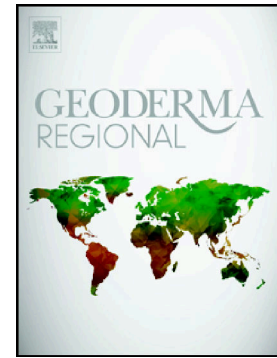


Identifying opportunities to improve digital soil mapping in India:
A systematic review

Prava Kiran Dash, Niranjana Panigrahi, Antaryami Mishra



PII: S2352-0094(21)00123-1

DOI: <https://doi.org/10.1016/j.geodrs.2021.e00478>

Reference: GEODRS 478

To appear in: *Geoderma Regional*

Received date: 22 October 2020

Revised date: 16 December 2021

Accepted date: 22 December 2021

Please cite this article as: P.K. Dash, N. Panigrahi and A. Mishra, Identifying opportunities to improve digital soil mapping in India: A systematic review, *Geoderma Regional* (2021), <https://doi.org/10.1016/j.geodrs.2021.e00478>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Identifying Opportunities to Improve Digital Soil Mapping in India: A Systematic Review

Prava Kiran Dash^{1,2}, Niranjana Panigrahi³, and Antaryami Mishra¹

¹Department of Soil Science and Agricultural Chemistry, College of Agriculture, Odisha University of Agriculture and Technology, Bhubaneswar, Odisha, India-751003

²Department of Agronomy, Iowa State University, USA

³School of Water, Energy and Environment, Cranfield University, Cranfield, MK43 0AL, UK

Abstract

Soaring food demand, population pressure, land degradation, small size of agricultural land holdings, and diversified soil types in India require advanced digital soil mapping (DSM) for sustainable land management. This paper systematically reviews the common trends of SCORPAN based DSM in India to identify the important research gaps and opportunities to improve in future. A systematic literature search from 2000 to October 2021 has yielded 35 numbers of peer reviewed articles, which have performed DSM in India following the SCORPAN approach. The increased number of published articles from 2017 onwards suggests that there is a growing interest for DSM in India over the past few years. However, only two articles have prepared digital soil maps at a national extent. Moreover, the local and regional extent DSM are being limited to only a few parts of the country. There still remains 50% of the states and Union Territories of the country where no DSM studies have been performed except the national or global level interventions. Among the target variables, soil carbon related attributes have been predicted most frequently, whereas soil classes have been rarely predicted. Environmental covariates representing organism (O) and relief (R) have been widely included for DSM, whereas use of other covariates has been limited. Among

different machine learning (ML) algorithms, regression kriging has been adopted most frequently followed by random forest and quantile regression forest. Most articles have adopted data splitting (76%) as the map evaluation approach, whereas independent validation has been limited only to 5% of the articles. Only 34% of the articles have presented the uncertainty maps. Major research gaps identified by this review include lack of standardized digital soil databases, poor sampling density, coarse resolution, limited use of environmental covariates, insufficient comparative studies among ML algorithms, inadequate independent validation, and undersupply of uncertainty maps. Key evidences from this review will be helpful for improving future DSM activities by scientists and practitioners involved with DSM in India and abroad.

Key words: Legacy soil data; SCORPAN; Environmental covariates; Machine learning

1. Introduction

Over the past few decades, population pressure, climate change, and devastating land management practices have led to a widespread decline in soil quality and productivity (Brevik et al., 2015). Land management has emerged as a major challenge at present (Garg et al., 2020). Hence, there is a worldwide demand for a systematic global soil resource database to address ‘soil security’ (McBratney et al., 2014). Moreover, a wide range of disciplines and sub-disciplines also require precise soil information for dealing with various global issues related to food, water, energy, environment, biodiversity, and human health (Grunwald et al., 2011; Hartemink, 2015).

Mapping soils is the preliminary step in decision making such as delineation of suitable crop growing areas or identification of affected areas (Dharumarajan et al., 2017). Conventional soil mapping (CSM) is typically based on the mental model (Hudson, 1992) of the soil surveyor who puts homogenous areas into soil polygons or classes. However, the polygons

drawn this way ignore the within-class variations and are hence difficult to interpret. Further, the tedious field survey and lengthy analytical steps make the mapping process intensive and time consuming (Lagacherie & McBratney, 2006). Moreover, these hard copy maps often remain inaccessible and unused by the end-users (Santra et al., 2021). Fortunately, with the conceptualisation of the SCORPAN approach and Digital Soil Mapping (DSM) techniques by McBratney et al. (2003), a widespread digital revolution in soil mapping has been witnessed in the last two decades. DSM technique provides quick measures to predict soil properties at locations, which are not sampled based on the observed values at the sampled sites and other environmental covariates (predictive factors). Unlike the conventional soil maps, digital soil maps provide us with estimates of accuracy and uncertainty associated with the prepared soil maps (Minasny & McBratney, 2016). Moreover, DSM is a strong supporting tool to the present-day precision agriculture (Minasny & Hartemink, 2011). Keeping the essentiality of DSM in view, DSM Working Group (WG) of the International Union of Soil Sciences (IUSS) has taken up the ‘GlobalSoilMap’ (<https://www.isric.org/projects/globalsoilmapnet>) initiative to prepare fine grid digital soil maps (~100 m spatial resolution) of the world (Arrouays et al., 2014, 2017). Recently, initial versions of global extent digital soil maps notably SoilGrids 2.0 (<https://soilgrids.org/>) have also been prepared at a 250 m spatial resolution using the World Soil Information Service (WoSIS) database developed by ISRIC (International Soil Reference and Information Centre; Poggio et al., 2021). The Global Soil Partnership (<http://www.fao.org/globalsoilpartnership/>) initiative by the Food and Agriculture Organization (FAO) also focuses on strengthening the GLOIS (Global Soil Information System) for promoting sustainable soil management across the world.

India is the seventh largest country of the world having a total geographical area of 328.73 M ha and a gross cropped area of 197.05 M ha (Government of India, 2020). With an 1.37

billion population (by 2019), India is currently the second most populous country of the world. Current projections by United Nations, Department of Economic and Social Affairs, Population Division (2019) suggests that India will surpass China and become the world's most populous country around 2027. Pressure on the land resources, land degradation (~64 M ha), shrinkage of agricultural lands, and soaring food demand for a projected 1.65 billion population by 2050 (Vision 2050, 2015) are the burning issues in India. Desertification, poor soil quality, multi-nutrient deficiency, and stagnant yield are few of the major aggravating soil-related issues in India warranting rapid soil assessments (Sahrawat et al., 2010; Singh and Chatterji, 2018). Specifically, as we transit from a rural to an urban India with smart cities, the assessment of soil properties at finer resolutions will provide much of our land management decisions in the upcoming years. However, highly diversified soil groups across the country (owing to the widely varying climatic conditions, geology, and topography) make soil management reasonably difficult (Bhattacharyya et al., 2013). Moreover, 86.08% of India's agricultural land holdings are small and marginal having land size less than 2 hectares with an average size being 1.08 ha (Agricultural census Division, 2015-16). Hence, the existing small scale (large mapping extent) conventional soil maps are not sufficient to manage the small and marginal land holdings of the country. Under these circumstances, adopting DSM methodologies for preparing fine resolution digital soil maps holds great relevance. In this context, this paper performs a systematic review of the published SCORPAN based DSM literatures in India to understand the common trends, research gaps, and future opportunities of DSM in India.

2. Evolution of soil mapping in India

2.1. Conventional soil mapping

Conventional soil maps are prepared by taking field observations and other auxiliary information such as geology, topography, land use, and land cover, which are then integrated with the expert or conceptual knowledge of the soil surveyor to group homogenous areas into classes or polygons (Minasny & McBratney, 2016). The first soil map of India was prepared by Schokalskaya in 1932 (Bhattacharyya et al., 2013) followed by which, several soil maps were prepared. Soil map of India at 1: 6.3 million cartographic scale was published in 1983 based on USDA soil classification system (Murthy and Pandey, 1983).

