperature (Tmax) (℃)	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 (℃)	WorldClim2	10 min

Random forest model:

Random Forest model (RFM) is an ensemble of decision trees which considers each individual tree's decision to predict and classify the soil properties in addition, it can rank the variables based on their significance in prediction (Breiman 2001). RF model tuned depending on the number of trees to be grown (ntree), number of variables used to split the nodes (mtry) and minimum number of samples at each node (nmin). Random forest model is insensitive to noisy, huge and missing dataset which can handle both quantitative and categorical dataset by regression and classification algorithm (Dharumarajan et al. 2017). Statistical software R of randomForest package (Version R 4.0.5) was used to establish the relationship between environmental variables (R Development Core Team 2012).

Model prediction performance:

For calibration of the prediction model, 80% of soil depth and CF datasets were used and 20% remaining dataset was used for validation dataset and based on the 50 iterations the performance of the models was evaluated. Model prediction performance was evaluated using coefficient of determination (R^2), Lin's concordance correlation coefficient (CCC), root mean square error (RMSE) and mean error (bias) of the validation datasets (Mitran et al., 2018). The value of R^2 and CCC nearing 1 and RMSE to 0, indicates the good prediction performance of the model (Dharumarajan et al. 2017).

$$R2 = 1 - \frac{\sum_{i=1}^{n}(pi - oi)^2}{\sum_{i=1}^{n}(oi - \bar{oi})^2}$$

$$ME = \frac{1}{n}\sum_{i=1}^{n}(oi - pi)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(oi - pi)^2}$$

Where, pi and oi are predicted and observed values, pi and oi are means of predicted and observed values.

Lin's Concordance correlation coefficient

$$CCC\ (\rho_c) = \frac{2\rho\sigma_0\ \sigma_p}{\sigma_0^2 + \sigma_p^2 + (\mu_0 - \mu_p)^2}$$

μ0 and μp are the means of observed and predicted values and σ^2_0 and σ^2_p are corresponding variance ρ is Pearson correlation coefficient between observed and predicted values.

Performance evaluation of RF classifier algorithm:

The performance of the RF classifier determined based on overall accuracy, and kappa index.

Overall accuracy:

Percentage of data correctly classified in the validation set is calculated as Overall accuracy. Overall accuracy (OA) is used to measure the effectiveness of the model which is shown as below,

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

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 (Byrt et al. 1993).

Kappa index:

Kappa index is the degree to which classification outcome is superior than the outcome of random classification (Cohen 1960). Kappa index >0.80, 0.4-0.8 and <0.4 indicates the strong, moderate and poor agreements of classification, respectively (Congalton and Green 1998).

Uncertainty analysis:

The lack of reality assurance in prediction is considered as uncertainty (Heuvelink, 2013) and its estimation along with prediction is important for assessing the risk in decision making. The uncertainty of the prediction was quantified by computing upper (95% quantile) and lower confidence (5% quantile) interval limits on the prediction map (Viscarra Rossel et al. 2014) which indicates that 90% of the time, an observed value will fall with in this prediction interval (Malone et al. 2014). The uncertainty for classifier algorithm is evaluated by error/confusion matrix (Boroujeni et al. 2020) which compares the measured and predicted soil classes and checks whether the categorization is satisfactory (Story and Congalton 1986). Confusion index/ error classification values vary from 0 to 1, of which larger values indicate higher uncertainty.

