



Remote Sensing of Land

Homepage: www.gathacognition.com/journal/gcj1
<http://dx.doi.org/10.21523/gcj1>

Online Edition

Remote
Sensing of
Land

Original Research Paper

Terrain Variables Based Spatial Mapping of Soil Properties in a Watershed of Himalayan Landscape Using Random Forest Model



Justin George Kalambukattu*^{ORCID}, Suresh Kumar^{ORCID}

Agriculture and Soils Department, Indian Institute of Remote Sensing, Dehra Dun -248001, Uttarakhand, India.

Abstract

Soil information, including their spatial variability is vital for devising various soil and land management policies and strategies as well as assessing the environmental impact of different land use changes. Among the various soil forming factors, terrain plays a vital role in determining the spatial distribution of soil properties especially in hilly terrains due to its impact on soil processes as well as the redistribution of soil materials. The study was carried out for assessing the applicability and efficiency of various terrain parameters for prediction as well as spatial mapping of different soil properties using Random Forest (RF) modeling approach, in a mid-Himalayan watershed located in Tehri Garhwal district, Uttarakhand. Field data collection was done by collecting soil samples from 0-15cm depth (surface layer) at 68 different locations in the study area, following a transect survey method. Collected samples were pre-processed and analyzed in the laboratory for estimating various soil properties such as sand, silt and clay contents as well as soil nutrients such as Soil Organic Carbon (SOC) and nitrogen (N). Fifteen terrain parameters derived by digital terrain analysis of CartoDEM of 10m spatial resolution were used as predictor variables during RF modeling. Pearson correlation analysis revealed poor linear relationship of soil properties with various terrain parameters. Variable importance ranking revealed that compound topographic index, elevation, heat load index, integrated moisture index and mean curvature are the most important variables, explaining maximum variability in different properties. The performance of RF models for predicting various soil properties were evaluated based on the coefficient of determination (R^2), Root Mean Square Error (RMSE) as well as Lin's Concordance Correlation Coefficient (CCC). The models were found to perform well with R^2 and CCC values of 0.47 and 0.50 for sand, 0.52 and 0.55 for silt, 0.44 and 0.57 for clay, 0.41 and 0.36 for organic carbon and 0.55 and 0.34 for nitrogen, respectively. The validated models were further employed for generating the spatial distribution maps of various soil properties. Such attempts combining various environmental variables with machine learning based models may aid us in mapping the spatial distribution of various soil properties in the hilly mountainous terrains with considerable accuracies.

© 2022 GATHA COGNITION® All rights reserved

Article History

Received: 11 February 2022

Revised: 08 June 2022

Accepted: 09 June 2022

Keywords

Digital Soil Mapping;
Himalaya;
Random Forest;
Soil Properties;
Terrain Variables;
Watershed.

Editor(s)

Vijay Bhagat

1 INTRODUCTION

Soil information, including their spatial variability is vital for soil management, devising agricultural policies as well as tracking environmental impacts due to changes in land use patterns. Deficiency of such quality information could result in implementation of poor policies, thus increasing the risk of environmental

degradation as well as carbon emissions into the atmosphere (Mulder *et al.*, 2011). Thus, in-depth knowledge about various soil properties at regional scale is essential for proper soil use and its effective management for sustainable future. However, the traditional soil mapping techniques are more of a bane

* Author address for correspondence

Agriculture and Soils Department, Indian Institute of Remote Sensing, Dehra Dun -248001, Uttarakhand, India.

Tel.: + 91 135-2524139

Emails: justinagri@gmail.com (J. G. Kalambukattu -Corresponding author), suresh_kumar@iirs.gov.in (S. Kumar)

<https://doi.org/10.21523/gcj1.2022060102>

© 2022 GATHA COGNITION® All rights reserved.

than facilitator for such kind of data at finer scales. This difficulties gain more prominence in soil mapping of the hilly and mountainous terrains. Mapping of soil properties in hilly and mountainous regions has been one of the most challenging tasks for surveyors because of the difficulties associated with it. Moreover, the soils in such areas have a high degree of local variability owing to the differences in various soil forming factors (Funnell and Parish, 2005). The spatial heterogeneity of various soil properties at finer scale is generated due to the complex interaction among these soils forming factors viz. climate, topography, parent materials, vegetation, etc. Of all the factors, topography plays a major role in determining the inconsistency of soil properties in the hilly terrains. Capturing these, finer variability at local/watershed scales in the mountainous regions possess a serious challenge due to the restricted accessibility and scarce availability of soil data. Only very few studies have been focused on detailed mapping of soil properties in the regions with hilly terrains (Chagas *et al.*, 2016; Jeong *et al.*, 2017). However, studies focusing entirely on hilly terrains are rarely been reported in literature.

Digital Soil Mapping (DSM) technique using remote sensing data derived covariates is an effective alternative to the traditional technique and is gaining prominence among researchers (Ben-Dor *et al.*, 2008; Boettinger *et al.*, 2008) as it helps to get away with expensive and prolonged field surveys (Mulder *et al.*, 2011). DSM aids in generation of spatial soil information systems utilizing the relationships between quantifiable soil properties and different environmental variables/predictors, which helps us to understand the spatial variability of different soil properties (Lagacherie *et al.*, 2007). Thus, DSM provides us with an effective alternative to map the properties of mountain soils and capture the variability at finer scale much accurately. Digital Elevation Model (DEM) derived terrain parameters provide us with an opportunity to characterize the various aspects of terrain and study how they impact the distribution of soil properties across the landscape and have been acting as vital inputs in DSM activities across the globe (Akpa *et al.*, 2014; Dharumarajan *et al.*, 2017). Several statistical as well as machine learning techniques/algorithms such as multiple linear regressions (Ben-Dor *et al.*, 2002), partial least squares regression (Stevens *et al.*, 2008), geo-statistical and hybrid methods (Lark *et al.*, 2012), boosted regression trees (Ciampalini *et al.*, 2014), Random Forest Method (RFM) (Chagas *et al.*, 2016), etc. have been used for predicting the soil properties using different remote sensing data as environmental

covariates. Several researchers have used remote sensing data and DEM derived terrain parameters for the prediction and spatial mapping of soil properties.

Among the different techniques, random forest has been used by several researchers to map soil properties. It has several benefits over many statistical models as identified by Breiman (2001) and Liaw and Wiener (2002). These models are known for their high capability to efficiently model nonlinear relationships, ability to resist 'over-fitting', comparative robustness even when using noisy data, usage of continuous and categorical variables, an unbiased error rate measurement, ability to estimate relative importance of various variables used, and requirement of less parameters for execution. Several researchers have used RFM for spatial prediction and mapping of different soil properties such as soil organic carbon (Grimm *et al.*, 2008), particle size distribution (Chagas *et al.*, 2016), soil nutrients (Jeong *et al.*, 2017), etc.

In mid-Himalayan region characterised by steep slopes with elevation reaching up to more than 2500m, topography may have a strong influence in governing the material distribution and thus the soil properties. In this study, Carto-DEM (10m) derived terrain parameters were used for spatial mapping of different soil properties namely texture, organic C and N employing Random Forest (FR) modelling approach in a mid-Himalayan watershed. The prime objective of the study was to test the applicability of different terrain variables for digital mapping of soil properties in hilly terrain using RF technique. Such studies addressing the use of terrain variables for spatial mapping of soil properties in Indian Himalayan region has not been carried out earlier to the best of our knowledge.

