Digital Soil Mapping: concepts, methods, and applications –Remote sensing and GIS perspectives

Chapter · December 2023

DOI: 10.1016/B978-0-443-18773-5.00018-1

Bhabani Prasad Mondal
Bihar Agriculture University

Rabi Narayan Sahoo
Indian Agricultural Research Institute

Bappa Das
ICAR-Central Coastal Agricultural Research Institute

Debashis Chakraborty
Indian Agricultural Research Institute

Chapter 13

Digital Soil Mapping: concepts, methods, and applications - Remote sensing and GIS perspectives

Bhabani Prasad Mondal[1,2,5], Rabi Narayan Sahoo[2], Bappa Das[3], Debashis Chakraborty[2], N. Kumar[4] and G Bhupal Raj[1]

[1]School of Agriculture, SR University, Warangal, Telangana, India, [2]Division of Agricultural Physics, ICAR-Indian Agricultural Research Institute, New Delhi, India, [3]ICAR-Central Coastal Agricultural Research Institute, Old Goa, Goa, India, [4]ICAR-National Bureau of Soil Survey and Land Use Planning, Nagpur, Maharashtra, India, [5]Department of Agriculture, Brainware University, Kolkata, West Bengal, India

13.1 Introduction

Soil is considered the living skin of the earth, and it acts as a medium for plant growth and helps sustain animal and human activity on earth (Wu et al., 2019). A precise understanding of the spatial distribution of soil properties and their interrelationship is necessary for the sustainable management of soil resources. For such sustainable management, accurate and detailed soil mapping is an urgent need because detailed soil mapping is helpful for land use planning, sustainable land management, environmental protection, and sustainable agriculture (Meersmans et al., 2008; Forkuor et al., 2017). Unfortunately, the traditional soil mapping techniques are laborious, time-consuming, expensive, coarser in resolution, and cannot provide the real-time or ongoing situation of land resources (McBratney et al., 2003). Thus, these polygon-based soil maps are unsuitable for site-specific management of nutrients and other resources (Castaldi et al., 2016). Therefore, a new robust method is required for detailed and accurate mapping of soil properties considering the spatial variation of these attributes for precise management of land resources. Recent progress in remote sensing (RS)-based techniques (proximal, airborne and satellite-based RS techniques) has proved their potential in predictive and quantitative mapping of various soil properties with a higher resolution and better accuracies (Ma et al., 2019). In this recent decade, digital soil mapping (DSM) has become popular in accessing and mapping world soil resources (Kidd et al., 2020). Many published scientific methodologies and case studies currently describe how and why it is employed, the techniques used, their benefits and drawbacks in operationalization, successes, and comparisons to traditional soil mapping. The last ten years have seen DSM transition from academic-based research to operational mapping of soil physical, chemical, and biological properties and soil types at various spatial scales. Nowadays, federal, state, and territorial agencies of countries like Australia and the USA employ DSM as an acknowledged and used methodology to map soil and provide information for various land resource assessment operations.

India is known as the seventh largest country in the world. It possesses 328 million hectares of geographical area, 197 million hectares of grossed cropped area, and 141 million hectares of net sown area (Government of India, 2020). India currently has the second-largest population in the world (1.37 billion people). Such immense population pressure on land resources causes severe land degradation and desertification, gradually decreasing the area of crop-growing agricultural land and increasing food demand for a projected population of 1.65 billion by 2050 (Vision, 2015). Some of India's most severe soil-related problems are desertification, low soil quality, multi-nutrient deficiencies, and stagnated crop yields. Such fatal conditions warrant a rapid assessment of soil resources for proper management (Singh and Chatterji, 2018). However, quick soil assessment and management is somewhat challenging due to significantly varied soil types nationwide, climatic conditions, geology, and topography.

Additionally, the main challenge in Indian Agriculture is small and marginal land holdings having land size of less than 2 hectares (Agriculture Census Division, 2019). Hence, the existing small-scale conventional maps covering large

Remote Sensing of Soils. DOI: https://doi.org/10.1016/B978-0-443-18773-5.00018-1

spatial extent are insufficient for managing such small land holdings. Given these facts, India would greatly benefit from adopting DSM approaches for creating fine-resolution digital soil maps for better soil management.

13.2 Evolution of digital soil mapping

Mapping of soil resources is important, providing valuable information regarding formation, spatial distribution of soil properties, and types of soil present for implementation of sustainable land use management (Pereira et al., 2017). From the historical perspective, during the middle of the 18th century, the earliest soil maps were generated to identify the homogeneous areas of similar soil attributes for suitable land management (Minasny and McBratney, 2016). Later, agro-geologists used topographic maps to portray the maps of various soil types (Miller and Schaetzl, 2014). During the 19th century, Russian concepts stressed the genetic system, while the US system stressed on the intrinsic soil properties for soil mapping. The conceptual framework or equation consisting of five factors of soil formation [i.e., climate (cl), organism (o), relief (r), parent material (p), and time (t)], mentioned by Hans Jenny (1941) made the scientific foundation of traditional soil mapping. Dokuchaiev, the Father of soil science, produced a human map of Russia based on point observation (Hartemink and Minasny, 2014). Thus, soil property mapping has been traditionally continued from earlier times. However, those conventional soil maps are qualitative, and such traditional mapping generally depends on the soil surveyors' conceptual soil-landscape model or mental model (Hudson, 1992). They have utilized their past work experience on similar types of landscapes, various landscape features, aerial photographs, digital elevation models (DEM), etc., to generate soil-type maps and databases containing descriptions of the soil profiles and the general interpretation of soil map units (Beaudette and O'Geen, 2009). The concepts of conventional soil mapping have become modernized by applying quantitative techniques in the past decade (Boettinger et al., 2010). Proximal soil sensing is one of the quantitative techniques developed in the complete phase in the early 2000s, and it needs spatial and temporal data of high resolution; thus, it may become an essential component of high-resolution soil mapping (McBratney et al., 2011). Such modern soil mapping with quantitative techniques is called DSM (McBratney et al., 2003). Various factors like the availability of digital spatial data from DEM or satellite imagery or aerial photography, high computing or supercomputing and significant processing efficiency, advancement in geographic information system (GIS) and data mining techniques (machine and deep-learning algorithms), the attraction of new spatial analysis tools by the new generation soil scientists, etc., produce a good influence for development and advancement of DSM.

In this study, we have provided a table (Table 13.1), showing the historical development of DSM (Source: Minasny and McBratney, 2016).

The evolution of soil mapping and monitoring systems in other countries like France has occurred since the 1960s. The 1960s and 1980s were defined from a soil science and pedometric perspective by so-called "conventional mapping." The first comprehensive guide, that is, "Bases et techniques d'une cartographie des sols" (meaning: Fundamentals and techniques of soil mapping), debuted at the end of the 1960s and was written by Marcel Jamagne (Jamagne, 1967). Scientists of France have developed the Commission on Soil Science and Soil Classification's collaborative effort, started traditional soil mapping in other countries predominantly African countries, and generated many databases, which were later incorporated into worldwide databases. Therefore, indirectly, the French activities during this period contributed to continental assessments of soil capability and condition, assisting in the realization of continental and worldwide DSM. (Leenaars et al., 2018). After a foundational work on DSM was published in the 2000s, that is, by McBratney et al. (2003), France hosted the First Global Workshop on DSM in Montpellier in 2004. Since then, France has assumed a more significant role in global DSM-related projects (Arrouays et al., 2020). Apart from France, the European Union also generated global soil maps earlier, like the European Soils Bureau, a Joint Research Center that created the continent's first geographical database and standardized soil maps (Arrouays et al., 2021).

13.3 Principles and concepts of DSM

The abovementioned Jenny's equation is useful for qualitative soil mapping but is unsuitable for quantitative mapping of soil classes or properties. Therefore, in the recent decade, (McBratney et al., 2003) proposed the *SCORPAN* model for expressing the quantitative relationship between soil and environmental covariates. The mathematical expression of this model is:

$$S_{c,a} = f(\mathrm{s, c, o, r, p, a, n})$$

where $S_{c,a}$ refers to soil class (S_c) or attribute (S_a) at a specific point of space in a specified time and is an empirical or quantitative function of seven environmental factors or covariates such as soil (s), climate (c), organism (o), relief (r),

TABLE 13.1 Historical development of digital soil mapping.

Period (decade)	Major developments	References
1800s	P = f (K, O, G, B)	Dokuchaev (1899)
1900s	S = f (cl, o, r, p, t)	Jenny (1941)
	S = F (addition, removal, translocation, transformation)	Simonson (1959)
1960s	Topofunctions	Walker et al. (1968)
	Integrated climate, organism, relief, parent material, time (CLORPT) model	Jenny et al. (1968)
1970s	Remote sensing application on soil	Cipra (1973)
	Soil information system (SIS)	Sadovski and Bie (1978)
	Digital soil cartography	Webster et al. (1979)
1980s	Application of geostatistics in soil	Burgess and Webster (1980)
	Application of GIS in soil	Burrough (1986)
1990s	Use of pedometrics in soil science	Odeh et al. (1992)
	Use of pedotransfer functions in soil	Wösten et al. (1995)
	Digital terrain modeling	Gallant and Wilson (1996)
2000s	Implication of data mining tools in soil	Bui and Moran (2003)
	Digital soil mapping (DSM)	McBratney et al. (2003)
	Advances in proximal soil sensing	Rossel et al. (2010)
	Digital soil morphometrics	Hartemink and Minasny (2014)

topography (t), age (a) and spatial location (n). This SCORPAN equation or model facilitates predicting any soil class or property in a spatial context by quantifying the empirical relationship between that soil class or attribute and environmental attributes. It also enables us to estimate the prediction error or uncertainties associated with the prediction model of DSM. Therefore, DSM refers to creating geographically referenced soil databases or soil maps based on quantitative relationships between spatially explicit environmental covariates and the observation and measurements made at the field level and laboratory (McBratney et al., 2003). It is also known as "predictive soil mapping," "computer-assisted soil cartography," etc. DSM consists of three main components viz., the *input*, *process* and *output*.

- The *input* of DSM generally refers to the legacy soil data or map, soil data obtained from the field laboratory-based measurements, etc.
- The *process* denotes the construction of statistical modeling for correlating the soil measurements with environmental variables, that is, spatial or nonspatial soil inference system.
- The *output* represents the spatial soil information system, illustrating the spatial distribution of soil classes or properties in the raster and prediction uncertainties.

The DSM encompasses several stages and processes, viz.