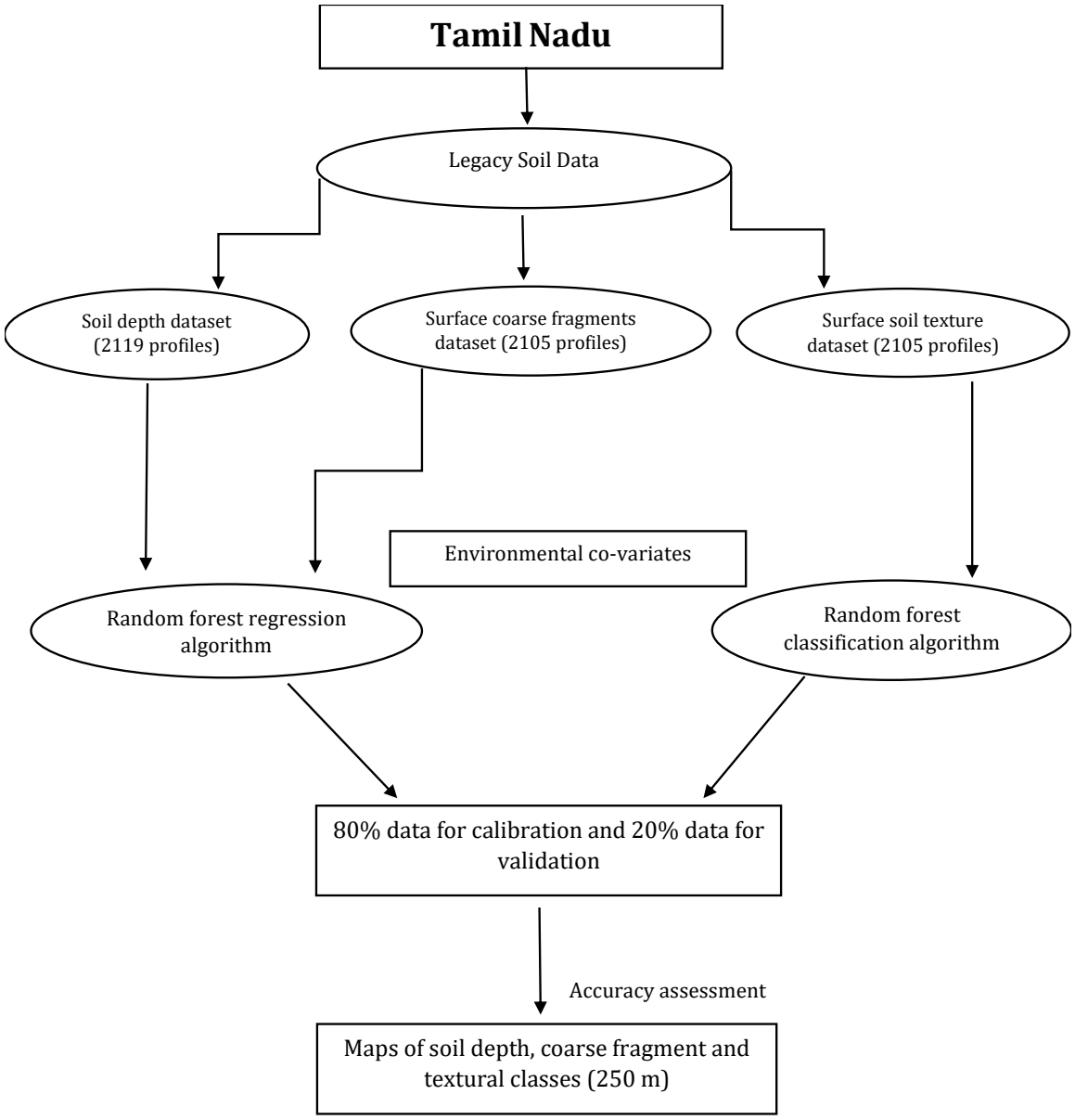


Fig. 2. Methodology Flow Chart

Result and Discussion:

Descriptive statistics:

Summary statistics of soil depth and coarse fragments are given in Table. 2. Soil depth varied from 9-217 cm with mean of 119±51 cm. Soil depth was skewed positively with negative kurtosis and the co-efficient of variation was 42.5 percent which indicates the high variability of soil depth in the study area. Similar results are in concordance with Kuriakose et al. (2009) and Dharumarajan et al. (2020a). Coarse fragments varied from non-gravelly (<15%) to very gravelly (80%) with positive skewness and kurtosis which shows the assymetrical distribution of CF. The very high variability (Wilding 1985) of coarse fragments (CV-135%) was recorded due to differences in landscape, transportability of soil particles and geographical dissimilarity (Tola et al. 2017). The variability of soil properties might attributed to the variability in topography and soil formation (Grimm et al. 2008). Correlation analysis has registered significant relationship between soil depth and coarse fragments with environmental variables (Table. 3). Soil depths were positively correlated with MRRTF (r=0.10**), MRVBF (r=0.20**), precipitation (r=0.24**) and minimum temperature (Tmin) (r=0.15**). It was negatively correlated with elevation (r=-0.14**), TWI (r=-0.13**) and with satellite band data. In contrast to soil depth, coarse fragments shown significant positive relationship with elevation (r=0.06**), TWI (r=0.13**) and band data and negative correlation with MRRTF (r=-0.08**), MRVBF (r=-0.15**), maximum (Tmax) and minimum temperature (Tmin). With increasing slope and elevation, the soil depth and coarse fragments shown negative and positive relationship respectively due to erosional soil loss on slope gradients (Tsai et al. 2001; Florinsky et al. 2002). This correlation study shows indirectly that coarse fragments of

valley (alluvial plains) are mere neglible as the depositional soils are devoid of coarse fragments (Dharumarajan et al. 2021). Soil textural classes were recorded from 2105 surface samples in Tamil Nadu and classified into clay (464), loam (641) and sandy (1000) textural classes. Among the studied soils, predominant soil textural class is sandy soil that occupies about 48% followed by loamy soils (31%).

Performance of RF prediction model:

The performance of the model was evaluated with average performance over 50 iterations using the accuracy indicators like R^2, RMSE and bias given in Table 4. The result shown that the soil depth was predicted with the explained variability of R^2=0.43±0.04 and RMSE of 38 cm. The concordance value for the prediction was 0.612±0.03 that confirms that the moderately good performance of the model (Dharumarajan et al. 2017). The performance of the Random Forest model was comparatively better than the QRF model used in Karnataka soil depth mapping which could only explain 17% variability (R^2) (Dharumarajan et al. 2020a). In African continent, Hengl et al. (2021) predicted soil depth to bedrock suing 30 m resolution co-variates using 28054 sample data and predicted depth with 0.43 R^2; 41.3 cm RMSE and 0.73 CCC with ensemble of RF, XG boost, deep net, Cubist and GLM net. The prediction of coarse fragments shown that the model could find the variability of 21 per cent (R^2=0.21) with 13% RMSE. Compared to soil depth, the bias is higher for coarse fragment prediction. The prediction of CF by RFM is quite low which might be due to high terrain variability and minimal dataset of gravelly soils (Mosleh et al. 2016). Hengl et al. (2004) predicted coarse fragments using SVM with medium level of validity. With regard to soil surface textural classification, the random forest classifier classified the textural class with overall accuracy and kappa index of 63.8% and 41.5%,

315 respectively with confidence interval of 0.59-0.68. This short range of prediction interval
316 indicates the better performance of RF classifier. The performance of RF model is similar to
317 Dharumarajan et al. (2020b) who mapped spatial and vertical (GlobalSoilMap depth
318 specification) textural class map of Andhra Pradesh by predicting PSF and could classify
319 the textural classes with overall accuracy and kappa coefficient of 52-65% and 42-47%
320 respectively. Similar results were obtained from soil textural classification using Landsat-5
321 TM images reported with over all accuracy of around 58-64% (Dematte et al. 2016). Zhang
322 et al. 2020) signified the performance of RF in soil textural classification with profound
323 improvement in reduction of error with increased kappa index of 0.21 by comparing 5
324 different ML models. Boroujeni et al. (2020) classified WRB soil classes using RF model and
325 resulted with 0.73 OA at first level with less confusion index. The mean of confusion matrix
326 obtained by running 50 RF models given in Table 6. In which, only 187 soils only correctly
327 classified as clay out of 464 soil, similarly 275 soils only classified as loamy out of 641.
328 Despite, about 60 per cent of the soils were classified to the correct class (Sandy texture).
329 The error matrix shows (Table 6) that the classification of sandy classes (error= 0.261) is
330 much better than other class (clay= 0.497; loam= 0.464) with respect to error estimates.
331 The results are in concordance with the previously documented studies that classes with
332 lower sampling frequencies were modeled less accurately (e.g., Lagacherie et al. 2019;
333 Boroujeni et al. 2020).