Before the 1980s, Indian soil survey systems remained confined to field and laboratory based methods rarely being associated with photointerpretation of aerial photographs (Sharma & Kalia, 1987). Subsequently, digital satellite imageries such as LANDSAT TM (Manchanda and Hilwig, 1981), IRS-1A LISS II (Kudrat et al., 1992), and IRS-1C/1D LISS-III were used to delineate physiographic boundaries of small cartographic scale soil resource maps (Manchanda et al., 2002). Attempts were also made to map salt-affected soils, waterlogged areas, and degraded lands (Dwivedi and Sreenivas, 1998). Topographical maps from the Survey of India were in use for obtaining information related to topography (Saha and Singh, 1991). In the 2000s, Digital Elevation Models (DEMs) from satellite imageries such as SPOT-1, ASTER, IRS-1C/1D, Cartosat- 1 were assessed for extracting terrain features (Ahmed et al., 2007). Besides the DEMs, the physiography-land use (PLU) maps prepared through digital image processing of RS products also assisted the CSM (Velmurugan & Carlos, 2009).

The institutional setups consisting of premier organisations such as All India Soil & Land Use Survey (Presently, Soil & Land Use Survey of India (SLUSI)) in 1958, Indian Photointerpretation Institute (presently, Indian Institute of Remote Sensing (IIRS)) in 1966, National Remote Sensing Agency (Presently, National Remote Sensing Centre (NRSC)) in 1974, National Bureau of Soil Survey and Land Use Planning (NBSS&LUP) in 1976 played

vital roles in soil resource mapping activities across India. USDA soil taxonomy based family level soil map of the entire country has been completed at a 1:2,50,000 cartographic scale by the NBSS&LUP (Bhattacharyya et al., 2009). Soil and terrain digital database of Indo-Gangetic plains and black soil regions of India has been prepared under the SOTOR framework systems (Chandran et al., 2014). So far, NBSS&LUP has prepared soil resource maps at 1:2,50,000 and 1: 50,000 cartographic scales for 300 M ha and 200 M ha land of the country, respectively. However, the detailed soil maps (1:4,000/ 1:10,000/ 1:15,000 cartographic scale) prepared by the year 2018 have been limited only to ~23 M ha of land (Dharumarajan et al., 2019). Most of the conventional soil maps in India have been digitized later to produce digitized soil resource maps. These digitized versions of the conventional soil maps can be considered as ‘digitized soil maps’, but they cannot be considered as ‘digital soil maps’ (Santra et al., 2017a, 2021). Digitizing the conventional soil maps or making the conventional soil maps available in digital platforms doesn’t qualify to be considered as digital soil maps. Rather, soil maps need to be prepared using DSM methodologies (McBratney et al., 2003) in order to be considered as digital soil maps.

2.2. Digital soil mapping

The concept of DSM originated from Jenny’s CLORPT theory (Jenny, 1941) and was later extended by McBratney et al. (2003) in form of SCORPAN approach to describe quantitative relationships between soil property or classes and environmental covariates to prepare continuous soil maps. Three major components of DSM are input, process, and output (Lagacherie and McBratney, 2006). The ‘input’ refers to data generated from field and laboratory observational methods, which are ‘processed’ through spatial soil inference systems (predictive models) to finally produce spatially continuous soil maps as ‘outputs’. The SCORPAN refers to the set of environmental covariates (also known as predictive factors or auxiliary variables), which can essentially be used for establishing quantitative

relationships with the target soil property or class to develop regression equations for prediction purposes. The SCORPAN approach as proposed by McBratney et al. (2003) is equated as:

$$S_a \text{ (or) } S_c = f(S, C, O, R, P, A, N) + e \quad (1)$$

Where, S_a (soil property) or S_c (soil class) is represented as a function (f) of 7 different environmental covariates viz. S (soil), C (climate), O (organism), R (relief), P (parent material), A (age of weathering), and N (spatial position). Here, 'e' is an error component.

Since 2000s, many studies in India have prepared digital soil maps using conventional DSM techniques such as geostatistics based spatial interpolation methods (Behera et al., 2018; Shukla et al., 2016). One of the popular geostatistical methods is ordinary kriging, which uses the principle of spatial autocorrelation for making predictions at unsampled locations based on the observed values at the sampled locations (Oliver, 1987). However, considering the limitations associated with the geostatistical methods such as assumptions for data distribution, need of higher sample size, data transformation, and incompetence to deal with non-linear relationships, SCORPAN approach to prepare digital soil maps using different machine learning (ML) tools have been widely adopted as advanced DSM methodologies all over the world (Wadoux et al., 2020a, 2020b). ML involves the process of regression and classification to build prediction models using principles of spatial association. ML models have been preferred over the geostatistical methods as they are capable of handling large datasets having complex relationships among different cross-correlated covariates and soil properties (Wadoux et al., 2020a). Recently, SCORPAN based DSM using different ML tools have started to draw more attention in the DSM scenario of India. In this paper, we will be discussing in detail the current status of SCORPAN based DSM in India.

3. Materials and Methods

For this systematic review, we followed a systematic mapping approach (James et al., 2016) to compile relevant information from the existing scientific literatures. Scopus, Google Scholar, ScienceDirect, and WoS (Web of Science) databases were searched from the year 2000 up to 31st October, 2021 to access all the published peer reviewed journal articles, which followed the SCORPAN based approach to prepare DSM in India. The literature search was performed on the titles, abstracts, and key words with strings including ‘soil mapping’, ‘digital mapping’, ‘digital soil mapping’, ‘map’, ‘predict’, ‘soil properties’, ‘soil classes’, ‘predictive soil mapping’, ‘India’, ‘machine learning’, ‘SCORPAN’ with ‘AND’ and ‘OR’ operators. All the search outcomes were screened and duplicates were removed first. Then, the articles, which did not perform DSM for soil property or classes were eliminated. The articles, which adopted only geostatistical methods of DSM without considering any environmental covariates were also removed. After the exclusion process, only the articles, which have performed DSM in India based on SCORPAN approach were targeted for this systematic review (last archived on 31st October 2021). Thirty five articles were identified according to the above criteria. Relevant information from these articles including year of publication, location, geographical area, mapping extent, sampling density, method of sampling, maximum depth of sampling, target variables, environmental covariates, sources of environmental covariates, prediction models, evaluation approaches, mapping of uncertainty etc. were extracted and presented in tabular formats (Table 1 and 2). Suitable figures and graphs were also presented for better understanding of the common trends and research gaps.

4. Results and Discussion

Relevant information from the selected 35 articles is being presented in Table 1. Only 6 articles were published by year 2016. However, a growing trend has been observed in the number of published articles from 2017 onwards with 5, 7, 6, 5 numbers of articles being published in 2017, 2018, 2019, and 2020, respectively. Also, 6 articles have already been

published by October, 2021 indicating that there is a growing interest for DSM in India over the past few years. Few of the publication outlets include Geoderma regional (20%), Geoderma (14%), Journal of the Indian Society of Remote sensing (9%), and Catena (9%). Key findings from the systematic review have been discussed in the following sub-sections.

4.1. Geographic distribution of DSM activities, mapping extent, and resolution

Fig. 1 presents state and Union Territory (UT) wise geographic distribution of DSM activities in India excluding those performed at a national or global extent. Out of the total 36 entities (28 states and 8 UTs) of India, 33% of the entities are observed with only one DSM study each, 14% with 2 to 5, and only 3% of the entities with more than 5 numbers of DSM studies. However, the remaining 50% of entities are observed with no DSM studies so far. It indicates that DSM is still at a primitive stage in India and the vast extents of digitally unmapped areas are yet to be targeted for DSM in future.