2 STUDY AREA

The study was conducted in a small watershed of mid-Himalaya in Tehri Garhwal district of Uttarakhand state in India (Figure 1). The watershed is located at longitude of 78° 25' 55" E to 78° 24' 49" E and latitude of 37° 32' 49" N to 30° 21' 31" N covering an area of 196 ha. The elevation in the watershed varied from 1275 to 1925 m above MSL. The area is a micro-watershed falling within the Maniyar watershed, which drains to the Tehri dam reservoir. The watershed is dominantly covered with forest (especially pine, followed by agriculture and little scrub area). The climate, parent material and LULC were considered uniform throughout the study area due to very small spatial extent of the watershed.

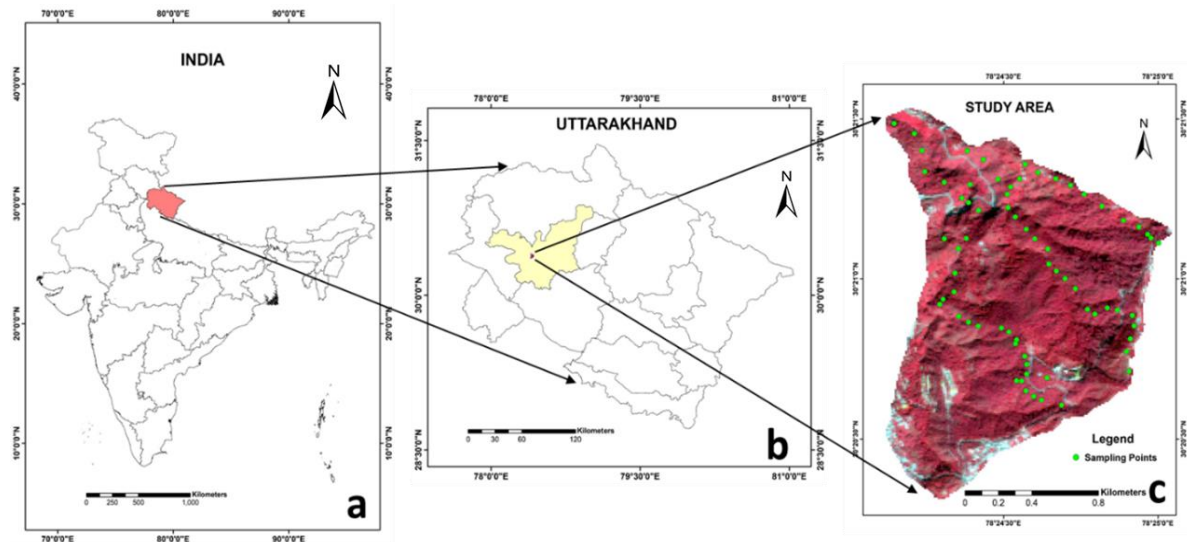


Figure 1. Study area: (a) India, (b) Tehri Garhwal district of Uttarakhand and (c) soil sampling locations on a LISS IV false color composite (FCC) image

3 MATERIALS AND METHODS

3.1 Data

CartoDEM version 2.0 having a spatial resolution of 10m, generated using Cartosat-1 stereo data was used in the study for generation of various terrain parameters (Table 1). The data was obtained from National Remote Sensing Centre, ISRO, Hyderabad. The DEM is reported to be high quality with vertical accuracy assessment yielding an RMSE value in the range of 3.6m to 4.4m and LE90 value of 7.2m (NRSC, 2014). The LISS IV image was used for generation of False Color Composite (FCC) image of the study area.

3.2 Methodology

The watershed boundary was delineated from the DEM, using outlet location, employing the hydrology toolbox in ArcGIS. The DEM was pre-processed for the removal of pits. The pre-processed DEM was used to derive various terrain parameters i.e., elevation, slope(degrees), aspect, cos(aspect), sin(aspect), average curvature, plan curvature, profile curvature, flow direction, flow accumulation, heat load index, integrated moisture index, solar radiation, compound topographic index as well as stream power index of the watershed area.

3.3 Terrain Variables

Slope, a widely known primary terrain attribute, indicates the change in elevation or steepness of the surface with respect to a certain linear distance (Burrough, 1986). It plays an important role in defining the erosion rate as well as influences the pace of soil-forming processes at a particular location and thus contribute significantly in defining the spatial distribution of soil properties. Aspect indicates the direction of the steepest slope/downslope of a particular

location. It is mainly measured clock wisely from north (0° - 360°). The aspect plays a significant role in determining the vegetative growth in the area as well as the rate at which pedogenic processes and soil erosion operates. Sin(aspect) and Cos(aspect) are transformations of aspect done mainly to avoid use of circular data. Instead, the transformations will help us to have a measure of northerness (cosine transformation) and easternness (sine transformation). Flow direction is primary terrain parameter, which indicates the direction in which flow will occur from each cell in a raster. The direction is determined based on the steepest descent, or maximum drop, from each cell, to its neighboring or surrounding cells. Flow accumulation indicates accumulated flow of a cell as the accumulated weight of all cells flowing into each downslope cell. The flow accumulation value denotes the number of cells that flow into each cell. Compound Topographic Index (CTI) a widely used secondary terrain index was computed as per equation (1) making use of average upstream contributing area (A_s) and slope gradient (β) (Moore *et al.*, 1993). It helps to quantify the involvement of topography on soil erosion occurring in a specific location. CTI also referred as Terrain Wetness Index (TWI) describes the spatial arrangement and extent of soil saturation zones as well as variable sources for run-off generation. Lower CTI values generally depict upper catenary/slope positions whereas lower positions are represented by larger values (Jenson, 1991).

$$CTI = \ln \left(\frac{A}{\tan \beta} \right) \quad (1)$$

where, A indicates the upstream contributing area in m^2 and β , the local slope gradient.

Table 1. Details of remote sensing data

Data	Spatial resolution (m)	Acquisition
CartoDEM version 2.0 from NRSC, ISRO, India	10	April 2015
LISS IV	5.8	18 May 2014

Stream Power Index (SPI) is a commonly adopted secondary terrain parameter especially in hilly terrains. It reflects the erosive power of a terrain and provides an idea about the rate of dissipation of energy by flowing water on the channel bed and basins and aids to determine potential strength of flowing water to stimulate erosion and sedimentation. Higher value of SPI indicates chances of higher erosion rates in an area. SPI computation as per equation (2) is based on the idea that discharge quantity is directly related to its contributing catchment area (Jenson, 1991).

$$SPI = \ln(As * \tan \beta) \quad (2)$$

where, A indicates the upstream contributing area in m² and β , the local slope gradient.

