ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/360890892

Modelling and Prediction of Soil Organic Carbon using Digital Soil Mapping in the Thar Desert Region of India

Article *in* Journal of the Indian Society of Soil Science · May 2022

DOI: 10.5958/0974-0228.2022.00009.3

CITATIONS
0

READS
63

14 authors, including:

Pravash Chandra Moharana
National Bureau of Soil Survey and Land Use Planning
67 PUBLICATIONS **515** CITATIONS
SEE PROFILE

Dharumarajan s
National Bureau of Soil Survey and Land Use Planning
99 PUBLICATIONS **436** CITATIONS
SEE PROFILE

Nirmal Kumar
National Bureau of Soil Survey and Land Use Planning
89 PUBLICATIONS **286** CITATIONS
SEE PROFILE

Roomesh Kumar Jena
Indian Institute of Water Management
52 PUBLICATIONS **141** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project 2nd Joint workshop for Digital Soil Mapping and GlobalSoilMap IUSS WGs View project

Project CRP On CA View project

All content following this page was uploaded by Pravash Chandra Moharana on 31 May 2022.

The user has requested enhancement of the downloaded file.

Journal of the Indian Society of Soil Science, Vol. 70, No. 1, pp 86-96 (2022)
DOI: 10.5958/0974-0228.2022.00009.3

Modelling and Prediction of Soil Organic Carbon using Digital Soil Mapping in the Thar Desert Region of India

P.C. Moharana[1]*, S. Dharumarajan[2], Nirmal Kumar[1], R.K. Jena[3], U.K. Pradhan[4], R.S. Meena, S. Sahoo[5], M. Nogiya, Sunil Kumar[6], R.L. Meena, B.L. Tailor, R.S. Singh, S.K. Singh[7] and B.S. Dwivedi[1]

ICAR-National Bureau of Soil Survey and Land Use Planning, Regional Centre, Udaipur, 313001, Rajasthan, India

In the present study, the distribution of soil organic carbon (SOC) was investigated using digital soil mapping for an area of ~29 lakhs ha in Bikaner district, Rajasthan, India. To achieve this goal, 187 soil profiles were used for SOC estimation by Quantile regression forest (QRF) model technique. Landsat data, terrain attributes and bioclimatic variables were used as environmental variables. 10-fold cross-validation was used to evaluate model. Equal-area quadratic splines were fitted to soil profile datasets to estimate SOC at six standard soil depths (0-5, 5-15, 15-30, 30-60, 60-100 and 100-200 cm). Results showed that the mean SOC concentration was very low with values varied from 1.18 to 1.53 g kg^{-1} in different depths. While predicting SOC at different depths, the model was able to capture low variability (R^2 = 1–7%). Overall, the Lin's concordance correlation coefficient (CCC) values ranged from 0.01 to 0.18, indicating poor agreement between the predicted and observed values. Root mean square error (RMSE) and mean error (ME) were 0.97 and 0.16, respectively. The values of prediction interval coverage probability (PICP) recorded 87.2–89.7% for SOC contents at different depths. The most important variables for predicting SOC concentration variations were the annual range of temperature, latitude, Landsat 8 bands 2, 5 and 6. Temperature-related variables and remote sensed data products are important for predicting SOC concentrations in arid regions. We anticipate that this digital information of SOC will be useful for frequent monitoring and assessment of carbon cycle in arid regions.

Key words: Digital soil mapping, quantile regression forest, soil organic carbon, desert regions of India

Digital soil mapping (DSM) has now been widely used globally for mapping soil classes and properties (Arrouays *et al.* 2014). In particular, DSM has been used to map soil organic carbon (SOC) efficiently around the world. The DSM methodology has been adopted by Food and Agriculture Organization (FAO) and Intergovernmental Technical Panel on Soils (ITPS) (2020) so that digital soil maps can be produced reliably for sustainable land management. Traditional soil mapping techniques mostly depend on ground based surveys and rarely provide information about the spatial distribution of soil properties at the desired resolution over the landscape (McBratney *et al.* 2003; Minasny *et al.* 2013). Furthermore, mapping soil spatial variations by traditional field surveys is time-consuming and expensive, especially at national, regional or global scales (Dharumarajan *et al.* 2019). Therefore, it is necessary to have robust methods and models to predict soil properties at a given location or scale. Considerable advances in remote sensing techniques and machine learning approaches have allowed

*Corresponding author (Email: pravashiari@gmail.com)
Present address
[1]ICAR-National Bureau of Soil Survey and Land Use Planning, Nagpur, 440033, Maharashtra, India
[2]ICAR-National Bureau of Soil Survey and Land Use Planning Regional Centre, Hebbal, Bangalore, 560024, Karnataka, India
[3]ICAR-Indian Institute of Water Management, Bhubaneswar, 751023, Odisha, India
[4]ICAR-Indian Agricultural Statistics Research Institute, New Delhi, 110012, India
[5]ICAR-Central Inland Fisheries Research Institute, Regional Centre, Bengaluru, 560089, Karnataka, India
[6]ICAR-National Bureau of Soil Survey and Land Use Planning, Regional Centre, New Delhi, 110012, India
[7]ICAR-Central Coastal Agricultural Research Institute, Goa, 403402, India

accurate prediction of soil properties with new methods like digital soil mapping (McBratney *et al.* 2003; Hengl *et al.* 2015). In recent years, DSM techniques have been used to map soil properties using environmental variables. These methods were designed to overcome the limitations of the conventional soil mapping approach and to estimate soil properties based on relationships between soil and environmental variables obtained from terrain attributes and satellite imagery (McBratney *et al.* 2003; Minasny and Hartemink 2011). The *GlobalSoilMap* consortium was established in response to such a soaring demand for up-to-date and relevant soil information (Arrouays *et al.* 2014). This consortium has undertaken the task of producing soil property maps at a fine resolution using DSM techniques. Many countries have abundant legacy soil data that include soil maps at a variety of scales, soil point data collected over decades, environmental covariate information and a network of partners that have contributed to building the soil information over many years.

Digital soil mapping methodology works are based on SCORPAN model where soil class or soil property (s) at particular location are derived indirectly from the environmental variables: climate (c), organisms (o), relief (r), parent material (p), age (a), and spatial location (n) (McBratney *et al.* 2003). Following the ideas of Dokuchaev (1883) and Jenny (1941), they described the SCORPAN model as the empirical quantitative relationship of a soil attribute and its spatially implicit forming factors.

$$S = f(s, c, o, r, p, a, n) \qquad \ldots(i)$$

In the SCORPAN model, soil, either as point observational data, existing soil maps, or remotely sensed spectral properties, can be used as input data. Environmental covariates are digital and spatially explicit data in a raster that is processed using a geographic information system (GIS). The SCORPAN model facilitates the quantification of the relationships between spatially explicit digital environmental covariates and the soil classes or attributes to be predicted in a spatial context. It also facilitates the estimation of error or uncertainty of the spatial prediction of soil classes or properties.