Plot (0-1 km²), local (> 1 km²-10⁴ km²), and regional (>10⁴-10⁷ km²) extent soil maps constitute 3%, 54%, and 43 % of the DSM studies in India. However, out of all the regional extent soil maps, only two articles (Reddy et al., 2021; Sreenivas et al., 2016) constituting 5% of the total number of articles have conducted DSM at the national extent covering the entire country. Vastness of the country with widely varying soil groups, diversified ranges of physiography, climate, and parent materials might have challenged the digital soil mappers to uptake national extent DSM attempts. Some of the global DSM products notably SoilGrids 2.0 (<https://soilgrids.org/>) and OpenLandMap (<https://openlandmap.org/>) also provide soil information for India each at a resolution of 250 m (Hengl et al., 2017). Recently, Dharumarajan et al. (2021a) have compared a locally prepared SOC stock map for a 56,763 km² area of South-West India with the global DSM product SoilGrids and observed the locally prepared soil map to be more accurate ($R^2=0.46$) compared to the global DSM

product ($R^2=0.11$). In this case, poor sampling density associated with the global DSM product could be a reason for its lesser accuracy against the local DSM product (de Sousa et al. 2020; Dharumarajan et al., 2021a). It indicates that irrespective of the huge potential of larger extent (such as national or global) DSM products, local extent studies are still essential for local level applications. Conducting more such comparison studies among DSM products prepared at different mapping extents could provide more insights about their practical utilities in different aspects.

Spatial resolution is a key factor in soil mapping that determines the applicability of the soil maps. Reddy et al. (2021) and Sreenivas et al. (2016) have prepared national extent soil maps at 500 m and 250 m resolutions, respectively. Bhuvan geoportal (<https://bhuvan.nrsc.gov.in/>) of ISRO-NRSC provides country-wide soil information for specific soil properties at a 5000 m spatial resolution. Regional extent soil maps have been prepared by Sreenivas et al. (2014), Santra et al. (2017a), and Dharumarajan et al. (2020a) at 1000 m, 1000 m and 250 m spatial resolutions, respectively. Similarly, Mitran et al. (2018) and Hinge et al. (2018) have prepared regional extent soil maps at 90 m spatial resolution. Kalambukattu et al. (2018) and Mitran et al. (2019) have prepared local extent soil maps at 30 m resolution. Xu et al. (2018 a, 2018b) have prepared fine resolution soil maps (<10 m) at local extents. Considering the average Indian agricultural landholding size being 1.08 ha, digital soil maps need to be prepared at fine resolutions for making farm level agricultural management decisions (Reddy et al., 2021). Fine resolution remote sensing (RS) products and DEMs with sufficient sampling density should be adopted for developing digital soil maps at fine resolutions (Dharumarajan et al., 2019; Santra et al., 2021).

4.2. Legacy soil data

Legacy soil data is also known as historic soil data (Grunwald et al., 2009) referring to geo-referenced soil databases generated in the past. Typically, these data are available in form of soil profile databases or polygon-based soil maps (Zhang et al., 2017). Indian legacy soil information is primarily available in form of soil series and soil survey reports of NBSS&LUP (Santra et al., 2017b). Hard copies of these soil survey reports and soil resource maps can be obtained from NBSS&LUP publications catalogue. Recently, Reddy et al. (2021) have published nationwide digital soil maps using the Indian legacy soil database consisting of 1,707 soil profiles, most of which have been retrieved from soil series publications by NBSS&LUP available at the Krishikosh public domain (<https://krishikosh.egranth.ac.in/>). Similarly, Dharumarajan et al. (2019) has reported about the availability of Indian legacy soil information in the digital platform ‘Bhoomi geoportal’ (www.bhoomigeoportal-nbsslup.in) of NBSS&LUP. So far, the 1:1M, 1:2,50,000, 1:50000 cartographic scale legacy soil maps on the digital platform offer only polygon boundary information with limited options to download the digitized soil maps. Moreover, the polygon boundary information is often confusing among the adjacent polygons and is also insufficient to provide point values. Although the 1:10,000 cartographic scale soil resource maps offer point values, they are being limited only to a few selected blocks and districts of the country. Nevertheless, the digital availability of analytical values of soil profiles is also being limited to a few regions of the country with limited options to download them. Even though Indian legacy soil data is huge, they need to be organised properly (Chandran et al., 2014; Dharumarajan et al., 2019). Bringing the enormous soil information of such a vast country into digital formats might be challenging, but is supposed to boost the legacy data based DSM in India. Nevertheless, a harmonized and searchable digital soil database is yet to be developed for the whole country by gathering information generated by different local, state, and national level institutes in the past (Chandran et al., 2014; Reddy et al., 2021).

As per this review, 28% of the studies have produced digital soil maps using legacy soil data (Table 1) and the rest (72%) have conducted soil sampling. Okonkwo et al. (2018) used legacy data of 34 soil profiles from NBSS&LUP to prepare soil pH and Electrical Conductivity (EC) maps in the state of Punjab. Similarly, Santra et al. (2017a) retrieved the legacy soil data for the states of Rajasthan and Gujarat for preparing digital soil maps of sand content at multiple soil depths. Dharumarajan et al. (2020a) have used a legacy soil dataset consisting of 5174 soil profile information to prepare a digital soil depth map of Karnataka state. The ‘GlobalSoilMap’ initiative necessitates organising the soil data at 6 specific depth intervals, i.e., 0-5 cm, 5-15 cm, 15-30 cm, 30-60 cm, 60-100 cm and 100-200 cm for preparing harmonized digital soil maps of the world (Arrouays et al., 2014). Recently, NBSS&LUP has also initiated a project called “IndianSoilGrids” in collaboration with Institut National de la Recherche Agronomique (INRA), France for preparing legacy data based digital soil maps in India using GlobalSoilMap specifications (Dharumarajan et al., 2020b). Nevertheless, one should be careful while using legacy soil data as these have been collected from multiple year-long projects conducted years back, which might have been generated using varying sampling strategies and outdated analytical techniques. Also, the old data may not truly represent the present status of soil, specifically for the non-static soil properties. Hence, improvements in specific aspects such as legacy data rescue, data harmonization, and uncertainty assessments are essential. Moreover, instead of being confined only to legacy soil data, regular soil profile sampling should be carried out using standard sampling strategies (Hengl et al., 2017).

4.3. Soil sampling strategies

Soil sampling is the preliminary step in soil mapping. The sampling techniques followed in India include random (58%), stratified random (31%), and grid (11%). Most of the soil studies (80%) have tried to map only for the surface soils and the remaining 20% of studies

have performed for multiple soil depths. Some of the recent studies (Santra et al., 2017a; Dharumarajan et al., 2020b; Reddy et al., 2021) constituting 14% of the total number of articles have generated digital soil maps at 6 GlobalSoilMap specified soil depths (up to 2m). Equal area quadratic spline function has been used in different studies (Dharumarajan et al., 2020b; Reddy et al., 2021; Santra et al., 2017a) for harmonizing the soil depths. Sampling density is observed to be varying greatly among different mapping extents. The mean sampling density (number of samples per square km) at national, regional, and local extent digital soil maps have been observed to be 0.00045, 0.034, and 10.07, respectively. The global DSM product SoilGrids 2.0 has even lesser sampling density (0.067 profiles per 1000 km²) for India (Batjes et al. 2020; de Sousa et al. 2020). Thus, larger mapping extents have often been associated with coarser sampling densities. Similar observation is also reported by Wadoux et al. (2020a). However, it is essential to maintain an optimum sampling density for producing quality soil maps (Minasny et al., 2013). Chhipa et al. (2019) have used an optimal sampling technique by selecting sample points, which provided minimum kriging variance. Over the years, many concepts on soil sampling have been developed. However, geostatistical and probability based samplings are considered better for model calibration and validation, respectively (Zhang et al., 2017). Nevertheless, since it is not always practically possible to collect soil samples from the model-derived points, more flexible and efficient sampling strategies need to be explored.