- Defining a project area
- Identifying the associated physical features of interest in the study area
- Collection of data from multiple sources and data exploration after preprocessing
- Prediction model development
- Uncertainty estimation
- Application.

A digital soil map is a graphic representation of raster data made up of pixels, and each pixel contains a specific geographic location and soil-based information. These digital maps can refine or update the existing soil survey map (Carré et al., 2007). However, digital soil maps are different from the digitized conventional soil maps because in traditional mapping, the output product is digital, but the preparation of such conventional digitized map is not related to any statistical inference system (Minasny and McBratney, 2016).

13.4 Methods of DSM

The methods of DSM include the processes of building mathematical or statistical models using several statistical algorithms, hence known as predictive mapping or modeling methods. Predictive modeling is making statistical or mathematical models that help approximate the proper relationship between soil properties or classes with environmental covariates for prediction. Predictive modeling techniques may be classified into two major categories: classification and regression. Classification methods are generally used for predicting soil classes, while the regression techniques are applied for predicting soil properties. Various statistical algorithms (predictive modeling approaches) are utilized to establish the quantitative relationship between predictor and target variables. Different machine learning (ML) based algorithms have been applied to build up predictive models for DSM using advanced programming languages like R, MATLAB®, Python, etc. A simple regression model predicts one target or dependent variable depending on the relationship between several independent or predictor variables (Zeraatpisheh et al., 2020). A linear regression model fits a linear equation in the dataset, whereas tree-based regression models are used where a nonlinear relationship exists among the variables. The multiple linear regression (MLR) model is another good option to model and predict any variable, and it has been applied to predict soil texture and other soil properties like soil organic carbon (SOC) and calcium carbonate equivalent (CCE) by several researchers (Mitran et al., 2019; Zeraatpisheh et al., 2019). Regression kriging is one of the popular methods of DSM, which refers to the hybrid model developed by combining any regression model with residuals of kriging. Kriging is also a good geostatistical approach used to predict and map various soil properties by constructing semivariogram to consider the spatial variation of those properties.

In recent years, tree-based classification models using ML algorithms have gained much more popularity than geostatistical techniques (Heung et al., 2016). The widely used tree-based model is the classification and regression tree (CART), which constructs a single tree from the given input variable to make a prediction. The advanced form of CART is the random forest (RF) model, which builds several decision trees called an ensemble of trees (forest) from the predictor variables and produces a good final prediction (Heung et al., 2016). RF is a very robust ML model capable of handling nonlinear, noisy, and large data and simultaneously providing fast computation efficiency. Another extension of the RF model is quantile random forest (QRF), which provides nonparametric estimates of the median predicted value and prediction quantiles (Dharumarajan et al., 2019). The most common hybrid approach combines RF regression and residual kriging methods to predict several soil properties through DSM (Hinge et al., 2018). Artificial neural network (ANN) is another good ML technique to handle nonlinear and complex datasets for DSM, and this neural technique consists of input layers (input variables), hidden layers, and output layers (target or output variable) (Khaledian and Miller, 2020). In recent years, the improved version of neural techniques, deep-learning (DL) techniques like convolutional neural network, has also been utilized to some extent in DSM. The best performance of predictive modeling depends on selecting the appropriate sampling techniques and adequate knowledge of the soil system (Kuhn and Johnson, 2013). The predictive model's prediction accuracy, the difference between predicted and measured value, needs to be appropriately assessed (Brus et al., 2011). Producer's accuracy, user accuracy, and overall accuracy are generally measured for estimating the prediction accuracy of soil class. Various statistical indices like the coefficient of determination (R^2), root mean square error (RMSE), concordance correlation coefficient (CCC), etc., are generally employed for measuring the accuracy of soil property prediction.

Hence, from the above discussion, it was evident that several researchers have explored models ranging from simple regression models to advanced DL-based models in DSM for quantifying the quantitative relationship between soil properties and environment covariates. Those models are MLR (Triantafilis et al., 2009), ordinary kriging (Bishop and McBratney, 2001), kriging with eternal drift (Santra et al., 2017), regression kriging (Hengl et al., 2004), CART (Breiman et al., 1984), k-nearest neighbor model (Taghizadeh-Mehrjardi et al., 2015), multinomial logistic regression (Kempen et al., 2012), SVM regression (Priori et al., 2014), and RF model (Wiesmeier et al., 2011; Dharumarajan et al., 2017; Sreenivas et al., 2016), ANN (Mosleh et al., 2016). Zeraatpisheh et al., 2019 used several machine learning-based models like cubist, RF, Regression Tree, and MLR for spatial prediction and mapping of various soil properties like SOC, CCE, and clay content from the environmental covariates derived from DEM and LANDSAT-ETM imagery. They reported that cubist and RF showed best prediction performances in predicting CCE ($R^2 = 0.30$, RMSE = 9.52) and clay content ($R^2 = 0.15$, RMSE = 7.86), while both cubist and RF performed better for accurate prediction of SOC. Digital mapping of SOC content was done by applying several statistical techniques or models like MLR (Meersmans et al., 2008), PCR and PLSR (Amare et al., 2013), kriging (Worsham et al., 2010), and regression kriging (Yigini and Panagos, 2016). The decision tree algorithm was utilized to map various soil properties in Australia (Henderson et al., 2005). In India, Santra et al. (2017) reported that the validation R^2 ranged from 0.21 to 0.30 for predicting the percent sand content using RF model in the arid parts of India, particularly for Rajasthan and parts of Gujarat states. Another study showed that the validation R^2 values varied from 0.11 to 0.44, 0.16 to 0.24, 0.35 to 0.40, and 0.66 to 0.73 for predicting SOC percent sand content, percent clay content and soil pH, respectively, for different soil layers of India using national scale digital soil maps (Reddy et al., 2021).

13.5 Uncertainty estimation in DSM

In DSM, uncertainty analysis is essential for determining whether the forecasted soil map is trustworthy enough for soil management systems or decision-making processes. Uncertainty associated with the prediction model needs to be adequately addressed. Prediction uncertainties related to DSM generally arise from some sources like positional inaccuracy, vertical anticipation of DEM, and manual error in soil property estimation. There are mainly two types of uncertainties: "aleatoric" and "epistemic". Aleatoric delays are associated with the model error, whereas epistemic uncertainties are associated with the model parameters' uncertainties (Wadoux et al., 2020). Quantile regression techniques are mostly used to quantify aleatoric uncertainty, whereas bootstrapping methods are utilized to quantify epistemic uncertainty. With techniques like the QRF, the delta or Bayesian approaches, and the mean plus variance estimate for neural network algorithms, both aleatoric and epistemic uncertainty can be quantified to offer prediction intervals (PI) (Wadoux et al., 2020). Confidence interval (CI) is usually calculated for determining the prediction uncertainty of soil classes, while PI is generally computed for measuring the prediction uncertainty of soil property. PI is generally represented by three companion maps such as lower prediction limit, mean, and upper prediction limit maps of predicted soil properties (Malone et al., 2011). Malone et al. (2011) proposed a methodological framework for quantifying uncertainties of several soil attributes like SOC, available water content, etc. Knowledge on spatial variability of soil property is required for calculating the PI. Generally, the triangulation method is applied in soil science for representing the spatial variation of soil properties based on fuzzy membership function (Odgers et al., 2015). Odgers et al. (2015) generated digital maps of soil pH, calculated the uncertainties at 90% PI, and finally showed the increment of prediction uncertainties with increasing soil depth. Previously, most of the studies ignored the analysis of uncertainties. Recent studies incorporating uncertainty assessments such as Vaysse and Lagacherie (2017) and Wadoux (2019) have both been successful in reporting PIs for RF and neural network models, respectively. CI are described in various publications and created using bootstrapped samples of the original data to train several disjoint models (Gomes et al., 2019). In Denmark, Adhikari et al. (2014) generated uncertainty maps to demonstrate the accuracy of the tree model's predictions of the study area's soil groups or categories. The confidence levels were between 0.2 and 1, with 1 being the most confident while 0.2 is the least confident. Podzols were predicted with the most confidence, with a mean value of roughly 0.72, while Podzoluvisols and Alfisols were predicted with the lowest confidence, according to pixel-by-pixel analysis of the predicted maps for the entire research region (0.48). However, we are practically lagging in assessing the uncertainty associated with DSM. Typically, businesses and commercial software do not' disclose metrics of the uncertainty related to the maps.

13.6 Conventional soil mapping versus digital soil mapping

The primary difference between conventional soil mapping (CSM) and DSM is that CSM is based on a conceptual soil-landscape model, which is qualitative. In contrast, the DSM is based on a quantitative soil-landscape model (Kempen et al., 2012). Additionally, such qualitative mental models are inflexible for quantitative studies, rarely stated straightforwardly, suffer from personal bias, and are difficult to duplicate (Jafari et al., 2012). Traditional maps have many flaws, including inadequate spatial detail and concerns with the accuracy of soil properties (Hartemink et al., 2010). The main limitation of CSM includes its irreproducibility and lack of quantification of prediction uncertainties. However, the DSM is reproducible; the digital maps can be updated at any time with new data and information availability.

Along with this, DSM quantifies uncertainties associated with the prediction model of DSM (Lark and Lapworth, 2012). With a predicted uncertainty interval, DSM can create quantitative, 2.5-D and 3-D gridded maps of soil characteristics. Maps obtained from CSM are called general-purpose maps, while the maps generated from DSM are called specific-purpose maps. Instead of producing categorical estimates of soil property values as is customary in traditional mapping methods, DSM-generated grids can depict the gradational spatial variations in soil attributes, and uncertainty ranges can give a level of confidence in the estimated attribute values. However, there are some potential drawbacks in

DSM. DSM is not a standardized technique, while the CSM is a standardized technique (NRCS US, 1993). DSM cannot represent the complex soil forming processes, while CSM can do it easily. Accurate DSM is still now challenging with the availability of limited datasets, while such limited data availability cannot hamper the accuracy of CSM.

