

Chapter 9

Digital Soil Mapping: The Future Need of Sustainable Soil Management



Priyabrata Santra, Mahesh Kumar, N. R. Panwar, and R. S. Yadav

Contents

9.1	Introduction	321
9.2	Digital Soil Mapping Methodology	322
9.3	Legacy Soil Data and Digital Soil Mapping	323
9.4	Scale Issue in DSM	324
9.5	Geostatistical Approach of Digital Soil Mapping	327
9.5.1	Semivariogram	329
9.5.2	Kriging	331
9.5.3	Co-kriging	334
9.5.4	Other Variants of Kriging	336
9.6	State-Factor (Clorpt) Approach of DSM	337
9.6.1	Covariates on Terrain Attributes	337
9.6.2	Covariates on Bioclimatic Variables	338
9.6.3	Machine Learning Algorithms in DSM	339
9.6.4	Application of Hyperspectral and Remote Sensing Signature in DSM	342
9.7	Pedotransfer Function (PTF) Approach for Digital Soil Mapping	345
9.8	Accuracy and Uncertainty Analysis of Digital Soil Maps	347
9.9	DSM Applications: Soil Information System	350
9.10	Conclusion	351
	References	352

Abstract Digital soil mapping (DSM) involves in development of a statistical or mathematical model to estimate soil class or properties at unsampled locations using information on spatial variation of soil properties and different covariates affecting soil formation process. There are three main approaches followed in DSM, and these are geostatistical approach, state-factor (clorpt) approach, and pedotransfer function (PTF) approach. In the geostatistical approach, spatial variation parameters (nugget, sill, and range) are identified from a spatial soil database using semivariogram followed by making unbiased estimate of soil properties at unsampled location through kriging. In the state-factor (clorpt) approach, the soil formation theory is the backbone. In this approach soil is considered to be influenced by five major

P. Santra (✉) · M. Kumar · N. R. Panwar · R. S. Yadav
ICAR-Central Arid Zone Research Institute, Jodhpur, Rajasthan, India

factors: climate (cl), organism (o), relief (r), parent material (p), and time (t). Therefore, abundantly available information on these factors in different digital platforms are exploited to develop model to estimate soil properties at unsampled location. The PTF approach is used to develop digital soil maps of complex soil properties and difficult to measure soil properties. In this approach digital soil map of basic soil properties is first developed using the first two approaches, which are then combined to map of complex soil properties through PTF model. All these three approaches of DSM are discussed in detail along with assessment of its accuracy and uncertainty. Through the DSM approaches, available legacy soil data may be converted to digital products for its better accessibility and utility, e.g., through development of soil information system.

Keywords Digital soil mapping · Geostatistical modeling · Hyperspectral soil signatures · Machine learning · Pedotransfer functions · Soil information systems

Abbreviations

DEM	Digital Elevation Model
DSM	Digital Soil Mapping
DSMM	Digital Soil Mapping and Modeling
ANN	Artificial Neural Network
AWI	Arid Western India
IK	Indicator Kriging
KED	Kriging with External Drift
k-NN	k-nearest neighborhood
LCCC	Lin's Concordance Correlation Coefficient
LK	Lognormal Kriging
MLR	Multiple Linear Regression
OK	Ordinary Kriging
PK	Probability Kriging
PTF	Pedotransfer Function
RF	Random Forest
RK	Regression Kriging
SVM	Support Vector Machine
SVR	Support Vector Regression
UK	Universal Kriging
VIS-NIR-SWIR	Visible-Near-Infrared-Shortwave-Infrared

9.1 Introduction

Mapping soil properties has been conventionally done through surveying efforts followed by laboratory analysis. Soil maps developed by the conventional approach are generally hard copy maps and therefore are not easily accessible to end users. Moreover, mapping units of these maps are delineated based on soil profile data and surveyor's field experience. These mapping units sometimes represent quite a large area in the field, and thus soil properties of interest vary considerably within a unit. With the advancement of geostatistics and abundant availability of digital information on earth features, there is a possibility to map soil properties utilizing available soil data and auxiliary information on earth features and environmental variables. Through this approach, available legacy soil data may be converted to digital products for its better accessibility and utility. Moreover, in the context of digital India and soil health missions, it is timely and apt to prepare the digital soil maps for different regions of the country.

In a conventional approach, mapping of soil is done by expert surveyor based on his field observations and few auxiliary information, e.g., aerial photographs, remote sensing imageries, geological maps, vegetation pattern maps, etc. Information on different soil properties is attached to each polygon or mapping unit of the map. These labeled polygon maps are often called digital soil maps since the late 1970s. However, in a true sense, these maps cannot be called as digital soil maps; rather it can be called as digitized soil maps. In the DSM methodology, a statistical or mathematical model is developed to estimate soil class or properties at unsampled locations using the information on spatial variation of soil properties and different covariates affecting the soil formation process. DSM have recently gained importance in different parts of the world (McBratney et al. 2003; Lagacherie et al. 2006; Behrens and Scholten 2006; Grunwald 2009; Sanchez et al. 2009; Minasny and McBratney 2016). To get quantitative answers on the role of soil in carbon sequestration and its impact on biomass production and human health, the [GlobalSoilMap.net](http://www.globalsoilmap.net) project has been implemented by FAO and UNESCO in the year 2006. World Soil Information Centre (ISRIC, Netherland) has been working on several projects on DSM, e.g., Global Soil Information Facilities (GSIF), Africa Soil Information Service (AfSIS), World Inventory of Soil Emission Potentials (WISE), Soil and Terrain Database (SOTER), etc. Apart from these international programs, several countries have initiated their DSM programs, e.g., DIGISOL in Europe, OzDSM in Australia, NCSS DSM program of the USA, etc. Some of the digital soil products available in the WebGIS version are soil map of Scotland (<http://www.soilsscotland.gov.uk/data/soilsurvey25k.php>), soil information systems of California (<http://casoilresource.lawr.ucdavis.edu/soilwebapps/>), etc. All these DSM programs were based on legacy soil data available from different surveying efforts; however rapidly measurable soil spectral signatures have potential to improve the accuracy of the developed map (Shepherd and Walsh 2002; Brown et al. 2006; BenDor et al. 2009; Rossel et al. 2016; Katuwal et al. 2018).

Quantifying the spatial variation of soil properties for mapping purpose has been studied in India long ago by Agarwal and Gupta (1998) and Dahiya et al. (1998) followed by several researchers (Das 2007; Santra et al. 2008, 2012a, b, 2017a, b, c; Kamble and Aggrawal 2011; Chatterjee et al. 2015, Singh et al. 2016). Recently, Santra et al. (2017a, b, c) made a comprehensive review of DSM in India. Scattered efforts have been made by several researchers in India for soil spectral library generation (NBSS&LUP 2005; Saxena et al. 2003; Srivastava et al. 2004; Singh et al. 2014). Reflectance spectra of soil have also been used for the rapid characterization of soil properties. For example, (i) Santra et al. (2009) characterized soil hydraulic properties using proximal spectral reflectance, (ii) Gulfo et al. (2012) assessed soil moisture content using hyperspectral reflectance, (iii) Divya et al. (2014) characterized soil texture using hyperspectral reflectance, (iv) Kaduputiya et al. (2010) assessed soil nutrient contents using diffused reflectance spectra, etc. Apart from these, reflectance spectroscopy has been recently applied to estimate several soil properties in West Bengal, Rajasthan, Karnataka, etc. (Sharathjith et al. 2014; Santra et al. 2015, Mohanty et al. 2016; Gupta et al. 2016; Chakraborty et al. 2017). A detailed review of hyperspectral signature-based soil resource assessment is available in Das et al. (2015).

9.2 Digital Soil Mapping Methodology

DSM is the digital way of mapping soil properties. The digital way indicates the application of several computation methods and modeling approaches in the mapping procedure and finally presenting the soil maps in digital format rather than hard copy. A big advantage of the digital format of the soil map is the easy accessibility of it to end users. There are three main approaches followed in DSM or digital soil mapping and modeling (DSMM) as depicted in Fig. 9.1. These are the geostatistical approach, state-factor (clorpt) approach, and PTF approach. In the geostatistical approach, spatial variation parameters (nugget, sill, and range) are identified from a spatial soil database using semivariogram followed by making an unbiased estimate of soil properties at an unsampled location through kriging. In the state-factor (clorpt) approach, the soil formation theory proposed by Dokuchaev (1883) and Jenny (1941) is the backbone. In this approach, soil is considered to be influenced by five major factors: climate (cl), organism (o), relief (r), parent material (p), and time (t). Therefore, abundantly available information on these factors in different digital platforms are exploited to develop a model to estimate soil properties at an unsampled location. Later on, the clorpt approach is slightly modified by McBratney et al. (2003) including soil itself and the spatial locations of samples as a factor in the soil formation process, which is termed as “scorpan” approach. The pedotransfer approach is used to develop digital soil maps of complex soil properties and difficult to measure soil properties. In this approach, a digital soil map of basic soil properties is first developed using the first two approaches, which are then combined to map of complex soil properties through the PTF model. The PTF models are typically the

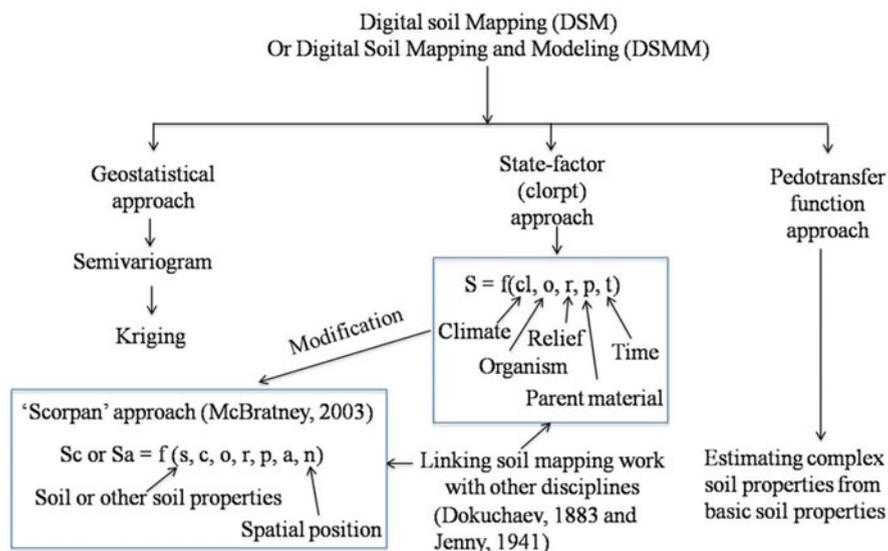


Fig. 9.1 Digital soil mapping approaches

regression-based models relating to complex soil properties with basic soil properties. Nowadays, PTFs are not only restricted to regression-based models; rather several advanced machine learning tools are applied.

9.3 Legacy Soil Data and Digital Soil Mapping

Surveying efforts during the past few decades have led to the development of large soil databases in different parts of the world but are often left as unused after achieving the primary goal of the survey. This large soil database is often called a legacy soil database. The legacy soil database of a country or a target ecosystem may be utilized to develop digital soil maps, which then can be reutilized by stakeholders in different land management decisions. However, these legacy soil data need to be harmonized both temporally and depth-wise. In the case of temporal harmonization, the time-dependent changes of soil properties, if any, are identified, and then soil properties at a particular time reference are computed. In the case of depth harmonization, the soil properties for standard soil depths are computed using the spline technique. Six standard soil depths as followed by FAO Global Soil Mapping protocol are 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm.

A schematic diagram to utilize legacy soil data in DSM is presented in Fig. 9.2. There are several sources of legacy soil data (e.g., soil series-level database, soil atlas, local soil archive, published soil data in journals, etc.), which may be collated together in a single platform. If the number of soil sampling points in the legacy soil

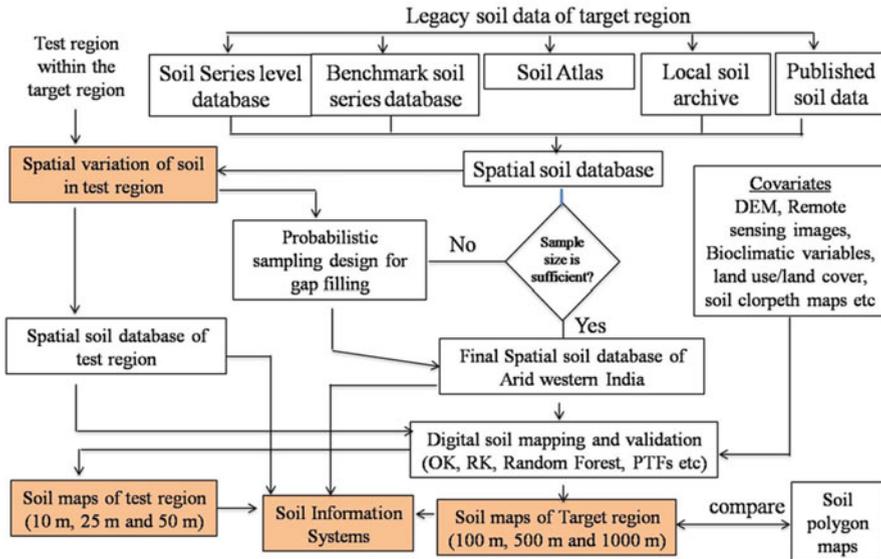


Fig. 9.2 A framework of DSM using legacy soil data

data is large enough to draw the semivariogram plot and to identify the spatial variation pattern, then we can straightway go for DSM using standard procedure. Otherwise, additional sampling effort may be required to fill the gap in sampling locations in order to compute the robust semivariogram model. For this purpose, a test region in the targeted spatial domain for DSM may be identified, and spatial variation pattern of the selected soil properties may be identified. Using this known spatial variation pattern, probabilistic sampling design may be formulated to increase the sampling density within the extent of the target area. Later on, information on several covariates may be used to develop accurate digital soil maps of the target area.