4.4. Predicted variables

Fig. 2a. represents frequency of predicted variables in different DSM articles. Twenty out of 35 articles (57%) have prepared digital soil maps for a single target soil variable, whereas the remaining 43% of articles have predicted for multiple target variables. For example, Dharumarajan et al. (2020b) have prepared digital soil maps for 8 different soil properties including soil pH, SOC, Cation Exchange Capacity (CEC), sand, silt, clay, Field Capacity

(FC), and Permanent Wilting Point (PWP). Soil carbon related soil properties such as SOC stock, SOC density, SIC density, and total organic carbon etc. constitute a majority of 32% of all the predicted variables. Other frequently predicted variables in India include soil depth (12%), soil separates (11%), macro nutrients (12%), and soil physicochemical properties such as pH and EC (12%). Other soil properties including FC, PWP, and micronutrients have been predicted less frequently (Fig. 2a.). On the other hand, only two articles (Shukla et al., 2018 and Gomez et al., 2019) constituting 4% of the total number of articles have created digital soil maps for soil classes. It indicates that the prediction for categorical variables (soil classes) has remained unexplored by DSM interventions in India. Similar trend has also been reported in other international articles (Grunwald et al., 2009; Zeraatpisheh et al., 2020). In this aspect, there is a need to expand DSM in India to other potential target variables including both categorical soil variables (such as textural class, soil suitability class, drainage class, soil erosion classes etc.) and quantitative soil properties. Locally important soil variables possessing relevance for multiple sectors such as agriculture, land use planning, construction, and engineering should be targeted for DSM on a priority basis (Zeraatpisheh et al., 2020). Besides soil property and classes, important soil functions also need to be mapped using DSM (Zhang et al., 2017). These soil functions maps will be helpful to assess soil specific functions that include biomass production, maintaining soil biodiversity, carbon sequestration, nutrient and water recycling, and other ecosystem services (Greiner et al., 2018). Nevertheless, GlobalSoilMap has specified 12 relevant target variables for DSM including soil depth, effective soil depth, sand, silt, clay contents, coarse fragments, pH, SOC, effective CEC, bulk density, available water capacity, and EC (Arrouays et al., 2014). Hence, besides focusing on the locally relevant soil properties, there is also a need to focus on the GlobalSoilMap specified target soil properties through which, Indian DSM can

potentially contribute towards enriching the global digital soil databases. *4.5. Environmental covariates*

Suitable environmental covariates (SCORPAN factors) can improve prediction accuracies of digital soil maps (McBratney et al., 2003; Zhang et al., 2017). A wide variety of environmental covariates including soil properties (S), climate (C), organisms (O), relief (R), parent material (P), age of weathering (A), and spatial position (N) can be used as environmental covariates for DSM. Legacy soil maps, climatic data, RS products, DEMs, and geology maps have been used as sources of environmental covariates for DSM in India (Table 2). The frequency of use of different environmental covariates in DSM activities in India has been presented in Fig. 2b. The environmental covariates representing organism (29%) and relief (29%) are the most frequently used covariates followed by covariates representing climate (18%) and soil (13%). However, the covariates reflecting parent material (8%) and spatial position (3%) have been included relatively less frequently.

Relevant soil information including those obtained from RS, proximal sensing, existing soil maps (legacy soil maps) can be used to represent ‘S’ in the SCORPAN (McBratney et al., 2003). In India, Sreenivas et al. (2014, 2016), Santra et al. (2017a), Okonkwo et al. (2018), Dharumarajan et al. (2021a, 2021b) have used legacy soil maps from NBSS&LUP as environmental covariates. Some of the commonly extracted environmental covariates from the legacy soil maps include order, sub-order, great group, soil type, depth, texture, drainage, physiography, and landform (Table 2). Indian legacy soil maps have been prepared at different cartographic scales by different national institutes including NBSS&LUP, SLUSI, and NRSC (Dharumarajan et al., 2019). Although extensive Indian legacy soil maps have been prepared, only 13% of the DSM articles have employed them as environmental covariates. Hence, more interventions are needed to explore the Indian legacy soil maps as potential environmental covariates for DSM. Bhoomi geoportal of NBSS&LUP and Bhuvan

geoportal of NRSC can be explored for soil resource maps. Also, greater efforts are desired to make the legacy soil maps available in digitally downloadable formats with suitable legal frameworks under which they can be used for DSM nationally and internationally. However, the spatial extent and cartographic scale of the legacy soil maps must be considered before using them for DSM (Santra et al., 2017b). For example, a legacy soil map prepared at a small cartographic scale (larger mapping extent) may not have sufficient data for plot-level decision-making. Hence, expanding the detailed soil surveys at 1: 4,000 and 1: 10,000 cartographic scales to more areas of the country will be helpful for obtaining more potential covariates for DSM in India (Dharumarajan et al., 2019). In addition to existing soil maps, there is also a need to adopt novel soil measurement methods such as electromagnetic induction, ground-penetrating radar, and spectroradiometers for rapid generation of soil information, which can also be used as environmental covariates in DSM (Zeraatpisheh et al., 2020).

Climatic factors are considered as important covariates at larger mapping extents, but might not have major impacts at a plot or local extent DSM (Zeraatpisheh et al., 2020). Most of the DSM activities in India have derived climate-related covariates from the Worldclim database (<http://worldclim.org/current>). Some of the frequently used covariates representing 'C' in the SCORPAN include temperature, precipitation, temperature seasonality, and precipitation seasonality (Table 2). Air temperature was observed to be an important predictor for SOC by Sreenivas et al. (2014, 2016). However, most of the DSM in India have been taken up at local or regional extents, which might be a reason that climate-related covariates have been used less frequently (18%) compared to relief and organism related covariates (Fig. 2b.). Relevant information from automatic weather stations can also be utilized for obtaining relevant covariates representing 'C' (Lamichhane et al., 2019).

Covariates representing ‘O’ and ‘R’ have been included very frequently for DSM in India (Fig. 2b.). Organism (O) in the SCORPAN refers to the covariates related to vegetation, soil fauna, and human activities such as varying land use and management practices. For example, vegetation and land use are important predictors of SOC (Minasny et al., 2013). In recent years, both conventional and RS derived products are in frequent use to extract organism related covariates. Some of the frequently used covariates in India to represent ‘O’ in the SCORPAN include land use land cover (LULC), Normalized Difference Vegetation Index (NDVI), and Enhanced Vegetation Index (EVI). Band reflectance values and band ratios of different satellite imageries such as MODIS, Landsat, and Sentinel have also been employed as potential covariates to represent ‘O’ (Table 2). Relief (topography) is also one of the important predictor for DSM, especially at a local or regional extent soil mapping (Zeraatpisheh et al., 2020). Relief affects soil microclimate and thereby influences soil characteristics (Dash et al., 2019). Most of the DSM studies in India have derived the terrain attributes from SRTM DEM, Cartosat DEM, ASTER DEM, and ALOS DEM (Table 2). Inclusion of more of such fine resolution RS products will be helpful for generating fine resolution digital soil maps (Santra et al., 2021).

Parent material or geology is a vital predictor of soil properties (Guo et al., 2015). Geological information has been considered important for explaining soil distributions across medium cartographic scale soil maps (Miller & Schaetzl., 2016). However, our review identifies that only a few (8%) DSM articles in India have included covariates representing ‘P’ (Fig. 2b). Hence, there is need to explore more options for including geological covariates for DSM in India. Sreenivas et al. (2016) and Mitran et al. (2018) have used geological information obtained from Geological Survey of India as covariates for DSM. Bhukosh geoportal of Geological Survey of India (<https://bhukosh.gsi.gov.in/>) can be explored for geological information. The use of spatial position (N) has also been limited only to a few (3%) studies

(Shukla et al., 2018; Xu et al., 2018a), whereas age of weathering (A) has never been used as a covariate for DSM in India. These results indicate that future DSM interventions should identify and include more potential covariates under the SCORPAN domain.