334 ***Importance of variables:***

335 The variables that contribute substantial for prediction of soil properties are
336 assessed by Random Forest model. Fig.3 shows the variable importance ranking of RFM for
337 prediction of soil depth, CF and textural classes. Derivatives of digital elevation models/

topographic variables such as elevation, MRVBF, MRRTF are the primary variables involves in prediction of soil depth and coarse fragment contents (Adhikari et al. 2013; Lu et al. 2019). Some climatic (Tmax) and bioclimatic variables (bio_4) are also bestowed for prediction of soil properties (Malone et al. 2020). Topographic variables were found to be generally more important than the land cover variables in predicting soil depth for this dataset (Tsai et al. 2001; Grimm et al. 2008). Complexity of the environmental variables resulting in variable soil depth (Chan et al. 2019). Similarly, MRVBF and elevation are the prime variables influencing the prediction of coarse fragments. For textural classification, remote sensing indices such as NDVI, EVI, topographical variables (elevation, MRVBF, MRRTF, TWI) (Mehrabi-Gohari et al. 2019) and climatic variables took part in classification of soil textural classes (minimum temperature (Tmin) and Precipitation) (Akpa et al. 2014).

Soil prediction maps:

Summary statistics of the predicted soil properties given in Table 5. Predicted soil depth and CF ranged between 90-195 cm and 0-40% respectively. Skewness and kurtosis of the predicted coarse fragments indicates the presence of extreme values which is reflected in coefficient of variation (CV%-142%), whereas, the predicted soil depth did not register much variability (CV%-11.4%). Predicted soil depth map (Fig. 4a) shows that the depth ranged between 46 and 200 cm and the deeper soils occurring on western ghat (western side) attributed to climate and vegetational influence on soil development (Dharumarajan et al. 2020a) and valley bottom (eastern and south-eastern part) of Tamil Nadu due to the deposition of soil particles from higher elevation to lower elevation (Bonfatti et al. 2016). Spatial variability of soil depth attributed to parent material,

relief/topography and climate (Shivaprasad et al. 1998). Rudiyanto et al. (2016) suggested that improved sampling strategies and machine learning approaches can increase the accuracy of soil depth prediction. Coarse fragments map varied with 1 to 42 per cent of CF (Fig. 4c). Higher CF content was noticed in western and northwestern part of Tamil Nadu (Western ghats and Tamil Nadu uplands). Contradict to depth distribution, CF recorded high in uplands and low in alluvial/coastal plains owing to runoff losses of soil finer particles. According to textural class map (Fig. 5), the 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). Clayey soils distributed in eastern and southern coastal part of Tamil Nadu, whereas, sandy soils distributed in northern and central part of Tamil Nadu. Variation of textural classes shows the transition of clay particles through fluvial movement from upland to lower elevation of eastern and south eastern part of Tamil Nadu with high clayey soils (Dharumarajan et al. 2020b). Coarser soil particles found in upper elevation than low lying areas due to the decreased transportability of sand particles (Mashalaba et al. 2020).

Uncertainty maps of soil depth (Fig. 4b) manifest that high uncertainty of soil depth was noticed in lowland (coastal) and western ghats region which might be due to local landscape variation. Whereas, the low uncertainty of soil depth was noticed in uplands attributed to the similarity in the landscape and climate. The highest uncertainty of coarse fragments (Fig. 4d) was observed in the Tamil Nadu uplands which might be due to high elevation, erosional variability (Kalaiselvi et al. 2019) and granitic gneiss parent material whereas, the lowest uncertainties were observed in the coastal and low lying lands attributed to the non-gravelliness of depositional soils (Dharumarajan et al. 2021).

Regardless, the predicted soil maps can help in assessing the physical constraints for crop growth (Vagen et al. 2016), and used to find the suitable crops to cultivate, optimal fertilizer and other management aspects.

Conclusion:

In the present study, digital soil depth, coarse fragments and textural class maps were prepared using RF model. The result shown that the interrelationship between covariates could explain 43 and 21 per cent of soil depth and coarse fragments variability respectively. Terrain attributes such as elevation and MrVBF are the main predictors of soil depth and CF, whereas, Remote sensing indices (NDVI and EVI) are the important predictors of soil textural classes. Uncertainty map shows that Tamil Nadu upland soil has high uncertainty of depth prediction than the coastal soils. Surface soil textural mapping also acceptable with overall accuracy of 63.8%. Faster run time and reliable prediction performance of RF model helps in prediction and classification of soil properties. The prediction performance could be improved by incorporating additional covariates such as lithological data etc, increasing sampling density and stratification of dataset for different landscape units.

Declarations

Conflicts of interest/Competing interests: The authors declare that they have no potential competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of data and material: The data that utilized for the study available from the corresponding-author upon reasonable request.