Here, we provide a few national or regional examples that provide a good comparison between CSM and DSM and how CSM can be utilized to evaluate the predictive performance of DSM. For example, Denmark, a nation with a wealth of knowledge, is a wonderful example of how to use digital tools. The existence of a conventional soil class map for Denmark, despite its construction at a very coarse cartographic scale, offers the chance to assess the accuracy of the DSM model's predictions (Adhikari et al., 2014). They make a good comparison between the existing CSM and newly predicted DSM for different soil groups. Most of the soil mapping units' (SMUs') calculated areas for the expected soil groups were comparable to the corresponding estimated areas those same soil groups occupied on the current map. For instance, SMU 7 had an estimated 9405 km^2 area coverage, of which Luvisols were thought to occupy 75%, Cambisols 15%, and Arenosols and Gleysols each 5%. The calculated area covered by Luvisols, Cambisols, Arenosols, and Gleysols for that SMU was 70%, 12%, 5%, and 4%, respectively, demonstrating an excellent match between the two maps, that is, CSM and DSM (Adhikari et al., 2014). Although the visual or qualitative comparison was done extensively between CSM and DSM products, quantitative comparison is still not done widely. Rossiter et al. (2021) utilized a set of quantitative indicators to compare conventional and digital soil maps, but those indicators were not related to the visual comparison of the products. In France, Lemercier et al. (2022) recently evaluated the accuracy of DSM maps utilizing local scale maps. They compared local soil maps with global, national, and regional scale digital maps of various soil attributes. 'Quantitative indicators like Lin's CCC and mean squared error skill score were used to check if the soil patterns displayed by the DSM products and those shown by the existing soil maps were consistent. Since there was good agreement with conventional evaluation using spatial sets of punctual soil property measurement, evaluating DSM products with soil maps using quantitative indicators representing the closeness of projected soil patterns was relevant.

13.7 Applications of DSM

DSM has numerous applications in agriculture, from land resource management to water resource management. Various developed countries adopted many ambitious projects on DSM; the United States Department of Agriculture-Natural Resources Conservation Services (USDA-NRCS) has launched a new program called "Soils2026" to prepare legacy soil data-based digital map of 30 m resolution for the country for developing a new inventory of soil and for providing ecological sites for all parts of the United States (Thompson et al., 2020). Odgers et al. (2015) have already developed a raster-based digital map of individual soil properties at a high-resolution scale (90 m) using a polygon-based map as a legacy map from the State Soil Geographic database for the United States. In Australia, DSM has reached an operational stage, and this approach is now widely used by different agencies for real-world applications such as land suitability mapping and classification, erosion potential assessment, and ecological modeling from field scale to continental scales (Kidd et al., 2020). The most prominent use of DSM is the predictive mapping of soil properties. Across the globe, several soil scientists employed the DSM technique to predict and map several soil properties like SOM (Byrne and Yang, 2016), soil pH (Pahlavan-Rad and Akbarimoghaddam, 2018), electrical conductivity (Ranjbar and Jalali, 2016), available phosphorus (Wilson et al., 2016), soil particle size distribution (Pahlavan-Rad and Akbarimoghaddam, 2018). Mponela et al. (2020) generated high-resolution (10 m) digital maps of SOC and three major nutrients such as N, P, and K at farm scale for their effective management. The United States, Australia, China, and France also utilize their national-scale digital soil maps for various practical purposes. In France, government bodies use digital maps to assess contamination levels, agricultural policy making, private sector selling their fertilizer products, and farmers and their associated organizations for formulating fertilizer recommendations and irrigation water management (Arrouays et al., 2014). The "*Global-Soil-Map*" program and its requirements, widely used to create digital soil maps, were a significant step in improving soil data, which is now readily accessible (Poggio et al., 2021). The "*Global-Soil-Map*" format complies with the requirements of most biophysical modeling programs and fulfills the demands of many end users for ready-to-use data (Voltz et al., 2020). Nowadays, many regions of the world are covered by DSM products with similar resolution and identical targeted soil attributes due to the development of DSM applications at various scales (Minasny and McBratney, 2016; Poggio et al., 2021). For example, for the USA, three DSM items were inventoried by Rossiter et al. (2021). Similarly, in France, DSM products, that is, "*Global-Soil-Map*" products, were developed for multiple scales such as from regional, national to global scales like regional "*Global-Soil-Map*" products (Vaysse and Lagacherie 2015) to national "*Global-Soil-Map*" products (Mulder et al., 2016) and globally Soil Grids version 2.0 (Poggio et al., 2021). In central France, using around 3,200 hand-feel soil texture data, Richer-de-Forges et al. (2022) evaluated four DSM products of soil textural class prediction created at various scales (global, continental, national, and subnational scales). Adhikari et al. (2014) generated a national-level soil class map of Denmark employing information gathered from point soil observations and environmental data as predictors from the FAO-UNESCO Revised Legend in Denmark. Thy utilized a boosted decision tree-based DSM model and showed that classification error of DSM could be reduced by boosting, the tree-based model, and simultaneously the overall prediction accuracies, especially for validation dataset, could be increased from 60% to 76% for similar soil groups. The SOC, texture, pH, bulk density, and soil depth maps for the Czech Republic were developed utilizing DSM techniques in conjunction with an extensive collection of historical and present-day soil samples (Žížala et al., 2022). They showed that the combination of mapped coordinates within a buffered distance and covariates based on a mosaic of bare soils using RS data proved useful for predictive mapping, especially when training datasets maintained adequate spatial coverage.

Although DSM is a potential tool for managing soil health, natural resources, and ecosystem services, very few studies have been conducted so far on DSM of various soil properties in India (Sreenivas et al., 2014). Santra et al. (2017) have used the legacy soil map of the National Bureau of Soil Survey and Land Use Planning (NBSS&LUP) for preparing the digital map of sand content of arid western India, employing a geostatistical approach. Reddy et al. (2021) have prepared the first national-scale digital soil map of four soil properties like SOC, pH, sand, and clay content using legacy soil data and quantifying complex relationships among soil properties DEM-derived and bioclimatic variables. They have mentioned that the elevation, topographic wetness index, temperature, and rainfall were the major environmental covariates influencing the spatial distributions of the abovementioned soil properties. Such national-scale digital soil maps are useful in delineating dominant soil types, suitable crop-growing areas, affected lands, and management zones for planning purposes (Dharumarajan et al., 2017). There are several other potential applications of DSM, like land capability classification, which can be helpful for policymakers in land use planning, crop suitability mapping, and identifying major limiting factors of soil (Dharumarajan et al., 2019).

Moreover, in order to identify prospective carbon sequestration zones, one can use the spatial distribution maps (digital maps) of carbon stocks, which are important for managing global carbon emissions, carbon trading, and mitigating climate change (Dharumarajan et al., 2021). DSM could be used to make assessments of land degradation, desertification, drought risk, and soil erosion for creating appropriate region-specific action strategies (Arrouays et al., 2014). DSM can also monitor the large portions of the nation's problematic soils, including acidic soils, soils affected by salts, waterlogged soils, and degraded lands (Mitran et al., 2021). Fine-scale digital maps of soil hydraulic properties such as field capacity, permanent wilting point, and hygroscopic coefficient can assist farmers in adopting a site-specific water management strategy to optimize water use for crop plants (Santra et al., 2021). Estimation of soil moisture is another crucial aspect of scheduling irrigation of crops, and this recent digital map of surface soil moisture (SSM) of agricultural research farm of Indian Agricultural Research Institute (IARI) for different dates has been prepared by our Indian scholars using optical-thermal-microwave RS synergies. (Das et al., 2022). Digital maps can potentially be utilized for soil fertility characterization, site-specific nutrient management, and fertilizer recommendations. The major practical implication of DSM in India is its potential application in assisting the Soil Health Card mission program for the fast preparation of soil health cards for many crop growers in India.

13.8 Case study: digital soil moisture mapping

Soil moisture (SM) is a crucial parameter for water management and the hydrological cycle on the earth's surface. In agriculture, it is a critical parameter for scheduling irrigation, monitoring drought, and predicting yield. So, the estimation of SM is highly essential for agricultural production purposes. However, accurate estimation of SM is difficult due to its higher spatial and temporal variability; therefore, it is not possible through traditional techniques. The currently available RS-based global SM products are coarser in resolution (3−15 km) and unsuitable for field-level moisture mapping and monitoring application. Recently, the advancement in RS techniques has made the dynamic assessment of SM possible for various practical applications.

13.8.1 Methodology

13.8.1.1 Study area

Considering the abovementioned points, the present study was conducted at the experimental research farm of IARI, New Delhi, to produce digital maps of SSM for the IARI farm at a resolution of 30 m using the synergies of optical, thermal, and microwave RS techniques. The study area is located at 28° 37″ N to 77° 16″ E. Soils of the study site show a light loamy soil texture. The area receives mainly monsoon rainfall (June to September months). Wheat is a dominant crop in the IARI farm. The other crops are maize, pulses, vegetables, etc. The SM retrieval has been done by direct fusion of all three RS data above with the help of ML algorithms.

13.8.1.2 RS data sources

The optical, thermal, and microwave RS data products were obtained from Landsat-8 (OLI data), Landsat-8 (TRIS data), and Sentinel-1 [synthetic Aperture Radar (SAR) data], respectively. All RS-based data acquisitions were done on the same date. Landsat-8, level-1 data product, was downloaded from the United States Geological Survey website using the Earth Explorer engine (https://earthexplorer.usgs.gov/). Then, the data product was atmospherically corrected following the Dark Object Subtraction method in R software version 4.0.1. (Leutner et al., 2019). Mono window algorithm was utilized to retrieve land surface temperature (LST) using band 10 of Landsat-8 (Qin et al., 2001). The SAR data of Sentinel-1 was obtained from Copernicus Open Access Hub (https://scihub.copernicus.eu/dhus/#/home). Then, the data product of Sentinel-1 was preprocessed using SNAP (Sentinel Application Platform) for further usage. The preprocessing steps involved noise removal, radiometric correction, range doppler correction, and speckle filtering (Filipponi, 2019). These RS data were used to compute various covariates like vegetation indices, LST, backscatter coefficient, etc., which are useful for the digital mapping of SSM. For example, Landsat-8 data were utilized to derive vegetation indices like Normalized Difference Vegetation Index, Normalized Difference Water Index with the help of "SpectralIndices" function of R package.

13.8.1.3 In-situ soil moisture measurement

The ML3 Theta-Probe (Delta-T Devices, UK) was utilized to measure the SSM (5 cm top soil layer). This thermal probe is a hand-held device consisting of four rods. It needs to be inserted into the soil for direct measurement of SSM using the principle of detecting transmitted electromagnetic waves under the influence of the permittivity of soil, which influences SM. The in-situ or field level moisture measurement was done in synchronization with the date of pass (October 2020 to March 2021) of Landsat-8 and Sentinel-1 satellites.