Different combinations of environmental covariates used in India has been presented in Fig. 2c. The combination of O+R (organism and relief) covariates has been used most frequently (29%) followed by S+C+O+R (11%). ‘O’ and ‘R’ related covariates are used in most of the combinations. It indicates that the open-source availability of RS products and DEMs have triggered their use as environmental covariates in DSM. Similar pattern of use of environmental covariates have also been reported by Lamichhane et al. (2019) while reviewing DSM of SOC across the world. Nevertheless, in case where multiple covariates are being used, potential feature selection methods should be explored for selecting relevant covariates for prediction modelling (Chandrashekar and Sahin, 2014).

4.6. Predictive models

Different algorithms (predictive models) are used to establish quantitative relationships between input (predictor) and output (target) variables by subjecting a training dataset to a regression and/or classification process (Heung et al., 2016). High-level computer-based programming languages such as R and Python have frequently been used to apply different ML models for DSM. A simple regression model refers to prediction of a target variable ‘y’ based on the relationships between several predictor variables ($x_1, x_2, \dots, x_k$) (Zeraatpisheh et al., 2020). Here, ‘y’ is the dependent variable and $x_1, x_2, \dots, x_k$ are the independent variables. Linear models tend to fit a linear equation to the measured data. Mitran et al. (2019) have used Multiple Linear Regression (MLR) to predict particle size distribution in the state of Karnataka. However, linear regression models are not competent in terms of dealing with non-linear relationships among the target variable and predictors. Any form of regression

model when used with kriging of residuals develops into a hybrid model such as regression kriging (RK). As per this review, 11% of the articles have used MLR and a majority of 37% of the articles have used RK as predictive models.

Tree models are robust ML techniques gaining popularity in the recent years (Heung et al., 2016). The most basic form of tree models is the classification and regression tree (CART), which produces a tree-based structure of its predictor variables for the decision purpose. A better version of CART is the random forest (RF), which follows same concepts as other tree-based models, but, instead of a single tree, RF produces multiple decision trees from the input variables. The final result or decision is the outcome of an ensemble of all the trees (Heung et al., 2016). RF is considered to be capable of handling large datasets (including numerical, categorical, and noisy datasets), establishing non-linear relationships, and computing at a faster speed (Khaledian & Miller, 2020). A hybrid approach of RF regression and kriging of residuals has been performed in some of the articles including Hinge et al. (2018) and Dharumarajan et al. (2021b). Other potential tree-based models include Boosted Regression Tree (BRT) and cubist. An extension of the RF model is Quantile Regression Forest (QRF), which has been adopted in several DSM studies in India (Dharumarajan et al., 2019, 2020b). Artificial Neural Network (ANN) is also one of the robust ML methods for DSM. Three layers of neurons viz. input neurons (predictors), hidden neurons, and output neurons (target variable) form the major components of ANN. ANN is also capable of establishing non-linear relationships among covariates and handling complex datasets (Khaledian & Miller, 2020). Deep Learning (DL) techniques like Convolutional Neural Network (CNN) are improved versions of neural networks used for DSM in the recent years. Support Vector Machine (SVM) is another potential ML tool, which optimizes the hyperplane to avoid the overlap of data between the non-separable classes (Heung et al., 2016). When used for regression, it is termed as Support Vector Regression (SVR). RF has been used in 20% of the DSM articles in

India, whereas other ML methods such as QRF (13%), ANN (7%), and SVM (4%) have been used less frequently (Fig. 2d.). However, use of only single models for DSM by a majority of 86% of the articles create a major research gap. Only 14% of the articles have used multiple predictive models, which is not enough to identify the competent models. Hence, there is a need to consider multiple ML models for identifying the competent ones in terms of prediction accuracy and robustness (Heung et al., 2016). Nevertheless, field experience of the local pedologists should be linked with the mathematical modelling for strengthening the decision-making process.

4.7. Evaluation approaches

Evaluation (validation) of the prediction maps is necessary to assess the level of accuracy and uncertainty (Dharumarajan et al., 2020b). Minasny et al. (2013) have discussed three different types of evaluation approaches viz. cross-validation, data splitting, and external validation (independent validation). Data splitting technique involves splitting of the input dataset into training and testing datasets, which are used for model calibration and validation, respectively. Cross-validation refers to omitting one (leave-one-out method) or a subset (n-fold cross-validation method) of observed values and the rest of the values are trained to predict the omitted values for evaluation purposes. External validation needs independent sampling (collecting additional samples) for evaluation purposes. In each case the correspondence between observed and predicted values are compared using suitable metrics for determining the prediction accuracy. This review identifies that 76% of the DSM studies in India have adopted the data splitting technique for model evaluation. Cross-validation and independent validation methods have been adopted in 19% and 5% of the articles, respectively (Fig. 2e.). Lack of sufficient data points, requirements for extra effort, budget, and time might have been the reasons for the fewer adoption of independent validation in India. However, data-splitting technique is considered as an internal assessment unless and

until the samples are collected using a probability sampling method (Brus et al., 2011). Also, the evaluation estimates of cross-validation are often under-optimistic (Wadoux et al., 2020a). Hence, independent validation is encouraged for obtaining unbiased error estimates (Lamichhane et al., 2019). Coefficient of determination (R^2), Concordance Correlation Coefficient (CCC), and Root Mean Squared Error (RMSE) are some of the frequently considered evaluation parameters in DSM. Accuracies can vary across soil properties, soil depths, sample sizes, prediction models, and mapping strategies. For example, in DSM studies in India, the R^2 for soil carbon is observed to be varying from 0.08 to 0.95 across different articles (not presented in the table). Hence, suitable strategies need to be figured out for obtaining soil maps with higher accuracy. Another research gap arises as 66% of DSM articles in India have not presented the uncertainty maps (Fig. 2f.). Uncertainty maps are necessary for reducing the risk of failure of projects as well as for convincing the funding agencies for additional field works where map uncertainties are higher (Arrouays et al., 2020a, 2020b).

5. DSM applications

In the recent years, DSM is considered as an essential tool for land resource management and it has resulted in numerous DSM interventions across the world (Wadoux et al., 2020a). Apart from the global initiatives like GlobalSoilMap and Global Soil Partnership, many nations have initiated their DSM focus groups such as USDA-NRCS DSM focus team in USA and OzDSM in Australia. Recently, USDA-NRCS has formulated an ambitious plan viz. 'Soils2026' to prepare legacy data based digital soil maps for the USA by 2026 at a 30m resolution (Thompson et al., 2020). In Australia, DSM has already moved to an operational stage, where digital soil maps are now used by the federal and state agencies for various real-world applications such as land suitability classification and erosion assessments (Kidd et al., 2020). Similarly, national digital soil maps in France have also been effectively used by

different stakeholders including farmers for increasing SOC content of their farms, governmental bodies for assessing site contaminations, private companies for selling fertilizers, and farmer development organisations for irrigation management purposes (Arrouays et al., 2020c).

DSM products in India can essentially be utilized for tackling two major issues of the country viz. land degradation and food security (Dharumarajan et al., 2019). DSM is also an effective tool for managing soil health, environment, natural resources, and different ecosystem services (Sreenivas et al., 2014). National extent digital soil maps will be helpful in delineating dominant soil types, suitable crop growing areas, affected lands, and management zones for planning purposes (Dharumarajan et al., 2017). DSM based land capability classes can guide policy makers and land use planners for adopting proper land use plans (Dharumarajan et al., 2021c, 2021d). Spatial distribution of carbon stocks can be used for identifying potential carbon sequestration zones, which holds relevance in terms of mitigating climate change, global carbon management, and carbon trading (Dharumarajan et al., 2021a). Assessments of land degradation, desertification, drought risk, and soil erosion could be performed using DSM for designing suitable region-specific action plans (Arrouays et al., 2014). DSM can also be targeted for supervising vast areas of problematic soils of the country including acid soils, salt-affected soils, waterlogged areas, and degraded lands (Mitran et al., 2021).