Soil organic carbon is more important in Indian desert regions where soils are inherently low in organic carbon content and the production system is fragile (Singh *et al.* 2005; Kumar *et al.* 2009). The desert regions are characterized by sparse and highly variable precipitation and high evaporation. Drought is characteristic of deserts due to lack of moisture. Vegetation is sparse due to limited availability of water and adverse climatic conditions. However, information on changes in soil organic matter (SOM) after agricultural interventions in arid regions is meagre. Compared with other ecosystems, deserts have received much less attention in this regard because of their lower plant biomass and soil carbon storage (Singh *et al.* 2005). Indeed, profound variances in the vegetation pattern and large spatial heterogeneity in available resources in desert ecosystems contribute to significant uncertainty in evaluating their roles as carbon sources or sinks which can exhibit significant spatial variation.Therefore, efficient modeling of SOC is important in arid regions because it provides information about soil fertility, water conservation, carbon sequestration, climate change, and the effects of land use practices. The digital mapping of SOC at fine resolution is a challenging task and the mapping is also a high priority for SOC assessment and monitoring. Several researchers applied splines to model the vertical distribution of SOC in the soil profiles and predicted SOC at a landscape scale using data-mining tools and environmental variables as predictors (Akpa *et al.* 2016). Recently, in several studies, soil properties such as soil pH, SOM, electrical conductivity (EC), phosphorus (P), and particle size distribution have been predicted and mapped. Numerous prediction models have been developed and introduced to correlate ancillary variables and soil properties through the DSM framework suggested by McBratney *et al.* (2003).

Procedures and methodologies to produce this information vary depending on the types and amount of available data, but all information should meet the *GlobalSoilMap* standards and specifications. In this communication, we review the specifications and the state of progress of *GlobalSoilMap* products delivery. We focus on information on SOC and related soil information useful for mapping and modelling SOC and their changes over large areas. *GlobalSoilMap* uses legacy soil data collected over many decades. Data for any point reflect the state of the soil at the time the point was sampled and analyzed. A major challenge for SOC is to reconcile differences in SOC reported for different times under different land uses. In India, ICAR-National Bureau of Soil Survey and Land Use Planning (ICAR-NBSS&LUP), Nagpur has recently launched an ambitious program called "*IndianSoilGrids*" with the objective to develop soil properties map as per *GlobalSoilMap* specifications. As there is a lack of information about the spatial variability of SOC in the Thar desert, the main aims of this research is to spatially predict SOC through Quantile Regression Forest (QRF) Model techniques. The objectives of this study were to (1) predict the spatial distribution of SOC using quantile regression forest techniques and (2) to find the main environmental covariates that control the SOC in the desert regions of India. We expect that our outputs would improve and update the current SOC Information System with new fine-resolution SOC maps that could be useful to end users and stakeholders.

Materials and Methods

Study area

This study was conducted in Bikaner district, located in the western plains of Rajasthan, India (71°50′52″ to 74°24′27″ E longitudes and 27°09′20″ to 29°05′14″ N latitudes) (Fig. 1). This landscape is dominated by eolian erosional and depositional landforms. The study area is relatively flat with a slope of 1–2% and covers an area about 29 lakhs ha. Elevation ranges from 120 m to 312 m above sea level. The climate of the area is hot arid, erratic rainfall (100–450 mm yr^{-1} ~90% occurs during July–September), extreme temperatures (often > 45 °C in the peak of summer and sub-zero in winter), and high summer winds (>30 km h^{-1} during sandstorms in summer). The land use is mainly cropland including mustard (*Brassica juncea*), moth bean (*Vigna aconitifolia*), cluster bean (*Cyamopsis tetragonoloba*), groundnut (*Arachis hypogaea*), chickpea (*Cicer arietinum*), wheat (*Triticum aestivum*), green gram (*Vigna radiata*) and pearl millet (*Pennisetum glaucum*) but abandoned lands, saline lands, and sand dunes also occur across the study area. Cereal and legume based cropping systems are practised more than last 30 years when Bikaner lift canal network was started during late 1980. Much of natural vegetation includes babul (*Acacia Arabica*), khejri (*Prosopis spicigera*), neem (*Azadirachta indica*), vilayati babul (*Prosopis juliflora*) and ber (*Zizyus jijuba*). Common soils include Entisols, and Aridisols according to US Soil Taxonomy. The dominant soils are deep to very deep, either calcareous or non-calcareous and sandy in nature. The area is characterized by strong south westerly winds during summer, which causes frequent sandstorm and sand movement. Gypsite rich beds are found in shallow depression surrounded by sand dunes. The soils of this region have low SOC because of sparse natural vegetation resulting from very low precipitation and high evaporation, as well as management practices that do not encourage SOC retention within the soil.

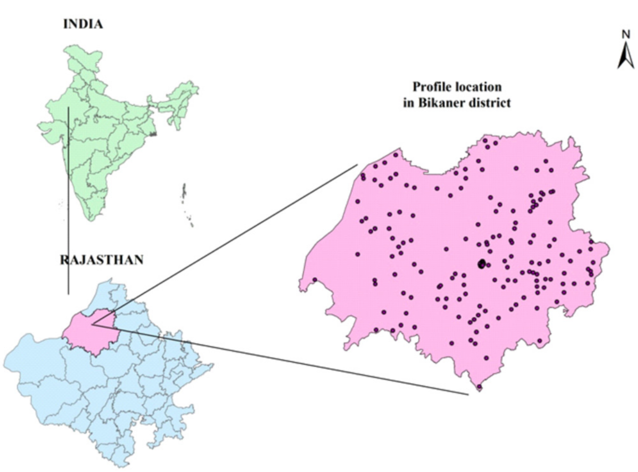

Fig. 1. Spatial distribution of profile sites in the study area of desert regions of India

Soil data source

A total of 187 soil profiles were used in this study (Fig. 1). They were obtained by a recent project of "Land resource inventory of Bikaner district, Rajasthan at 1:10000 scale for optimal agricultural land use planning using geo-spatial technique" conducted from 2017 to 2019. Typical soil profiles representing main soil-landscapes were collected. The geographical coordinates of each soil profile location were recorded with a GPS receiver. The soil pits were generally dug to a depth of 1.5–2 m or until a lithic or paralithic contact. All soil profiles were divided vertically into different pedological horizons according to specific profile morphology, and soil samples were collected from each horizon. In laboratory, samples were air-dried at room temperature and then passed through a 2 mm sieve. The SOC was determined by the Walkley and Black (1934) method.