406 **Table 2. Descriptive statistics of observed soil depth and CF**

Parameter	Depth (cm)	CF (%)
Mean	119	11
Minimum	9	0
Maximum	217	80
Std. Dev	50.78	15.0
Skewness	0.02	1.80
Kurtosis	-1.09	3.09
CV (%)	42.5	135

407

408 **Table 3. Correlation analysis between covariates and soil properties**

Variables	**Correlation co-efficient (R)**	
	Soil depth	**Coarse fragments**
Elevation	-0.14**	0.06**
EVI	0.02	-0.00
NDVI	0.01	0.02
TWI	-0.13**	0.13**
LS_Factor	-0.04	0.00
MRRTF	0.10**	-0.08**
MRVBF	0.20**	-0.15**
Plan curvature	0.05*	-0.04
Slope	-0.01	0.01
TPI	-0.01	0.04
Aspect	0.00	-0.02

Prec	0.24**	-0.19**
Minimum temperature (Tmin)	0.15**	-0.07**
Maximum temperature (Tmax)	0.00	0.02
bio_7	-0.24**	0.14**
bio_8	0.02	0.02
bio_6	0.16**	-0.07**
bio_5	-0.01	0.03
bio_4	0.08**	-0.06**
bio_3	-0.20**	0.12**
bio_2	-0.43**	0.25**
bio_19	0.29**	-0.23**
bio_10	0.08**	-0.02
bio_11	0.07**	-0.01
bio_12	0.24**	-0.20**
bio_13	0.24**	-0.20**
bio_14	0.10**	-0.08**
bio_15	0.13**	-0.11**
bio_16	0.30**	-0.25**
bio_17	0.17**	-0.13**
bio_1	0.08**	-0.02
bio_18	-0.08**	0.04
bio_9	0.06**	-0.01
b1_LISS (0.52-0.59 µm)	-0.19**	0.11**
b2_LISS (0.62-0.68 µm)	-0.12**	0.07**
b3_LISS (0.77-0.86 µm)	-0.04*	0.00
b4_LISS (1.55-1.70 µm)	-0.31**	0.18**
*Correlation is significant at the 0.05 level (2-tailed) **Correlation is significant at the 0.01 level (2-tailed)		

409

410 **Table 4. Performance of Random Forest model for prediction of soil properties**

Soil properties	R^2	Concordance C	RMSE	Mean error
Soil depth (cm)	0.43	0.61	38	0.19
Coarse fragments (%)	0.21	0.42	13	0.44

411

412

413

414 **Table 5. Descriptive statistics of predicted soil depth and CF**

Parameter	Depth (cm)	CF (%)
Mean	148	2.5
Minimum	90	0
Maximum	195	40
Std. Dev	16.8	3.5
Skewness	0.01	1.66
Kurtosis	-0.31	4.16
CV (%)	11.4	142

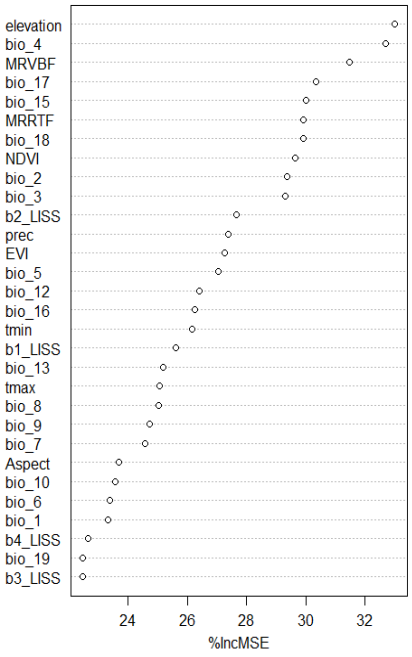
415

416 **Table 6. Confusion matrix of Soil textural classification**

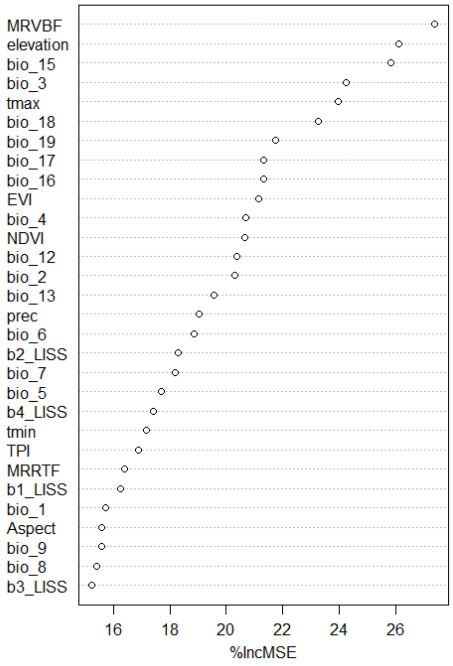
		Observed			**Classification error**
Predicted	**Soil textural class**	**Clay soils**	**Loamy soils**	**Sandy soils**	
	Clay soils	187	79	106	0.50
	Loamy soils	63	275	175	0.46
	Sandy soils	76	133	591	0.26

417

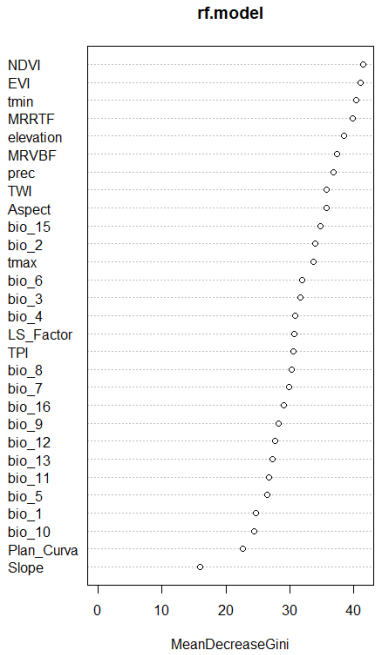
418



(a) Soil depth



(b) Coarse fragments



(c) Soil texture

Fig. 3. Importance variables for prediction of soil depth (a), coarse fragment (b) and soil textural classes (c)