Fine resolution (<10 m) digital soil maps of hydraulic properties such as field capacity, permanent wilting point, and infiltration rate can be employed for improving site-specific agricultural water management by the farmers (Dharumarajan and Hegde, 2020; Santra et al., 2021). Utilizing DSM for recommending site-specific (precise) nutrient recommendations to the farmers are also of major interest for researchers, crop planners, and policy makers (Xu et al., 2018b). However, better strategies might be required while mapping for nutrient elements

in order to account for the small and marginal sized agricultural land holdings of the Indian farms, diversified cropping systems, variable doses of nutrient applications, dissimilar agronomic management practices, and non-static nature of nutrient elements. DSM in India can also be potentially exploited for assisting the Soil Health Card (SHC) project (<https://soilhealth.dac.gov.in/>) for rapid generation of soil health cards for the farmers. Nevertheless, researchers all around the world can use the Indian digital soil maps for research oriented modelling and simulation approaches.

6. Conclusion and outlook

Current alarming situations of land degradation and soaring food demand for the growing population necessitate adoption of a fully-fledged DSM in India. In this paper, we have systematically reviewed the common trends of DSM in India by gathering relevant information from published peer reviewed journal articles, which have prepared digital soil maps in India following SCORPAN approach. Only 35 publications have been published, most of which are being limited only to a few parts of the country and only two national extent DSM interventions have been carried out indicating that DSM is still at a primitive stage in India. However, an increased number of publications from the year 2017 onwards suggest that there is a growing interest for DSM in India over the past few years. DSM has been performed at various mapping extents with very few fine resolution DSM attempts. But, considering the small and marginal sized agricultural land holdings of the country, the future DSM attempts in India should focus on preparing digital soil maps at fine spatial resolutions (<10 m). As part of preparation of fine resolution maps, DSM efforts should also adopt higher sampling density, optimized sampling strategies, and fine resolution RS products as environmental covariates. Considering the vastness of the country, national extent DSM might be challenging. Hence, strategic plans are needed to support the local level institutes for performing state or agroclimatic zone based DSM using globally consistent models under

the supervision of national level institutes. Moreover, the scope of DSM needs to be expanded to soil classes, GlobalSoilMap specified soil properties, and other locally relevant target variables having importance for multiple sectors.

Previous soil surveys have generated huge legacy soil data for India, most of which are available as printed books and reports. Lack of digitally downloadable harmonized soil profile database is a major block to DSM in India. A national digital soil database is yet to be developed with complete descriptions regarding source of soil information, sampling strategies, date of collection, analytical methods, and legal framework under which they can be utilized nationally and internationally. This requires a strong collaborative effort involving different organizations such as NBSS&LUP, SLUSI, ISRO-NRSC, IIRS, Indian Institute of Technologies (IITs), Indian Council of Agricultural Research (ICAR) institutes, Space Applications Centre (SAC), State Agricultural Universities (SAUs), Regional Research and Technology Transfer Stations (RRTTSs), Krishi Vigyan Kendras (KVKs), and other Non-governmental organizations along with technological support from different internationally focused organisations such as FAO and ISRIC.

Our review identifies that covariates representing relief, organism, and climate have been used most frequently for DSM in India, whereas the use of other potential covariates such as legacy soil maps, geology maps has been limited. Hence, future DSM should try to incorporate more possible covariates under the SCORPAN domain with suitable feature selection methods. RK has been the most frequently employed predictive model followed by RF and QRF. However, only 14% of articles have included multiple prediction models, which is insufficient to draw relevant conclusions regarding competence and robustness of the predictive models. Hence, future DSM attempts are expected to include more numbers of ML models for conducting comparative studies. This review also identifies that only 5 % of articles have performed independent validation and a majority of 66% of articles have not

presented the uncertainty maps. Therefore, improvements are also needed in terms of evaluation strategies and communicating uncertainty maps to the end users. Besides improving the technical aspects of DSM, sufficient training and workshops on geospatial soft wares and coding skills can support extensive adoption of DSM in India. Nevertheless, an Indian soil information system need to be developed to make all DSM products available at a wide range of digital platforms including websites, web map services, mobile, and desktop applications to reach all kinds of end users including the scientists and practitioners involved with DSM in India and abroad.

Acknowledgement

We acknowledge the Department of Science and Technology (DST), Government of India for providing financial assistantship (INSPIRE fellowship) to the first author. We greatly acknowledge the anonymous reviewers for their valuable suggestions to improve the manuscript.

References

- Agriculture Census Division, 2019. All India Report on Number and Area of Operational Holdings, Agricultural Census (2015-16), Government of India, New Delhi.
- Ahmed, N., Mahtab, A., Agrawal, R., Jayaprasad, P., Pathan, S.K., Ajai, Singh, D.K., Singh, A.K., 2007. Extraction and validation of Cartosat-1 DEM. *J. Indian Soc. Remote Sens.* 35 (2), 121–127. <https://doi.org/10.1007/BF02990776>
- Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M. d. L., Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P.A., Thompson, J.A., Zhang, G.L., 2014. GlobalSoilMap. toward a fine-resolution global grid of soil properties. In:

Advances in Agronomy. Academic Press Inc., pp. 93–134. <https://doi.org/10.1016/B978-0-12-800137-0.00003-0>

Arrouays, D., Lagacherie, P., Hartemink, A.E., 2017. Digital soil mapping across the globe. *Geoderma Reg.* 9, 1–4. <https://doi.org/10.1016/j.geodrs.2017.03.002>

Arrouays, D., McBratney, A., Bouma, J., Libohova, Z., Richer-de-Forges, A. C., Morgan, C. L. S., Roudier, P., Poggio, L., Mulder, V. L., 2020a. Impressions of digital soil maps: The good, the not so good, and making them ever better. *Geoderma Regional* 20, e00255. <https://doi.org/10.1016/j.geodrs.2020.e00255>

Arrouays, D., Poggio, L., Salazar Guerrero, O. A., Mulder, V. L., 2020b. Digital soil mapping and GlobalSoilMap. Main advances and ways forward. *Geoderma Regional* 21, e00265. <https://doi.org/10.1016/j.geodrs.2020.e00265>

Arrouays, D., Richer-de-Forges, A.C., Héliès, F., Mulder, V.L., Saby, N.P.A., Chen, S., Martin, M.P., Román Dobarco, M., Follain, S., Jolivet, C., Laroche, B., Loiseau, T., Cousin, I., Lacoste, M., Ranjard, L., Toutain, B., le Bas, C., Eglin, T., Bardy, M., Antoni, V., Meersmans, J., Ratié, C., Bispo, A., 2020c. Impacts of national scale digital soil mapping programs in France. *Geoderma Regional* 23, e00337. <https://doi.org/10.1016/J.GEODRS.2020.E00337>

Bangroo, S. A., Najar, G. R., Achin, E., Truong, P. N., 2020. Application of predictor variables in spatial quantification of soil organic carbon and total nitrogen using regression kriging in the North Kashmir forest Himalayas. *Catena* 193, 104632. <https://doi.org/10.1016/j.catena.2020.104632>

Batjes, N.H., Ribeiro, E., van Oostrum, A., 2020. Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). *Earth System Science Data* 12, 299–320. <https://doi.org/10.5194/ESSD-12-299-2020>

Behera, S. K., Mathur, R. K., Shukla, A. K., Suresh, K., Prakash, C., 2018. Spatial variability of soil properties and delineation of soil management zones of oil palm plantations grown in a hot and humid tropical region of southern India. *Catena* 165, 251–259. <https://doi.org/10.1016/j.catena.2018.02.008>

Bhattacharyya, T., Sarkar, D., Sehgal, J. L., Velaytham, M., Gajbhiye, K. S., Nagar, A. P. and Nimkhedkar, S. S., 2009. Soil Taxonomic Database of India and the States (1: 250,000 scale), NBSS&LUP NBSSLUP Publ. 143, National Bureau of Soil Survey and Land Use Planning, Nagpur, India.