9.4 Scale Issue in DSM

Accuracy and uncertainty of digital soil maps largely depend on the scale of spatial data. The scale is defined by scale triplets (Blöschl and Sivapalan 1995), which are spacing, support, and extent (Fig. 9.3). The spacing is defined as the distance between a pair of sampling points, the extent is defined as the maximum distance between two sampling points in two-dimensional space of a spatial data, and the support is defined as the ground area from where the sample is collected and analyzed in the laboratory to represent it as a point data in a spatial database. These three scale triplets uniquely specify the scale of a spatial soil database and generally help to identify the pattern in the data. For example, the spatial extent of a

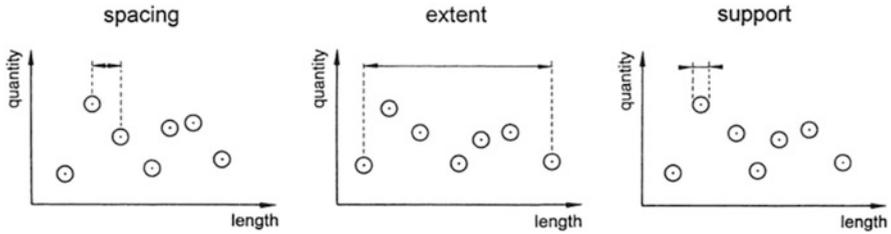


Fig. 9.3 The scale triplet (spacing, extent, and support). (Adopted, Blöschl and Sivapalan 1995)

spatial soil database in a farmer's field may be a maximum 100–200 m in India, whereas for a state-level spatial soil database, the extent is about 600–800 km. The support for measurement of bulk density in the field is about 5 cm, whereas if we take multiple samples from a field and then composting it to a single value for that field, then the support of that measurement will be total field block. Spacing is another important scale parameter in the DSM approach specifically in the geostatistical approach. If the minimum spacing between a sampling pair in the spatial database is large, then it will not be able to capture the spatial variation parameter. In the case of large spacing, the sampling density is low, whereas in closed spaced sampling points, the sampling density is high. It is always desirable to have large sampling density in the spatial database; however, the cost and time involved in achieving this optimistic sampling density is also needed to be looked into.

Therefore, the scale issue of a spatial soil database needs to be resolved first, and it depends on the soil properties on which we are interested to identify the spatial variation. The effect of the sampling scale on hydrological processes is beautifully depicted schematically by Blöschl and Sivapalan (1995), which is presented here in Fig. 9.4. This explanation for the hydrological process in the figure is also true for the spatial pattern of soil properties. In the figure, the solid line represents the natural variation of soil properties, whereas the small circles represent sampling locations. In general, it is not possible to collect soil samples from all possible locations to capture the full natural variability of the target soil property. Rather, we collect soil samples from a subset of all possible locations. For example, if we take soil samples following Fig. 9.4a, we fail to capture the microscale variation in the data because the spacing is too large to capture this small-scale variation. This type of spatial pattern may be observed for soil nutrient content which is highly influenced by specific land management practices followed in fragmented land units in an area. Under such a situation, the spacing between sampling pairs needs to be decreased or the sampling density needs to be increased, and this change in sampling scale is again to be optimized with sampling budget and time constraints. In another case, as shown in Fig. 9.4b, if the extent of the data is too small, then it will not capture the large-scale variability. Therefore, we need to increase our survey area to include large-scale variations of the soil property in the data. In this case, we may ignore excessive sampling with small separation distance; rather we may shift our focus to

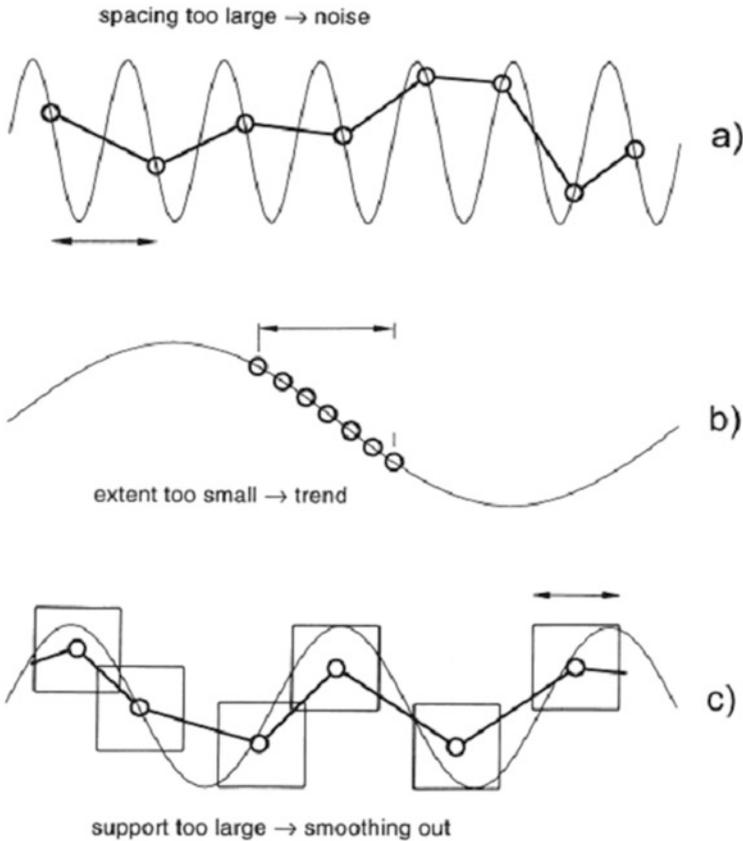


Fig. 9.4 The effect of measurement scale on capturing the “true” spatial pattern. The circles are the measurements, and the thin line is the “true” spatial pattern. (Adopted, Blöschl and Sivapalan 1995)

include sampling pairs with large separation distance. Such type of large-scale variation may be observed in soil properties which are more influenced by parent material, which has large spatial continuity. In the third case as depicted in Fig. 9.4c, if the support is quite large, then most of the variability in the data will be smoothed out. Measurement of soil properties is generally done based on samples collected from the field using an auger or sampling core with a cross-sectional area having 4–5 cm diameter. Therefore, support of measurements of soil properties is about 4–5 cm, which is often assumed as point support considering the large extent of the field as compared to the support. Sometimes, in situ measurement of soil properties is carried out with large support, e.g., about 20–40 cm for infiltration measurements, about 2–4 m for in situ measurement of soil water retention, etc. Often, we composite the soil samples from multiple locations in a field and in the process we increase the support of measurements to rule out the random variation in the field. In a regional-scale soil mapping, the support of measurements may be increased to 1 ha by making

multiple measurements in 1 ha field and then averaging it to a single value. However, in a field-scale soil mapping, such an increase in support is not desirable. From the above discussion, we understand that issue of sampling scale can be resolved after gathering knowledge on natural variation of soil properties. The question may be asked that “how we can know the natural variation of a soil property so that we can optimize the sampling strategy?” Natural variation of a soil property in an area may be approximated from previously identified spatial variation parameters of the target soil property from a nearby place. Otherwise, it may also be approximated from a preliminary survey. Later on, sampling efforts may be designed through probabilistic sampling theory with prior knowledge on spatial variation identified in the preliminary survey.

9.5 Geostatistical Approach of Digital Soil Mapping

In geostatistics, soil property at a particular location [$Z(x)$] is considered as a set of values following a probability distribution and not just a single value. Therefore, at each possible location x , a soil property, $Z(x)$, is considered as a random variable with a mean, μ , and a variance, σ^2 . This description of random variable of $Z(x)$ applies to infinitely many locations in space. At each possible location in space (x_i , $i=1, 2, 3, \dots$), it has its own probability distribution. Therefore, a range of possible values exists at a particular location following the probability distribution, and this is called an ensemble. One member from this ensemble for a particular location is called as realization of the property and is represented as $Z(x_i)$. A set of random variables or multiple realizations, $Z(x_1), Z(x_2), \dots, Z(x_i)$, is called as a random function, a random process, or a stochastic process. The set of true values of Z at each possible location that comprise the true realization of the random function is known as a regionalized variable.

Following the regionalized variable theory, values of the variables which are located near to each other are expected to be similar, whereas values of the variables which are separated from each other by a large distance are expected to be dissimilar. This relation of regionalized variables may be described by covariance. In classical statistics, covariance of two variables z_1 and z_2 for n pair of observations can be written as

$$C(z_1, z_2) = \frac{1}{n} \sum_{i=1}^n (z_{i,1} - \bar{z}_1)(z_{i,2} - \bar{z}_2) \quad (9.1)$$

Likewise, in geostatistics, the covariance of a regionalized random variable, Z , for two locations (x_1 and x_2) can be written as

$$C(x_1, x_2) = E[\{z(x_1) - \mu(x_1)\}\{z(x_2) - \mu(x_2)\}] \quad (9.2)$$

However, we cannot simply calculate the covariance because we do not know exactly the value of $\mu(x_1)$ and $\mu(x_2)$. We have measured only one value at each location x_1 and x_2 . To solve this problem, assumption of stationarity comes into picture.

The stationarity rule of geostatistics implies that the distribution of a random process has certain parameters that are stationary across all possible locations in two-dimensional space. The first-order stationarity states that the expected value of a regionalized variable at any location is constant for all x , which is mathematically written as $E[Z(x)] = \mu$. Assuming the first-order stationarity rule, we can replace mean of the regionalized variable at all possible location $[\mu(x_1), \mu(x_2), \dots, \mu(x_i)]$ by a single value μ . The value of μ can be estimated from arithmetic averaging of measured values at multiple locations. The second-order stationarity rule states that the squared deviation of the value from μ at all possible locations is also constant and equals to square of standard deviation. Mathematically, the second-order stationarity is written as $E[\{Z(x) - \mu\}^2] = \sigma^2$. Another rule of second-order stationarity defines the stationarity of covariance, which states that covariance of the regionalized variables located at two locations x_i and x_j depends only on their separation distance and not on their absolute positions. By applying this stationarity rule, it can be stated that for any pair of observation points x_i and x_j separated by a lag distance h , $E[\{Z(x_i) - \mu\}\{Z(x_j) - \mu\}] = C(x_i, x_j)$ and is constant for any given h . Therefore, the constancy of mean, variance, and covariance as discussed above are called as the second-order stationarity or weak stationarity.

After considering the stationarity rule, the auto-covariance function can be rewritten as

$$\begin{aligned} COV[Z(x), Z(x+h)] &= E[\{Z(x) - \mu\}\{Z(x+h) - \mu\}] \\ &= E[\{Z(x)\}\{Z(x+h)\} - \mu^2] \\ &= C(h) \end{aligned} \quad (9.3)$$

The above covariance is also called as auto-covariance since it represents the covariance of Z with itself but at different locations. To remove the dependence of auto-covariance on scale, i.e., h , it is often represented as dimensionless parameter autocorrelation, $\rho(h)$:

$$\rho(h) = \frac{C(h)}{C(0)} \quad (9.4)$$

where $C(0)$ is the covariance at lag 0, which is actually σ^2 .

After assuming the stationarity rules, problem arises again to consider μ to be constant within the sampling domain. Generally, μ changes in field as we keep on increasing the extent of sampling domain, and variances also increase with increase

in area of interest. Here, Matheron (1965) identified the problem and proposed the intrinsic hypothesis, which states that for a short separation distance at least, the difference between $Z(x)$ and $Z(x+h)$ is zero and the term covariance is replaced by variance of the difference in $Z(x)$ and $Z(x+h)$. In mathematical formula, these two hypotheses are written as

$$E[Z(x) - Z(x+h)] = 0 \quad (9.5)$$

$$\text{VAR}[Z(x) - Z(x+h)] = E\left[\{Z(x) - Z(x+h)\}^2\right] = 2\gamma(h) \quad (9.6)$$

where $\gamma(h)$ refers to semivariance, which is obviously the half of the variance.