Deriving sample data at predefined depths

For each profile of a SOC, we used equal-area quadratic splines to fit a continuous depth function to original horizon sample data. The *GobalSoilMap* programme follows six standard depth horizons for better comparison between profile characteristics *e.g.* 0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm and 100–200 cm (Arrouays *et al.* 2014). Therefore, soil properties of each profile were calculated for the above mentioned standard depths using mass preserving spline method (Bishop *et al.* 1999). This was done using equal-area spline functions implemented in the GSIF R-package.

Environmental covariates

Environmental covariates are the variables characterizing soil forming factors of climate, parent material, topography, vegetation, human activities and time. In the last decade, many efforts have been made in sifting and developing effective environmental covariates according to the targeted soil properties and landscapes. Climate and terrain factors have been widely used in DSM. A digital elevation model (DEM) acquired with shuttle radar topographic mission (SRTM) has a spatial resolution of 30 m and processed using ArcGIS10 data management tool box. The primary and secondary derivates of DEM like elevation, slope, aspect, curvatures (plan and profile), topographic wetness index (TWI) and topographic position index (TPI), LS factor, multi-resolution ridge top flatness (MrRTF) and multi-resolution index of valley bottom flatness (MrVBF) were derived by using Saga-GIS 6.3.0 version (Table 1). Along with DEM attributes, all the bands of landsat-8 imagery (11 bands), normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI) (MOD13Q1) were used as covariates for prediction of SOC. Raster data on bioclimatic variables at 30 sec resolution were downloaded from http://worldclim.org/current for the whole world and the respective grids for study area were extracted from these world grids. Nine bioclimatic variables were selected for use as covariates in digital soil mapping study and these variables are mean annual temperature (MAT), temperature seasonality (TS), mean diurnal range of temperature (MDRT), annual range of temperature (ART), isothermality, annual precipitation (AP), precipitation seasonality (PS), precipitation of wettest month (PWM) and precipitation of driest month (PDM).

Spatial prediction of SOC using quantile regression forest model

Quantile regression forest (QRF) model was used for prediction of SOC and uncertainty estimates in the study area. The QRF is an extension of random forest model (RFM) and the advantage of QRF over RFM is for each node in each tree, RFM keeps only the mean of the observations that fall into this node and neglects all other information whereas QRF keeps the value of all observations in this node, and assesses the conditional distribution based on the information (Meinshausen 2006). For the present study, ranger package was used for running the QRF algorithmin in R environment (R Core Team 2019). Ranger package helps to identify the best RF properties for running the model. Ten folds cross validation techniques with

Table 1. Different covariates used for Quantile Regression Forest model

Predictor	Source	Resolution
Elevation (m)	SRTM DEM	30 m
Slope	SRTM DEM	30 m
Aspect	SRTM DEM	30 m
Topographic position index (TPI)	SRTM DEM	30 m
Terrain ruggedness index (TRI)	SRTM DEM	30 m
Plan curvature	SRTM DEM	30 m
Profile curvature	SRTM DEM	30 m
Multi-resolution index of valley bottom flatness (MrVBF)	SRTM DEM	30 m
Multi-resolution ridge top flatness (MrRTF)	SRTM DEM	30 m
Normalized difference vegetation index (NDVI)	MOD13Q1	250 m _16 days
Enhanced vegetation index (EVI)	MOD13Q1	250 m _16 days
Landsat data	11 bands	30 m
Annual mean temperature (MAT) (°C)	Worldclim2	30 sec
Temperature seasonality (TS)	Worldclim2	30 sec
Mean diurnal range of temperature (MDRT)	Worldclim2	30 sec
Annual range of temperature (ART)	Worldclim2	30 sec
Isothermality	Worldclim2	30 sec
Annual precipitation (AP)	Worldclim2	30 sec
Precipitation seasonality (PS)	Worldclim2	30 sec
Precipitation of wettest month (PWM)	Worldclim2	30 sec
Precipitation of driest month (PDM)	Worldclim2	30 sec

20 times repetition was used to evaluate the performance of QRF model.

Model accuracy assessment

Prediction performance of QRF was evaluated based on three parameters *viz*. coefficient of determination (R^2) which is defined by percentage of variation explained by the model, mean error (ME), root mean square error (RMSE) and Lin's concordance correlation coefficient (CCC) which is a measure of agreement between predicted and observed values. The good models have coefficient of determination and concordance correlation coefficient is equal or close to 1 and root mean square error close to 0.

$$\text{Coefficient of determination } (R^2) = 1-\frac{\sum_{i=1}^{n}(\text{pi}-\text{oi})^2}{\sum_{i=1}^{n}(\overline{\text{pi}}-\overline{\text{oi}})^2} \quad \text{...(ii)}$$

$$\text{Mean error (ME)} = \frac{1}{n}\sum_{i=1}^{n}(\text{pi}-\text{oi}) \quad \text{...(iii)}$$

$$\text{Root mean squared error (RMSE)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\text{pi}-\text{oi})^2} \quad \text{...(iv)}$$

where, pi and oi are predicted and observed values, $\overline{\text{pi}}$ and $\overline{\text{oi}}$ are means of predicted and observed values.

$$\text{Lin's concordance correlation coefficient (CCC)} = \frac{2\rho\sigma_0\sigma_p}{\sigma_0{}^2+\sigma_p{}^2+(\mu_0-\mu_p)^2} \quad \text{...(v)}$$

where, μ_0 and μ_p are the means of observed and predicted values, σ_0^2 and σ_p^2 are corresponding variances ρ is pearson correlation coefficient between observed and predicted values.

Uncertainty estimation

Uncertainty representation is a crucial aspect of DSM (Arrouays *et al.* 2014). The DSM models are not only expected to deliver accurate soil predictions at a given location but their suitability to deliver maps should encompass ability to predict the uncertainty (Vaysse and Lagacherie 2017). Prediction interval coverage percentage (PICP) was used to evaluate the uncertainty of prediction. The PICP is simply the proportion of observations at each depth that are encapsulated by the corresponding prediction interval. In this study, the uncertainties of the predicted maps of SOC contents were estimated using QRF (Vaysse and Lagacherie 2017). If the uncertainty estimates have been reasonably defined, the PICP should result in an estimate of 90% for a 90% prediction interval. The estimation produced by differentiation of 0.05 and 0.95 quantiles for SOC in different depth and the map produced for every pixel location of the study area and every depth.