Curvature is the second derivative of a surface and is computed as slope of the slope. The curvature outputs are helpful in describing the physical characteristics of a drainage basin with the aim of understanding runoff and erosion processes. Profile curvature is calculated in a direction parallel to the maximum slope, whereas plan curvature is in a direction perpendicular to it. The values indicate whether the surface is upwardly convex, upwardly concave or flat/linear at cell depending on the values (negative, positive or zero). Profile curvature values influence the acceleration and deceleration of flow and thereby influence erosion and deposition processes, whereas convergence and divergence of flow across a surface is regulated by plan curvature. Heat Load Index (HLI) generated using the method given by McCune and Keon (2002) signify temperature variations between different slope directions by means of 'folding' the aspect. In doing so, the highest values of the index are assigned to southwest and the lowest to northeast aspects. The index is based on the concept that while Potential Direct Incident Radiation (PDIR) is found to be symmetrical along the north-south axis, temperatures are not. This is because a slope experiencing afternoon sun was found to have higher maximum temperatures than a similar slope with morning sun (McCune and Keon, 2002). As a result, slopes facing southwest direction would be much warmer compared to a southeast facing slope, even though both receive equivalent amount of solar radiation. In addition, this index takes into account steepness of slope, which many other aspect-rescaling equations fail to do (McCune and Keon, 2002). Integrated Moisture Index (IMI) helps to integrate the GIS-derived topographic and soil features of the landscape into a single index. A wide variety of ecological processes operating at landscape levels are found to be statistically related to IMI (Iverson *et al.*, 1997).

$$IMI = (HS. 0.40) + (FA. 0.30) + (C. 0.10) + (TWHC. 0.20) \quad (3)$$

The value of IMI ranges between 0 and 100, where higher values indicate high relative moisture status. It offers a relative rating of moisture using four inputs namely hill-shade, flow accumulation, curvature and total water holding capacity. The values of all input variables are scaled; scores brought to the range of 0 to 100 and further multiplied with their corresponding weights as per equation (3) to estimate IMI. The solar radiation parameter calculates insolation across specific locations or a landscape, adopting procedures from the hemispherical view-shed algorithm (Rich *et al.*, 1994; Fu and Rich., 2000; Fu and Rich, 2002). The analysis yields a solar radiation product (Wh/m²), indicating the total amount of radiation calculated for a particular location or area. The solar radiation analysis facilitates analysis and mapping of solar effects over a geographic area for specific periods. It can account for atmospheric effects, elevation, slope, aspect, latitude, daily and seasonal shifts of sun angle, and shadow effects by neighboring topography. The resultant outputs can assist in physical and biological process' modeling due to their easy integration with various GIS data (ESRI, 2011).

3.4 Soil Sampling

Transect sampling approach was used for soil sample collection from the study area, as the steep sloping terrain was difficult to traverse and ruled out the possibility of grid sampling. Transect sampling was done taking special care to cover the entire elevation range within the watershed. Field data collection in the study area was done by collecting soil samples from 0-15cm depth (surface layer) at 68 different locations following a transect survey method. Handheld GPS (Trimble Juno SD) was during field sampling for recording precise geographic coordinates of sampling locations in the various transects covering entire elevation range in the watershed. The collected soil samples were securely packed and correctly labeled in polythene bags, and transported. The samples were then subjected to air-drying in the laboratory and sieved using a 2-mm sieve for removing small pebbles as well as root/litter particles.

3.5 Estimation of Soil Properties

TOC Analyzer (Shimadzu TOC-VCPH) was used for determination of total soil carbon in the soil samples, using 0.2mm sieved samples, after air-drying and homogenization. Similarly, total nitrogen content of the samples were estimated using CHNS analyzer

(Elementar vario MACRO cube). Soil textural composition (percentages of sand, silt and clay) was estimated using Bouyoucos hydrometer method (Bouyoucos, 1962) after dispersing the soil samples using sodium hexametaphosphate in solution.

3.6 Random Forest Modeling

The random forest regression model (Breiman, 2001), a non-parametric technique was developed by extending the Classification and Regression Trees (CART) program with the aim of improving the model's prediction performance. It uses randomness and bootstrap samples in the tree building procedure. The model is a combination of several trees simulating the concept of forest which act as predictors and the results derived from the model represents the mean of predictions/results obtained from all trees forming/comprising the forest (Breiman, 2001).

Random Forest (RF) package in R software (R Development Core Team, 2007) was employed for implementation of RF model in the study. Model fitting was decided by three parameters viz., the number of trees in the forest (n_{tree}), the minimum number of predictors used per tree (m_{try}), and the minimum number of samples per terminal node (node size; n_{min}) (Liaw and Wiener, 2002). The standard n_{tree} value of 500 defined in the package was used in this study. Node size value of five and m_{try} value of four was used. Out-of-bag (OOB) mean square error and the percentage of variance explained by the model were used for evaluating the efficiency of RF modeling for spatial prediction of various soil properties.

The developed models' performance is evaluated by means of an independent dataset referred as validation data set, which was not involved in calibration or model development. The soil dataset of 68 points were split randomly into two, one used for calibration (46 points) whereas validation was done using the other (22 points). The model performance was evaluated based on coefficient of determination (R^2), Root Mean Square Error (RMSE) as well as Lin's Concordance Correlation Coefficient (CCC) values generated using validation dataset, based on observed and estimated values of each soil property. Lower the RMSE value, higher the accuracy of prediction. The validated models were further used for mapping the spatial distribution of different soil properties in the study area.

4 RESULTS AND DISCUSSION

4.1 Soil Properties

Descriptive statistics of various soil properties addressed in this study (both calibration and validation datasets) are shown in Table 2 and Table 3. The results indicated that both calibration and validation datasets are having very similar mean values for all the soil properties with no significant difference in the various soil properties at 95% confidence interval based on the results of analysis of variance (ANOVA). The total data set had mean values of 38.63%, 39.93%, 21.62%, 2.37 gkg⁻¹ and 0.18% corresponding to sand, silt, clay, total organic carbon and total nitrogen, respectively (Table 2).

Table 2. Summary statistics of the soil properties (N = 68)

Parameters	Sand*	Silt*	Clay*	Org. C [#]	Nitrogen*
Mean	38.63	39.93	21.62	2.37	0.18
Min	5.88	15.64	10.76	0.31	0.02
Max	67.6	59.64	36.48	7.78	0.77
SD	11.26	8.33	5	1.07	0.11
CV	29.15	20.86	23.14	45	59.08

SD- Standard deviation; CV – Coefficient of variation (%); * - Values in %, [#] - Values in g kg⁻¹

Table 3. Descriptive statistics of the calibration (N=46) and validation (N=22) data sets used for prediction of soil properties

Soil properties	Calibration			Validation		
	Mean	SD	CV	Mean	SD	CV
Sand*	38.44	9.51	24.73	39.04	14.51	37.18
Silt*	40.39	7.24	17.92	38.96	10.38	26.64
Clay*	21.43	4.27	19.92	22.02	6.37	28.92
Carbon [#]	2.4	0.88	36.9	2.33	1.4	60.24
Nitrogen*	0.18	0.08	45.25	0.18	0.15	80.62

SD- Standard deviation; CV- Coefficient of variation (%); * Values in %, [#] - Values in g kg⁻¹

Both the sample sets were found to be highly heterogeneous, indicated by higher values of coefficient of variation (CV). The CV values varied from 17.92 to 45.25 % in case of calibration dataset, whereas it varied from 26.64 to 80.62% in case of validation dataset. The nitrogen content exhibited highest CV values followed by carbon in both the sets, indicating the increased variability of these soil properties in the watershed. This higher variability may help us in getting better prediction results as reported by [Dematte et al. \(2007\)](#).