9.5.1 Semivariogram

Semivariance as a function of h is called the semivariogram. From field measurements of soil properties at multiple locations, experimental semivariograms $\hat{\gamma}(h)$ for different lag distances h are calculated as follows (Goovaerts 1998):

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i+h)]^2 \quad (9.7)$$

where $N(h)$ is the number of data pairs within a given lag class, $Z(x_i)$ is the value of the variable at the location x_i , and $Z(x_i+h)$ is the value of the variable at a lag of h from the location x_i . Experimental semivariograms $[\hat{\gamma}(h)]$ as obtained from Eq. (9.7) are generally fitted in standard models so as to obtain the spatial variation parameters: nugget (C_0), sill ($C + C_0$), and range (a). Weighted least square technique is generally followed in fitting procedure, and the weight to semivariogram value at each lag is assigned in such a way that it is inversely proportional to the number of pairs for that particular lag. Sometimes, the semivariogram values at smaller lags are assigned with higher weights than the semivariogram values at large lag distance. During semivariogram calculation, maximum lag distance is generally taken as half of the minimum extent of sampling area so as to minimize the border effect. We are not discussing here the isotropic and anisotropic semivariogram. For general purpose, omnidirectional or isotropic semivariogram is followed if there is no trend of direction on the data. However, if there is strong trend of x - and y -direction on the data, the anisotropic semivariogram may also be calculated. Best-fit semivariogram model is selected with the lowest value of fitting error. Four commonly used semivariogram models are spherical, exponential, Gaussian, and linear, and mathematical expressions of these models are given below:

$$\begin{aligned} \text{Spherical model : } \gamma(h) = C_0 + C \left[1.5 \frac{h}{a} - 0.5 \left(\frac{h}{a} \right)^3 \right] & \text{ if } 0 \leq h \\ & \leq a; \text{ otherwise } C_0 + C \end{aligned} \quad (9.8)$$

$$\text{Exponential model : } \gamma(h) = C_0 + C_1 \left[1 - \exp \left\{ -\frac{h}{a} \right\} \right] \quad \text{for } h \geq 0 \quad (9.9)$$

$$\text{Gaussian model : } \gamma(h) = C_0 + C \left[1 - \exp \left\{ -\frac{h^2}{a^2} \right\} \right] \quad \text{for } h \geq 0 \quad (9.10)$$

$$\text{Linear model : } \gamma(h) = C_0 + C_1 \left[\frac{h}{a} \right] \quad \text{if } h < a; \text{ otherwise } = C_0 + C_1 \quad (9.11)$$

Apart from these four standard models, the Matern model is also quite often used. The parameter a in all these semivariogram models indicates range up to which spatial correlation between a pair of observation exists, beyond which a pair of observations is not spatially correlated. However, in case of exponential and Gaussian models, a represents the theoretical range, whereas practical range for these two semivariogram models is the lag distance at which semivariogram value reaches to 95% of sill. In all the above semivariogram models, nugget is expressed as C_0 , which actually quantifies microscale variation and measurement error for the respective soil property, whereas partial sill (C) indicates the amount of variation which can be defined by spatial correlation structure.

All these semivariogram models as discussed above are called as bounded semivariogram models or authorized semivariogram models. Apart from bounded semivariograms, there are some unbounded semivariograms also. The major feature of unbounded semivariogram is continuous increase in semivariogram values with lag distance and is generally expressed by the formula $\gamma(h) = wh^\alpha$, where $\gamma(h)$ is the semivariogram for a lag distance h , w describes the intensity of variation, and α describes the curvature. For a value of $\alpha=1$, the semivariogram is unbounded liner and w is simply the gradient. The parameter α has lower limit 0 and upper limit 2. In case of $\alpha < 1$, the semivariogram looks like convex upward, whereas for $\alpha > 1$, the semivariogram looks like concave upward. It is very strange to obtain unbounded variation or infinite variation of a feature on earth surface. However, often we observe infinite variation on this planet. This is so because we have been encountering more variation as long as we have been incorporating new regions into survey. This type of unbounded variation is observed if the environmental variables and parent material have influence on soil property. Thus, the unbounded variation is often neglected; rather a bounded semivariogram model is generally tried to fit in the experimental variograms. For this purpose, we generally detrend the influence of

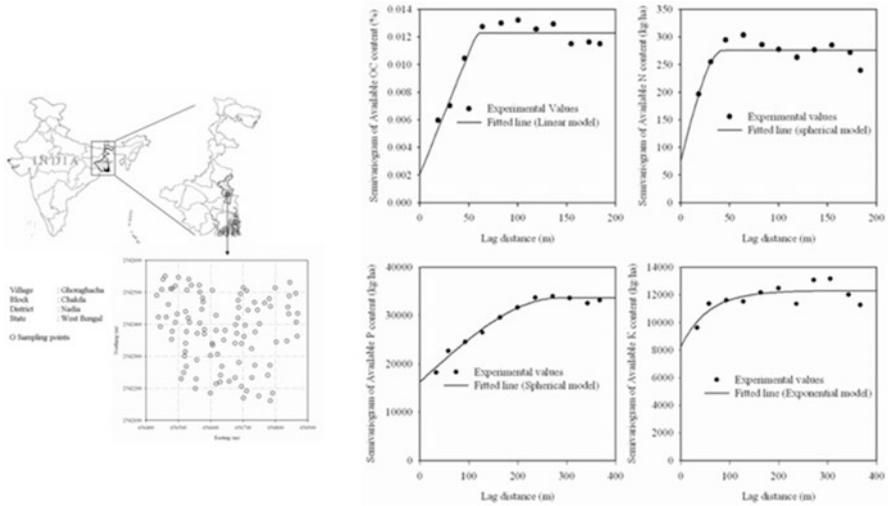


Fig. 9.5 Semivariogram of SOC and major soil nutrient contents (N, P, and K) in an intensively cultivated village in West Bengal, India. (Adopted, Chatterjee et al. 2015)

direction (*x*- and *y*-direction), environmental covariates, and other earth features on soil properties. After detrending, the residual value is again fitted in standard models.

Experimental and fitted semivariogram of soil organic carbon (SOC) and major nutrient (N, P, and K) contents in an intensively cultivated village at Gayeshpur, West Bengal, is presented in Fig. 9.5 (Chatterjee et al. 2015). From these semivariogram structures, it is observed that spatial variation pattern is different for different soil properties. Spherical semivariogram model was found best fitted for N and P content, whereas the linear model is best fitted for SOC content, and the exponential model is best fitted for K content. If we look at the range of spatial variation, it is higher in P content and less in N content. It indicates that spatial variation of P content shows more spatial continuity than other soil properties in this case. The variation of N content is highly random since it is highly influenced by external inputs of nitrogenous fertilizer. These semivariograms also show that how much the proportion of total sill is contributed by the nugget component. The more is the nugget component, the less is the spatial variation component and the more is the randomness. In a pure nugget model, the total variation is contributed by nugget and there is no spatial component.

9.5.2 Kriging

Once the semivariogram parameters are identified, it is possible to estimate the soil property at unsampled location through kriging approach, and ordinary kriging (OK) is mostly followed for this purpose. In OK, the unbiased estimates of soil

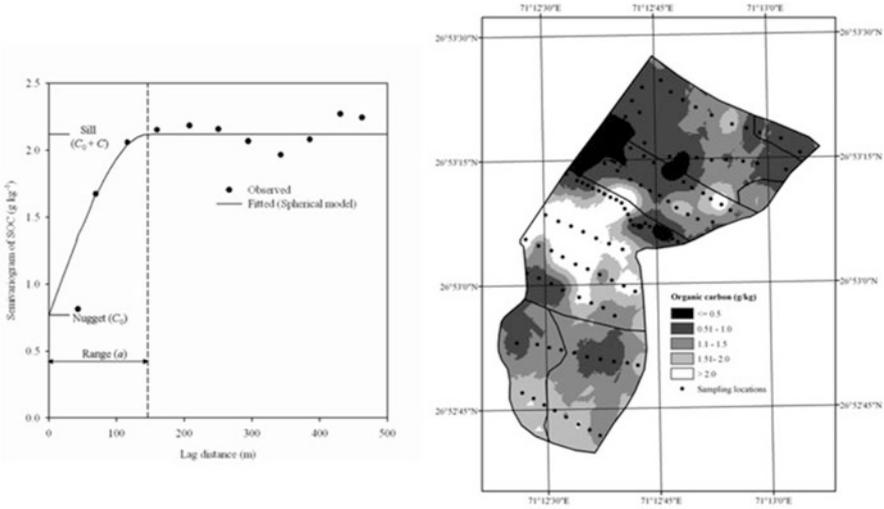


Fig. 9.6 Semivariogram and ordinary kriging map of SOC content in a farmer’s field at Jaisalmer. (Adopted, Santra et al. 2012a, b)

properties at unsampled locations, $\hat{z}(u)$, are computed through weighted linear combinations of measured soil attributes at neighbor points $z(u_\alpha)$ located within a neighborhood centered around u :

$$\hat{z}(u) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha z(u_\alpha) \tag{9.12}$$

where λ_α is the weight assigned to the measured data points $z(u_\alpha)$ located within a given neighborhood, $W(u)$ centered on u . Weights for n number of neighbor points are chosen in such a way so that error variance, $\sigma_E^2(u) = Var\{z * (u) - z(u)\}$, is minimized under the constraint of no bias of the estimator. Figure 9.6 shows a map of soil organic content in a farmer’s field at Jaisalmer, Rajasthan, prepared through OK approach (Santra et al. 2012a). From Fig. 9.6, it may be noted that the range of SOC content is around 150 m, which indicates that soil sampling locations that are apart by 150 m or less are spatially correlated with each other beyond which it shows a random pattern. In the case of a random pattern, which is generally observed for the pure nugget model, the arithmetic average of all sampling points could be a simple approach to obtain an estimate. However, still we prefer nugget model semivariogram because in this case, we get an estimate at an unsampled location along with error variance, which helps to judge the reliability of the estimate. Later on, we will discuss the accuracy and uncertainty issue of a digital soil map. Such a digital soil map may be quite helpful for the management of organic manure applications in a farmer’s field.

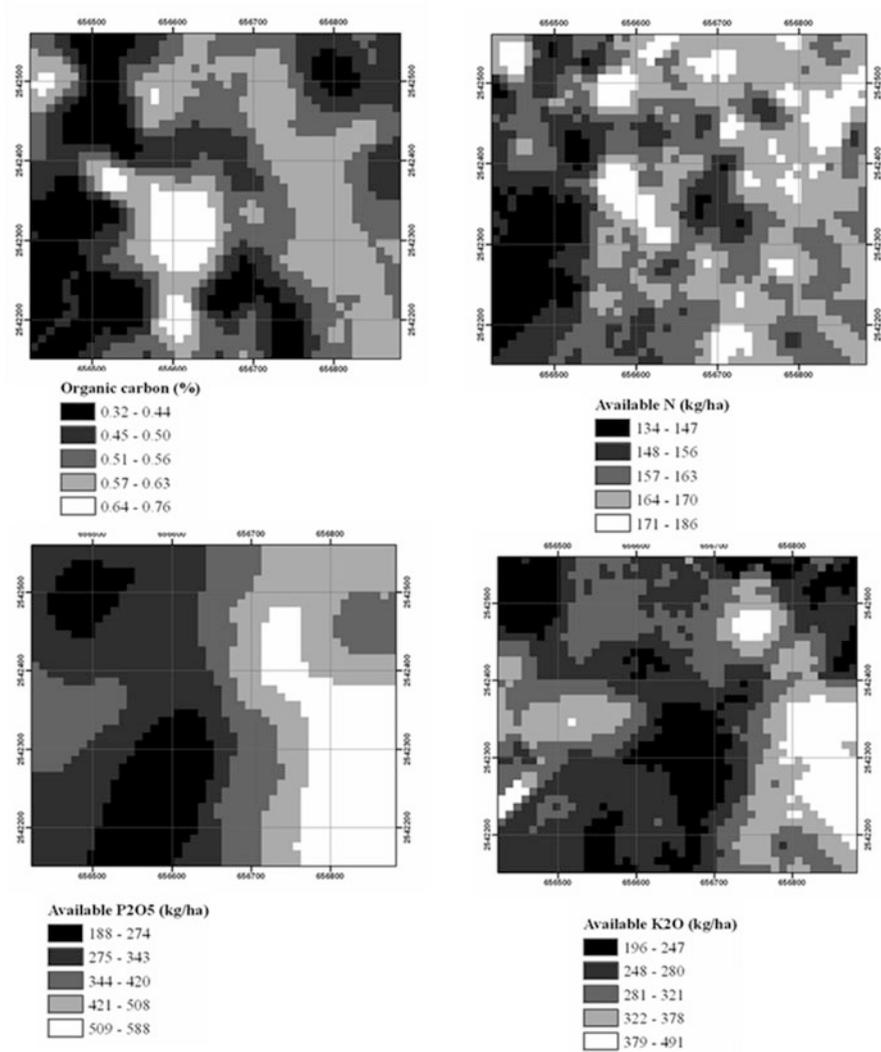


Fig. 9.7 Ordinary kriging map of SOC and major soil nutrient (N, P, and K) content in an intensively cultivated village in West Bengal, India. (Adopted, Chatterjee et al. 2015)

Apart from the SOC content map, digital maps of soil nutrient content prepared through the OK approach are depicted in Fig. 9.7. The spatial continuity in P content as we observed in the semivariogram of this property (Fig. 9.5) is also clearly visible on the map. The patchy variation in N content is quite understood from the short range in semivariogram. These maps of soil nutrient content will be quite helpful for nutrient management in the agricultural field. However, it is quite difficult to obtain a good spatial variation structure of nutrient content since it is largely influenced by

the application of fertilizer. Therefore, it is suggested to include the past history of fertilizer doses applied at different locations in the field to detrend the influence of external fertilizer application and then modeling the spatial variation using the standard semivariogram.