Results and Discussion

Statistical description of SOC

Table 2 lists statistical description of the splines-fitted SOC percentages at different depths based on the soil profiles. The SOC in the study area ranged from 0.13 to 6.3 g kg^{-1} with a mean of 1.53 g kg^{-1} in 0-5 cm. The SOC in 5-15 cm ranged from 0.05 to 6.32 g kg^{-1} with a mean and standard deviation of 1.52 g kg^{-1} and 0.98 g kg^{-1}, respectively. The spline-predicted SOC content for the first three depths, *i.e.*, 0-5, 5-15, and 15-30 cm, for the profiles remained more or less the same due to a 'slicing' pretreatment applied to the splines to take account of the homogeneity of the plough layer. The SOC content decreases with increasing depth. The SOC of all layer had high levels of variation (CV > 36%) according to the guidelines provided by Wilding (1985). The SOC content had the higher CV with depths and values ranging from 65.1% for 0–15 cm to 83.9% for 100–200 cm. The SOC content skewed positively with skewness values to 1.51 to 2.29 for different depths. Correlation analysis between SOC content in different

Table 2. Statistical description of the splines-soil organic carbon at different depths in the study area of desert regions of India

Parameters	Soil organic carbon (g kg^{-1})					
	0-5 cm	5 -15cm	15-30 cm	30-60 cm	60-100 cm	100-200 cm
Minimum	0.13	0.05	0.11	0.00	0.01	0.11
Maximum	6.33	6.32	7.28	6.01	5.22	4.35
Mean	1.53	1.52	1.44	1.31	1.22	1.18
SE	0.073	0.072	0.07	0.069	0.065	0.098
SD	1.00	0.98	0.95	0.94	0.89	0.99
CV (%)	65.1	64.5	66.3	71.7	72.4	83.9
Skewness	1.58	1.61	2.29	2.2	1.76	1.51
Kurtosis	3.67	3.90	8.80	6.24	3.90	1.77

depths and covariates used in the model (Table 3) showed that TRI, MDRT, ART and landsat 8 band 1, 2, 3 and 8 have a significant correlation with SOC content in the surface soil (0-5 and 5-15 cm) whereas TS, MDRT, ART, latitude and landsat 8 band 1, 2, 3, 5 and 8 have a significant correlation with SOC content in the 15-30 cm soil depth. At below 100 cm soil depth, TS, ART, AP and PWM have a significant correlation with SOC. Bioclimatic covariates have significant negative correlation with SOC content. These results revealed that the mean SOC concentration was extremely low. The main reasons for low SOC concentration in the study area which part of Thar Desert include the high temperature, high evaporation rates, lack of precipitation, intense wind erosion, and spare natural vegetation. The lower SOC mainly found in the sand dunes area which may be due to low vegetation and exposure of soils to the scorching sun causes oxidation of SOC (Singh *et al.* 2009). The highest SOC concentration in our study may be due to adoption of different soil management practices including variation in fertilizer application and other crop management practices (Sharma and Singh 2015; Moharana *et al.* 2017).

Performance of model prediction and uncertainty estimation

The performance of quantile regression forest model was evaluated by calculating statistical indicators *viz.*, coefficient of determination (R^2), mean error (ME) and root mean square error (RMSE). The cross validation results showed that the combination

Table 3. Correlation analysis between covariates and soil organic carbon with soil depths in the study area of desert regions of India

Covariates	Soil organic carbon					
	0-5 cm	5 -15 cm	15-30 cm	30-60 cm	60-100 cm	100-200 cm
Elevation	0.013	0.021	0.040	0.110	0.164*	0.054
Slope	-0.102	-0.096	-0.086	-0.072	-0.115	-0.089
Aspect	-0.001	-0.007	-0.033	-0.011	-0.028	-0.036
Plan_Curva	-0.002	-0.003	-0.005	-0.007	-0.002	0.069
Profile_Cu	-0.059	-0.062	-0.067	-0.043	-0.010	0.113
TPI	-0.076	-0.070	-0.026	-0.035	-0.013	0.069
TRI	-0.152*	-0.146*	-0.124	-0.075	-0.082	-0.079
TWI	0.057	0.057	0.062	0.009	-0.003	0.211**
MrVBF	0.063	0.053	0.056	0.045	-0.012	0.001
MrRTF	-0.111	-0.112	-0.126	-0.118	-0.024	-0.174*
NDVI	-0.017	-0.021	-0.056	-0.094	-0.069	-0.071
EVI	-0.017	-0.020	-0.049	-0.089	-0.067	-0.055
B1_landsat	0.202**	0.205**	0.191**	0.166*	0.073	0.037
B2_landsat	0.192**	0.195**	0.187*	0.169*	0.076	0.040
B3_landsat	0.168*	0.170*	0.163*	0.164*	0.100	0.076
B4_landsat	0.134	0.138	0.139	0.151*	0.087	0.054
B5_landsat	0.120	0.124	0.149*	0.186*	0.104	0.215**
B6_landsat	0.111	0.115	0.118	0.151*	0.122	0.062
B7_landsat	0.080	0.085	0.100	0.136	0.101	0.056
B8_landsat	0.148*	0.151*	0.146*	0.142	0.063	0.048
B9_landsat	0.080	0.080	0.073	0.045	-0.054	-0.027
B10_landsat	0.016	0.016	-0.001	0.065	0.167*	0.070
B11_landsat	0.005	0.007	0.003	0.082	0.175*	0.059
MAT	0.042	0.042	0.034	-0.022	-0.055	-0.030
TS	-0.121	-0.130	-0.145*	-0.199**	-0.207**	-0.166*
MDRT	-0.214**	-0.216**	-0.178*	-0.145*	-0.137	-0.032
ART	-0.163*	-0.170*	-0.166*	-0.185*	-0.180*	-0.071
Isothermal	-0.125	-0.118	-0.060	0.023	0.030	0.051
AP	0.047	0.053	0.062	0.109	0.157*	-0.037
PS	-0.001	0.003	0.027	0.090	0.106	0.101
PWM	0.081	0.086	0.087	0.128	0.177*	0.003
PDM	0.030	0.032	0.069	0.117	0.101	0.012
Longitude	-0.076	-0.075	-0.059	-0.024	0.009	-0.181*
Latitude	-0.125	-0.134	-0.148*	-0.199**	-0.226**	-0.236**