4.2 Relationship with Terrain Variables

The correlation analysis revealed the nature of relationship between different soil properties and various terrain parameters ([Table 4](#)). Flow direction, heat load index and solar radiation were found to be significantly correlated with sand content in the study area. The above mentioned three parameters along with an aspect were found to be significantly correlated to silt content. Elevation and aspect were found to have significant inverse relationships with TOC content as well as nitrogen content ([Table 4](#)). [Dharumarajan et al., 2017](#) have also reported similar inverse relationships. The absence of significant relationships indicates the nonlinear nature of relationship various soil properties are having with terrain variables.

4.3 Performance of Random Forest Modeling

The performance of random forest modeling for prediction of various soil properties using terrain variables were evaluated using various uncertainty indicators viz. R^2 , RMSE and Lin's CCC. The results revealed that models generated using various terrain parameters could explain 47%, 52%, 44%, 41% and 55% of the variation for sand, silt, clay, carbon and nitrogen, respectively ([Table 5](#); [Figure 2](#)). Lin's concordance correlation test revealed that clay exhibited the highest CCC value (57%) whereas nitrogen exhibited the lowest CCC value (34%). The RMSE values were found to be 11.02, 7.51, 4.75, 0.73 and 0.12

for sand, silt, clay, carbon and nitrogen, respectively ([Table 5](#)). The models developed were found to be low to moderately satisfactory for predicting various properties.

Among the various models generated, RF models for prediction of particle size fractions were found to have higher predictive performance (indicated by LCCC values >0.5) in comparison to organic carbon and nitrogen contents. Attempts to predict spatial distribution of soil properties solely based on terrain parameters is very rare and less reported. [Ließ et al. \(2012\)](#) has reported the use of DEM derived terrain parameters as covariates for RF modeling. They reported prediction models that can explain 30-40% of the variation in soil texture, which was lower than the results obtained in this study ([Table 5](#)). Similarly, [Akpa et al. \(2014\)](#) reported Random Forest Models (RFM), which explained 48-49%, 26-27% and 53-56% of variance for sand, silt and clay respectively in surface layer. Their outcomes are in agreement with our findings for sand, poor in comparison to silt and better in case of clay. Similar or less accurate prediction results with respect to sand, silt and clay have also been reported by various researchers ([Hitziger and Ließ, 2014](#); [Vaysee and Lagacherie, 2015](#); [Chagas et al., 2016](#)).

The developed models were able to account for around 50% variation only of different soil parameters. The inability to account for the rest 50% variance may be due to the different factors like tree height, storied growth, boulders, coarse fragments, embedded deposits etc., which cause intra terrain variations in flow and transport of materials. The intra-terrain variability may have caused lower accuracy as the slopes here are not characterized as smooth slopes, which cause limitations in accounting for the impact of terrain on material transport (comparative erosion/deposition processes operating in slopes at small scale).

Table 4. Results of correlation analysis between soil properties and terrain variables

Terrain variables	Sand	Silt	Clay	TOC	Nitrogen
Flow accumulation	-0.065	0.053	0.050	0.075	0.045
Flow direction	-0.370 ^a	0.333 ^a	0.264 ^a	-0.001	-0.001
HLI	0.303 ^a	-0.312 ^a	-0.166	-0.022	-0.042
Profile curvature	0.035	-0.096	0.111	0.079	0.005
Plan curvature	0.070	-0.091	-0.039	0.093	0.164
SPI	-0.061	0.048	0.061	0.007	-0.076
Mean curvature	0.018	0.008	-0.094	0.003	0.092
Solar radiation	0.201 ^b	-0.196 ^b	-0.091	0.023	0.078
CTI	-0.039	0.044	0.021	0.149	0.052
Slope	0.050	-0.037	-0.060	-0.066	-0.057
Aspect	0.181	-0.272 ^a	0.013	-0.182 ^b	-0.232 ^a
IMI	0.071	-0.079	-0.071	-0.034	0.096
Elevation	-0.112	0.063	0.154	-0.301 ^a	-0.314 ^a
Sin(aspect)	0.186	-0.176	-0.118	0.110	0.167
Cos(aspect)	-0.063	0.082	0.002	-0.064	-0.108

^a - Correlation is significant at 5 % level; ^b - Correlation is significant at 10 % level.

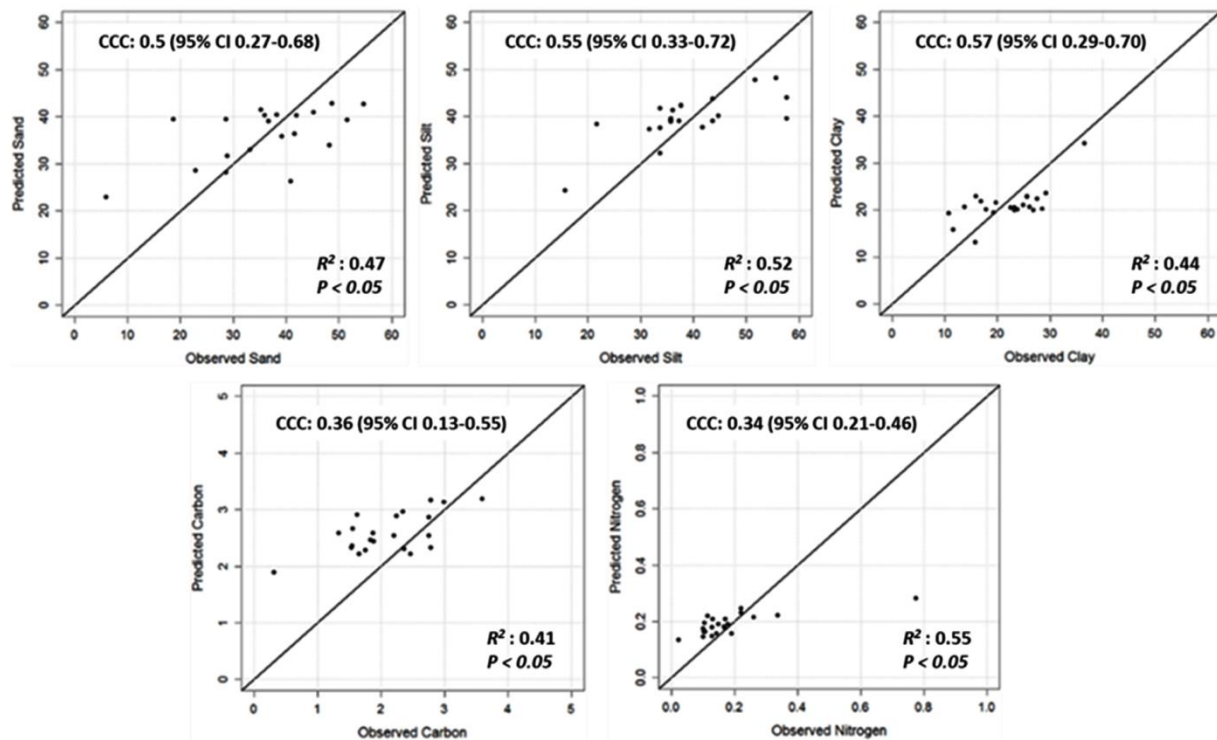


Figure 2. Concordance correlation coefficient plots of different soil properties

Table 5. Performance of RF models for prediction of soil properties

Parameters	RMSE	R ²	LCCC
Sand (%)	11.02	0.47	0.5
Silt (%)	7.51	0.52	0.55
Clay (%)	4.75	0.44	0.57
Carbon (%)	0.73	0.41	0.36
Nitrogen (%)	0.12	0.55	0.34