9.5.3 Co-kriging

Sometimes, the target soil property, which we want to estimate spatially, is very costly and time-consuming to measure at multiple locations in the field. In those cases, we use the information about surrogate soil properties, which have an influence on target soil properties. Co-kriging may be a suitable solution under such a situation. For applying co-kriging, the data on co-variables may be available at the same locations where the measured value of the target variable is available (co-located points) or may be available at other locations or both. Generally, co-kriging is most appropriate if the co-variables can be measured cheaply and therefore a denser sampling of co-variables than of target variable can be done. A detailed description of the co-kriging method can be found in Webster and Oliver (2007) and Rossiter (2018). Here, the co-kriging system is mentioned in brief to understand the theory behind it. Co-kriging is an extension of the theory of single regionalized variable used for OK. Similar to semivariogram, the cross-semivariogram between the target variable and co-variable is calculated as follows:

$$\hat{\gamma}_{uv}(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} \{z_u(x_i) - z_u(x_i + h)\} \{z_v(x_i) - z_v(x_i + h)\} \quad (9.13)$$

where $\hat{\gamma}_{uv}(h)$ is the cross-semivariogram between target variable, z_u , and co-variable, z_v , and $m(h)$ is the number of data pairs of target variable and co-variable with a lag distance of h . The co-kriging system estimates z_u at unknown location x_0 with the following expression:

$$z_u(x_0) = \sum_{l=1}^V \sum_{i=1}^{n_l} \lambda_{il} z_l(x_i) \quad (9.14)$$

where V is the number of variables and among these one is target variable and λ_{il} is the weight, which is assigned in such a way that

$$\sum_{i=1}^{n_l} \lambda_{il} = \begin{cases} 1 & l = u \\ 0 & l \neq u \end{cases} \quad (9.15)$$

An example of cross-semivariogram of SOC with pH and EC is depicted in Fig. 9.8. It is to be noted here that this cross-semivariogram is calculated because of

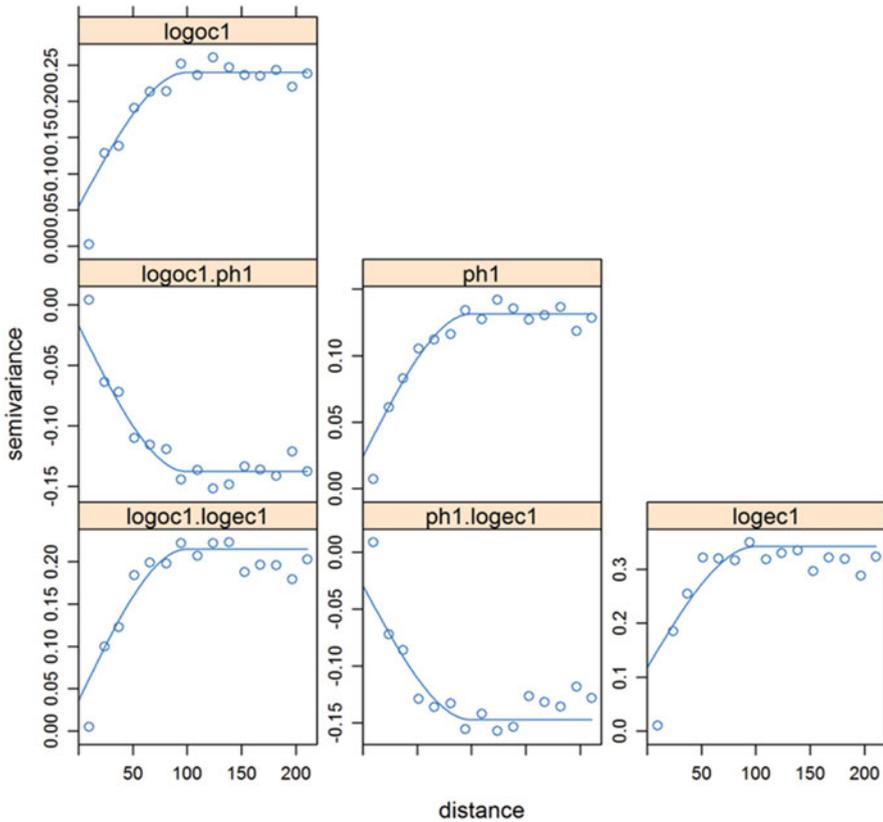


Fig. 9.8 Cross-variograms of SOC contents, pH, and electrical conductivity (EC)

significant correlation of SOC with pH and EC. From the cross-semivariogram, it is observed that range of target variable (OC) is almost similar with the co-variables (pH and EC) and this should be a major criterion to apply co-kriging.

If OK is compared with co-kriging, we generally observe that the prediction performance is improved in co-kriging. However, the uncertainty of prediction is higher in the case of co-kriging since multiple numbers of variables are employed in the prediction process. For example, the performance of OK and co-kriging in the prediction of SOC content in the above example is presented in Table 9.1. Mean error (ME) and root mean squared residual (RMSR) of predicted SOC are slightly lower in co-kriging than OK, and the performance is best when both the co-variables (pH and EC) are used in the co-kriging process. Such superior performance of co-kriging was also reported in literatures, e.g., Ersahin (2001), Carter et al. (2011), etc. However, the value of mean squared deviation ratio (MSDR), which generally quantifies the uncertainty, deviates from its desirable value of 1 in case of co-kriging. This suggests that co-kriging improves the prediction of the target variable but along with it also increases the uncertainty in predicted values;

Table 9.1 Cross-validation performances of ordinary kriging and co-kriging

Kriging method	Target variable	Covariates	Soil layer (cm)	ME	RMSR	MSDR
Ordinary kriging	Log[SOC]	-	0-15	0.0085	0.37	1.04
			15-30	0.0053	0.35	0.95
Co-kriging	Log[SOC]	pH	0-15	- 0.0018	0.28	1.13
			15-30	0.0005	0.31	1.07
	Log[SOC]	Log [EC]	0-15	0.0018	0.30	1.12
			15-30	- 0.0002	0.30	1.11
	Log [SOC (%)]	pH and Log [EC]	0-15	- 0.0005	0.26	1.15
			15-30	- 0.0006	0.28	1.14

therefore, it is strictly followed if only the target variable is very costly and time-consuming to measure in the field. Otherwise, it is always advisable to follow OK.

9.5.4 Other Variants of Kriging

The OK approach is mostly followed for providing a spatial estimate of soil property. When the OK approach provides the estimate for a point location, it is called punctual kriging. Otherwise, if the OK approach provides the estimate over block support, then it is called block kriging. Other than OK, simple kriging is also sometimes followed where the mean value of the target variable is known. Apart from OK and co-kriging, there are several variants of kriging approaches, e.g., regression kriging (RK), universal kriging (UK), kriging with external drifts (KED), probability kriging (PK), indicator kriging (IK), lognormal kriging (LK), etc. Regression kriging is followed if there is a presence of an external trend on data. Under such cases trend is predicted through regression model, and regression residual is predicted through OK, which is finally added to obtain RK prediction. RK is often confused with UK and KED since all these three kriging approaches model the trend (drift) in the data. However, there is little difference between RK and (UK and KED). In the case of RK, the trend and residuals are predicted separately, whereas in the case of UK and KED, trend and residual predictions are made simultaneously within the kriging system. In the case of UK, the trend of spatial coordinates is only considered, whereas in the case of KED, the trend of an external variable is modeled. The PK is able to provide an estimate with a probability to be near to a predefined threshold value and thus is often used to assess the risk associated with a target variable. The IK is a nonparametric and nonlinear approach of kriging where target variables are converted to a binary variable (indicator). In the case of LK, the target variable is first transformed logarithmically to fit it in a normal

distribution, and then OK is applied on the log-transformed variable. However, to understand better the prediction, it is to be back-transformed. The back-transformation of predicted log-transformed values needs to be done carefully following the standard procedure; otherwise, it will lead to wrong interpretation (Webster and Oliver 2007). Details of all these kriging approaches are available in Webster and Oliver (2007) and Santra et al. (2017b, c).

9.6 State-Factor (Clorpt) Approach of DSM

In the state-factor approach, statistical models are built between target soil property and the “clorpt” factors. The information on “clorpt” factor is now abundantly available in digital platforms, which are often called as covariates. Apart from the availability of data on covariates, several statistical and mathematical tools have been evolved in recent times, which have the capability to handle a huge amount of fine-resolution data on covariates and also are able to build model both linear and nonlinear relationship. Therefore, the state-factor approach of DSM methodologies has now been preferred over the other two approaches. In the following, we describe the data on covariates and the machine learning tools that are available to apply the DSM methodology.

9.6.1 Covariates on Terrain Attributes

Maps of terrain attributes provide information on the relief factor of the “clorpt” approach. Different terrain attributes can be calculated using the digital elevation model (DEM) of an area. Hydrology and spatial analysis tools of GIS software, e.g., ArcGIS, QGIS, SAGA, etc., may be used to determine these terrain attributes. For the processing of DEM of a targeted study area, the data acquired through Shuttle Radar Topography Mission (SRTM) with a spatial resolution of 90 m may be used which is available at <http://srtm.csi.cgiar.org/> (Rabus et al. 2003). In Fig. 9.9, examples of terrain attributes determined from SRTM DEM of arid western India are presented. These terrain attributes are altitude, slope, elevation, above channel network, hillshade, profile curvature, plan curvature, terrain ruggedness index (TRI), and topographic wetness index (TWI).

A detailed description of such terrain properties may be found in Santra et al. (2017a). All these derived terrain attributes have significant relation with sand content in arid western India as observed through stepwise regression analysis.

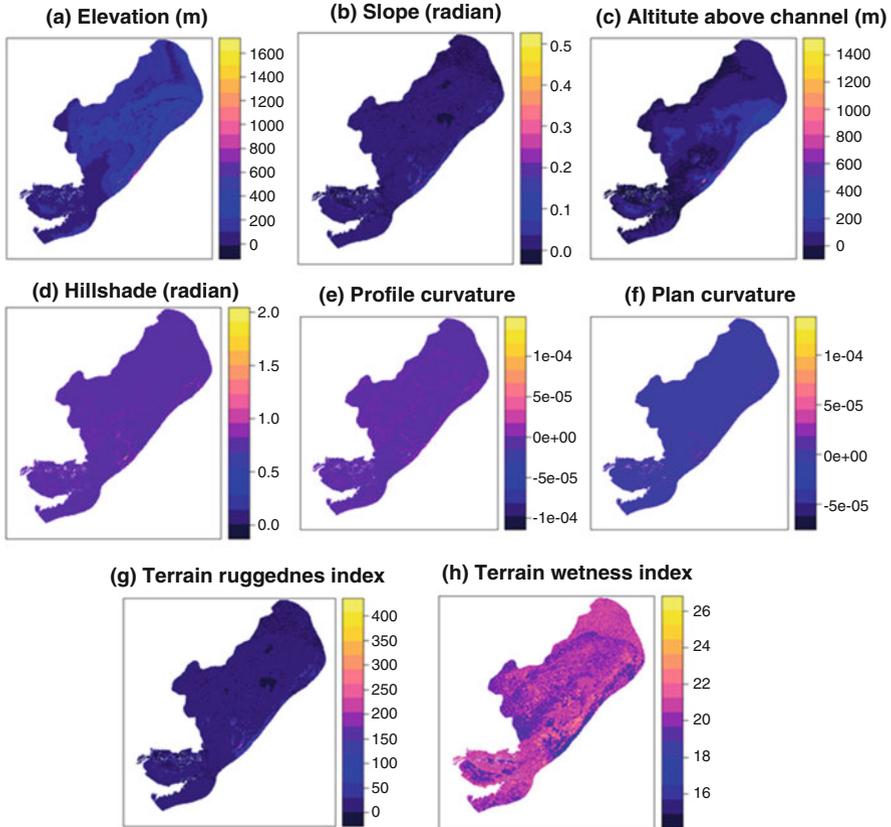


Fig. 9.9 Covariate maps of terrain attributes in arid western India. (Adopted, Santra et al. 2017a)

9.6.2 Covariates on Bioclimatic Variables

Bioclimatic variables provide information on the climate factor of the “clorpt” approach. The increasing availability of these bioclimatic variables in digital platforms makes it easy to apply these covariate data in the DSM approach. The raster data (30-second resolution) on bioclimatic variables can be downloaded from <http://worldclim.org/current> for its use in DSM. Hijmans et al. (2005) presented a detailed description of such bioclimatic variables. Examples of bioclimatic variables for arid western India, which were used in DSM of sand content by Santra et al. (2017a), are presented in Fig. 9.10.