*Correlation is significant at the 0.05 level; **Correlation is significant at the 0.01 level

Table 4. Performance of quantile regression forest model for prediction of soil organic carbon in the study area of desert regions of India

Parameters	Soil organic carbon					
	0-5 cm	5 -15 cm	15-30 cm	30-60 cm	60-100 cm	100-200 cm
Mean error	0.16±0.01	0.16±0.01	0.16±0.01	0.25±0.01	0.20±0.01	0.31±0.01
RMSE	0.97±0.01	0.96±0.01	0.94±0.01	0.96±0.00	0.91±0.00	1.02±0.01
R^2 (%)	7.0±2.0	6.0±2.0	5.0±2.0	2.0±1.0	1.0±1.0	3.0±2.0
PICP (%)	88.4±1.12	89.1±0.86	87.6±0.85	88.2±0.72	89.7±0.58	87.2±1.7

of different covariates explained the variability of predicted SOC (Table 4 and Fig. 2). The model could capture low variability (R^2 = 1–7%) while predicting SOC for different depths. Overall, the mean CCC values ranged from 0.01 to 0.18, indicating poor agreement between the predicted and observed values. The R^2 and RMSE values slightly decreased downward from the depth of 5 cm while the ME values increased, exhibiting a vertical decline of predictability of SOC. The mean ME values were very close to zero, suggesting overall unbiased predictions. The 0–5 cm depth interval had a better model performance than the 5-15 cm depth interval, with the higher R^2 values. The performance of organic carbon prediction is related to its dynamic nature and poor performance may be related to the low levels of SOC compared to soils having high organic carbon (Sharma and Singh 2015; Dharumarajan *et al.* 2021). The values of prediction interval coverage probability (PICP) recorded 87.2-89.7% for SOC contents at different depths. For a 90% prediction interval, we would expect 90% of observations to fall within the

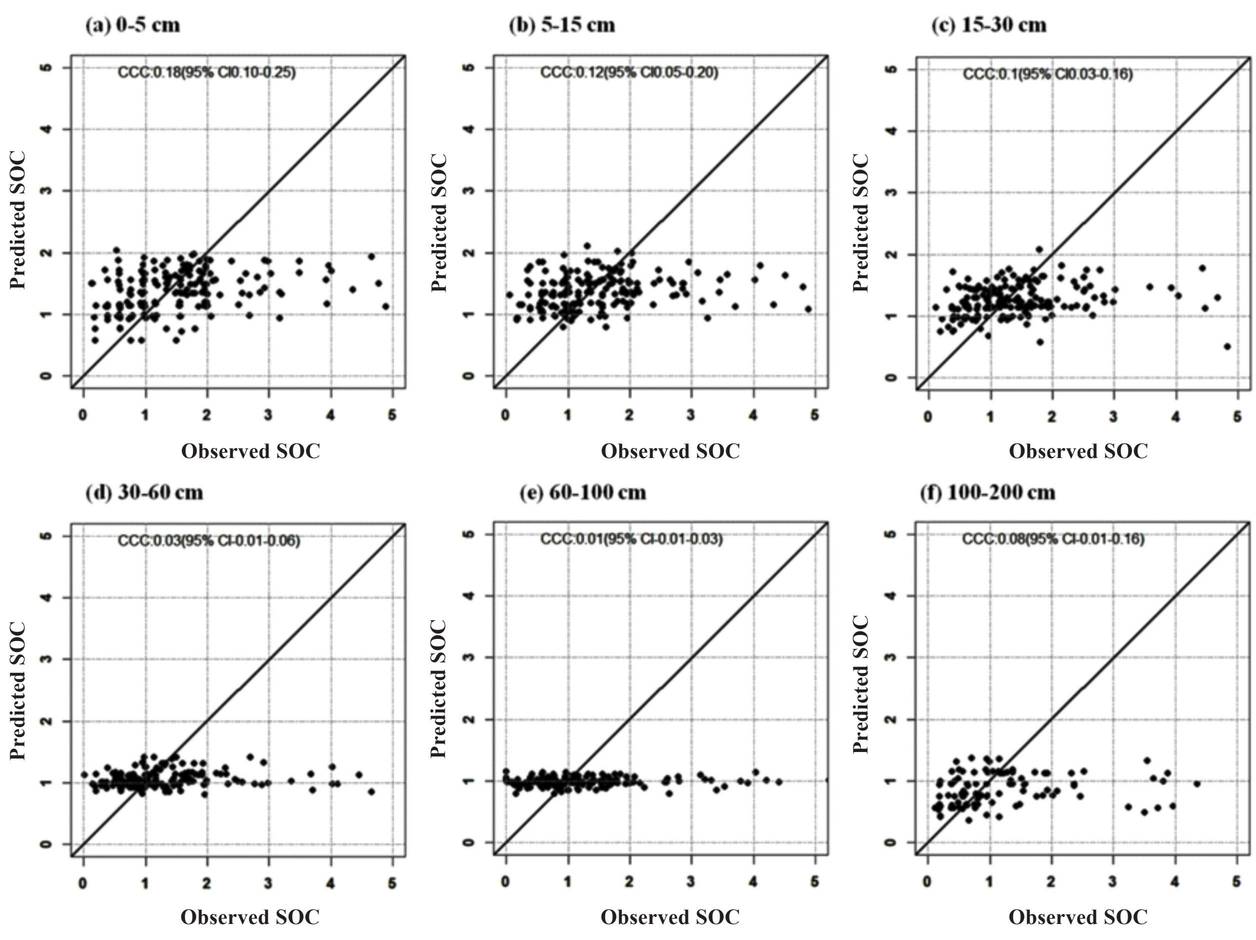

Fig. 2. Depth wise predicted *vs* observed soil organic carbon by quantile regression forest model in the study area of desert regions of India

lower and upper prediction limits. It can be seen that all the fractions achieved PICP values very close to 90%, suggesting that these lower and upper prediction limits estimated by the QRF method were of an appropriate magnitude. That is to say, the uncertainty estimations, to a large extent, were reliable. This can also be indicated by the small values of standard deviation of overall prediction accuracy indicators. Overall, the prediction performance of this model was low for SOC. Higher sample density is required for better findings in regions with a wide range of environmental covariates such as soil properties, climate variables, land cover and relief factors.