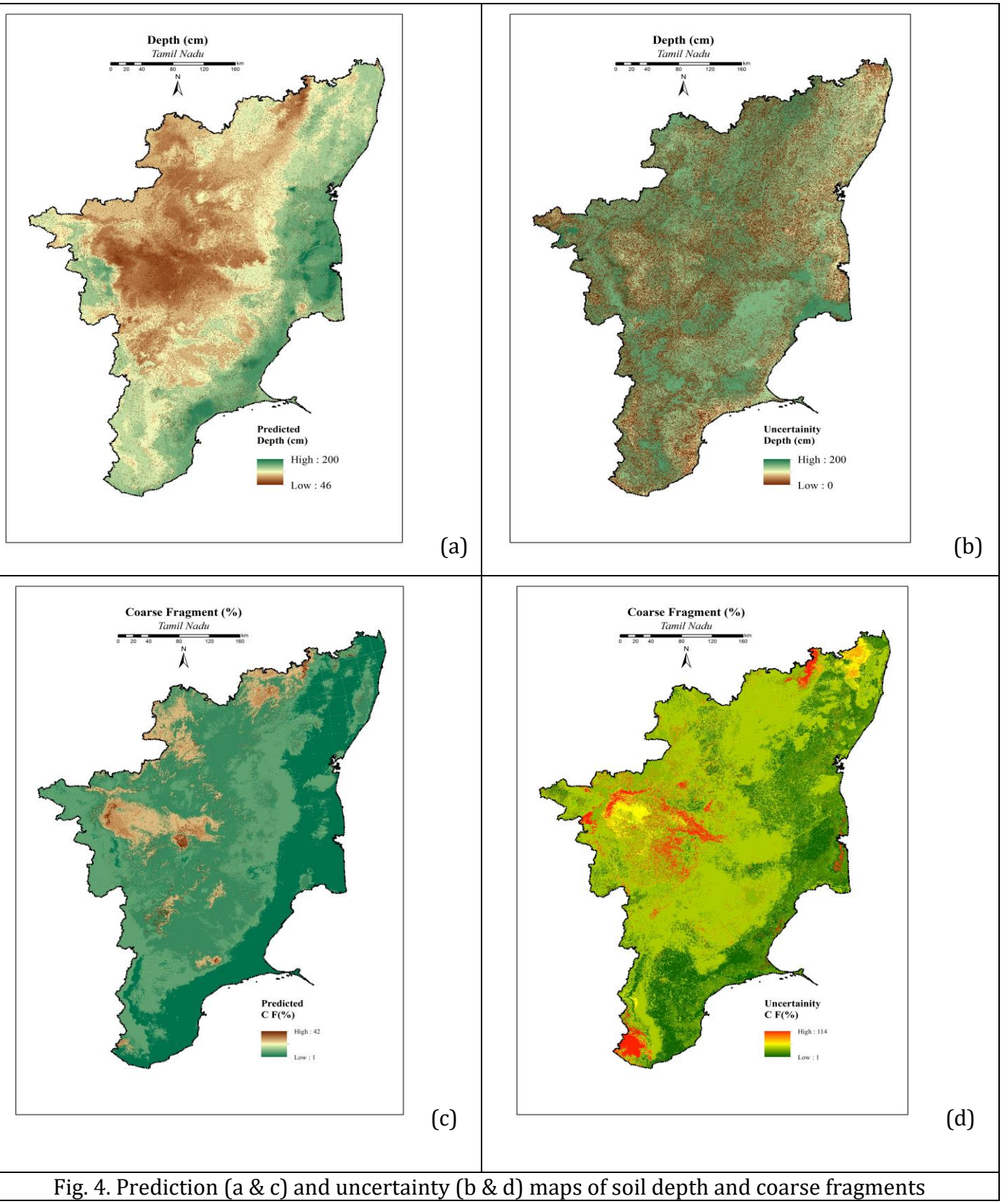


Fig. 4. Prediction (a & c) and uncertainty (b & d) maps of soil depth and coarse fragments