These bioclimatic variables are annual mean temperature and precipitation, seasonality of temperature and precipitation, annual range of temperature, mean diurnal range of temperature, and precipitation during the wettest quarter of the year.

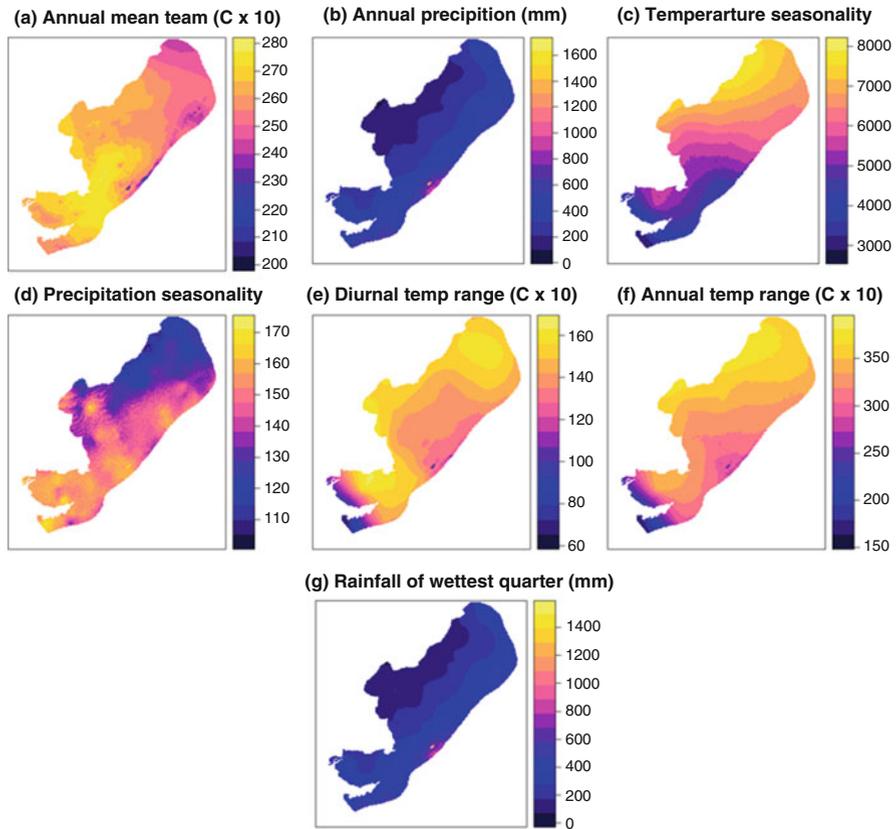


Fig. 9.10 Covariate maps of bioclimatic variables in arid western India. (Adopted, Santra et al. 2017a)

9.6.3 Machine Learning Algorithms in DSM

With the advancement of machine learning algorithms as a field of artificial intelligence, there is a possibility to build a model describing the relationship between soil parameters and the covariates affecting soil formation process (Muras 2000; Banerjee et al. 2018; Jha et al. 2019). The machine learning tools apply data mining techniques to identify the statistical relationship and then build the model. Different machine learning tools are now available to identify the relationship between soil properties and covariates. Few common machine learning tools are multiple linear regression (MLR), support vector machine/regression (SVM/SVR), random forest regression (RF), artificial neural network (ANN), k-nearest neighborhood (k-NN), cubist, etc. The machine learning tools are becoming popular since it requires less intervention of human brain and also learns input-output relationship in a better way. Increasing accessibility of high-level computer programming language, e.g., R,

Python, etc., makes it much easier to apply machine learning tools in DSM. Few commonly used machine learning tools in DSM are discussed below. However, a detailed description of these machine learning tools is available in Khaledian and Miller (2019).

The MLR approach builds the linear regression equations between soil properties and multiple covariates. It is the most simple and popular approach to machine learning tools. The coefficients of the regression equations are called the model parameters. A basic requirement of the MLR approach is that the covariates should not be correlated with each other, i.e., absence of collinearity in covariates. For this purpose, often stepwise regression analysis is carried out before building the MLR model to remove the collinearity in data. Several efforts have been made in the past to apply the MLR approach in DSM, e.g., Angelini et al. (2017), and are still widely used.

SVM/SVR is an ML algorithm that has gained popularity in recent times. The procedures adopted in SVR are complex in nature to understand; however, the outputs are very close to the real field situation. An SVM actually constructs a hyperplane or a set of hyperplanes in a high- or infinite-dimensional space, which are used for regression models. In the SVR approach, a margin of tolerance is defined for covariates using the observations (support vectors), and then data are separated and fitted linearly. The margin is actually the distance from the decision surface, which is maximally far from any observation. This decision surface ensures the high generalization ability of the algorithm and thus makes the results more applicable to the unseen data. In addition, the SVR approach applies kernel functions to map nonlinear vectors to a very high-dimensional space for solving nonlinear problems. The SVR algorithm requires the user to set the number of support vectors and the fraction of support vectors needed to maximize the margin, which is also called the hyperparameter of the algorithm. Application of SVR in the classification of soil types and estimation of soil properties may be found in Kovačević et al. (2010).

RF regression is a ML approach, which consists of an ensemble of randomized classification and regression trees (CART) (Breiman 2001). Predictions through RF regressions are made by generating numerous trees within the algorithm and finally aggregating them using the average of the individual tree outputs. There are three user-defined parameters on which RF regression is dependent, and these are the number of trees in the forest, the minimum number of data points in each terminal node, and the number of features tried at each node (mtry). A detailed description of the use of RF regression in DSM is available in Grimm et al. (2008). Here an example of preparing a sand content map by applying the RF regression algorithm is presented in Fig. 9.11. Covariates used in this example are soil category map, terrain attributes, and bioclimatic variables. A major advantage of RF regression-based digital soil maps is that the predicted data are available in the resolution of covariate maps. Therefore, the use of fine-resolution covariate maps results in digital soil maps with detailed information. Hence, the RF regression-based DSM is most suitable in case of sparsely available soil data, where it is difficult to build semivariogram models from limited data.

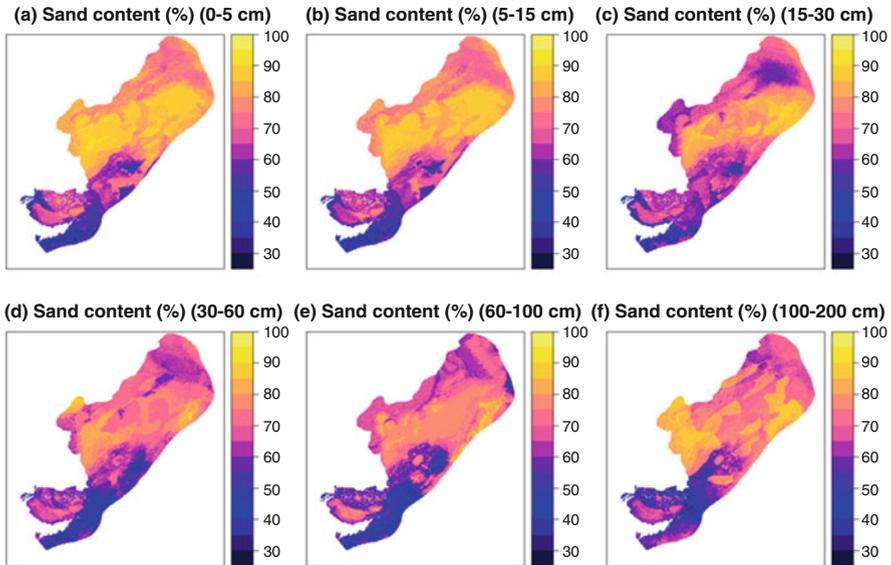


Fig. 9.11 Spatial maps of sand content in arid western India developed through random forest regression approach

ANNs are composed of artificial neurons that mimic biological neurons, which receive input, combine the input with their internal state, and produce output using an output function. The neurons are typically organized into multiple layers. Neurons of one layer are connected to neurons of nearby layers, which may be immediately preceding layer and immediately the following layer. The neuron layer that receives external data is called the input layer, whereas the layer that produces the output is called the output layer. In between the input and output layer, there may be hidden layers. Major hyperparameters of ANN are learning rate, the number of hidden layers, and batch size. Learning from the input data through ANN is done by adjusting the weights of the network so that the accuracy of the output is highest. The hyperparameter learning rate is defined as the number of the corrective steps to adjust for errors in each observation. A high learning rate shortens the training time, but with lower ultimate accuracy, whereas a lower learning rate takes longer time, but may lead to greater accuracy. Because of their ability to reproduce and model nonlinear processes, ANN has found applications in many disciplines. Details on the procedures of ANN application in DSM methodology may be found in Behrens et al. (2005).

k-NN algorithm applies a nonparametric method to provide an output based on the similarity concept, which assumes that similar things exist together in proximity. An estimate of soil property at an unknown location is obtained by averaging the values at k-nearest neighbors. Weights are assigned to each neighbor based on the distance; the higher is the distance of neighbor the lesser is the weight. The distance

metric is commonly calculated as Euclidean distance. Other distance metrics, e.g., Mahalanobis distance, Manhattan distance, Hamming distance, etc., are also used. The k-NN is an instance-based learning where the regression functions are approximated locally, and therefore a variety of regression curves are calculated based on the neighbors. The parameter k needs to identify optimally to obtain the best estimate of the target variable. An example of k-NN application in DSM may be found in Mansuy et al. (2014).

The cubist is a rule-based algorithm that is an extension of Quinlan's M5 model tree. Cubist generates a tree structure from a pool of covariates. The tree breaks through intermediate nodes to several final nodes using rules. A prediction is made using the linear regression model at the terminal node of the tree but is "smoothed" by taking into account the prediction from the linear model in the previous node of the tree. Besides, cubist as an ensemble model adds boosting to improve the prediction performance using two hyperparameters (i.e., committees and instances). Through the committee parameter, iterative model trees are created in sequence, and final prediction is obtained by simple averaging of the predictions from each model tree. The instance parameter adjusts the predictions from rule-based models (whether it is with a committee or without committee) using nearest neighbors. Thus, ensemble learning combines models produced by multiple repetitions of the same algorithm. This strategy usually obtains stronger predictive performance than results produced from any of the models individually. The application of the cubist model in DSM may be found in Akpa et al. (2016).

9.6.4 Application of Hyperspectral and Remote Sensing Signature in DSM

Quantifying the spectral reflectance of soil visible, near-infrared, and shortwave infrared (VIS-NIR-SWIR) region (350 to 2500 nm) and then relating it with soil properties has emerged as a rapid and noninvasive technique for estimation of soil properties (Ben-Dor et al. 2009). Hyperspectral signature of soil in 350–2500 nm region has been successfully used for estimating soil properties. A brief review of such applications of hyperspectral signature in estimating soil properties is available in Das et al. (2015) and Santra et al. (2015).