Bhattacharyya, T., Pal, D.K., Mandal, C., Chandran, P., Ray, S.K., Sarkar, D., Velmourougane, K., Srivastava, A., Sidhu, G.S., Singh, R.S., Sahoo, A.K., Dutta, D., Nair, K.M., Srivastava, R., Tiwary, P., Nagar, A.P., Nimkhedkar, S.S., 2013. Soils of India: historical perspective, classification and recent advances. *Current Science* 104, 1308–1323.

Bhunia, G. S., Kumar Shit, P., Pourghasemi, H. R., 2019a. Soil organic carbon mapping using remote sensing techniques and multivariate regression model. *Geocarto International* 34 (2), 215–226. <https://doi.org/10.1080/10106049.2017.1381179>

Bhunia, G. S., Shit, P. K., Pourghasemi, H. R., Edalat, M., 2019b. Prediction of soil organic carbon and its mapping using regression analyses and remote sensing data in GIS and R. In: *Spatial Modeling in GIS and R for Earth and Environmental Sciences* (pp. 429–450). Elsevier. <https://doi.org/10.1016/b978-0-12-815226-3.00019-3>

- Brevik, E.C., Cerdà, A., Mataix-Solera, J., Pereg, L., Quinton, J.N., Six, J., Van Oost, K., 2015. The interdisciplinary nature of SOIL. *Soil* 1, 117–129. <https://doi.org/10.5194/soil-1-117-2015>
- Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. *Eur. J. Soil Sci.* 62 (3), 394–407. <https://doi.org/10.1111/j.1365-2389.2011.01364.x>
- Chandran, P., Tiwary, P., Bhattacharyya, T., Mandal, C., Prasad, J., Ray, S. K., Sarkar, D., Pal, D. K., Mandal, D. K., Sidhu, G. S., Nair, K. M., Sahoo, A. K., Das, T. H., Singh, R. S., Srivastava, R., Sen, T. K., Chatterji, S., Patil, N. G., Obireddy, G. P., ... Thakre, S., 2014. Development of soil and terrain digital database for major food-growing regions of India for resource planning. *Current Science*, 107 (9): 1420-1430.
- Chandrashekar, G., Sahin, F., 2014. A survey on feature selection methods. *Computers & Electrical Engineering* 40, 16–28. <https://doi.org/10.1016/J.COMPELECENG.2013.11.024>
- Chhipa, V., Stein, A., Shankar, H., George K, J., Alidoost, F., 2019. Assessing and transferring soil health information in a hilly terrain. *Geoderma* 343, 130–138. <https://doi.org/10.1016/j.geoderma.2019.02.018>
- Dash, P. K., Mishra, A. Saren, S., 2019. Characterization and taxonomic classification of soils under a toposequence located in Eastern India. *Environment and Ecology* 37 (4), 1240-1249.
- de Sousa, L., Poggio, L., Batjes, N., Heuvelink, G., Kempen, B., Riberio, E., Rossiter, D., 2020. SoilGrids 2.0: producing quality-assessed soil information for the globe. *Soil Discussions*, 1–37. <https://doi.org/10.5194/soil-2020-65>

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. <https://doi.org/10.1016/j.geodrs.2017.07.005>

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

Dharumarajan, S., Hegde, R., 2020. Digital mapping of soil texture classes using Random Forest classification algorithm. *Soil Use and Management*. <https://doi.org/10.1111/SUM.12668>

Dharumarajan, S., Vasundhara, R., Suputhra, A., Lalitha, M., Hegde, R., 2020a. Prediction of Soil Depth in Karnataka Using Digital Soil Mapping Approach. *J. Indian Soc. Remote Sens.* 1–8. <https://doi.org/10.1007/s12524-020-01184-7>

Dharumarajan, S., Kalaiselvi, B., Suputhra, A., Lalitha, M., Hegde, R., Singh, S. K., Lagacherie, P., 2020b. Digital soil mapping of key GlobalSoilMap properties in Northern Karnataka Plateau. *Geoderma Regional* 20, e00250. <https://doi.org/10.1016/j.geodrs.2019.e00250>

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

Dharumarajan, S., Hegde, R., Lalitha, M., Vasundhara, R., 2021b. Predicting and Mapping of Soil Hydraulic Properties in Karnataka. *Journal of the Indian Society of Remote Sensing* 2021 49:7 49, 1623–1631. <https://doi.org/10.1007/S12524-021-01336-3>

Dharumarajan, S., Hegde, R., Lalitha, M., 2021c. Modelling of soil depth and hydraulic properties at regional level using environmental covariates- A case study in India. *Geoderma Regional* 27, e00439. <https://doi.org/10.1016/J.GEODRS.2021.E00439>

Dharumarajan, S., Kalaiselvi, B., Lalitha, M., Vasundhara, R., Hegde, R., 2021d. Defining fertility management units and land suitability analysis using digital soil mapping approach. *Geocarto International* <https://doi.org/10.1080/10106049.2021.1926553>

Dwivedi, R.S., Sreenivas, K., 1998. Delineation of salt-affected soils and waterlogged areas in the Indo-Gangetic plains using IRS-1C LISS-III data. *Int. J. Remote Sens.* 19 (14), 2739–2751. <https://doi.org/10.1080/014311698214488>

Garg, P.K., Garg, R.D., Shukla, G., Srivastava, H.S., 2020. Digital Mapping of Soil Landscape Parameters, *Studies in Big Data*. Springer Singapore, Singapore. <https://doi.org/10.1007/978-981-15-3238-2>

Gomez, C., Dharumarajan, S., Féret, J.-B., Lagacherie, P., Ruiz, L., Sekhar, M., 2019. Use of Sentinel-2 time-series images for classification and uncertainty analysis of inherent biophysical property: case of soil texture mapping. *Remote Sensing* 11 (5), 565. <https://doi.org/10.3390/rs11050565>

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.

Greiner, L., Nussbaum, M., Papritz, A., Zimmermann, S., Gubler, A., Grêt-Regamey, A., Keller, A., 2018. Uncertainty indication in soil function maps - Transparent and easy-to-use information to support sustainable use of soil resources. *SOIL* 4, 123–139. <https://doi.org/10.5194/SOIL-4-123-2018>

Grunwald, S., 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma* 152, 195–207. <https://doi.org/10.1016/j.geoderma.2009.06.003>

Grunwald, S., Thompson, J.A., Boettinger, J.L., 2011. Digital Soil Mapping and Modeling at Continental Scales: Finding Solutions for Global Issues. *Soil Sci. Soc. Am. J.* 75, 1201–1213. <https://doi.org/10.2136/sssaj2011.0025>

Guo, P. T., Li, M. F., Luo, W., Tang, Q. F., Liu, Z. W., Lin, Z. M., 2015. Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach. *Geoderma* 237–238, 49–59. <https://doi.org/10.1016/j.geoderma.2014.08.009>

Hartemink, A.E., 2015. On global soil science and regional solutions. *Geoderma Regional* 5, 1–3. <https://doi.org/10.1016/j.geodrs.2015.02.001>

Hengl, T., Jesus, J.M. de, Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE* 12, e0169748. <https://doi.org/10.1371/JOURNAL.PONE.0169748>

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. <https://doi.org/10.1016/j.geoderma.2015.11.014>

Hinge, G., Surampalli, R. Y., Goyal, M. K., 2018. Prediction of soil organic carbon stock using digital mapping approach in humid India. *Environmental Earth Sciences* 77 (5). <https://doi.org/10.1007/s12665-018-7374-x>

Hudson, B.D., 1992. The soil survey as paradigm-based science. *Soil Sci. Soc. Am. J.* 56, 836–841. <https://doi.org/10.2136/sssaj1992.03615995005600030027x>

James, K.L., Randall, N.P., Haddaway, N.R., 2016. A methodology for systematic mapping in environmental sciences. *Environ. Evid.* 5 (1), 1–13. <https://doi.org/10.1186/s13750-016-0059-6>

Jenny, H., 1941. *Factors of soil formation-a system of quantitative pedology*. McGraw- Hill, New York, pp. 281.