The performance of the developed models for organic carbon and nitrogen were found to be lower than that of the models for particle size fractions (Table 5). The prediction performance of models for organic carbon in this study was found to be high in comparison to the results obtained by de Carvalho *et al.* (2014) ($R^2 = 0.20$), Gastaldi *et al.* (2012), ($R^2 = 0.18$) and Dharumarajan *et al.* (2017), ($R^2 = 0.23$). The comparatively better results in this study may be attributed to the higher sampling density (34.69 samples per km²), coupled with the use of high spatial resolution (10m) covariates for modeling, which can capture the local scale variations occurring along the slopes in the hilly terrain more efficiently. Still the results were poor in comparison to the works reported by Wiesmeier *et al.* (2011), Guo *et al.* (2015), Vagen *et al.* (2016), Sreenivas *et al.* (2016) and Kalambukattu *et al.* (2018) who have reported R^2 values of 0.76, 0.69-0.77, 0.82 and 0.83, respectively for organic carbon predictions. This may be due to the inclusion of various continuous as well as

categorical environmental covariates corresponding to different factors of soil formation. In case of nitrogen even though R^2 value was above 0.5, the Lin's CCC value (0.34) was found to be the lowest among all the variables. The approach adopted in the study may be useful for soil properties prediction in areas with similar terrain conditions.

4.4 Importance of Variables

The relative importance of various input parameters for modeling can be estimated using RF model, an advantage over various other modeling approaches (Cutler *et al.*, 2009). In addition, the RF modeling aids us in avoiding the elimination of various predictive variables (even if they are correlated), which may be significant with respect to specific properties (Akpa *et al.*, 2014). In RF modeling, the variable importance is estimated based on how good or bad the model predicts, if one or few variables are removed (Prasad *et al.*, 2006). It also safeguards good predictor variables from

being eliminated and preserves them for model. Variable importance rankings of RF models generated for spatial prediction of sand, silt, clay, organic carbon and nitrogen are given in Figure 3. CTI was observed to be the most important variable for prediction of sand, followed by HLI, elevation and plan curvature. The most important variables for modeling and prediction of silt content were elevation > HLI > mean curvature > flow accumulation, whereas for clay, the most relevant variables were flow direction > CTI > elevation > flow accumulation. For organic carbon the important variables were IMI > HLI > CTI > mean curvature, and for nitrogen, IMI > mean curvature > elevation > aspect. These results clearly confirm that the relative importance of various predictor variables is dependent on soil property under consideration along with other variables. The lower percent increases in mean square error (% IncMSE) values may be improved by the inclusion of various other environmental covariates corresponding to different factors of soil formation, in addition to the terrain related covariates as well as more

detailed soil samples (Akpa *et al.*, 2014). Various terrain parameters governing the spatial distribution of soil properties could be identified during the study, which may be quite useful in future studies, for understanding the influence of terrain on soil properties.

4.5 Spatial Mapping of Soil Properties

Spatial distribution maps of various soil properties were generated using the validated RF models (Figure 4). The predicted sand content ranged from 18.86-55.74%, the clay content varied from 14.53-31.02%, whereas silt content varied from 28.27-51.81% in the area (Table 6). The soil carbon and nitrogen contents ranged from 1.24-4.88% and 0.08-0.46%, respectively (Table 6). The Skewness values of different properties indicated that the values are symmetrical (sand, clay, carbon) to moderately skewed (silt, nitrogen). Symmetrical distribution indicates that the predicted values followed a normal distribution, whereas the moderate Skewness indicates the predominance of predicted values in the lower range.

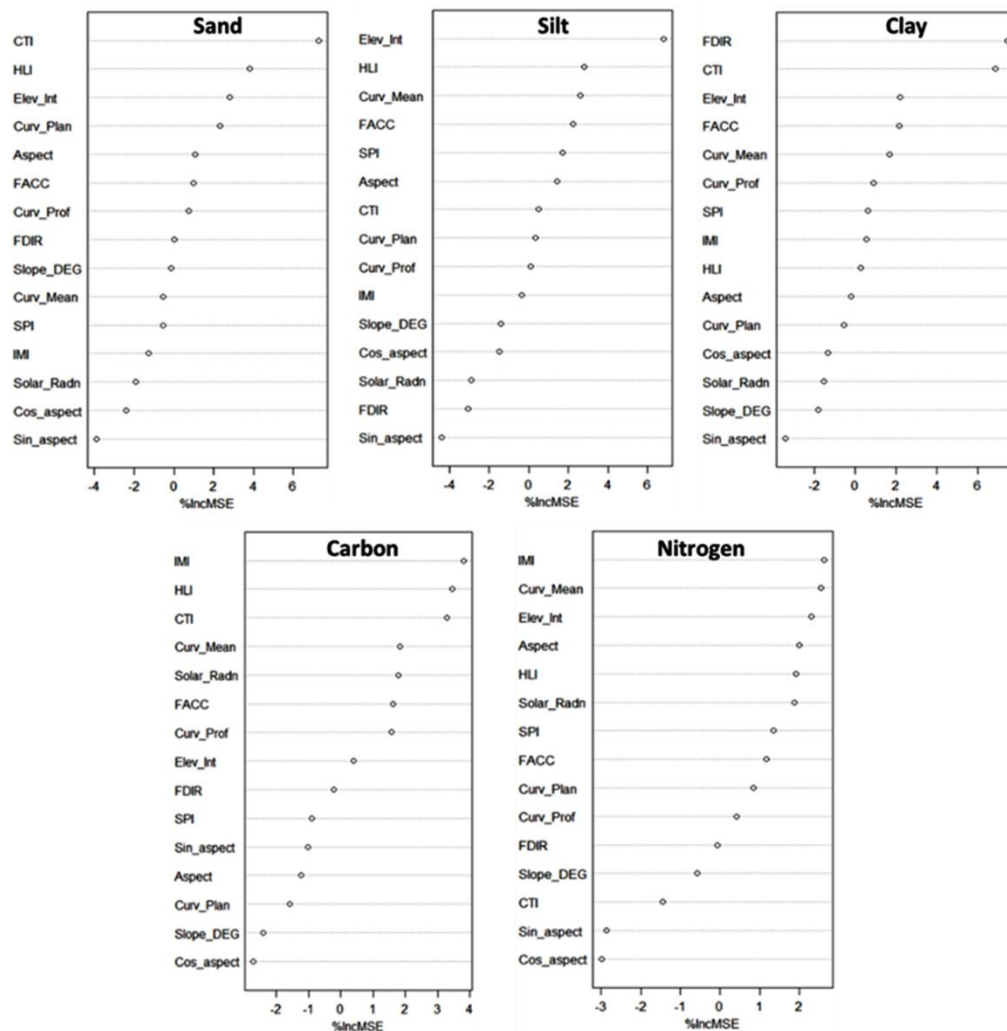


Figure 3. Variable importance plot/rankings of the environmental covariates generated using RF models for sand, silt, clay, organic carbon and nitrogen.