Figure 9.12 represents typical hyperspectral signatures in the VNIR region (350 to 2500 nm) for few arid soils of India. From the spectral curves, a wide range in spectral signatures is quite visible. The vibrational absorbance of the soil reflectance spectra is because of presence of various functional groups, namely, –OH in minerals and –OH, –CH, and –NH in soil organic matter (Rossel and McBratney 1998; Reeves et al. 1999). The soil reflectance spectra predominantly shows three absorption peaks at 1400, 1900, and 2200 nm as presented in Fig. 9.12. The absorption peaks at 1400 and 1900 nm correspond to water absorption (Leone and Sommer 2000), and 2200 nm indicates metal-hydroxyl stretching because of clay mineral (Chabrilat et al. 2002). The absorption features at 870 and 1000 nm and

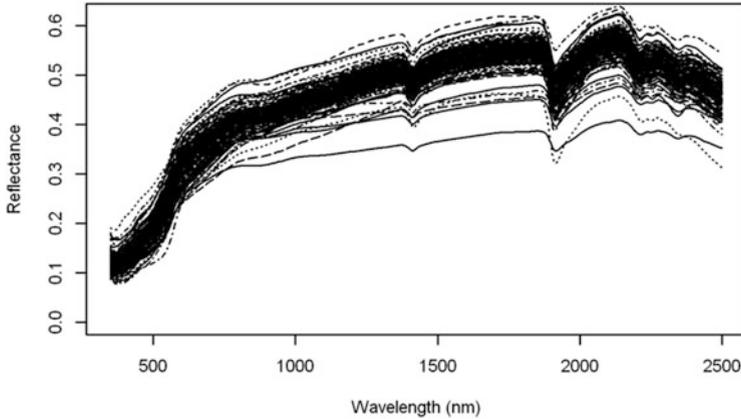


Fig. 9.12 Hyperspectral signature of soils from hot arid region of India

between 2200 and 2500 nm are mainly due to the presence of iron oxides and carbonates, respectively (Clark 1999; Chang and Laird 2002).

Soil properties and parameters influencing the reflectance at specific bands can be identified through the spectral data modeling approach. Band reflectance from these spectral data can be calculated to relate to soil properties. While calculating the band reflectance, it may also be kept in mind the bandwidth corresponds to the available spectral bands in operational or futuristic remote sensing satellites so that the algorithm can be translated to remote sensing platforms in the future. The overall brightness of spectra and the slope of the spectra at red to the near-infrared region can also be used to relate with soil properties. Spectral absorption features at specific wavelength region can also be analyzed in detail to capture the variation in spectral features so as to relate with content of a specific material in soil, which causes the absorption feature. Use of spectral signatures to estimate soil properties have been tried by different researchers throughout the world (Das et al. 2015). Here, an example of few spectral algorithms is presented in Table 9.2 from Santra et al. (2015).

Laboratory-based algorithms developed using relationship between soil properties and proximally measured spectral reflectance can be translated to the remote sensing platforms. However it depends on various factors like spectral resolution, spectral and spatial resolution, consistency of satellite images, atmospheric degradation of spectral behavior, land surface composition, soil moisture content, roughness of the surface, presence of gravels on surface, etc. An example of such demonstration is shown in Fig. 9.13, where sand content is estimated using the band reflectance of Landsat-8 data (path, 142; row, 49). The Landsat-8 data that corresponds to 19th of June 2013 was downloaded from the earth explorer website (<http://earthexplorer.usgs.gov/>). Finally, the reflectance-based models as shown in Table 9.2 were used to convert Landsat-8 data to map of sand content.

Table 9.2 Spectral algorithms soil properties estimation using principal components of soil reflectance spectra in VNIR-SWIR region, using Resourcesat-1 and Landsat-8 OLI band reflectance

Model type	Model equation	R ²
PCs of hyperspectral soil reflectance -based model	$OC = 0.192 + -0.0008 \times PC1 + 0.002 \times PC2 + 0.002 \times PC3$	0.12
	$Sand = 90.15 + 0.025 \times PC1 - 0.537 \times PC3$	0.41
	$Silt = 4.46 + 0.284 \times PC3$	0.27
	$Clay = 5.40 - 0.017 \times PC1 + 0.252 \times PC3$	0.43
^a Derived IRS- P6 band reflectance-based model	$OC = 1.11 + 3.82 \times B2 - 5.64 \times B3$	0.27
	$Sand = 66.3 - 304.5 \times B2 + 605.7 \times B3 - 366.3 \times B4 + 88.1 \times B5$	0.20
	$Silt = 11.53 + 157.52 \times B2 - 264.82 \times B3 + 102.10 \times B4$	0.17
	$Clay = 18.19 + 109.65 \times B2 - 255.16 \times B3 + 175.8 \times B4 - 49.42 \times B5$	0.16
^b Derived Landsat-8 OLI band reflectance-based model	$OC = 1.12 + 3.72 \times \text{Band } 3 - 5.56 \times \text{Band } 4$	0.27
	$Sand = 52.8 - 168.5 \times \text{Band } 3 + 316.1 \times \text{Band } 4 - 129.1 \times \text{Band } 5 - 434.9 \times \text{Band } 6 + 480.5 \times \text{Band } 7$	0.44
	$Silt = 22.54 + 102.21 \times \text{Band } 3 - 147.35 \times \text{Band } 4 + 266.63 \times \text{Band } 6 - 253.86 \times \text{Band } 7$	0.32
	$Clay = 23.67 - 45.91 \times \text{Band } 4 + 251.89 \times \text{Band } 6 - 252.77 \times \text{Band } 7$	0.44

Adopted, Santra et al. (2015)

^aDerived band reflectance corresponding to IRS-P6 bands of LISS-III, LISS-IV, and AWiFS camera: B2 = 520–590 nm, B3 = 620–680 nm, B4 = 770–860 nm, B5 = 1550–1700 nm

^bDerived band reflectance to Landsat-8 OLI bands: Band 3 = 530–590 nm, Band 4 = 640–670 nm, Band 5 = 850–880 nm, Band 6 = 1570–1650 nm, Band 7 = 2110–2290 nm

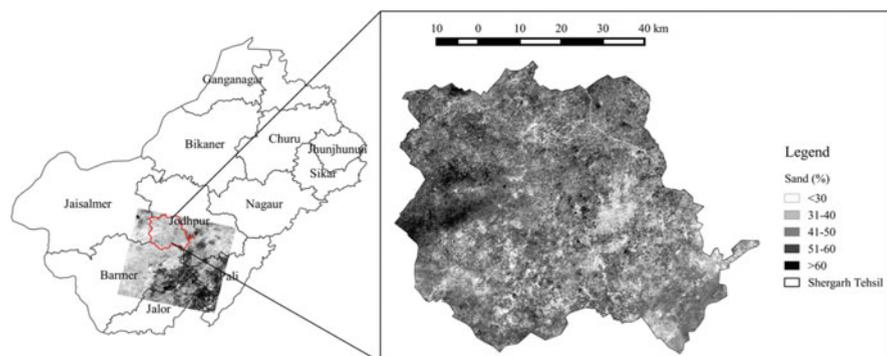


Fig. 9.13 Estimated sand content map of Shergarh Tehsil, Jodhpur, using spectral algorithm and Landsat data (OLI band)

9.7 Pedotransfer Function (PTF) Approach for Digital Soil Mapping

PTFs are models which help to estimate complex and difficult to measure soil properties using basic soil properties as input. Soil water retention behavior relating soil moisture content (θ) with pressure head (h) is generally tedious to measure at multiple locations in field and therefore is often estimated through PTF models (Santra and Das 2008; Santra et al. 2018). Soil physicochemical properties, e.g., cation exchange capacity (CEC), and soil thermal properties, e.g., specific heat capacity, conductivity, etc., have also been used as a target variable in PTF model. These established PTF models can be used to convert digital soil maps of basic soil properties to maps of complex soil properties. In the following, few examples on converting maps of basic soil properties to soil water retention behavior are given.

In the first example, spatial maps on water content at field capacity (FC) (θ_{FC}) and permanent wilting point (PWP) (θ_{PWP}) were prepared through linking soil maps on basic properties and PTFs (Santra et al. 2008). The PTFs for θ_{FC} and θ_{PWP} used in this example were developed from the available soil data in benchmark soils of India, and these PTF models are given below:

$$\theta_{FC}(\%, w/w) = 21.931 - 0.20564 \times \text{sand} + 0.175 \times \text{clay} + 4.6737 \times \text{OC} (R^2 = 0.89) \quad (9.16)$$

$$\theta_{PWP}(\%, w/w) = 8.7255 - 0.092946 \times \text{sand} + 0.15944 \times \text{clay} (R^2 = 0.78) \quad (9.17)$$

where sand is the % sand content (0.05–2 mm), clay is the % clay content (<0.002 mm), and OC is the % OC content in the soils. Using OK approach, maps of sand content, clay content, and OC content were first prepared, and then these three maps were joined together using above mentioned PTF models. The developed maps of θ_{FC} and θ_{PWP} are presented in Fig. 9.14. Another possible way to generate these maps is to predict θ_{FC} and θ_{PWP} at each location, where basic soil properties were measured and then OK is applied on estimated θ_{FC} and θ_{PWP} to generate the final maps. In the first approach, the error of spatial prediction associated with each map of basic soil property and the error of PTF model will be added on to the final map of soil water retention behavior. Therefore, the reliability of final map highly depends on the accuracy of spatial prediction methods as well as on the accuracy of PTF models. In the second approach, the limitation is to obtain good spatial trend to apply geostatistical methods for preparation of maps of complex soil properties which therefore are not commonly used. Moreover, in the first approach, we obtain the digital map of basic soil properties along with target map of complex soil properties, which together help in several land management decisions. These maps of soil water retention at farm level may help in applying right amount of irrigation water at right time.

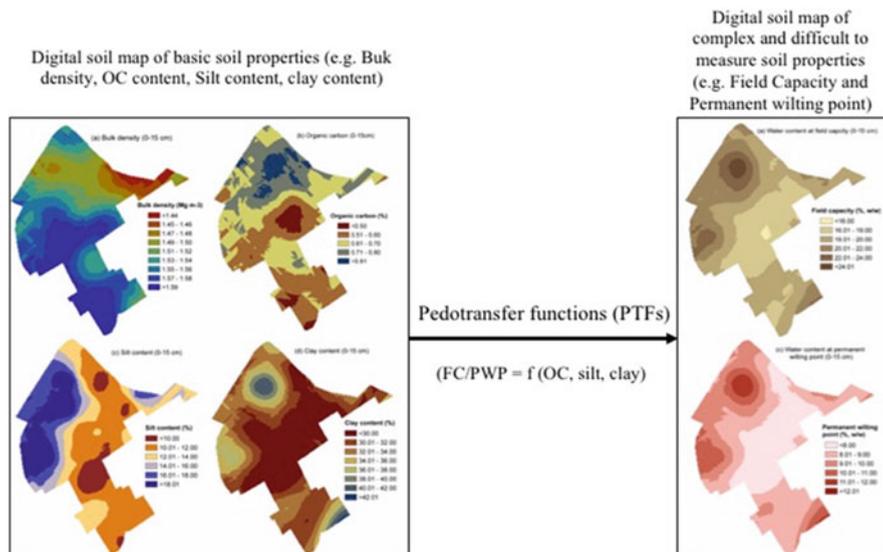


Fig. 9.14 Translating digital soil maps of basic soil properties to map of soil water retention using PTF model at experimental farm of ICAR-Indian Agricultural Research Institute, New Delhi. (Adopted, Santra et al. 2008)

Another example of converting maps of basic soil properties to maps of soil water retention at field capacity and the permanent wilting point is presented in Fig. 9.15. Here, maps of sand, silt, and clay content are converted to maps of FC and PWP content in the hot arid ecosystem of India. The PTFs used in this example are regression-based PTF models and are available in Santra et al. (2018). The developed maps of FC and PWP may be quite useful for the sustainable utilization of water resources in arid western India (AWI). From these maps, it is noted that soil water retention at FC was lowest ($\sim 10\%$) at the western part of the AWI, where sand dunes are dominant. The value of $\theta_{1/3\text{bar}}$ was around 25% for soils at coastal deltaic plain lying at the southern part of the AWI. The central and northern part of AWI, which covers a major portion of the region, has $\theta_{1/3\text{bar}}$ of 15%. Similarly, soil water retention at PWP was also very low ($\sim 4\text{--}6\%$) at the western and northern plain of the AWI and high ($\sim 8\text{--}10\%$) at the southern coastal plain of the AWI. Soil water retention at FC reaches 2–3 days after saturation, whereas to reach PWP it may require a long time to dry at which plants start to wilt. The amount of soil water available between these two critical soil moisture contents is called available water capacity (AWC), which is extracted by the plant for its growth and development. From the surface maps of $\theta_{1/3\text{bar}}$ and $\theta_{15\text{bar}}$, it is found that the plant available water content is about 6–9% for western, central, and northern part of the AWI. Therefore, growing crops with high water requirement in this region may not be feasible since it will require frequent application of irrigation water to maintain sufficient soil moisture regime for plant growth. In such situation, surface map of $\theta_{1/3\text{bar}}$ and

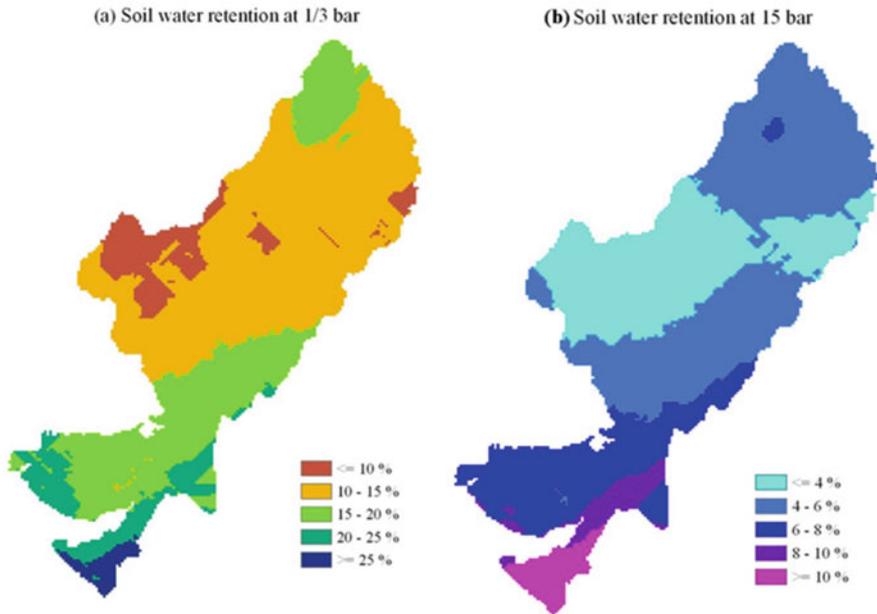


Fig. 9.15 Estimated surface map of soil water retention within the arid western India (AWI) for (a) 1/3bar ($\theta_{1/3\text{bar}}$) and 15 bar ($\theta_{15\text{bar}}$) using PTF model

$\theta_{15\text{bar}}$ may help the end users for judicious use of water, which is very scarce in the region.