13.8.1.4 Machine learning approaches

Four different ML-based models such as RF, Support Vector Regression (SVR), Gaussian Processes Regression (GPR), and Extreme Gradient Boosting (XGB) have been applied for digital mapping of SSM. RF is a supervised ML technique consisting of several decision trees called "forest" and trained with "bagging" method. The tuning parameters for RF are number of trees ("ntree"), number of independent variables ("mtry") at each subset, minimum node size ("min.node.size"). The study used the average of all constructed trees. The study uses the "ranger" package of R software (R 4.0.1) to develop the RF model. SVR is another ML approach that constructs the hyperplane that separates all datapoints (support vectors) into separate classes. The "kernlab" package of R studio (R 4.0.1) was utilized to develop SVR model in this study. GPR is a nonparametric, kernel-based probabilistic model, that uses the Bayesian approach to regression. It calculates the probability distribution of overall admissible functions that fit the data. "kernlab" package of R was utilized to build GPR model for soil property prediction. XGB is an ensemble ML-based model that can produce very robust and accurate results by sequentially adding weak learners in each step (Friedman, JH, 2001). It is very much flexible and customizable, allowing the optimization of several hyperparameters like "learning rate" (eta), that is, the contribution of each added learner, "number of iterations" of model (n_round or n_estimator), "depth of decision tree" (max_depth), etc., which are very much useful to reduce the bias of the model (Das et al., 2022). The "gbm" (Greenwell et al., 2019) package was used to develop this model. The statistical indices like coefficient of determination (R^2), mean biased error (MBE), RMSE, and ratio of performance to inter-quartile (RPIQ) distance were used to evaluate the predictive performance of the studied ML-based models.

13.8.2 Results and discussion

13.8.2.1 Descriptive statistics

The descriptive statistical analysis (Table 13.2) showed that SM (v/v) in both the calibration and validation dataset had nearly equal means (17.42% and 18.66%), medians (15.65% and 17.15%), standard deviations (7.37% and 7.48%), and similar ranges (3.2%−39% and 5.6%−39.2%).

TABLE 13.2 Summary of descriptive statistical analysis.

Statistical parameters	Calibration dataset	Validation dataset
Mean	17.42	18.66
Median	15.65	17.15
Standard deviation	7.37	7.48
Sample variance	59.77	56.00
Minimum	3.20	5.60
Maximum	39.00	39.20
Skewness	0.55	0.66

The datasets were normally distributed with nonzero skewness. Such commonality between calibration and validation datasets proved that both were randomly selected, well separated, and well represented the whole dataset. Such datasets are beneficial for building models using both datasets.

13.8.2.2 Feature selection

It is an important step that identifies the important variables having significant contributions to prediction and removes the redundant variables. It also reduces the complexity of the model and ensures better predictability (Hong et al., 2018). Several algorithms like Variable Importance for Projections (VIP) and Boruta algorithms can be employed to identify the variable of importance (Das et al., 2022). The present study used VIP technique to find the relative importance of covariates for different four models used in the study (Fig. 13.1A–D).

Several important variables like Sigma0_VV, normalized burn ratio index (NBRI), modified normalized difference water index (MNDWI), etc., were identified. Among all variables utilized in the study, the backscatter coefficient with vertical-vertical polarization (Sigma0_VV) is the most crucial variable for SM prediction because it is related to dielectric constant of soil, which is a function of SM (Dobson and Ulaby, 1986). The study showed that Sigma0_VV is more important than Sigma0_VH for SSM mapping, which remains consistent with earlier findings (Amazirh et al., 2018; Kumar et al., 2019).

13.8.2.3 Evaluation of model performance

The hyperparameters were optimized by following tenfold cross-validation with five times repetition of each model, and the optimization revealed that the optimized values of "ntree," "mtry," and "min-node-size" were 500, 15, and 5, respectively. For the XGB model, the final values of the tuning parameters used for modeling were "n_round," that is, number of iteration or number of trees boosted, "max_depth" (maximum tree depth), "eta" (learning rate), "gamma" (minimum loss reduction to do a split), "colsample_bytree" (fractions of the column selected for each tree), min_child_weight and "subsample" were 50, 1, 0.3, 0, 0.6, 1 and 1 respectively. For SVR model, the optimized values of the parameters like "sigma" (controls the distance of influence of a training point) and "C" (adds a penalty for misclassified point) were 0.09763221 and 1.0, respectively. The final GPR model had a sigma value of 0.1341819.

The RF exhibited the highest prediction accuracy in terms of R^2 (highest) and RMSE (least) both for calibration ($R^2 = 0.93$, RMSE = 2.76) and validation ($R^2 = 0.65$, RMSE = 5.48) datasets (Table 13.3). Such good predictability of the RF model can be attributed to the bagging technique associated with the model, which can significantly improve the predictability by considering the minute variation within the datasets (Shirzadi et al., 2018). Earlier studies also confirmed the supremacy of RF in the predictive mapping of SM (Adab et al., 2020; Liu et al., 2021; Das et al., 2022). The scatter plots of measured versus predicted showed the graphical representation of performances of all models in validation dataset (Fig. 13.2). Next to RF, SVR performed better than any other models, especially in the validation dataset ($R^2 = 0.58$, RMSE = 5.59, RPIQ = 1.84). The lowest predictability was achieved from the XGB model in terms of R2 for calibration ($R^2 = 0.61$) and validation ($R^2 = 0.44$) dataset. Interestingly, no model showed the overestimation of moisture because all the calculated MBE values for both datasets were negative. It reflects that all the models split the datasets well and trained well to predict moisture, creating multivariate models at each point (Das et al., 2022).

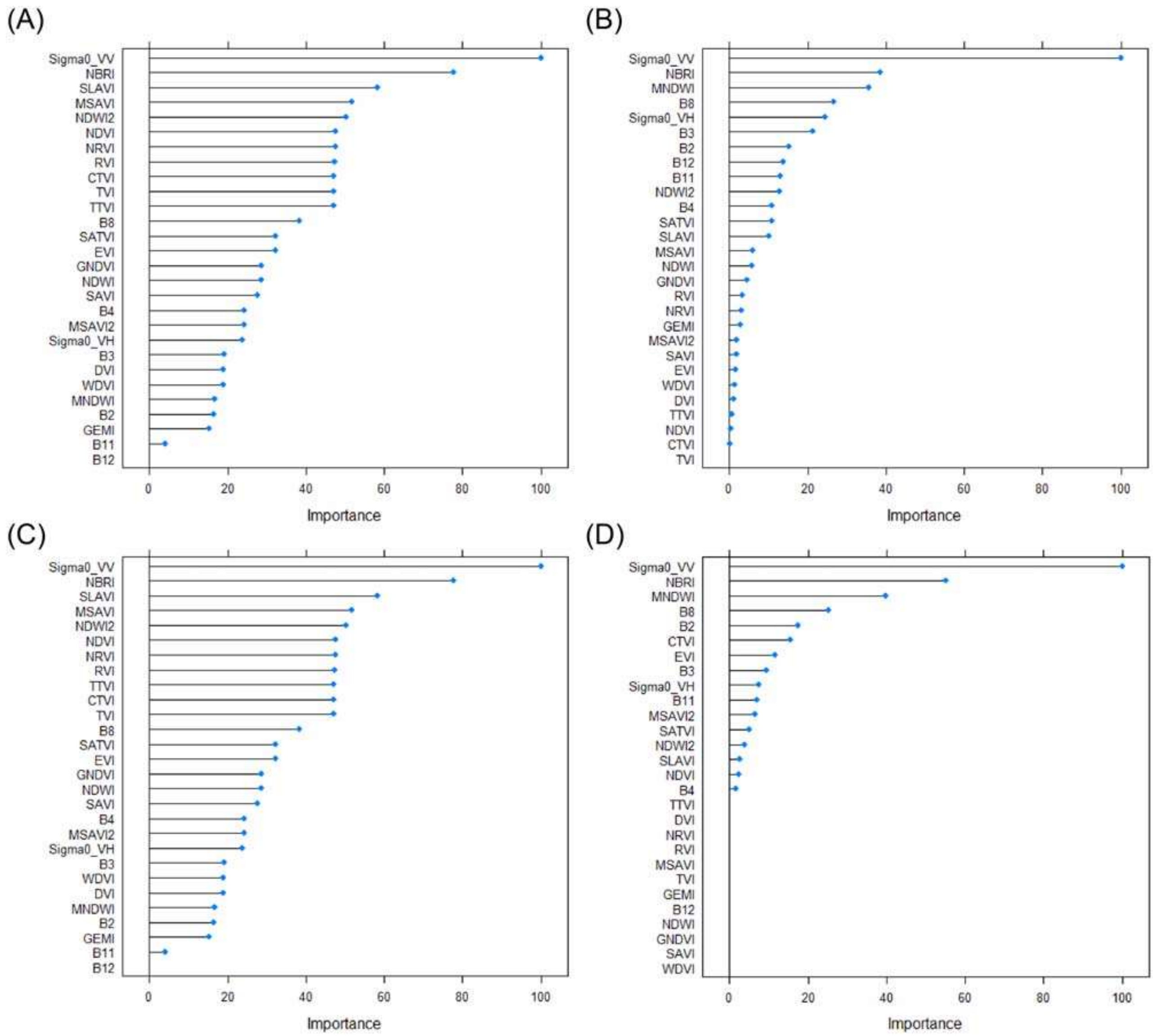

FIGURE 13.1 VIP plots (generated by RF, XGB, SVR and GPR models) showing the relative importance of variables for predictive mapping of Surface soil moisture.

TABLE 13.3 Comparing efficiency of four models for both calibration and validation datasets.