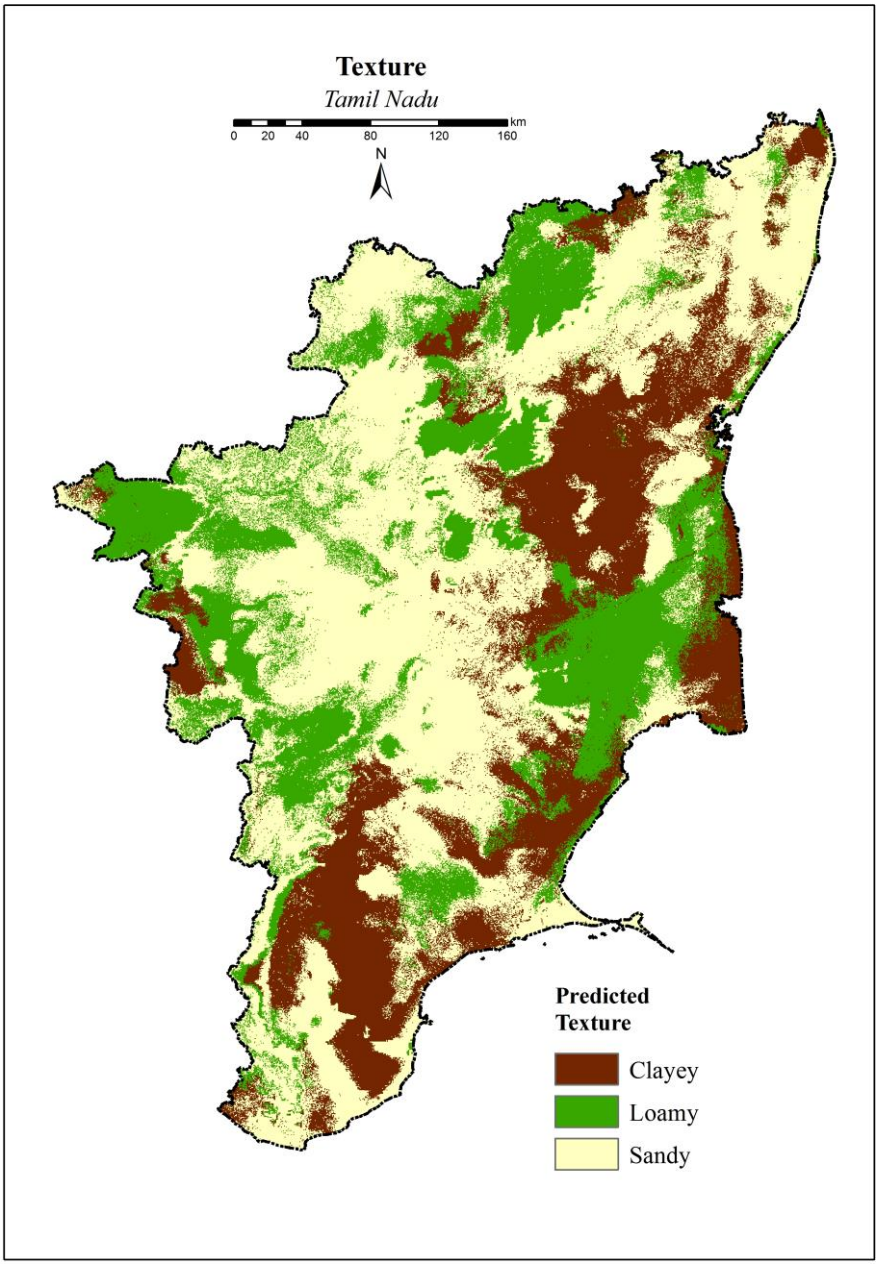


Fig. 5. Surface soil textural map of Tamil Nadu

References:

Abraham S, Huynh C, Vu H (2020) Classification of Soils into Hydrologic Groups Using Machine Learning. Data 5: 2. doi:10.3390/data5010002

Adhikari K, Kheir RB, Greve MB, Bøcher PK, Malone BP, Minasny B, McBratney AB, Greve MH (2013) High-resolution 3-d mapping of soil texture in Denmark. Soil Sci Soc Am J 77: 860–876

Akpa SIC, Odeh IOA, Bishop TFA, Hartemink AE (2014) Digital mapping of soil particle-size fractions for Nigeria. Soil Sci Soc Am J 78: 1953–1966

Arrouays D, Grundy MG, Hartemink AE, Hempel JW, Heuvelink GBM, Hong SY, Lagacherie P, Lelyk G, McBratney AB, McKenzie NJ, Mendonça-Santos MD, Minasny B, Montanarella L, Odeh IOA, Sanchez PA, Thompson JA, Zhang GL (2014) GlobalSoilMap: towards a fine-resolution global grid of soil properties. Advances in Agronomy 125: 93–134

Bonfatti BR, Hartemink AE, Giasson E, Tornquist CG, Adhikari K (2016) Digital mapping of soil carbon in a viticultural region of Southern Brazil. Geoderma 261: 204–221. http://dx.doi.org/10.1016/j.geoderma.2015.07.016

Breiman L (2001) Random forests. Machine Learn. https://doi.org/10.1023/A:1010933404324

Byrt T, Bishop J, Carlin JB (1993) Bias, prevalence and kappa. Journal of Clinical Epidemiology 46(5): 423-429

Camera C, Zomeni Z, Noller JS, Zissimos AM, Christoforou IC, Bruggeman A (2017) A high-resolution map of soil types and physical properties for cyprus: A digital soil mapping optimization. Geoderma 285: 35-49, https://doi.org/10.1016/j.geoderma.2016.09.019

447Carrizo ME, Alesso CA, Cosentino D, Imhoff S (2015) Aggregation agents and structural
448 stability in soils with different texture and organic carbon contents. Scientia Agricola
449 72: 75-82
450Chagas CDS, Junior WDC, Bhering SB, Filho BC (2016) Spatial prediction of soil surface texture
451 in a semi-arid region using random forest and multiple linear regressions. Catena 139:
452 232–240
453Chan HC, Chang CH, Chen PA, Lee JT (2019) Using multinomial logistic regression for
454 prediction of soil depth in an area of complex topography in Taiwan. Catena 176: 419–
455 429
456Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Meas 20: 37–46
457Congalton RG, Green K (1998) Assessing the Accuracy of Remotely Sensed Data: Principles and
458 Practices. CRC/Taylor & Francis, Boca Raton, USA
459Demattê JAM, Alves MR, Terra F da S, Bosquilia RWD, Fongaro CT, Barros PP da S (2016) Is It
460 Possible to Classify Topsoil Texture Using a Sensor Located 800 km Away from the
461 Surface? Revista Brasileira de Ciência do Solo, 40, e0150335
462Dharumarajan S, Hegde R, Singh SK (2017) Spatial prediction of major soil properties using
463 Random Forest techniques-a case study in semi-arid tropics of South India. Geoderma
464 Regional, 10:154–162
465Dharumarajan S, Kalaiselvi B, Lalitha M, Vasundhara R, Rajendra Hegde (2021) Defining
466 fertility management units and land suitability analysis using digital soil mapping
467 approach. Geocarto International. https://doi.org/10.1080/10106049.2021.1926553
468Dharumarajan S, Kalaiselvi B, Suputhra A, Lalitha M, Hegde R, Singh SK, Lagacherie P (2020)
469 Digital soil mapping of key GlobalSoilMap properties in Northern Karnataka Plateau.

Geoderma Regional, 20. https://www.x-mol.com/paperRedirect/1307836545166970880

Dharumarajan S, Vasundhara R, Suputhra A, Lalitha M, Hegde R (2020a) Prediction of soil depth in Karnataka using digital soil mapping approach. Journal of Indian Society of Remote Sensing 48(11): 1593–1600

Dharumarajan S, Hegde R (2020b) Digital mapping of soil texture classes using Random Forest classification algorithm. Soil Use and Management. https://doi.org/10.1111/sum.12668

Dharumarajan S, Hegde R, Janani N, Singh SK (2019) The need for digital soil mapping in India. Geoderma Regional, https://doi.org/10.1016/j.geodrs.2019.e00204

Boroujeni EI, Shahini-Shamsabadi M, Shirani H, Mosleh Z, Bagheri-Bodaghabadi M, Salehi MH (2020) Assessment of different digital soil mapping methods for prediction of soil classes in the Shahrekord plain, Central Iran. Comparison of error and uncertainty of decision. Catena 193, 104648

Florinsky I, Eilers R, Manning G, Fuller L (2002) Prediction of soil properties by terrain modeling. Environmental Modelling and Software 17(3): 295–311