Importance of predictor variables for predicting SOC

Fig. 3 shows relative importance of the 32 important covariates used in the predictions of SOC at multiple depths. In SOC predictions, the importance of the covariates did not have obvious changes with depth. The most important variables for predicting SOC variations were the annual range temperature (ART), latitude and landsat 8 bands 2, 5 and 6. Temperature-related variables (ART and TS) were the second important covariates in surface soil (0-5 cm). It appeared that changes of air temperature at diurnal and seasonal and even annual scales became more important than its mean status with the increase in depths. Relatively temperature changes and low moisture conditions in the hot arid region of India lead to relatively importance of predictor variables for predicting SOC. The MrRTF was the second important covariates as most important predictor for prediction in 100-200 cm soil. Different bands of landsat 8 imagery occupies in the top position for prediction of SOC in most of soil depths. Different researchers recorded usefulness of landsat 8 imageries in prediction of SOC (Dharumarajan *et al.* 2021). The NDVI and EVI were not important variables in this study. This could be because of the sandy arid nature of the study area or because of sparse natural

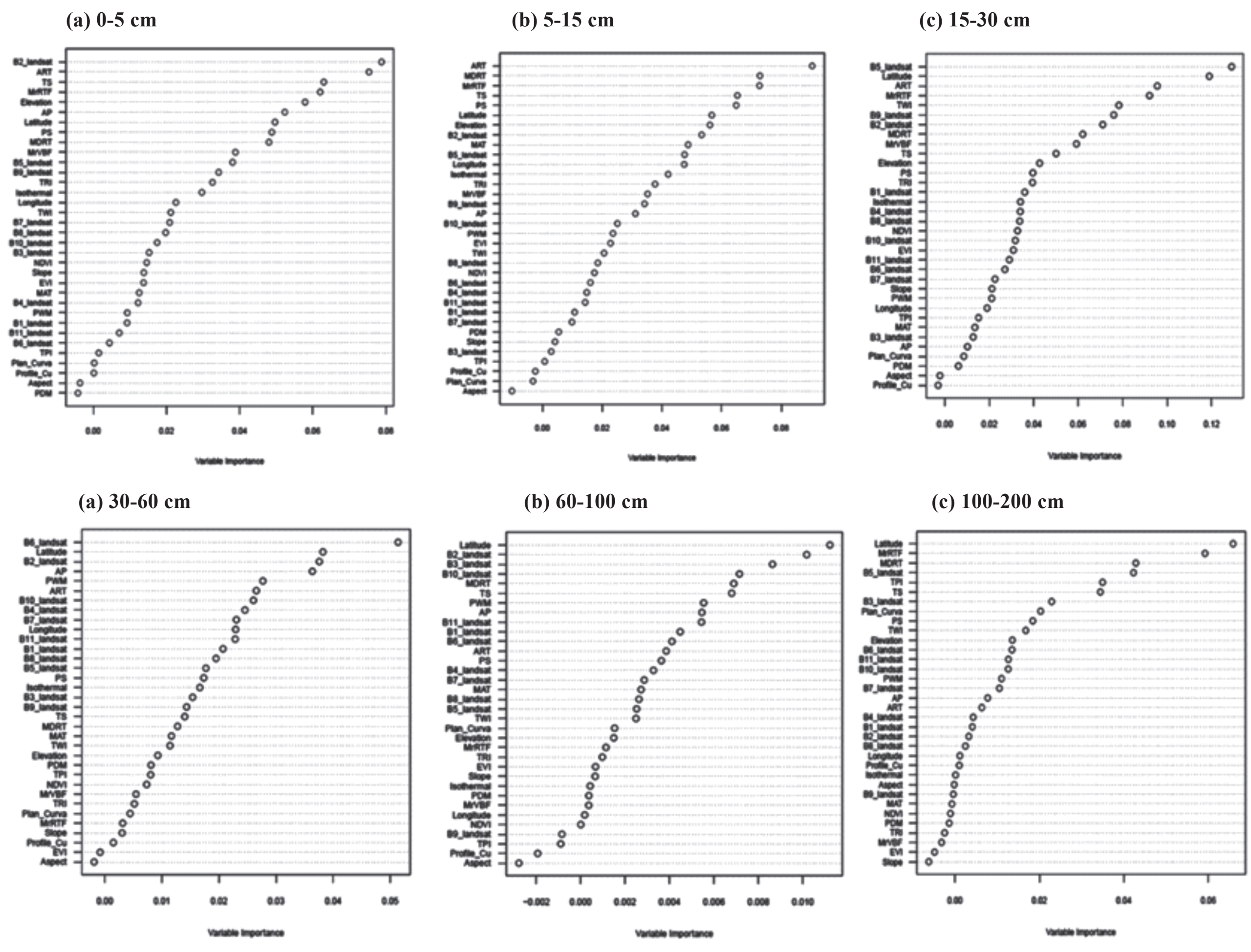

Fig. 3. Depth wise importance of variables for prediction of SOC by quantile regression forest model in the study area of desert regions of India

Table 5. Prediction of soil organic carbon by quantile regression forest model in the study area of desert regions of India

Parameters	Soil organic carbon (g kg^{-1})					
	0-5 cm	5 -15 cm	15-30 cm	30-60 cm	60-100 cm	100-200 cm
Minimum	0.58	0.58	0.40	0.83	0.80	0.19
Maximum	4.88	2.91	4.83	1.50	1.16	3.55
Mean	1.29	1.29	1.24	1.07	1.00	0.84
SE	0.001	0.0009	0.001	0.0005	0.0002	0.0009
SD	0.35	0.31	0.36	0.16	0.07	0.3
CV (%)	27.3	23.92	28.8	15.11	6.63	36.19
Skewness	0.7	0.19	2.27	0.90	0.71	0.58
Kurtosis	4.42	-0.93	-0.56	-0.36	-0.32	-0.32

vegetation and less agricultural practices. Overall these results indicate that both temperature-related variables and imagery products are important for predicting SOC concentrations in sandy desert regions.