Models	Calibration				Validation			
	R^2	MBE	RMSE	RPIQ	R^2	MBE	RMSE	RPIQ
RF	0.93	−0.04	2.76	4.13	0.65	−0.54	5.48	1.88
XGB	0.61	−0.16	5.10	2.23	0.44	−0.75	6.04	1.71
SVR	0.63	−0.63	4.92	2.31	0.58	−1.28	5.59	1.84
GPR	0.69	−0.17	4.84	2.35	0.51	−1.09	6.10	1.69

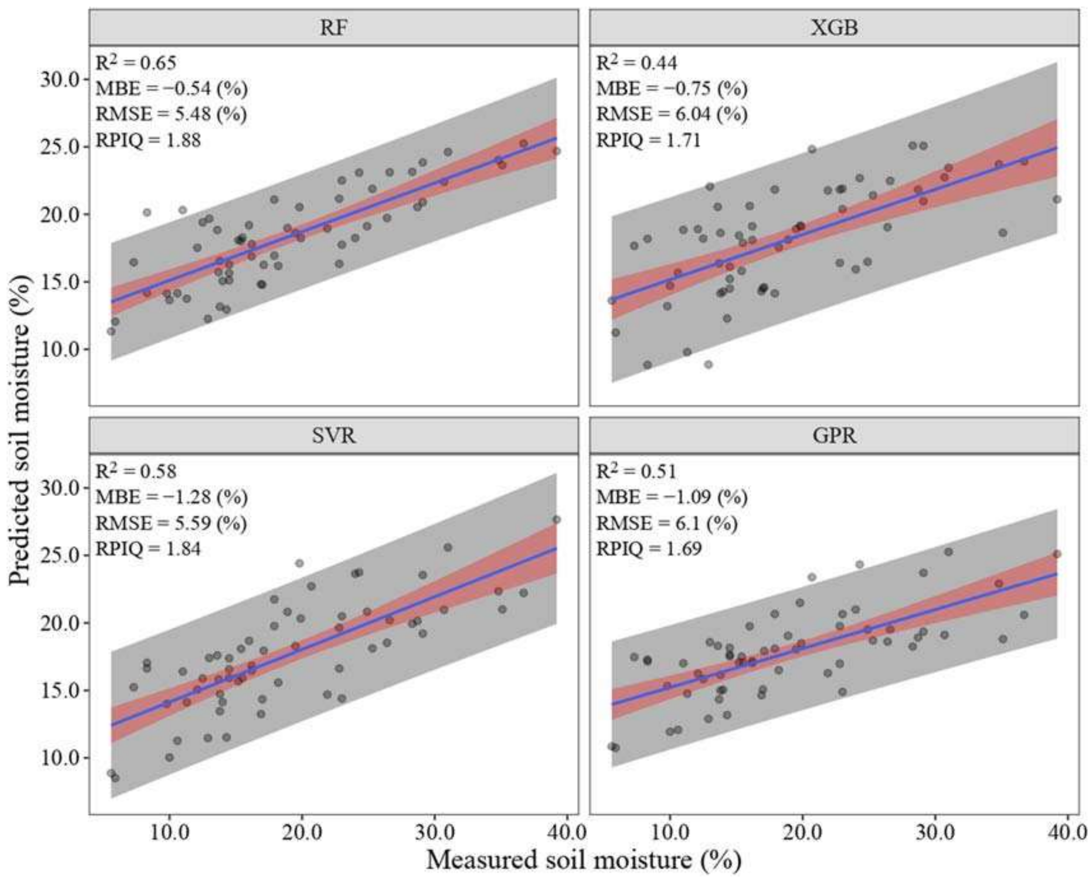

FIGURE 13.2 Scatter plots showing measured vs predicted soil moisture (%) in validation datasets for four different types of ML based models (RF, XGB, SVR and GPR models).

13.8.2.4 Spatial pattern of soil moisture

All ML models exhibited almost similar distribution of SSM in the study area (Fig. 13.3). The present study categorized the moisture classes of the study area into six categories, starting from “very dry” (less than 10% SM level) to “very wet” (more than 30% SM level). However, the earlier studies mentioned four categories of SSM such that “dry” (less than 8% SSM), “moderately wet” (8%−15% SSM), “wet” (15%−26% SSM), and “very wet” (more than 26% SSM) based on moisture content at field capacity and permanent wilting point (Santra et al., 2008; Das et al., 2022).

A major portion of the study area exhibited moderate SSM content in the range of 15%−25% (v/v), and it is mainly due to the applied irrigation to the wheat crop, the dominant crop in the study area, requiring 4−5 irrigations throughout the growth period (Das et al., 2022). As the study site is located under semiarid climatic conditions, irrigation is the main water source for crop cultivation. Such irrigation water and mild rainfall during rabi seasons, especially in January, keep the soil surface wet. As the soil has a light texture, some portions have less water-holding capacity. However, moisture uptake by high-yielding cultivars of the IARI farm is another reason for the fast depletion of SSM.

13.8.2.5 Uncertainty assessment in soil moisture mapping

The spatial distribution of uncertainty is depicted in the form of standard error (SE) maps, which are represented in Fig. 13.4.

The associated SE maps of SSM content were generated using four different ML models (GPR, RF, SVM, and XGB). The generated four SE maps of SSM content showed different trends for four different models. The SE maps of SSM measure moisture variability surrounding a coefficient of estimate. The highest variability of SSM in terms of SE was recorded for the SVM model in the range of 16%−18% of SE. The lowest SE (less than 14%), that is, less error in the prediction of SSM was observed in the case of RF model. GPR and XGB models also produced similar types of uncertainties in the prediction of SSM. It is very difficult to explain the performance of various ML models in terms of

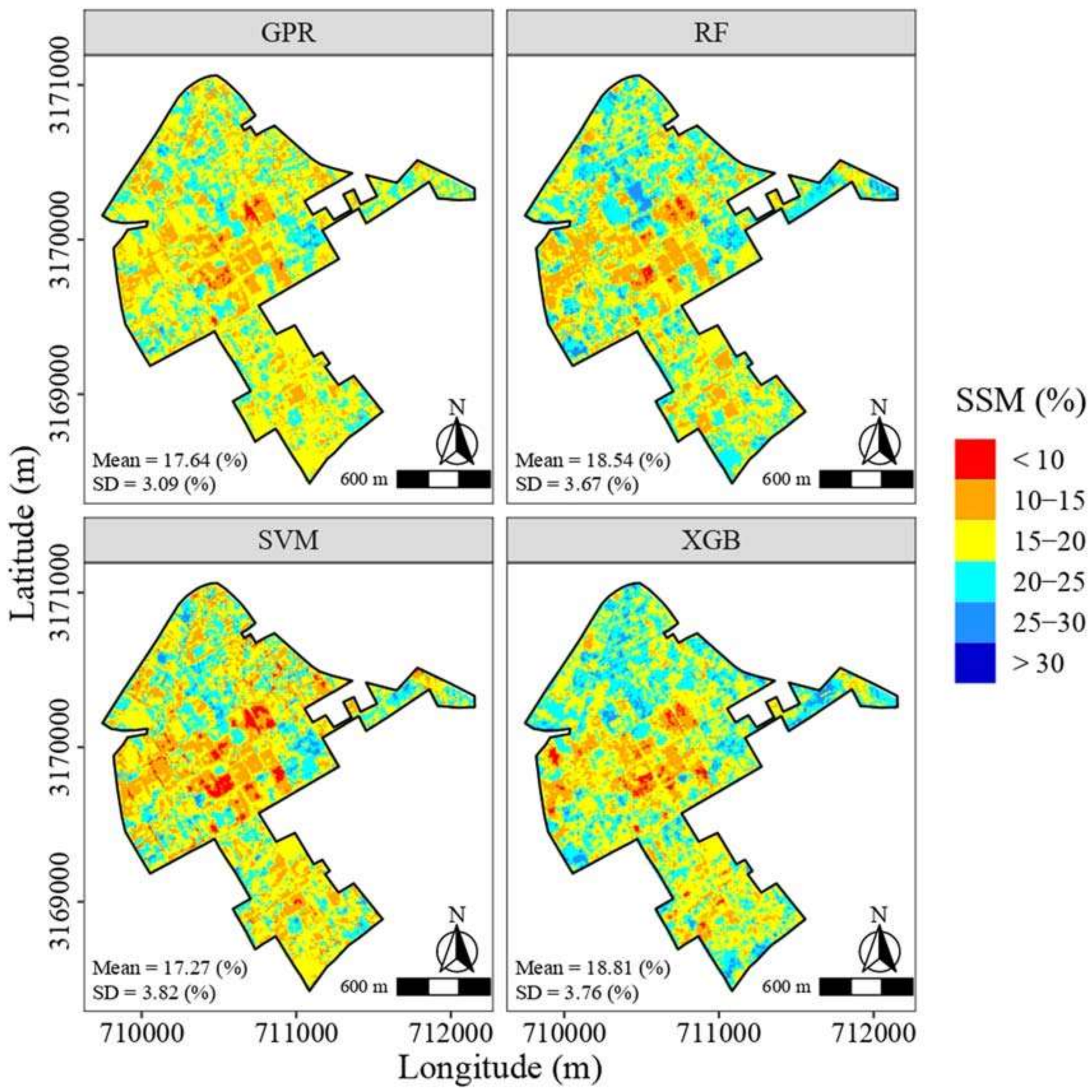

FIGURE 13.3 Spatial distribution of surface soil moisture in the study area prepared by using four ML models (GPR, RF, SVM and XGB).

accuracy and uncertainty assessment because ML-based models are empirical in nature and have region-specific applicability as compared to physical models having universal applicability requires detailed field-based observations like SM data, land surface parameters like surface roughness, vegetation indices, and other related parameters (Das et al., 2022). Another problem of ML-based models is that they cannot predict the extreme observed values of any parameter like SM. Generally, the ML-based models under estimate the very high values of observations and overestimate the very low values of observation of a parameter. Overfitting of the SM data by ML models was reported by several researchers (Bonfatti et al., 2016; Chatterjee et al., 2020). In the present study, most of the observation values were under-estimated, which was confirmed by negative bias, that is, negative MBE values. Overall, it was observed that the predicted error in the estimation of SM was comparatively higher for all ML-based models. Such error could be attributed to the high spatiotemporal variability of SM. It is well known that SM, a highly dynamic soil property, varies spatially and temporally. Moreover, there is a mismatch between the time when field observation was taken for SM in an area, and the time the satellite passes over the same area because the study area is irrigation. For example, a field-based reading of SM was taken during the morning time when the field was dry, but the satellite passed over the same area during the evening time when irrigation was given to that area; the satellite image was captured as a wet area, which is a serious mismatch for capturing the moisture data. Such mismatch in recording observation results in higher error or inaccuracies in predicting soil moisture.