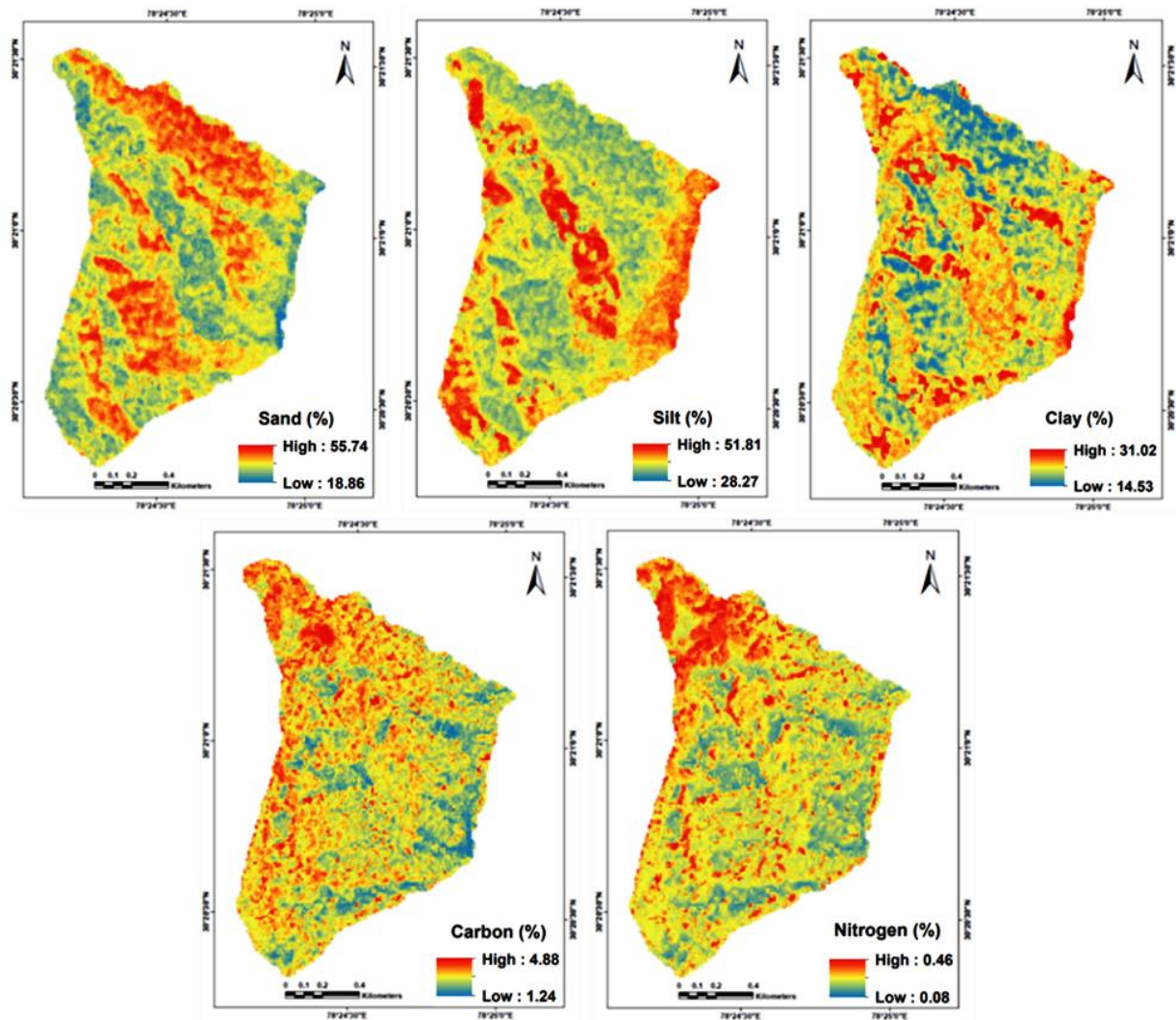


Figure 4. Predicted spatial distribution of different soil properties using RF model

Table 6. Descriptive statistics of spatially predicted soil properties

Parameters	Sand (%)	Silt (%)	Clay (%)	Org. C (%)	Nitrogen (%)
Min	18.86	28.27	14.52	1.24	0.08
Max	55.74	51.81	31.02	4.88	0.46
Mean	37.71	40.44	21.84	2.45	0.18
Standard deviation	4.37	3.45	1.7	0.29	0.03
Skewness	0.09	0.65	0.08	0.19	0.64
Kurtosis	-0.74	-0.12	0.43	0.53	0.72

The spatial distribution of primary soil particles (sand, silt and clay) as well as soil nutrients (soil carbon and nitrogen) showed large variation in their contents which is governed by the variations in terrain variables. The terrain characteristics play the most vital role in governing the transport and deposition of materials in these hilly mountainous ecosystems. The improved contribution of IMI as well as HLI in prediction of carbon as well as nitrogen may be attributed to their

significance in influencing the micro-variations in temperature as well as moisture distribution at landscape level. These moisture as well as temperature variations can influence the decomposition and other biochemical changes for these nutrients in addition to the physical material transport processes of erosion and deposition. Several researchers have reported similar role of moisture regimes in influencing distribution of various soil properties (Hairston and Grigal, 1994; Gessler *et al.*,

1995; Morris and Boerner 1998). Application of various advanced techniques suggested by various researchers such as wavelet transforms, neural networks, support vector machine, additive regression etc (Kalambukattu *et al.*, 2018; Pande *et al.*, 2022; Elbeltagi *et al.*, 2022) for soil properties prediction and mapping may also be explored for improving the accuracies in such difficult terrains. The maps of sand, silt and clay may be used for identifying the various zones of erosion and deposition, which will help in improved planning as well as implementation of various soil and water conservation measures in the watershed. Carbon and nitrogen maps may also help us in identifying areas with soil fertility constraints as well as degradation risks. This will aid in planning and implementation of various management practices ultimately leading to improvement in sustainable agricultural production.