Prediction of soil organic carbon

Spatial predictions of SOC contents are presented in table 5 and fig. 4. The predicted SOC content varied from 0.58–4.88 g kg^{-1} in 0-5 cm, 0.58–2.91 g kg^{-1} in 5-15 cm and 0.40–4.83 g kg^{-1} in 15-30 cm, respectively. The mean predicted SOC were 1.29, 1.29, 1.24, 1.07, 1.00 and 0.84 at 0-5, 5-15, 15-30, 30-60, 60-100 and 100-200 cm depth, respectively. Predicted SOC had the lower CV with the depths and values ranging from 6.63% for 60–100 cm to 36.19% for 100–200 cm. The SOC skewed positively with skewness values to 0.19 to 2.27 for different depths. The predicted SOC maps in different depths are presented in fig. 4. The SOC concentrations were predicted highest in the south part of the study area and lower in the north-east, east and south-east parts of the study area where sand dunes, eroded soils and wind-eroded sediments are located. The SOC concentration was also low in the north of study area where the sand dunes and sand deposits have high percentages of light textured soils with low organic matter. Eroded soils have lost topsoil where most of the SOC is presumed to exist. Fig. 5 shows maps of the uncertainties in predictions of SOC contents for

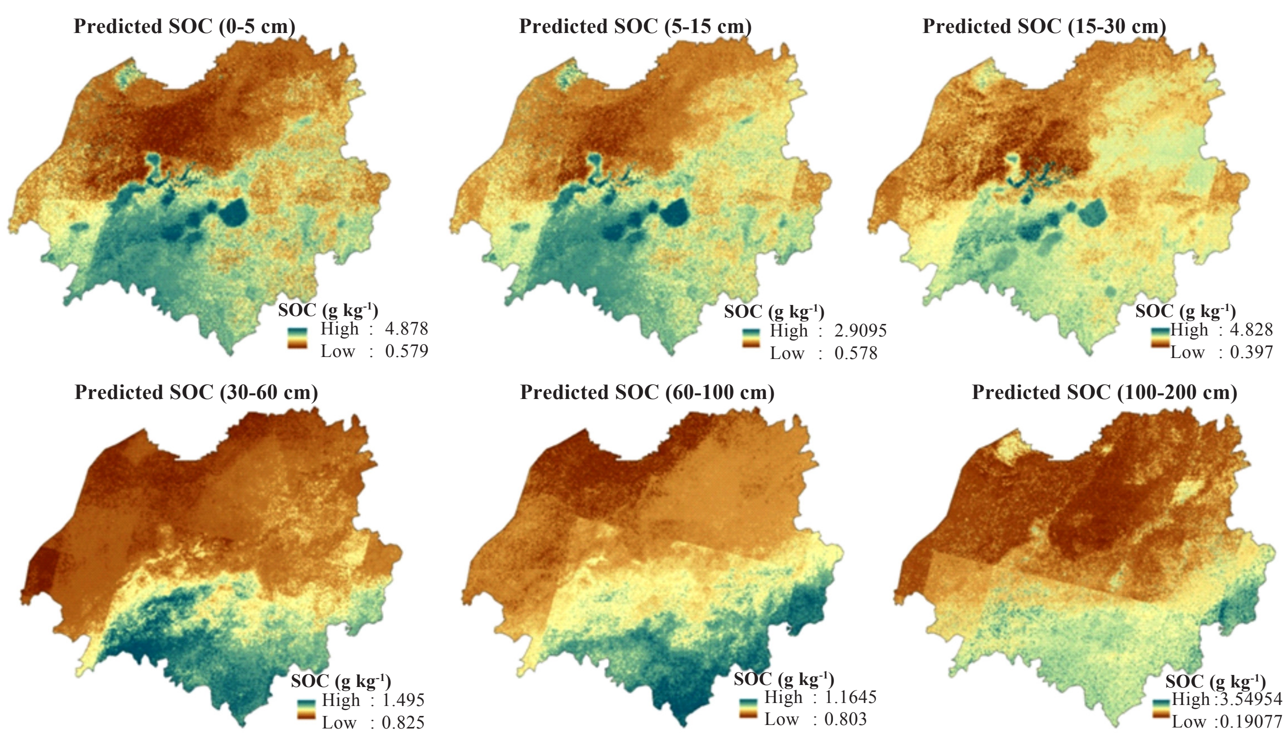

Fig. 4. Map of soil organic carbon spatial prediction by quantile regression forest model in the study area of desert regions of India

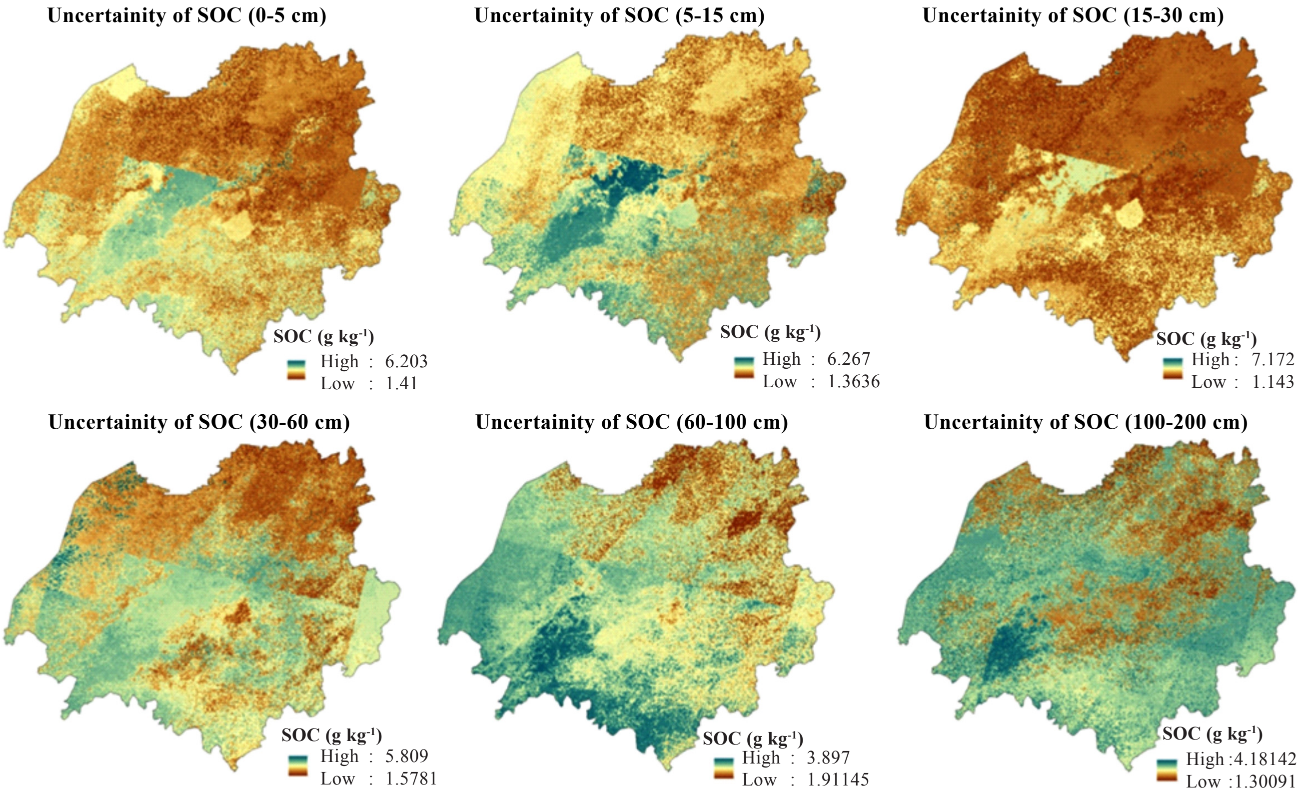

Fig. 5. Uncertainity map of soil organic carbon spatial prediction by quantile regression forest model in the study area of desert regions of India

each depth. The uncertainty was expressed as difference between lower and upper prediction limits at a 90% confidence interval. The wide range of uncertainty suggests that there is room to improve the current spatial predictions of SOC.