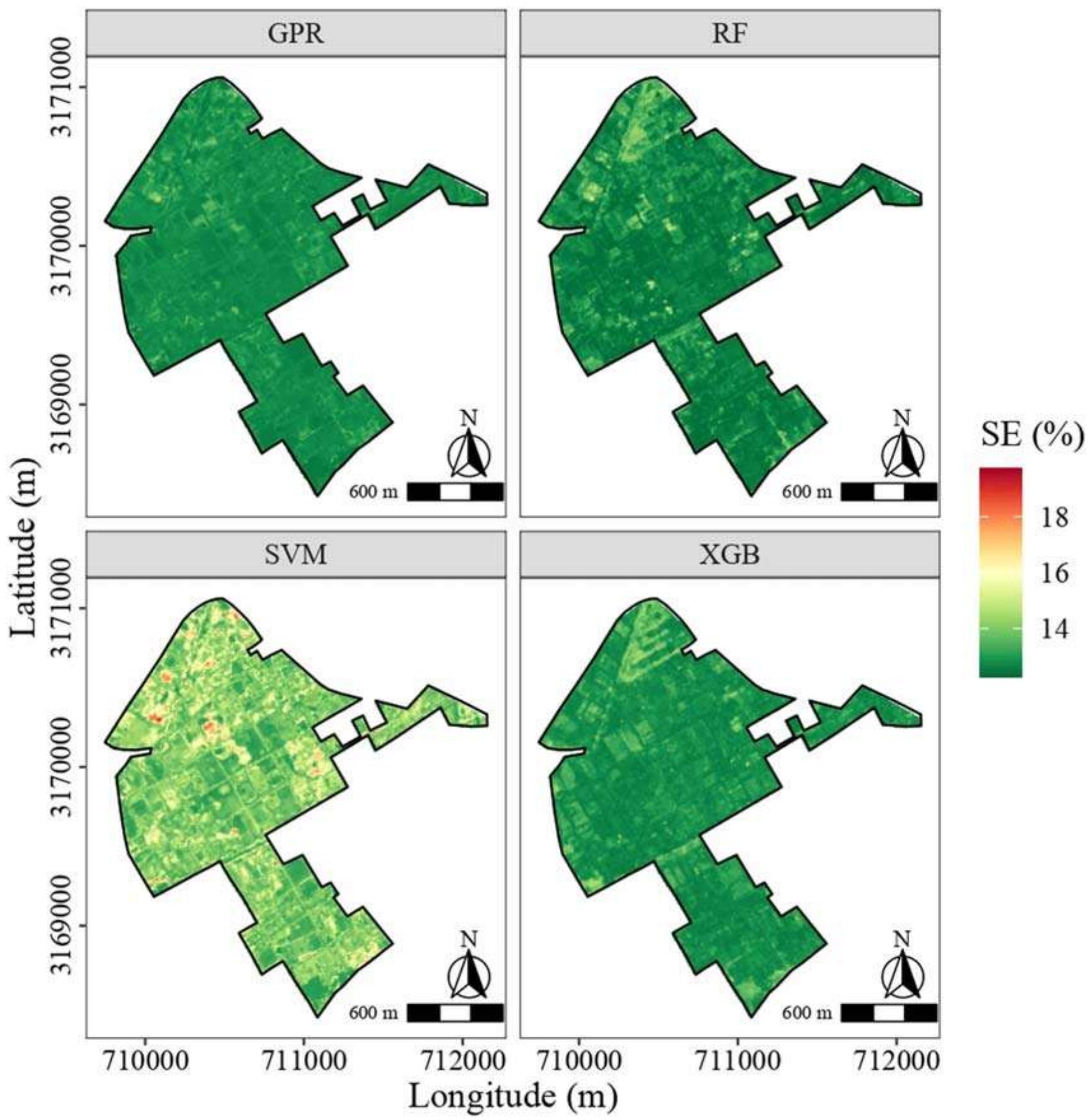

FIGURE 13.4 Uncertainty maps (standard error maps) associated with the prediction of surface soil moisture by four ML based models (GPR, RF, SVM and XGB)

13.8.3 Summary and key findings

The current study nice represented the integration of optical, thermal, and microwave RS technologies for producing digital maps of SSM of the study area using four different ML-based models. The key findings of the study are:

1. VIP technique identified the backscatter coefficient (Sigma0_VV), NBRI, and MNDWI were the key variables for SSM mapping.
2. RF model outperformed the other models (GPR, SVR, and XGB) for the predictive mapping of SSM in the study area.
3. The uncertainty analysis revealed that SVM produced more error in terms of SE as compared to the other three models. A higher error was encountered during the estimation of SSM because of the highly dynamic nature of SM in the study area. So, the spatiotemporal variability of SM must be considered before scheduling irrigation to crops.

However, the study might be limited due to the difficulty in getting all three types of RS data, especially cloud-free images and data, at the same date and time. Overall, the study demonstrated the potential of digital SSM mapping using ML techniques with better predictive accuracy. The generated high-resolution (30 m spatial resolution) SSM digital map will really be helpful for irrigation scheduling and better soil water management in the study area. Inclusion of more variables and the application of DL techniques may further improve the accuracy of digital SSM in future studies.

13.9 Limitations and future prospects of DSM

Based on the literature review, we tried to identify some of the challenges related to DSM or limitations of using ML in DSM. In the first paragraph, we pointed out the challenges associated with DSM, and then in the second paragraph, we outline some of the future aspects of DSM.

Regarding sampling designing, it is well known that the selection of a suitable sampling design plays a crucial role in obtaining a better predictive accuracy in DSM (De Gruijter et al., 2006). However, till now ML-compatible sample designs have not yet been found. One study demonstrated that when the sample size is small, it is optimal to utilize basic linear models. It also emphasized that a sufficiently high sample size is more crucial than a complicated ML method (Somarathna et al., 2017). For India, previous soil surveys have produced an enormous amount of legacy soil data, mostly available as printed books and papers. A significant barrier to DSM in India is the lack of standardized soil profile information that can be downloaded digitally. However, additional difficulties in categorizing mapping are brought on by the nature of the legacy soil data, crucial input of DSM. Balanced sets of units are necessary for the categorical mapping of soil classes or properties using ML techniques. However, the legacy soil samples create an imbalance among the classes, frequently leading to models biased toward the oversampled classes and very low predicted accuracy for the under-sampled classes (He and Garcia, 2009). Unless specifically stated, ML techniques do not consider the spatial autocorrelation present in the raw soil data. Recent studies showed that the geographically weighted regression technique can consider spatial autocorrelation in dataset for DSM (Georganos et al., 2019). The addition of pseudo-covariates can be detrimental since it prevents the study of the residuals and the development of new hypotheses from them (Hawkins, 2012). In the case of validation, because of the presence of autocorrelation in the observations and the covariates points, these procedures like random k-fold cross validation, simple random split produce noticeably overly optimistic estimates for the validation statistics (Meyer et al., 2018). The ML algorithms only utilize the information obtained from input data provided to draw digital maps of any soil properties or soil-based processes without utilizing a priori knowledge or conceptual framework of model (Koch et al., 2019). DSM techniques cannot be applied for mapping of complex soil-based processes. Hengl et al. (2014) do not provide soil maps in several under-sampled regions of the world, such as deserts and glaciers, for global mapping of several soil attributes in order to prevent extrapolation. The present case study's main flaw was the inability to get similar Sentinel-1 and cloud-free Landsat-8 launch dates, preventing SM maps with better temporal resolution. Another drawback of the present DSM of SM is the exclusion of terrain parameters, which have been reported as crucial input parameters for the digital mapping of SSM (Chatterjee et al., 2020). Suppose we again consider all the related issues in a single model to produce DSM. In that case, it will increase the complexity of a model and reduce our understanding of the basic processes occurring in soil. Therefore, an ML model that forecasts a number based on the correlation between covariates that, according to current knowledge, are uncertain should not be treated as seriously as a number that is anticipated by mechanistic processes or an accepted theory.

Based on the challenges, we tried to extract some solutions from the review of the literatures in order to draw some future perspectives on DSM. For example, most existing case studies confirmed that conditioned Latin Hypercube sampling or grid-based sampling could be used to get better prediction accuracies in DSM (Wadoux et al., 2020). To reduce the problem of overfitting or underfitting, it is suggested to calibrate a single model for predicting multiple soil properties or a single soil property at multiple soil depths (Wadoux, 2019). In the present scenario, the data-driven ML-based models provide better predictability in digital or predictive mapping of soil properties than mechanistic or geostatistical models. However, we need to improve our scientific understanding of the soil-related processes to choose a suitable ML technique to map the soil properties. Therefore, future research on DSM should include three fundamental components, that is, plausibility, interpretability, and explainability, as proposed by Roscher et al., 2020 and Lipton (2018). Plausibility denotes that models must be valid in light of scientific ideas and current knowledge in addition to being accurate. The ability to translate an abstract model or model output into human-understandable words is known as interpretability (Montavon et al., 2018). The three issues of "what is the modeled process," "how has it been modeled," and "why has this process been modeled" are the focus of explainability (Miller, 2019).

13.10 Conclusion

This chapter highlighted the basic concepts, principles, methods, and brief applications of DSM worldwide. Developed countries like Australia, USA, and France are now in leading positions for applying the DSM in solving real-world problems. Considering the alarming situation of land degradation and increasing food demand, India should adopt the DSM in its operational phase with the collaborative efforts of both government and nongovernmental bodies.

Although DSM was started in India at the research level, it is still in primitive condition to follow it in operation. So far, only two national-level digital maps have been prepared in India. However, the gradually increasing interest in DSM will show a path to shift DSM from the research to the operational stage. As the Indian agricultural land holdings are small and marginal, fine-resolution digital maps (10−30 m) are necessary to better manage soil fertility and water resources. For this, researchers in India should choose the optimum sampling strategy, sample size, and fine resolution RS product to make DSM more successful. The initiative should be taken by leading organizations like the Indian Council of Agricultural Research, NBSS&LUP, Indian Space Research Organisation, National Remote Sensing Centre, Indian Institute of Remote Sensing, Indian Institute of Technology, and other organizations in a collaborative effort to make the DSM data products easily available to the end users using online platforms for their successful application in real-world conditions.

References

Adab, H., Morbidelli, R., Saltalippi, C., Moradian, M., Ghalhari, G.A.F., 2020. Machine learning to estimate surface soil moisture from remote sensing data. Water 12, 3223.

Adhikari, K., Minasny, B., Greve, M.B., Greve, M.H., 2014. Constructing a soil class map of Denmark based on the FAO legend using digital techniques. Geoderma 214, 101−113.

Agriculture Census Division, 2019. All India Report on Number and Area of Operational Holdings, Agricultural Census (2015−16). Government of India, New Delhi.

Amare, T., Hergarten, C., Hurni, H., Wolfgramm, B., Yitaferu, B., Selassie, Y.G., 2013. Prediction of soil organic carbon for Ethiopian highlands using soil spectroscopy. Int. Sch. Res. Not. 2013.

Amazirh, A., Merlin, O., Er-Raki, S., Gao, Q., Rivalland, V., Malbeteau, Y., et al., 2018. Retrieving surface soil moisture at high spatio-temporal resolution from a synergy between Sentinel-1 radar and Landsat thermal data: a study case over bare soil. Remote. Sens. Environ. 211, 321−337.

Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., et al., 2014. GlobalSoilMap. pp. 93−134. Available from: https://doi.org/10.1016/B978-0-12−800137-0.00003-0.