5 CONCLUSION

The study was conducted to see the applicability of DEM derived terrain parameters for prediction and spatial mapping of different soil properties using random forest modeling in a watershed representing mountainous landscape of north west Himalayan region. The Pearson's correlation analysis revealed that only few terrain parameters like elevation, flow direction, HLI, solar radiation, aspect, elevation etc. had significant linear relationship with different soil properties while others have nonlinear relationships. Random forest modeling accounts for the nonlinear relationships and proved to be beneficial for spatial prediction. Assessment of the variable importance for RF modeling revealed that the covariates most crucial for modeling varied with the soil property under consideration and hence needs to be estimated independently. The random forest modeling using terrain parameters showed satisfactory results for the mapping of sand, silt, clay, organic carbon and nitrogen (r^2 values of 0.47, 0.52, 0.44, 0.44 and 0.41) explaining 47%, 52%, 44%, 41% and 55% of the variations, respectively. The different soil forming factors including parent material, climate and LULC were considered constant and hence not varying in the study watershed due to the small geographical area. The maps generated using predictive soil mapping technique could be used for delineation of broad spatial patterns of various soil properties, which will help us in planning as well as implementation of various soil and water conservation measures, nutrient management practices, etc. These interventions will help us towards achieving sustainable agricultural production, proper environment management as well as delivery of ecosystem services etc.

ACKNOWLEDGEMENTS

The authors sincerely thank Indian Space Research Organization (ISRO) for providing financial support under "Technology Development Plan" project. Authors place their sincere gratitude to all the anonymous

reviewers for their valuable comments and suggestions that improved the quality of this manuscript.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ABBREVIATIONS

CART: Classification and Regression Trees; **CCC**: Concordance Correlation Coefficient; **Cos_aspect**: Cos Transformed Aspect; **CTI**: Compound Topographic Index; **Curv_mean**: Mean Curvature; **Curv_plan**: Plan Curvature; **Curv_prof**: Profile Curvature; **CV**: Coefficient of variation; **DEM**: Digital Elevation Model; **DSM**: Digital Soil Mapping; **Elev_int**: Elevation; **FACC**: Flow Accumulation; **FCC**: False Color Composite; **FDIR**: Flow Direction; **FR**: Random Forest; **HLI**: Heat Load Index; **IMI**: Integrated Moisture Index; **IMI**: Integrated Moisture Index; **LULC**: Land Use Land Cover; **PDIR**: Potential Direct Incident Radiation; **RFM**: Random Forest Method; **RMSE**: Root Mean Square Error; **Sin_aspect**: Sin Transformed Aspect; **Slope_DEG**: Slope in Degrees; **SOC**: Organic Carbon; **Solar_radn**: Area Solar Radiation; **SPI**: Stream Power Index; **TWI**: Terrain Wetness Index.

REFERENCES

- Akpa, S. I. C., Odeh, I. O. A., Bishop, T. F. A. and Hartemink, A. E., 2014. Digital mapping of soil particle-size fractions for Nigeria. *Soil Sci. Soc. Am. J.* 78, 1953-1966. DOI: <https://doi.org/10.2136/sssaj2014.05.0202>
- Ben-Dor, E., Patkin, K., Banin, A. and Karnieli, A., 2002. Mapping of several soil properties using DAIS-7915 hyperspectral scanner data- A case study over clayey soil in Israel. *Int. J. Remote Sens.*, 23, 1043-1062. DOI: <https://doi.org/10.1080/01431160010006962>
- Ben-Dor, E., Taylor, R.G., Hill, J., Demattê, J. A. M., Whiting, M. L., Chabrilat, S., Sommer, S. and Donald, L. S., 2008. Imaging spectrometry for soil applications. *Adv. Agron.* 97, 321-392. DOI: [https://doi.org/10.1016/S0065-2113\(07\)00008-9](https://doi.org/10.1016/S0065-2113(07)00008-9)
- Boettinger, J. L., Ramsey, R. D., Bodily, J. M., Cole, N. J., Kienast-Brown, S., Nield, S.J., Saunders, A.M. and Stum, A. K., 2008. Landsat spectral data for digital soil mapping. In: Hartemink, A.E., Mcbratney, A.B., Mendonça-Santos, M.L. (Eds.), *Digital Soil Mapping with Limited Data*. Springer-Verlag, New York, pp. 192-202.
- Bouyoucos, G. J., 1962 Hydrometer method improved for making particle size analysis of soils. *Agron J.*, 54, 464-465. DOI: <https://doi.org/10.2134/agronj1962.00021962005400050028x>
- Breiman, L., 2001 Technical report for Version 3, 2001 (Accessed at: 12/28/14).
- Burrough, P. A., 1986. Principles of Geographic Information Systems for land resource assessment. Monographs on Soil and Resources Survey No. 12, Oxford Science Publications, New York.