9.8 Accuracy and Uncertainty Analysis of Digital Soil Maps

Accuracy and uncertainty of digital soil maps play a key role in the reliability of digital products. Accuracy is generally defined as how close is the estimated value to the true value. The more close is the estimated value toward the measured value, the higher will be the accuracy. It is generally calculated as an error, which is the difference in observed and predicted value. Several error indices are used to quantify the magnitude and distribution of error. Uncertainty indicates the fluctuations of the estimated value from its mean. Otherwise, it can also be quantified as a confidence interval. The narrow is the range of confidence interval, the less is the uncertainty and vice versa. The uncertainty of digital soil products is often neglected. Because in most of the classical spatial prediction approaches, the error variance of the predicted values at a particular location is not calculated, rather a single predicted value is obtained in most cases. However, in geostatistical approaches, kriging variance of prediction is always calculated along with the mean predicted values. Therefore, confidence interval may be calculated either at 90% significance level ($\mu \pm$

1.645×σ) or 95% significance level (μ±1.96×σ). Otherwise, repetitive stochastic simulations, e.g., sequential Gaussian simulation, are carried on a particular location to obtain the mean and standard deviation of predicted values, which helps to quantify the uncertainty.

In DSM approaches, accuracy is generally quantified through cross-validation approach. k-fold cross validation is generally followed. In this approach, the total dataset is randomly divided into k sets of data. Then the (k-1) sets of data are used as training data for building the model, and then the developed model is tested on kth fold dataset as validation data. The procedure is repeated till each set of data gets a chance to appear as validation data once in the total procedure. The k-fold cross-validation approach results into observed and predicted values of soil property at each measured location. These observed and predicted values are then used to calculate different cross-validation indices, few of which are given below:

$$r = \frac{\sum_{i=1}^n [z(s_i) - \bar{z}_{obs}] [\hat{z}(s_i) - \bar{z}_{pred}]}{\sqrt{\sum_{i=1}^n [z(s_i) - \bar{z}_{obs}]^2} \sqrt{\sum_{i=1}^n [\hat{z}(s_i) - \bar{z}_{pred}]^2}} \tag{9.18}$$

$$LCCC = \frac{2\rho\sigma_{obs}\sigma_{pred}}{(\bar{z}_{obs} - \bar{z}_{pred}) + \sigma_{obs}^2 + \sigma_{pred}^2} \tag{9.19}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [Z(s_i) - \hat{Z}(s_i)]^2} \tag{9.20}$$

$$bias = \frac{1}{n} \sum_{i=1}^n [Z(s_i) - \hat{Z}(s_i)] \tag{9.21}$$

where $z(s_i)$ is the measured values of the variable at the location s_i , $\hat{Z}(s_i)$ is the predicted values with variance σ^2 at the location s_i , and n is the number of sampling sites.

The R^2 indicates the precision of prediction, and it is actually measured as square of the Pearson correlation coefficient (r) between observed and predicted values. Both accuracy and precision of the prediction are evaluated by Lin’s concordance correlation coefficient (LCCC) (Lin 1989). LCCC is calculated as the orthogonal distance of values from the 1:1 line of observed vs predicted values and it ranges from -1 to +1.

A zero LCCC value indicates no agreement between measured and predicted values. However, values equal to 1 and -1 indicate perfect positive and negative agreement, respectively. The accuracy of the prediction can be measured using RMSE statistics. The larger RMSE value shows less prediction accuracy. Similarly the mean error of prediction can be estimated using bias, and a value of zero indicates unbiasedness of the prediction.

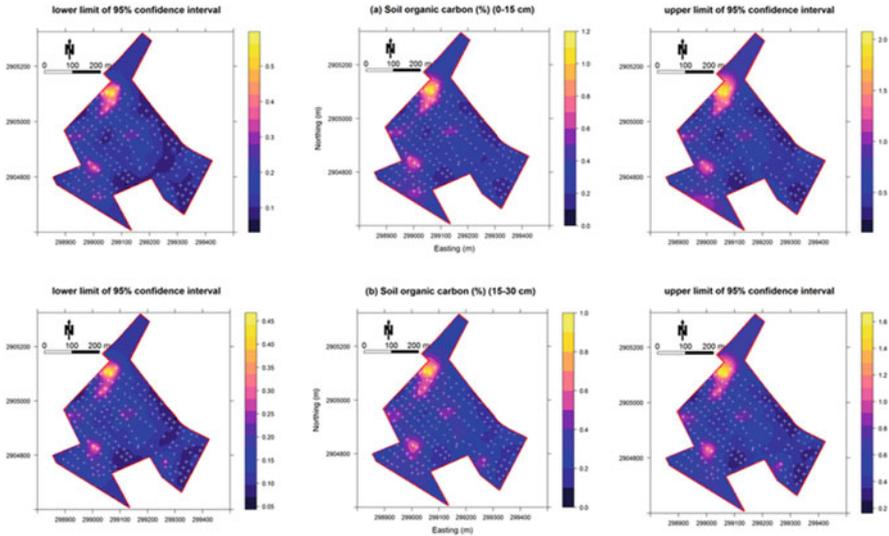


Fig. 9.16 Uncertainty of digital soil map of SOC content in horticultural orchard of ICAR-CAZRI, Jodhpur. (Adopted, Singh et al. 2016)

Apart from these indices, mean squared deviation ratio (*MSDR*) is also an important index to judge the goodness of fit in prediction (Bishop and Lark 2008), which is actually the transformation of *G* index and is calculated as follows:

$$MSDR = \frac{1}{n} \sum_{i=1}^n \left(\frac{\{z(x_i) - \hat{z}(x_i)\}^2}{\sigma_i^2} \right) \tag{9.22}$$

If the correct semivariogram model is used, the *MSDR* values should be close to 1 (Lark 2000).

Most commonly used approach to quantify the uncertainty of prediction is the calculation of 95% confidence interval maps as follows:

$$\text{Upper limit} = \text{kriged prediction map} + 1.96 \times \text{map of standard deviation of prediction}$$

$$\text{Lower limit} = \text{kriged prediction map} - 1.96 \times \text{map of standard deviation of prediction}$$

An example of such confidence interval map of SOC in a horticultural orchard is presented in Fig. 9.16. The left-hand side maps of the figure show the lower limit, and right-hand side maps show the upper limit of 95% confidence interval, whereas the central map shows the mean predicted SOC content of the orchard. From these maps, it is clearly visible that if we ignore the confidence interval maps, we remain unaware of the fluctuation in predictions. The more is the fluctuations (the wider is

the range of confidence interval), the less is the reliability of the map. It is like the wild guess on soil properties for a particular location and judges how much correct is the guess. The wider is the interval, the higher will be chance of correctness of the guess. Supposing that if the confidence interval is more than standard deviation of measured values, the predicted map is of little use because under such cases we can rely more on the arithmetic mean as the most probable value of any unsampled location.

9.9 DSM Applications: Soil Information System

The ultimate goal of the DSM is to make available the unutilized soil data (legacy data) to end users at a spatial scale. This helps a wide variety of users for different purposes, e.g., farmers for nutrient and water management in an agricultural field, decision-makers for adopting different land management decisions, researchers for modeling landscape processes, etc. The development of the soil information system leads to achieving the ultimate goal of DSM. Several countries have developed the national-level soil information system throughout the world. Here, as an example, the soil information system “SoilGrids250m” developed by ISRIC-World Soil Information is discussed. Soil organic stock map of the world as a snapshot from “SoilGrids250m” is presented in Fig. 9.17.

The “SoilGrids250m” is developed based on soil profile data of 240,000 locations. Global-level predictions of organic carbon concentration, total nitrogen



Fig. 9.17 SoilGrids250m: An example of soil information system

content, pH_{water} , cation exchange capacity (measured at pH 7), soil texture (proportion of sand, silt, and clay), and volume of coarse fragments are available in the “SoilGrids250m.” Predictions are available in six standard soil depths as specified by the Global Soil Map project (0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm, and 100–200 cm). The major features of the “SoilGrids250m” are:

- (a) Direct coupling with standardized soil profile (point) data provided by the ISRIC-World Soil Information Service (*WoSIS*)
- (b) Use of the modern map projection like homolosine that minimizes angular and distance distortions simultaneously
- (c) An improved selection of covariate layers using recursive feature elimination
- (d) Adoption of an improved and more realistic cross-validation procedure
- (e) Quantification of uncertainties in the soil predictions, using prediction intervals, through implementation of quantile regression forests

9.10 Conclusion

Soil plays a crucial role not only in the agricultural production system but also helps in taking many soil and land management decisions. For example, soil nutrients support plant growth and yield, soil hydraulic properties dictate partitioning rainfall into a runoff, soil water retention behavior governs soil moisture regime in an agricultural field, soil pollutants content helps in assessing the risk associated with handing polluted soils, etc. Therefore, soil survey or target-based soil sampling efforts have been done regularly to gather knowledge on soil properties to adopt suitable soil management practices. However, it is not always possible to collect soil samples from multiple locations from a target area. Therefore, estimates are tried at unsampled locations using the information at measured locations of the surveyed area. DSM provides the estimate of soil properties at unsampled locations in the most rational approach which includes geostatistics, state-factor (clorpt) approach, and PTF models. Here, we discussed, in brief, these three approaches with examples. Further, the accuracy and uncertainty of digital soil products help to judge the reliability of it to stakeholders, and thus the inclusion of this information in digital soil products should be an essential requirement. With the advancement of information technology (IT) applications, it will be more appropriate to make these digital soil products available in different IT platforms, e.g., android applications, WebGIS applications, spatial soil database management systems, etc., which together is called soil information system. Therefore, future efforts are required to apply DSM technology to available legacy soil data and to prepare soil maps and make it accessible to wide users as soil information systems.