Arrouays, D., Poggio, L., Salazar Guerrero, O.A., Mulder, V.L., 2020. Digital soil mapping and GlobalSoilMap. Main advances and ways forward. Geoderma Reg. 21, e00265.

Arrouays, D., Mulder, V.L., Richer-de-Forges, A.C., 2021. Soil mapping, digital soil mapping and soil monitoring over large areas and the dimensions of soil security−a review. Soil. Secur. 5, 100018.

Beaudette, D.E., O'Geen, A.T., 2009. Soil-Web: an online soil survey for California, Arizona, and Nevada. Comput. Geosci. 35, 2119−2128.

Bishop, T.F.A., McBratney, A.B., 2001. A comparison of prediction methods for the creation of field-extent soil property maps. Geoderma 103, 149−160.

Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S., 2010. Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. Springer Science & Business Media.

Bonfatti, B.R., Hartemink, A.E., Giasson, E., Tornquist, C.G., Adhikari, K., 2016. Digital mapping of soil carbon in a viticultural region of Southern Brazil. Geoderma 261, 204−221.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and RegressionTrees. Wadsworth & Brooks Wadsworth Statistics/ Probability Series.

Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. Eur. J. Soil. Sci. 62, 394−407.

Bui, E.N., Moran, C.J., 2003. A strategy to fill gaps in soil survey over large spatial extents: an example from the Murray−Darling basin of Australia. Geoderma 111, 21−44.

Burgess, T., Webster, R., 1980. Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging. J. Soil. Sci. 31, 315−331.

Burrough, P.A., 1986. Principles of Geographical Information Systems for Land Resources Assessment. Clarendon Press, Oxford.

Byrne, J.M., Yang, M., 2016. Spatial variability of soil magnetic susceptibility, organic carbon and total nitrogen from farmland in northern China. Catena 145, 92−98.

Carré, F., McBratney, A.B., Mayr, T., Montanarella, L., 2007. Digital soil assessments: beyond DSM. Geoderma 142, 69−79.

Castaldi, F., Palombo, A., Santini, F., Pascucci, S., Pignatti, S., Casa, R., 2016. Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon. Remote. Sens. Environ. 179, 54−65.

Chatterjee, S., Huang, J., Hartemink, A.E., 2020. Establishing an empirical model for surface soil moisture retrieval at the U.S. climate reference network using Sentinel-1 backscatter and ancillary data. Remote. Sens. 12, 1242.

Cipra, J.E., 1973. Mapping soil associations using ERTS MSS data. LARS Tech. Rep. 117.

Das, B., Rathore, P., Roy, D., Chakraborty, D., Jatav, R.S., Sethi, D., et al., 2022. Comparison of bagging, boosting and stacking algorithms for surface soil moisture mapping using optical-thermal-microwave remote sensing synergies. Catena 217, 106485.

De Gruijter, J.J., Brus, D.J., Bierkens, M.F.P., Knotters, M., 2006. Sampling for Natural Resource Monitoring. Springer Science & Business Media, Dordrecht, NL.

Dharumarajan, S., Hegde, R., 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 Reg. 10, 154–162.

Dharumarajan, S., Hegde, R., Janani, N., Singh, S.K., 2019. The need for digital soil mapping in India. Geoderma Reg. 16, e00204.

Dharumarajan, S., Hegde, R., Lalitha, M., 2021. Modelling of soil depth and hydraulic properties at regional level using environmental covariates—a case study in India. Geoderma Reg. 27, e00439. Available from: https://doi.org/10.1016/j.geodrs.2021.e00439.

Dobson, M.C., Ulaby, F.T., 1986. Active microwave soil moisture research. IEEE Trans. Geosci. Remote. Sens. 23–36.

Dokuchaev, V.V., 1899. Report to the Transcaucasian Statistical Committee on Soil Taxonomy in General and Especially for the Transcaucasia. Horizontal and Vertical Soil Zones. Office Press of the Civilian Affairs Commander-in-Chief in the Caucasus, Tifilis (in Russian).

Filipponi, F., 2019. Sentinel-1 GRD Preprocessing Workflow. Proceedings 18, 11. 10.3390/ECRS-3-06201.

Forkuor, G., Hounkpatin, O.K.L., Welp, G., Thiel, M., 2017. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. PLoS One 12, e0170478.

Friedman, JH, 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232. Available from: https://doi.org/10.1214/aos/1013203451.

Gallant, J.C., Wilson, J.P., 1996. TAPES-G: a grid-based terrain analysis program for the environmental sciences. Comput. Geosci. 22, 713–722.

Georganos, S., Grippa, T., Gadiaga, A.N., Linard, C., Lennert, M., Vanhuysse, S., et al., 2019. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. Geocarto Int. 1, 1–12.

Gomes, L.C., Faria, R.M., de Souza, E., Veloso, G.V., Schaefer, C.E.G.R., Fernandes Filho, E.I., 2019. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 340, 337–350.

Government of India, 2020. Agricultural Statistics at a Glance 2019. Department of Agriculture Cooperation and Farmers Welfare, Directorate of Economics and Statistics, Government of India.

Greenwell, B., Boehmke, B., Cunningham, J., Developers, G.B.M., Greenwell, M.B., 2019. Package ‘gbm.’ R Packag. version 2.

Hartemink, A.E., Hempel, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., MacMillan, R.A., et al., 2010. GlobalSoilMap. net—a new digital soil map of the world. In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. Springer Science, Dordrecht, pp. 423–428.

Hartemink, A.E., Minasny, B., 2014. Towards digital soil morphometrics. Geoderma 230, 305–317.

Hawkins, B.A., 2012. Eight (and a half) deadly sins of spatial analysis. J. Biogeogr. 39, 1–9.

He, H., Garcia, E.A., 2009. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 21, 1263–1284.

Henderson, B.L., Bui, E.N., Moran, C.J., Simon, D.A.P., 2005. Australia-wide predictions of soil properties using decision trees. Geoderma 124, 383–398.

Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120, 75–93.

Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., et al., 2014. SoilGrids1km|global soil information based on automated mapping. PLOS ONE 9, e105992.

Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E., Schmidt, M.G., 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. Geoderma 265, 62–77.

Hinge, G., Surampalli, R.Y., Goyal, M.K., 2018. Prediction of soil organic carbon stock using digital mapping approach in humid India. Environ. Earth Sci. 77, 1–10.

Hong, H., Liu, J., Bui, D.T., Pradhan, B., Acharya, T.D., Pham, B.T., et al., 2018. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). Catena 163, 399–413.

Hudson, B.D., 1992. The soil survey as paradigm-based science. Soil. Sci. Soc. Am. J. 56, 836–841.

Jafari, A., Finke, P.A., Van deWauw, J., Ayoubi, S., Khademi, H., 2012. Spatial prediction of USDA—great soil groups in the arid Zarand region, Iran: comparing logistic regression approaches to predict diagnostic horizons and soil types. Eur. J. Soil. Sci. 63 (2), 284–298.

Jamagne, M., 1967. Bases et techniques d’une cartographie des sols. Annales Agronomiques, hors-śerie 18, 142.

Jenny, H., 1941. Factors of soil formation. A System of Quantitative Pedology. McGraw-Hill Book Company, New York.

Jenny, H., Salem, A.E., Wallis, J.R., 1968. Interplay of soil organic matter and soil fertility with state factors and soil properties. Study Week Org. Matter Soil. Fertil. Pontif. Acad. Sci. Scr. Varia 32, 5236.

Kempen, B., Brus, D.J., Stoorvogel, J.J., Heuvelink, G.B.M., de Vries, F., 2012. Efficiency comparison of conventional and digital soil mapping for updating soil maps. Soil. Sci. Soc. Am. J. 76, 2097–2115.

Khaledian, Y., Miller, B.A., 2020. Selecting appropriate machine learning methods for digital soil mapping. Appl. Math. Model. 81, 401–418.

Kidd, D., Searle, R., Grundy, M., McBratney, A., Robinson, N., O’Brien, L., et al., 2020. Operationalising digital soil mapping−lessons from Australia. Geoderma Reg. 23, e00335.

Koch, J., Stisen, S., Refsgaard, J.C., Ernstsen, V., Jakobsen, P.R., Hojberg, A.L., 2019. Modeling depth of the redox interface at high resolution at national scale using random forest and residual gaussian simulation. Water Resour. Res. 55, 1451–1469.

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer.

Kumar, P., Prasad, R., Choudhary, A., Gupta, D.K., Mishra, V.N., Vishwakarma, A.K., et al., 2019. Comprehensive evaluation of soil moisture retrieval models under different crop cover types using C-band synthetic aperture radar data. Geocarto Int. 34, 1022–1041.

Lark, R.M., Lapworth, D.J., 2012. Quality measures for soil surveys by lognormal kriging. Geoderma 173, 231–240.

Leenaars, J.G.B., Claessens, L., Heuvelink, G.B.M., Hengl, T., Ruiperez Gonźalez, M., van Bussel, L.G.J., et al., 2018. Mapping rootable depth and root zone plant-available water holding capacity of the soil of sub-Saharan Africa. Geoderma 324, 18−36.

Lemercier, B., Lagacherie, P., Amelin, J., Sauter, J., Pichelin, P., Richer-De-Forges, A.C., et al., 2022. Multiscale evaluations of global, national and regional digital soil mapping products in France. Geoderma 425, 116052.

Leutner, B., Horning, N., Schwalb-Willmann, J. and Hijmans, R.J., 2019. RStoolbox: Tools for remote sensing data analysis (version 0.2. 4).

Lipton, Z.C., 2018. The mythos of model interpretability. Queue 16, 31−57.

Liu, Y., Qian, J., Yue, H., 2021. Combined Sentinel-1A with Sentinel-2A to estimate soil moisture in farmland. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 14, 1292−1310.

Ma, Y., Minasny, B., Malone, B.P., Mcbratney, A.B., 2019. Pedology and digital soil mapping (DSM). Eur. J. Soil. Sci. 70, 216−235.

Malone, B.P., McBratney, A.B., Minasny, B., 2011. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. Geoderma 160, 614−626.

McBratney, A., Mendonça Santos, M., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3−52. Available from: https://doi.org/10.1016/S0016-7061(03)00223-4.

McBratney, A.B., Minasny, B., Whelan, B., 2011. Defining proximal soil sensing. In: The Second Global Workshop on Proximal Soil Sensing—Montreal. pp. 15−18.