Gessler PE, Moore ID, McKenzie NJ, Ryan PJ (J 1995) Soil landscape modelling and spatial prediction of soil attributes. Int Geogr Inf Syst, 9: 421– 432. doi:10.1080/02693799508902047

Gribb M, Forkutsa I, Hansen AJ, Chandler D, McNamara J (2009) The effect of various soil hydraulic property estimates on soil moisture simulations. Vadose Zone J 8:321 – 331. doi:10.2136/vzj2008.0088

Grimm R, Behrens T, Märker M, Elsenbeer H (2008) Soil organic carbon concentrations and stocks on Barro Colorado Island—digital soil mapping using random forests analysis. Geoderma 146: 102–113. http://dx.doi.org/10.1016/j. geoderma. 2008.05.008

Hengl T, Heuvelink GBM, Stein A (2004) A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120: 75–93

Hengl T, Miller MAE, Križan J, Shepherd KD, Sila A, Kilibarda M, Antonijević O, Glušica L, Dobermann A, Haefele SM, McGrath SP, Acquah GE, Collinson J, et al (2021) African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. Sci Rep, 11: 6130. https://doi.org/10.1038/s41598-021-85639-y

Heung B, Bulmer CE, Schmidt MG (2014) Predictive soil parent material mapping at a regional-scale: a Random Forest approach. Geoderma 214-215, 141–154. http://dx.doi.org/10.1016/j.geoderma.2013.09.016.

Heuvelink GB (2013) Uncertainty quantification of GlobalSoilMap products. Glob Basis Glob Spat soil Inf Syst: 327–332

Hillel D (1980) Applications of Soil Physics. Academic Press Inc

Jenny H (1941) Factors of Soil Formation: A Quantitative System in Pedology, 281 pp, McGraw-Hill, New York

Kalaiselvi B, Lalitha M, Dharumarajan S, Srinivasan R, Rajendra Hegde, Anil Kumar KS, Singh SK (2019) Fertility capability classification of semi-arid upland soils of Palani block, Tamil Nadu for sustainable soil management. Indian Journal of Soil Conservation 47 (3): 143-147

513Katerji N, Mastrorilli M (2009) The effect of soil texture on the water use efficiency of irrigated
514 crops: Results of a multi-year experiment carried out in the Mediterranean region.
515 European Journal of Agronomy 30(2): 95–100. doi:10.1016/j.eja.2008.07.009
516Kettler TA, Doran JW, Gilbert TL (2001) Simplified method for soil particle-size determination
517 to accompany soil-quality analyses. Soil Science Society of America Journal 65: 849–
518 852
519Kong X, Dao TH, Qin J, Qin H, Li C, Zhang F (2009) Effects of soil texture and land use
520 interactions on organic carbon in soils in North China cities' urban fringe. Geoderma
521 154(1-2): 86–92. doi:10.1016/j.geoderma.2009.09.016
522Kuriakose SL, Devkota S, Rossiter DG, Jetten VG (2009) Prediction of soil depth using
523 environmental variables in an anthropogenic landscape, a case study in the Western
524 Ghats of Kerala, India. Catena 79(1): 27–38
525Lacoste M, Minasny B, McBratney A, Michot D, Viaud V, Walter C (2014) High resolution 3D
526 mapping of soil organic carbon in a heterogeneous agricultural landscape. Geoderma
527 213: 296–311
528Lagacherie P, Arrouays D, Bourennane H, Gomez C, Martin M, Saby NP (2019) How far can the
529 uncertainty on a Digital Soil Map be known? A numerical experiment using pseudo
530 values of clay content obtained from Vis-SWIR hyperspectral imagery. Geoderma 337:
531 1320–1328
532Lagacherie P, McBratney AB (2007) Spatial soil information systems and spatial Soil inference
533 systems: perspectives for digital soil mapping. In: Lagacherie P, McBratney AB, Voltz M
534 (Eds), Digital Soil Mapping: An Introductory Perspective. Elsevier BV. hlm, The
535 Netherlands, pp. 3–22

536Ließ M, Glaser B, Huwe B (2012) Making use of the World Reference Base diagnostic horizons
537 for the systematic description of the soil continuum – application to the tropical
538 mountain soil-landscape of southern Ecuador. Catena 97: 20–30
539Lu Y, Liu F, Zhao Y, Song X, Zhang G (2019) An integrated method of selecting environmental
540 covariates for predictive soil depth mapping. Journal of Integrative Agriculture 18: 301–
541 315. https://doi.org/10.1016/S2095-3119(18)61936-7
542Mashalaba L, Galleguillos M, Seguel O, Poblete-Olivares J (2020) Predicting spatial variability of
543 selected soil properties using digital soil mapping in a rainfed vineyard of central Chile.
544 Geoderma Regional 22, e00289
545Malone B, Minasny B, Odgers N, McBratney AB (2014) Using model averaging to combine soil
546 property rasters from legacy soil maps and from point data. Geoderma 232-234, 34–44
547Malone BP, Luo Z, He D, Rossel RAV, Wang E (2020) Bioclimatic variables as important spatial
548 predictors of soil hydraulic properties across Australia's agricultural region. Geoderma
549 Regional, e00344. doi:10.1016/j.geodrs.2020.e00344
550McBratney AB, Mendonca-Santos ML, Minasny B (2003) On digital soil mapping. Geoderma
551 117: 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4
552Mehnatkesh A, Ayoubi S, Jalalian A, Sahrawat KL (2013) Relationships between soil depth and
553 terrain attributes in a semi-arid hilly region in western Iran. J Mt Sci, 10 (1): 163–172
554Mehrabi-Gohari E, Matinfar HR, Jafari A, Taghizadeh-Mehrjardi R, Triantafilis J (2019) The
555 Spatial Prediction of Soil Texture Fractions in Arid Regions of Iran. Soil System 3: 65
556Minasny B, McBratney AB (2016) Digital soil mapping: a brief history and some lessons.
557 Geoderma 264: 301–311