References

- Aggarwal P, Gupta RP (1998) Two dimensional Geosta-tistical analysis of soils. In: Gupta RP, Gnildyal BP (eds) Theory and practice in agrophysics measurements. Allied Publishers, Buffalo, pp 253–263
- Akpa SI, Odeh IO, Bishop TF, Hartemink AE, Amapu IY (2016) Total soil organic carbon and carbon sequestration potential in Nigeria. *Geoderma* 271:202–215. <https://doi.org/10.1016/j.geoderma.2016.02.021>
- Angelini ME, Heuvelink GBM, Kempen B (2017) Multivariate mapping of soil with structural equation modelling. *Eur J Soil Sci* 68(5):575–591. <https://doi.org/10.1111/ejss.12446>
- Bannerjee G, Sarkar U, Das S, Ghosh I (2018) Artificial Intelligence in Agriculture: A Literature Survey. *Int J Sci Res Comput Sci Appl Manage Stud* 7(3):1–6
- Behrens T, Scholten T (2006) Digital soil mapping in Germany-a review. *J Plant Nutr Soil Sci* 169:434–443
- Behrens T, Förster H, Scholten T, Steinrücken U, Spies ED, Goldschmitt M (2005) Digital soil mapping using artificial neural networks. *J Plant Nutr Soil Sci* 168(1):21–33. <https://doi.org/10.1002/jpln.200421414>
- Ben Dor E, Chabrilat S, Dematte JAM, Taylor GR, Hill J, Whiting ML, Sommer S (2009) Using imaging spectroscopy to study soil properties. *Remote Sens Environ* 113(1):S38–S55
- Bishop TFA, Lark RM (2008) A comparison of parametric and non-parametric methods for modelling a coregionalization. *Geoderma* 148:13–24
- Blöschl G, Sivapalan M (1995) Scale issues in hydrological modelling – a review. *Hydrol Process* 9:251–290
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Brown DJ, Shepherd KD, Walsh MG, Dewayne Mays M, Reinsch TG (2006) Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma* 132:273–290
- Carter GP, Miskewitz RJ, Isukapalli S, Mun Y, Vyas V, Yoon S, Georgeopoulos P, Uchirin CG (2011) Comparison of kriging and cokriging for the geostatistical estimation of specific capacity in the Newark Basin (NJ) aquifer system. *J Environ Sci Health A Tox Hazard Subst Environ Eng* 46(4):371–377
- Chabrilat S, Goetz AFH, Krosley S, Olsen HW (2002) Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. *Remote Sens Environ* 82:431–445
- Chakraborty S, Li B, Deb S, Paul S, Weindorf DC, Das BS (2017) Predicting soil arsenic pools by visible near infrared diffuse reflectance spectroscopy. *Geoderma* 296:30–37
- Chang CW, Laird DA (2002) Near-infrared reflectance spectroscopic analysis of soil C and N. *Soil Sci* 167:110–116
- Chatterjee S, Santra P, Majumdar K, Ghosh D, Das I, Sanyal SK (2015) Geostatistical approach for management of soil nutrients with special emphasis on different forms of potassium considering their spatial variation in intensive cropping system of West Bengal, India. *Environ Monit Assess* 187(4):183
- Clark RN (1999) Spectroscopy of rocks and minerals, and principles of spectroscopy. In: Rencz AN (ed) Remote sensing for the earth sciences: manual of remote sensing. American Society for Photogrammetry and Remote Sensing, New York, pp 3–58
- Dahiya IS, Kalta B, Agrawal RP (1998) Kriging for interpolation through spatial variability analysis of data. In: Gupta RP, Ghildyal BP (eds) Theory and practice in agrophysics measurements. Allied Publishers, Buffalo, pp 242–252
- Das M (2007) Spatial variability analysis of soil hydraulic conductivity in an irrigation command. *J Indian Soc Soil Sci* 55(1):10–13
- Das BS, Sarathjith MC, Santra P, Sahoo RN, Srivastava R, Routray A, Ray SS (2015) Hyperspectral remote sensing: opportunities, status and challenges for rapid soil assessment in India. *Curr Sci* 10:860–868

- Divya Y, Sanjeevi S, Ilamparuthi K (2014) Studies on textural and compositional characteristics of sand and clay mixtures using hyperspectral radiometry. *J Indian Soc Remote Sens* 42 (3):589–600
- Dokuchaev VV (1883) The Russian chernozem report to the free economic society. Imperial University of St. Petersburg, St. Petersburg, in Russian
- Ersahin S (2001) Comparing ordinary kriging and cokriging to estimate infiltration rate. *Soil Sci Soc Am J* 67(6):1848–1855
- Goovaerts P (1998) Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. *Biol Fertil Soils* 27:315–334
- Grimm R, Behrens T, Märker M, Elsenbeer H (2008) Soil organic carbon concentrations and stocks on Barro Colorado Island—digital soil mapping using random forests analysis. *Geoderma* 146 (1–2):102–113. <https://doi.org/10.1016/j.geoderma.2008.05.008>
- Grunwald S (2009) Multicriteria characterization of recent digital soil mapping and modelling approaches. *Geoderma* 152:195–207
- Gulfo E, Sahoo RN, Sharma RK, Khanna M (2012) Soil moisture assessment using hyperspectral remote sensing. In: Proceedings of the second national workshop on challenges and opportunities of water resources management in Tana Basin, Upper Blue Nile Basin, Ethiopia. Blue Nile Water Institute, Bahir Dar University, Ethiopia 2012 Oct, pp 69–77
- Gupta A, Das BS, Kumar A, Chakraborty P, Mohanty B (2016) Rapid and non-invasive assessment of Atterberg limits using diffuse reflectance spectroscopy. *Soil Sci Soc Am J* 80(5):1283–1295
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25:1965–1978
<http://casoilresource.lawr.ucdavis.edu/soilwebapps/>
<http://earthexplorer.usgs.gov/>
<http://earthexplorer.usgs.gov/>
<http://srtm.csi.cgiar.org/>
<http://worldclim.org/current>
<http://www.soilsscotland.gov.uk/data/soilsurvey25k.php>
- Jenny H (1941) Factors of soil formation, a system of quantitative pedology. McGraw-Hill, New York
- Jha K, Doshi A, Patel P, Shah M (2019) A comprehensive review on automation in agriculture using artificial intelligence. *Artif Intell Agric* 2:1–12
- Kadupitiya HK, Sahoo RN, Ray SS, Chakraborty D, Ahmed N (2010) Quantitative assessment of soil chemical properties using visible (VIS) and near-infrared (NIR) proximal hyperspectral data. *Trop Agric* 158:41–60
- Kamble KH, Aggrawal P (2011) Geostatistical analyst for deciding optimal interpolation strategies for delineating compact zones. *Int J Geosci* 2(04):585
- Katuwal S, Hermansen C, Knadel M, Moldrup P, Greve MH, de Jonge LW (2018) Combining X-ray computed tomography and visible near-infrared spectroscopy for prediction of soil structural properties. *Vadose Zone J* 17(1)
- Khaledian Y, Miller BA (2019) Selecting appropriate machine learning methods for digital soil mapping. *Appl Math Model*. <https://doi.org/10.1016/j.apm.2019.12.016>
- Kovačević M, Bajat B, Gajić B (2010) Soil type classification and estimation of soil properties using support vector machines. *Geoderma* 154(3–4):340–347. <https://doi.org/10.1016/j.geoderma.2009.11.005>
- Lagacherie P, McBratney AB, Voltz M (2006) Digital soil mapping: an introductory perspective. Elsevier, Amsterdam
- Lark RM (2000) A comparison of some robust estimators of the variogram for use in soil survey. *Eur J Soil Sci* 51:137–157
- Leone AP, Sommer S (2000) Multivariate analysis of laboratory spectra for the assessment of soil development and soil degradation in the Southern Apennines (Italy). *Remote Sens Environ* 72:346–359

- Lin LI (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45:255–268
- Mansuy N, Thiffault E, Paré D, Bernier P, Guindon L, Villemaire P, Poirier V, Beaudoin A (2014) Digital mapping of soil properties in Canadian managed forests at 250 m of resolution using the k-nearest neighbor method. *Geoderma* 235:59–73. <https://doi.org/10.1016/j.geoderma.2014.06.032>
- Matheron G (1965) Les variables régionalisées et leur estimation. Une application de la théorie des fonctions aléatoires aux sciences de la nature. Masson, Paris
- McBratney AB, MendonçaSantos ML, Minasny B (2003) On digital soil mapping. *Geoderma* 117:3–52
- Minasny B, McBratney AB (2016) Digital soil mapping: a brief history and some lessons. *Geoderma*. <https://doi.org/10.1016/j.geoderma.2015.07.017>
- Mohanty B, Gupta A, Das BS (2016) Estimation of weathering indices using spectral reflectance over visible to mid-infrared region. *Geoderma* 266:111–119
- Murase H (2000) Artificial intelligence in agriculture. *Comput Electron Agric* 29:1/2
- NBSS & LUP (2005) Reflectance libraries for development of soil sensor for periodic assessment of state of soil resources. NATP project report (NBSS No. 835). National Bureau of Soil Survey/Land Use Planning, Nagpur
- Rabus B, Eineder M, Roth A, Bamler R (2003) The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. *ISPRS J Photogramm Remote Sens* 57 (4):241–262
- Reeves JB, McCarty GW, Meisenger JJ (1999) Near-infrared diffuse reflectance spectroscopy for the analysis of agricultural soil. *J Near Infrared Spectrosc* 7:179–193
- Rossel RV, McBratney A (1998) Laboratory evaluation of a proximal sensing technique for simultaneous measurement of soil clay and water content. *Geoderma* 85:19–39
- Rossel RV, Behrens T, Ben-Dor E, Brown DJ, Dematté JA, Shepherd KD, Shi Z, Stenberg B, Stevens A, Adamchuk V, Aïchi H (2016) A global spectral library to characterize the world's soil. *Earth Sci Rev* 155:198–230
- Rossiter DG (2018) Technical note: co-kriging with the gstat package of the R environment for statistical computing. http://www.css.cornell.edu/faculty/dgr2/_static/files/R_PDF/CoKriging.pdf
- Sanchez PA, Ahamed S, Carré F, Hartemink AE, Hempel JW, Huising J, Lagacherie P, McBratney AB, McKenzie NJ, Mendonça-Santos ML, Minasny B, Montanarella L, Okoth P, Palm CA, Sachs JD, Shepherd KD, Vagen TG, Vanlauwe B, Walsh MG, Winowiecki LA, Zhang GL (2009) Digital soil map of the world. *Science* 325(5941):680–681
- Santra P, Das BS (2008) Pedo-transfer functions for soil hydraulic properties developed from a hilly watershed of Eastern India. *Geoderma* 146(3–4):439–448
- Santra P, Chopra UK, Chakraborty D (2008) Spatial variability of soil properties and its application in predicting surface map of hydraulic parameters in an agricultural farm. *Curr Sci* 10:937–945
- Santra P, Sahoo RN, Das BS, Samal RN, Pattanaik AK, Gupta VK (2009) Estimation of soil hydraulic properties using proximal spectral reflectance in visible, near-infrared, and shortwave-infrared (VIS–NIR–SWIR) region. *Geoderma* 152(3–4):338–349
- Santra P, Kumawat RN, Mertisa RS, Mahla HR, Sinha NK (2012a) Spatial variation of soil organic carbon stock in a typical agricultural farm of hot arid ecosystem of India. *Curr Sci* 10:1303–1309
- Santra P, Das BS, Chakravarty D (2012b) Spatial prediction of soil properties in a watershed scale through maximum likelihood approach. *Environ Earth Sci* 65(7):2051–2061
- Santra P, Singh R, Sarathjith MC, Panwar NR, Varghese P, Das BS (2015) Reflectance spectroscopic approach for estimation of soil properties in hot arid western Rajasthan, India. *Environ Earth Sci* 74(5):4233–4245
- Santra P, Kumar M, Panwar N (2017a) Digital soil mapping of sand content in arid western India through geostatistical approaches. *Geoderma Reg* 9:56–72

- Santra P, Kumar M, Panwar NR, Das BS (2017b) Digital soil mapping and best management of soil resources. In: Rakshit A, Abhilash PC, Singh HB, Ghosh S (eds) Adaptive soil management: From theory to practice. Springer, Singapore, pp 3–38. https://doi.org/10.1007/978-981-10-3638-5_1
- Santra P, Kumar M, Panwar NR, Pandey CB (2017c) Soil resources and its mapping through geostatistics using R and QGIS. New India Publishing Agency, New Delhi, p 340
- Santra P, Kumar M, Kumawat RN, Painuli DK, Hati KM, Heuvelink G, Batjes N (2018) Pedotransfer functions to estimate soil water retention at field capacity and permanent wilting point in hot arid western India. *J Earth Syst Sci* 127:35. <https://doi.org/10.1007/s12040-018-0937-0>
- Sarathjith MC, Das BS, Wani SP, Sahrawat KL (2014) Dependency measures for assessing the covariation of spectrally active and inactive soil properties in diffuse reflectance spectroscopy. *Soil Sci Soc Am J* 78(5):1522–1530
- Saxena RK, Vermal KS, Srivastava RA, Av AK, Shiwalkar AA, Londhel SL (2003) Spectral reflectance properties of some dominant soils occurring on different altitudinal zones in Uttarakhand Himalayas. *Agropedology* 13(2):35–43
- Shepherd KD, Walsh MG (2002) Development of reflectance spectral libraries for characterization of soil properties. *Soil Sci Soc Am J* 66:988–998
- Singh M, Srivastava R, Sethi M, Sood A (2014) Development of spectral reflectance methods and low cost sensors for real-time application of variable rate inputs in precision farming (final report). National Agricultural Innovation Project (ICAR)/Punjab Agricultural University, Ludhiana
- Singh A, Santra P, Kumar M, Panwar N, Meghwal PR (2016) Spatial assessment of soil organic carbon and physicochemical properties in a horticultural orchard at arid zone of India using geostatistical approaches. *Environ Monit Assess* 188(9):529
- Srivastava RA, Prasad JA, Saxena R (2004) Spectral reflectance properties of some shrink-swell soils of Central India as influenced by soil properties. *Agropedology* 14:45–54
- Webster R, Oliver MA (2007) Geostatistics for environmental scientists. Wiley, Chichester