Meersmans, J., De Ridder, F., Canters, F., De Baets, S., Van Molle, M., 2008. A multiple regression approach to assess the spatial distribution of Soil Organic Carbon (SOC) at the regional scale (Flanders, Belgium). Geoderma 143, 1−13. Available from: https://doi.org/10.1016/j.geoderma.2007.08.025.

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., Nauss, T., 2018. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Env. Model. Softw. 101, 1−9.

Miller, B.A., Schaetzl, R.J., 2014. The historical role of base maps in soil geography. Geoderma 230, 329−339.

Miller, T., 2019. Explanation in articial intelligence: insights from the social sciences. Artif. Intell. 267, 1−38.

Minasny, B., McBratney, A.B., 2016. Digital soil mapping: a brief history and some lessons. Geoderma 264, 301−311.

Mitran, T., Meena, R.S., Chakraborty, A., 2021. Geospatial technologies for crops and soils: an overview. Geospatial Technologies for Crops and Soils. Springer, Singapore, Singapore, pp. 1−48. Available from: https://doi.org/10.1007/978-981-15-6864-0_1.

Mitran, T., Solanky, V., Janakirama Suresh, G., Sujatha, G., Sreenivas, K., Ravisankar, T., 2019. Predictive mapping of surface soil texture in a semi-arid region of India through geostatistical modeling. Model. Earth Syst. Environ. 5, 645−657.

Montavon, G., Samek, W., Muller, K.-R., 2018. Methods for interpreting and understanding deep neural networks. Digit. Signal. Process. 73, 1−15.

Mosleh, Z., Salehi, M.H., Jafari, A., Borujeni, I.E., Mehnatkesh, A., 2016. The effectiveness of digital soil mapping to predict soil properties over low-relief areas. Environ. Monit. Assess. 188, 1−13.

Mponela, P., Snapp, S., Villamor, G., Tamene, L., Le, Q.B., Borgemeister, C., 2020. Digital soil mapping of nitrogen, phosphorus, potassium, organic carbon and their crop response thresholds in smallholder managed escarpments of Malawi. Appl. Geogr. 124, 102299.

Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Arrouays, D., 2016. GlobalSoilMap France: high-resolution spatial modelling the soils of France up to two meter depth. Sci. Total. Environ. 573, 1352−1369.

NRCS US, 1993. Soil survey division staff (1993) soil survey manual. Soil conservation service. US Dep. Agric. Handb. 18, 315.

Odeh, I.O.A., Chittleborough, D.J., McBratney, A.B., 1992. Fuzzy-c-means and kriging for mapping soil as a continuous system. Soil. Sci. Soc. Am. J. 56, 1848−1854.

Odgers, N.P., McBratney, A.B., Minasny, B., 2015. Digital soil property mapping and uncertainty estimation using soil class probability rasters. Geoderma 237, 190−198.

Pahlavan-Rad, M.R., Akbarimoghaddam, A., 2018. Spatial variability of soil texture fractions and pH in a flood plain (case study from eastern Iran). Catena 160, 275−281.

Pereira, P., Brevik, E., Munoz-Rojas, M., Miller, B., 2017. Soil Mapping and Process Modeling for Sustainable Land Use Management. Elsevier.

Poggio, L., de Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., et al., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil 7, 217−240.

Priori, S., Bianconi, N., Costantini, E.A.C., 2014. Can γ-radiometrics predict soil textural data and stoniness in different parent materials? A comparison of two machine-learning methods. Geoderma 226, 354−364.

Qin, Z., Karnieli, A., Berliner, P., 2001. A mono-window algorithm for retrieving land surface temperature from Landsat TM data and its application to the Israel-Egypt border region. Int. J. Remote. Sens. 22, 3719−3746.

Ranjbar, F., Jalali, M., 2016. The combination of geostatistics and geochemical simulation for the site-specific management of soil salinity and sodicity. Comput. Electron. Agric. 121, 301−312.

Reddy, N.N., Chakraborty, P., Roy, S., Singh, K., Minasny, B., McBratney, A.B., et al., 2021. Legacy data-based national-scale digital mapping of key soil properties in India. Geoderma 381, 114684.

Richer-de-Forges, A.C., Arrouays, D., Poggio, L., Chen, S., Lacoste, M., Minasny, B., et al., 2022. Hand-feel soil texture observations to evaluate the accuracy of digital soil maps for local prediction of particle size distribution. A case study in central France. Pedosphere 32.

Roscher, R., Bohn, B., Duarte, M.F., Garcke, J., 2020. Explainable machine learning for scientific insights and discoveries. Ieee Access 8, 42200−42216.

Rossel, R.A.V., McBratney, A.B., Minasny, B., 2010. Proximal Soil Sensing. Springer Science & Business Media.

Rossiter, D.G., Poggio, L., Beaudette, D., Libohova, Z., 2021. How well does predictive soil mapping represent soil geography? SOIL, under discussion an investigation from the USA 10.5194/soil-2021-80.
Sadovski, A.N., Bie, S.W., 1978. Developments in soil information systems. In: Proceedings of the Second Meeting of the ISSS Working Group on Soil Information Systems, Varna/Sofia, Bulgaria, 30 May−4 June 1977. Pudoc.
Santra, P., Chopra, U.K., Chakraborty, D., 2008. Spatial variability of soil properties and its application in predicting surface map of hydraulic parameters in an agricultural farm. Curr. Sci. 95, 937−945.
Santra, P., Kumar, M., Panwar, N., 2017. Digital soil mapping of sand content in arid western India through geostatistical approaches. Geoderma Reg. 9, 56−72.
Santra, P., Kumar, M., Panwar, N.R., Yadav, R.S., 2021. Digital soil mapping: the future need of sustainable soil management. Geospatial Technologies for Crops and Soils. Springer, Singapore, Singapore, pp. 319−355. Available from: https://doi.org/10.1007/978-981-15-6864-0_9.
Shirzadi, A., Soliamani, K., Habibnejhad, M., Kavian, A., Chapi, K., Shahabi, H., et al., 2018. Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. Sensors 18, 3777.
Simonson, R.W., 1959. Outline of a generalized theory of soil genesis. Soil. Sci. Soc. Am. J. 23, 152−156.
Singh, S.K., Chatterji, S., 2018. Land resource inventory for agricultural land use planning using geospatial techniques. Geospatial Technologies in Land Resources Mapping, Monitoring and Management. Springer, pp. 163−183.
Somarathna, P.D.S.N., Minasny, B., Malone, B.P., 2017. More data or a better model? guring out what matters most for the spatial prediction of soil carbon. Soil. Sci. Soc. Am. J. 81, 1413−1426.
Sreenivas, K., Dadhwal, V.K., Kumar, S., Harsha, G.S., Mitran, T., Sujatha, G., et al., 2016. Digital mapping of soil organic and inorganic carbon status in India. Geoderma 269, 160−173.
Sreenivas, K., Sujatha, G., Sudhir, K., Kiran, D.V., Fyzee, M.A., Ravisankar, T., et al., 2014. Spatial assessment of soil organic carbon density through random forests based imputation. J. Indian Soc. Remote. Sens. 42, 577−587. Available from: https://doi.org/10.1007/s12524-013-0332-x.
Taghizadeh-Mehrjardi, R., Nabiollahi, K., Minasny, B., Triantafilis, J., 2015. Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region, Iran. Geoderma 253−254, 67−77. Available from: https://doi.org/10.1016/j.geoderma.2015.04.008.
Thompson, J.A., Kienast-Brown, S., D'Avello, T., Philippe, J., Brungard, C., 2020. Soils2026 and digital soil mapping−A foundation for the future of soils information in the United States. Geoderma Reg. 22, e00294.
Triantafilis, J., Lesch, S.M., La Lau, K., Buchanan, S.M., 2009. Field level digital soil mapping of cation exchange capacity using electromagnetic induction and a hierarchical spatial regression model. Soil. Res. 47, 651−663.
Vaysse, K., 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.
Vaysse, K., Lagacherie, P., 2017. Using quantile regression forest to estimate uncertainty of digital soil mapping products. Geoderma 291, 55−64.
Vision, C., n.d. 2050: Indian Council of Agricultural Research, New Delhi; 2015.
Voltz, M., Arrouays, D., Bispo, A., Lagacherie, P., Laroche, B., Lemercier, B., et al., 2020. Possible futures of soil-mapping in France. Geoderma Reg. 23.
Wadoux, A.M.J.C., 2019. Using deep learning for multivariate mapping of soil with quantied uncertainty. Geoderma 351, 59−70.
Wadoux, A.M.C., Minasny, B., McBratney, A.B., 2020. Machine learning for digital soil mapping: applications, challenges and suggested solutions. Earth-Sci. Rev. 210, 103359.
Walker, P.H., Hall, G.F., Protz, R., 1968. Relation between landform parameters and soil properties. Soil. Sci. Soc. Am. J. 32, 101−104.
Webster, R., Harrod, T.R., Staines, S.J. and Hogan, D.V., 1979. Grid sampling and computer mapping of the Ivybridge area, Devon (No. Tech Monograph No 12Mongograph).
Wiesmeier, M., Barthold, F., Blank, B., 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.
Wilson, H.F., Satchithanantham, S., Moulin, A.P., Glenn, A.J., 2016. Soil phosphorus spatial variability due to landform, tillage, and input management: a case study of small watersheds in southwestern Manitoba. Geoderma 280, 14−21.
Worsham, L., Markewitz, D., Nibbelink, N., 2010. Incorporating spatial dependence into estimates of soil carbon contents under different land covers. Soil. Sci. Soc. Am. J. 74, 635−646.
Wösten, J.H.M., Finke, P.A., Jansen, M.J.W., 1995. Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. Geoderma 66, 227−237.
Wu, T., Luo, J., Dong, W., Sun, Y., Xia, L., Zhang, X., 2019. Geo-object-based soil organic matter mapping using machine learning algorithms with multi-source geo-spatial data. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 12, 1091−1106.
Yigini, Y., Panagos, P., 2016. Assessment of soil organic carbon stocks under future climate and land cover changes in Europe. Sci. Total. Environ. 557, 838−850.
Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., Finke, P., 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. Geoderma 338, 445−452.
Zeraatpisheh, M., Jafari, A., Bodaghabadi, M.B., Ayoubi, S., Taghizadeh-Mehrjardi, R., Toomanian, N., et al., 2020. Conventional and digital soil mapping in Iran: past, present, and future. Catena 188, 104424.
Žížala, D., Minařík, R., Skála, J., Beitlerová, H., Juřicová, A., Rojas, J.R., et al., 2022. High-resolution agriculture soil property maps from digital soil mapping methods, Czech Republic. Catena 212, 106024.