Conclusions

We developed baseline data for SOC map for desert regions of India using quantile regression forest (QRF) model. The prediction of SOC and uncertainty by QRF model was reasonable with R^2 of 0-7%. Higher sample density is required for better results in desert regions where soil and climatic pattern are complex in nature. The most important variables for predicting SOC concentration variations were the ART, latitude and landsat 8 bands 2, 5 and 6. Temperature-related variables and remote sensed data products are important for predicting SOC concentrations in arid regions. The vast range of uncertainty in the desert environment shows that spatial modelling should be improved. It's possible that the current QRF model's ability to capture low variability is attributable to the covariate selection. We anticipate that these predictions can be used to assist comprehensive planning for arid regions to increase SOC content. This raster based digital information of SOC will be useful for frequent monitoring and assessment of carbon cycle, carbon trading, soil health sustainability and environmental stability. Several countries have already produced maps according to the *GlobalSoilMap* specifications and the project is rejuvenating soil survey and mapping in many parts of the world. We believe that *GlobalSoilMap* constitutes the best available framework and methodology to address global issues about SOC mapping.

Acknowledgements

This work is part of IndianSoilGrid project supported by ICAR-National Bureau of Soil Survey and Land Use Planning. The authors are also thankful to laboratory staff, Regional Centre, ICAR-NBSS&LUP, Udaipur for their assistance during the analysis of soil samples.

References

Akpa, S.I.C., Odeh, I.O.A., Bishop, T.F.A., Hartemink, A.E. and Amapu, I.Y. (2016) Total soil organic carbon and carbon sequestration potential in Nigeria. *Geoderma* **271**, 202–215.

Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M.D.L., Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P.A., Thompson, J.A. and Zhang, G.L. (2014) Globalsoilmap: toward a fine-resolution global grid of soil properties. *Advances in Agronomy* **125**, 93–134.

Bishop, T.F.A., McBratney, A.B. and Laslett, G.M. (1999) Modeling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* **91**, 27–45.

Dharumarajan S., Hegde, R., Janani, N. and Singh. S.K. (2019) The need for digital soil mapping in India, *Geoderma Regional* **16**, e00204. https://doi.org/10.1016/j.geodrs.2019.e00204.

Dharumarajan, S., Kalaiselvi, B., Suputhra,A., Lalitha, M., Vasundhara, R., Anil Kumar, K.S., Nair, K.M., Hegde,R., Singh, S.K. and Lagacherie, P. (2021) Digital soil mapping of soil organic carbon stocks in Western Ghats, South India. *Geoderma Regional* **25**, e00387. https://doi.org/10.1016/j.geodrs.2021.e00387.

Dokuchaev, V.V. (1883) Russian Chernozems (Russkii Chernozems). Israel Program for Scientific Translations, Jerusalem, *Translation* from Russian by N. Kaner (1967). Available from. US Department of Commerce, Springfield, VA.

FAO and ITPS (2020) *Global Soil Organic Carbon Map V1.5*. Technical Report, Rome, FAO. http://www.fao.org/3/ca7597en/ca7597en.pdf.

Hengl, T., Heuvelink, G.B., Kempen, B., Leenaars, J.G., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., de Jesus, J.M., Tamene, L. and Tondoh, J.E. (2015) Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. *PLoS One* **10**, 1–26.

Jenny, H. (1941) *Factors of Soil Formation*. McGraw Hill, New York.

Kumar, M., Singh, S.K. and Sharma, B.K. (2009) Characterization, classification and evaluation of soils of Churu district, Rajasthan. *Journal of the Indian Society of Soil Science* **57**, 253-261.

McBratney, A.B., Mendonca-Santos, L. and Minasny, B. (2003) On digital soil mapping. *Geoderma* **117**, 3–52.

Meinshausen, N. (2006) Quantile regression forests. *Journal of Machine Learning Research* **7**, 983–999.

Minasny, B. and Hartemink, A.E. (2011) Predicting soil properties in the tropics. *Earth-Science Reviews* **106**, 52–62.

Minasny, B., McBratney, A.B., Malone, B.P. and Wheeler, I. (2013) Digital mapping of soil carbon. *Advances in Agronomy* **118**, 1–47.

Moharana, P.C., Naitam, R.K., Verma, T.P., Meena, R.L., Kumar, S., Tailor, B.L., Singh, R.S., Singh, S.K. and Samal S.K. (2017) Effect of long term cropping systems on soil organic carbon pools and soil quality in western plain of hot arid India. *Archives of Agronomy and Soil Science* **63**, 1661–1675.

R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sharma, R.P. and Singh, R.S. (2015) Carbon stock and its management in soils of Bhilwara ddistrict, Rajasthan. *Journal of the Indian Society of Soil Science* **63**, 304-309.

Singh, S.K., Kumar, M. and Sharma, B.K. (2009) Changes in soil properties in hot arid region of India. *Journal of the Indian Society of Soil Science* **57**, 24-30.

Singh, S.K., Singh, R.S., Shyampura, R.L. and Narain Pratap (2005) Organic and inorganic carbon stock in soils of Rajasthan. *Journal of the Indian Society of Soil Science* **53**, 281-287.

Vaysse, K. and Lagacherie, P. (2017) Using quantile regression forest to estimate uncertainty of digital soil mapping products. *Geoderma* **291**, 55–64.

Walkley, A. and Black, I.A. (1934) An examination of Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. *Soil Science* **37**, 29-37.

Wilding, L.P. (1985) Spatial variability: its documentation, accommodation, and implication to soil surveys. In *Soil Spatial Variability* (D.R. Nielsen and J. Bouma, Eds.), Pudoc, Wageningen pp. 166-194.

Received 2 June 2021; Accepted 4 March 2022

View publication stats