558Mitran T, Mishra U, Lal R, Ravisankar T, Sreenivas K (2018) Spatial distribution of soil carbon
559 stocks in a semi-arid region of India. Geoderma Regional, e00192
560 https://doi.org/10.1016/j.geodrs.2018.e00192
561Mosleh Z, Salehi MH, Jafari A, Borujeni IE, Mehnatkesh A (2016) The effectiveness of digital soil
562 mapping to predict soil properties over low-relief areas. Environmental Monitoring and
563 Assessment, 188 (195). http://dx.doi.org/10.1007/s10661-016-5204-8
564Natarajan A, Reddy PSA, Sehgal J, Velayutham M (1997) Soil Resources of Tamil Nadu for Land
565 Use Planning. NBSS Publ 46b (Soils of India series) National Bureau of Soil Survey and
566 Land Use Planning, Nagpur, India p. 88 +4 sheet soil map 1:500,000 scale
567Nussbaum M, Spiess K, Baltensweiler A, Grob U, Keller A, Greiner L, Schaepman ME, Papritz A
568 (2018) Evaluation of digital soil mapping approaches with large sets of environmental
569 covariates. Soil 4:1–22. https://doi.org/10.5194/soil-4-1-2018, 2018
570Odeh IOA, Chittleborough DJ, McBratney AB (1991) Elucidation of soil-landform
571 interrelationships by canonical ordination analysis. Geoderma 49:1-32
572Reddy NN, Chakraborty P, Roy S, Singh K, Minasny B, McBratney AB, Biswas A, Das BS (2021)
573 Legacy data-based national scale digital mapping of key soil properties in India.
574 Geoderma. https://doi.org/10.1016/j.geoderma.2020.114684
575Rudiyanto, Minasny B, Setiawan BI, Arif C, Saptomo SK, Chadirin Y (2016) Digital mapping for
576 cost-effective and accurate prediction of the depth and carbon stocks in Indonesian
577 peatlands. Geoderma 272: 20–31
578Senagi K, Jouandeau N, Kamoni P (2017) Using parallel random forest classifier in predicting
579 land suitability for crop production. Journal of Agricultural Informatics 8(3): 23–32.
580 https://doi.org/10.17700/jai.2017.8.3.390

581Shivaprasad CR, Reddy RS, Sehgal J, Velayutham M (1998) Soils of Karnataka for optimizing
582 land use. NBSS&LUP publication No.47b, Nagpur
583Silva BPC, Silva MLN, Avalos FAP, Menezes MDD, Curi N (2019) Digital soil mapping including
584 additional point sampling in Posses ecosystem services pilot watershed, southeastern
585 Brazil. Sci Rep Nat 9, 13763
586Soil Survey Staff (2014) Soil taxonomy: a basic systems of soil classification for making and
587 interpreting soil surveys (Twelfth ed). USDA, NRCS
588Sreenivas K, Dadhwal VK, Kumar S, Harsha GS, Mitran T, Sujatha G, Suresh GJR, Fyzee MA,
589 Ravisankar T (2016) Digital mapping of soil organic and inorganic carbon status in
590 India. Geoderma 269: 160–173. https://doi.org/10.1016/j.geoderma.2016.02.002
591Story M, Congalton R (1986) Accuracy assessment: A user's perspective. Photogrammetric
592 Engineering and Remote Sensing 52:397–399
593Tesfa TK, Tarboton DG, Chandler DG, McNamara JP (2009) Modelling soil depth from
594 topographic and land cover attributes. Water Resources Research 45, W10438.
595 https://doi.org/10.1029/2008WR007474
596Thompson JA, Roecker S, Grunwald S, Owens PR (2012) Digital soil mapping: Interactions with
597 and applications for hydropedology, In Hydropedology, edited by: Lin H, Academic
598 Press, Boston, 665-709
599Tola E, Al-Gaadi KA, Madugundu R, Zeyada AM, Kayad AG, Biradar CM (2017) Characterization
600 of spatial variability of soil physicochemical properties and its impact on Rhodes grass
601 productivity. Saudi Journal of Biological Sciences 24(2): 421–429.
602 doi:10.1016/j.sjbs.2016.04.013

603Tsai CC, Chen ZS, Duh CT, Horng FW (2001) Prediction of soil depth using a soil landscape
604 regression model: a case study on forest soils in southern Taiwan. National Science
605 Council of the Republic of China. Part B- Life Sciences 25(1): 34–39
606Vågen TG, Winowiecki LA, Tondoh JE, Desta LT, Gumbricht T (2016) Mapping of soil properties
607 and land degradation risk in Africa using MODIS reflectance. Geoderma 263: 216–225
608Vaysse K, Lagacherie P (2015) Evaluating digital soil mapping approaches for mapping
609 GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France).
610 Geoderma Regional 4: 20–30
611Vaysse K, Lagacherie P (2017) Using quantile regression forest to estimate uncertainty of
612 digital soil mapping products. Geoderma 291: 55–64.
613 https://doi.org/10.1016/j.geoderma.2016.12.017
614Viscarra Rossel RA, Webster R, Bui EN, Baldock JA (2014) Baseline map of organic carbon in
615 Australian soil to support national carbon accounting and monitoring under climate
616 change. Glob Chang Biol 20: 2953–2970
617Wilding L (1985) Spatial variability: its documentation, accommodation and implication to soil
618 surveys. In Soil Spatial Variability Workshop, 166–194
619Zeraatpisheh M, Ayoubi S, Jafari A, Tajik S, Finke P (2019) Digital mapping of soil properties
620 using multiple machine learning in a semiarid region, central Iran. Geoderma 338: 445-
621 452. https://doi.org/10.1016/j.geoderma.2018.09.006
622Zhang M, Shi W, Xu Z (2020) Systematic comparison of five machine-learning models in
623 classification and interpolation of soil particle size fractions using different transformed
624 data. Hydrology and Earth System Science 24: 2505–2526.
625 https://doi.org/10.5194/hess-24-2505-2020

626

627