- Chagas, C. S., Carvalho Junior, W., Bhering, S. B. and Filho, B. C., 2016. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. *Catena*, 139, 232-240. DOI: <https://doi.org/10.1016/j.catena.2016.01.001>
- Ciampalini, R., Martin, M., Saby, N., de Forges, A.C.R., Nehlig, P., Martelet, G. and Arrouays, D., 2014. Modelling soil particle-size distribution in the region "Centre" (France). In: Arrouays, D., McKenzie, N., Hempel, J., de Forges, A.C.R., McBratney, A. (Eds.), *GlobalSoilMap: Basis of the Global Spatial Soil Information System*. CRC Press, 325-331.
- Cutler, D. R., Edwards Jr., T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J. and Lawler, J. J., 2009. Random forests for classification in ecology. *Ecology*, 88, 2783-2792. DOI: <https://doi.org/10.1890/07-0539.1>
- de Carvalho J., W., Lagacherie, P., da Silva Chagas, C., Filho, C., B., and Bhering S. B., 2014. A regional-scale assessment of digital mapping of soil attributes in a tropical hillslope environment. *Geoderma*, 479-486. DOI: <https://doi.org/10.1016/j.geoderma.2014.06.007>
- Demattê, J. A. M., Galdos, M. V., Guimarães, R. V., Genú, A. M., Nanni, M. R. and Zullo Jr., J., 2007. Quantification of tropical soil attributes from ETM+/LANDSAT-7 data. *Int. J. Remote Sens.* 28, 3813-3829. DOI: <https://doi.org/10.1080/01431160601121469>
- Dharumarajan, S., Rajendra Hegde and Singh, S. K., 2017. Spatial prediction of major soil properties using Random Forest techniques- A case study in semi-arid tropics of South India. *Geoderma Regional*, 10, 154-162. DOI: <https://doi.org/10.1016/j.geodrs.2017.07.005>
- Elbeltagi, A., Pande, C. B., Kouadri, S., and Islam, A. R.T., 2022. Applications of various data-driven models for the prediction of groundwater quality index in the Akot basin, Maharashtra, India. *Environ Sci Pollut Res.*, 29, 17591-17605. <https://doi.org/10.1007/s11356-021-17064-7>
- ESRI [Environmental Systems Research Institute], 2011. ArcGIS Desktop: Release 10. Redlands, CA.
- Fu, P., and P. M. Rich. 2000. The Solar Analyst 1.0 Manual. Helios Environmental Modelling Institute (HEMI), USA.
- Fu, P. and Rich, P. M., 2002. A Geometric Solar Radiation Model with Applications in Agriculture and Forestry. *Computers and Electronics in Agriculture*, 37, 25-35. DOI: [https://doi.org/10.1016/S0168-1699\(02\)00115-1](https://doi.org/10.1016/S0168-1699(02)00115-1)
- Funnell, D. and Parish, R., 2005. *Mountain Environments and Communities*. Routledge, London.
- Gastaldi, G., Minasny, B. and McBratney, A. B., 2012. Mapping the occurrence and thickness of soil horizons within soil profiles. In: Minasny, B., Malone, B.P., McBratney, A.B. (Eds.), *Digital Soil Assessments and Beyond*. CRC Press/Balkema, London, pp. 145-148.
- Gessler, P. E., Moore, I. D., McKenzie, N. J. and Ryan, P. J., 1995. Soil-landscape modelling and spatial prediction of soil attributes. *Int. J. Geographical Information Systems*, 9, 421-432. DOI: <https://doi.org/10.1080/02693799508902047>
- Grimm, R., Behrens, T., Märker, M. and Elsenbeer, H., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island- Digital soil mapping using Random Forests analysis. *Geoderma*, 146, 102-113. DOI: <https://doi.org/10.1016/j.geoderma.2008.05.008>
- Guo, P. T., Li, M. F., Luo, W., Tang, Q. F., Liu, Z. W. and Lin, Z. M., 2015. Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach. *Geoderma*, 237, 49-59. DOI: <https://doi.org/10.1016/j.geoderma.2014.08.009>
- Jeong, G., Oeverdieck, H., Park, S. J., Huwe, B. and Ließ, M., 2017. Spatial soil nutrients prediction using three supervised learning methods for assessment of land potentials in complex terrain. *Catena*, 154, 73-84. DOI: <https://doi.org/10.1016/j.catena.2017.02.006>
- Hairston, A. B. and Grigal, D. F., 1994. Topographic variation in soil water and nitrogen for two forested. *Geoderma*, 64(1-2), 125-138. DOI: [https://doi.org/10.1016/0016-7061\(94\)90093-0](https://doi.org/10.1016/0016-7061(94)90093-0)
- Hitziger, M. and Ließ, M., 2014. Comparison of three supervised learning methods for digital soil mapping: Application to a complex terrain in the Ecuadorian Andes. *Appl. Environ. Soil Sci.*, 1-12. DOI: <https://doi.org/10.1155/2014/809495>
- Iverson, L. R., Dale, M. E., Scott, C. T. and Prasad, A., 1997. A GIS-derived integrated moisture index to predict forest composition and productivity in Ohio forests. *Landscape Ecology*, 12, 331-348.
- Jenson, S. K., 1991 Applications of hydrological information automatically extracted from digital elevation models. *Hydrol Process*, 5, 31-44. DOI: <https://doi.org/10.1002/hyp.3360050104>
- Kalambukattu, J. G., Kumar, S. and Arya Raj, R., 2018. Digital soil mapping in a Himalayan watershed using remote sensing and terrain parameters employing artificial neural network model. *Environmental Earth Sciences*, 77, 203. DOI: <https://doi.org/10.1007/s12665-018-7367-9>
- Lagacherie, P., McBratney, A. B. and Voltz, M., 2007. *Digital Soil Mapping: An Introductory Perspective*. Elsevier, Amsterdam.
- Lark, R. M., Dove, D., Green, S. L., Richardson, A. E., Stewart, H. and Stevenson, A., 2012. Spatial prediction of seabed sediment texture classes by cokriging from a legacy database of point observations. *Sediment. Geol.* 281, 35-49.
- Liaw, A. and Wiener, M., 2002. Classification and regression by randomForest. *R News*, 2(3), 18-22.
- Ließ, M., Glaser, B. and Huwe, B., 2012. Uncertainty in the spatial prediction of soil texture: Comparison of regression tree and Random Forest models. *Geoderma*, 170, 70-79. DOI: <https://doi.org/10.1016/j.geoderma.2011.10.010>
- McCune, B. and Keon, D., 2002. Equations for potential annual direct incident radiation and heat load. *J. Veg. Sci.*, 13, 603-606. DOI: <https://doi.org/10.1111/j.1654-1103.2002.tb02087.x>
- Moore, I. D., Gesslewr, P. E., Nielsen, G. A. and Peterson, G. A., 1993. Soil attribute prediction using terrain analysis. *Soil Sci. Soc. Am. J.*, 57, 443-452. DOI: <https://doi.org/10.2136/sssaj1993.03615995005700020026x>
- Morris, S. M and Boerner, R. E. J., 1998. Landscape patterns of nitrogen mineralization and nitrification in southern Ohio hardwood forests. *Landscape Ecology*, 13, 215-224. DOI: <https://doi.org/10.1023/A:1007967630020>
- Mulder, V. L., De Bruin, S., Schaepman, M.E. and Mayr, T. R., 2011. The use of remote sensing in soil and terrain mapping- A review. *Geoderma*, 162, 1-19. DOI: <https://doi.org/10.1016/j.geoderma.2010.12.018>
- NRSC [National Remote Sensing Centre]. 2014. Evaluation of Indian National DEM (Version-2) from Cartosat-1 data. NRSC, ISRO.
- Pande, C. B., Kadam, S. A., Jayaraman, R., Gorantiwar, S. and Shinde, M., 2022. Prediction of soil chemical

- properties using multispectral satellite images and wavelet transforms methods. *Journal of the Saudi Society of Agricultural Sciences*, 21(1), 21-28. DOI: <https://doi.org/10.1016/j.jssas.2021.06.016>
- Prasad, A. M., Iverson, L. R. and Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, 9, 181-199. DOI: <http://dx.doi.org/10.1007/s10021-005-0054-1>
- R Development Core Team, 2007. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, Accessed on 05/08/2015.
- Rich, P. M., Dubayah, R., Hetrick, W. A. and Saving, S. C., 1994. Using View shed Models to Calculate Intercepted Solar Radiation: Applications in Ecology. *American Society for Photogrammetry and Remote Sensing Technical Papers*, 524-529.
- Sreenivas, K., Dadhwal, V. K., Kumar, S., Sri Harsha, G., Mitran, T., Sujatha, G., Janaki Rama Suresh, G., Fyzee, M. A. and Ravisankar, T., 2016. Digital organic and inorganic carbon mapping of India. *Geoderma*, 269, 160-173. DOI: <https://doi.org/10.1016/j.geoderma.2016.02.002>
- Stevens, A., Van Wesemael, B., Bartholomeus, H., Rossillon, D., Tychon, B. and Ben-Dor, E., 2008. Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. *Geoderma*, 144, 395-404. DOI: <https://doi.org/10.1016/j.geoderma.2007.12.009>
- Vågen, T. G., Winowiecki, L. A., Tondoh, J. E., Desta, L. T. and Gumbricht, T., 2016. Mapping of soil properties and land degradation risk in Africa using MODIS reflectance. *Geoderma*, 263, 216-225. DOI: <https://doi.org/10.1016/j.geoderma.2015.06.023>
- Vaysse, K. and Lagacherie, P., 2015. Evaluating Digital Soil Mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). *Geoderma Reg.*, 4, 20-30. DOI: <https://doi.org/10.1016/j.geodrs.2014.11.003>
- Wiesmeier, M., Barthold, F., Blank, B. and Kögel-Knabner, I., 2011. Digital mapping of soil organic matter stocks using random Forest modeling in a semi-arid steppe ecosystem. *Plant Soil*, 340, 7-24. <http://dx.doi.org/10.1007/s11104-010-0425-z>