Kalambukattu, J.G., Kumar, S., Arya Raj, R., 2018. Digital soil mapping in a Himalayan watershed using remote sensing and terrain parameters employing artificial neural network model. *Environemntal Earth Sciences* *Environ. Earth Sci.* 77 (5), 1–14. <https://doi.org/10.1007/s12665-018-7367-9>

Khaledian, Y., Miller, B.A., 2020. Selecting appropriate machine learning methods for digital soil mapping. *Appl. Math. Model* 81, 401–418. <https://doi.org/10.1016/j.apm.2019.12.016>

Kidd, D., Searle, R., Grundy, M., McBratney, A., Robinson, N., O'Brien, L., Zund, P., Arrouays, D., Thomas, M., Padarian, J., Jones, E., Bennett, J.M.L., Minasny, B., Holmes, K., Malone, B.P., Liddicoat, C., Meier, E.A., Stockmann, U., Wilson, P., Wilford, J., Payne, J., Ringrose-Voase, A., Slater, B., Odgers, N., Gray, J., van Gool, D., Andrews, K., Harms, B., Stower, L., Triantafilis, J., 2020. Operationalising digital soil mapping – Lessons from Australia. *Geoderma Regional* 23, e00335. <https://doi.org/10.1016/J.GEODRS.2020.E0033>

Krishnan, P., Bourgeon, G., Seen, D. L., Nair, K.M., Prasanna, R., Srinivas, S., Muthusankar, G., Dufy, L., Ramesh, B. R., 2007. Organic carbon stock map for soils of Southern India: A multifactorial approach. *Current Science* 93 (5), 706-710.

- Kudrat, M., Tiwari, A.K., Saha, S.K., 1992. Soil resource mapping using IRS-1A-LISS II digital data-a case study of Kandi area adjacent to Chandigarh-India. *Int. J. Remote Sens.* 13 (17), 3287–3302. <https://doi.org/10.1080/01431169208904119>
- Kuriakose, S. L., Devkota, S., Rossiter, D. G., Jetten, V. G., 2009. Prediction of soil depth using environmental variables in an anthropogenic landscape, a case study in the Western Ghats of Kerala, India. *Catena* 79 (1), 27–38. <https://doi.org/10.1016/j.catena.2009.05.005>
- Lagacherie, P., Bui, S., Constantin, J., Dharumarajan, S., Ruiz, L., Sekhar, M., 2022. Evaluating the impact of using digital soil mapping products as input for spatializing a crop model: The case of drainage and maize yield simulated by STICS in the Berambadi catchment (India). *Geoderma* 406, 115503. <https://doi.org/10.1016/J.Geoderma.2021.115503>
- Lagacherie, P., McBratney, A.B., 2006. Spatial Soil Information Systems and Spatial Soil Inference Systems: Perspectives for Digital Soil Mapping. *Dev. Soil Sci.* 3–22. [https://doi.org/10.1016/S0166-2481\(06\)31001-X](https://doi.org/10.1016/S0166-2481(06)31001-X)
- Lalitha, M., Dharumarajan, S., Suputhra, A., Kalaiselvi, B., Hegde, R., Reddy, R., Prasad, C.S., Harindranath, C., Dwivedi, B., 2021. Spatial prediction of soil depth using environmental covariates by quantile regression forest model. *Environmental Monitoring and Assessment* 193. <https://doi.org/10.1007/S10661-021-09348-9>
- Lamichhane, S., Kumar, L., Wilson, B., 2019. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications : A review. *Geoderma* 352, 395–413. <https://doi.org/10.1016/j.geoderma.2019.05.031>
- Manchanda, M.L., Hilwig, F.W., 1981. Visual interpretation of computer transformed Landsat imagery for salt affected areas of part of Haryana. *J. Indian Soc. Photo-Interpretation Remote Sens.* 9 (2), 1–11. <https://doi.org/10.1007/BF02991458>

Manchanda, M.L., Kudrat, M., Tiwari, A.K., 2002. Manchanda, Kudrat, 2002. Soil survey and mapping using remote sensing. *Trop. Ecol.* 43 (1), 61–74.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)

McBratney, A., Field, D. J., Koch, A., 2014. The dimensions of soil security. *Geoderma* 213, 203–213. <https://doi.org/10.1016/j.geoderma.2013.08.013>

Miller, B.A., Schaetzl, R.J., 2016. History of soil geography in the context of scale. *Geoderma* 264, 284–300. <https://doi.org/10.1016/J.Geoderma.2015.08.041>

Minasny, B., Hartemink, A.E., 2011. Predicting soil properties in the tropics. *Earth-Science Rev.* 106, 52–62. <https://doi.org/10.1016/j.earscirev.2011.01.005>

Minasny, B., McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. *Geoderma* 264, 301–311. <https://doi.org/10.1016/j.geoderma.2015.07.017>

Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Digital mapping of soil carbon. In: Sparks, D.L. (Ed.), *Advances in Agronomy*. Academic Press, pp. 1–47. <https://doi.org/10.1016/B978-0-12-405942-9.00001-3>

Mitran, T., Mishra, U., Lal, R., Ravisankar, T., Sreenivas, K., 2018. Spatial distribution of soil carbon stocks in a semi-arid region of India. *Geoderma Regional* 15, e00192. <https://doi.org/10.1016/j.geodrs.2018.e00192>

Mitran, T., Solanky, V., Janakirama Suresh, G., Sujatha, G., Sreenivas, K., Ravisankar, T., 2019. Predictive mapping of surface soil texture in a semiarid region of India through geostatistical modeling. *Modeling Earth Systems and Environment* 5 (2), 645–657. <https://doi.org/10.1007/s40808-018-0556-4>

- Mitran, T., Meena, R.S., Chakraborty, A., 2021. Geospatial Technologies for Crops and Soils: An Overview. *Geospatial Technologies for Crops and Soils* 1–48. https://doi.org/10.1007/978-981-15-6864-0_1
- Mondal, A., Khare, D., Kundu, S., Mondal, S., Mukherjee, S., Mukhopadhyay, A., 2017. Spatial soil organic carbon (SOC) prediction by regression kriging using remote sensing data. *The Egyptian Journal of Remote Sensing and Space Science* 20 (1), 61–70. <https://doi.org/10.1016/j.ejrs.2016.06.004>
- Murthy, R. S., Pandey, S., 1983. Soil Map of India (1: 6.3 m). National Bureau of Soil Survey and Land Use Planning, Nagpur, India.
- Okonkwo, E. I., Corstanje, R., Kirk, G. J. D., 2018. Digital soil assessment for quantifying soil constraints to crop production: a case study for rice in Punjab, India. *Soil Use and Management* 34 (4), 533–541. <https://doi.org/10.1111/sum.12446>
- Oliver, M.A., 1987. Geostatistics and its application to soil science. *Soil Use and Management* 3, 8–20. <https://doi.org/10.1111/J.1475-2743.1987.TB00703.X>
- Poggio, L., de Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL* 7, 217–240. <https://doi.org/10.5194/SOIL-7-217-2021>
- Reddy, N.N., Chakraborty, P., Roy, S., Singh, K., Minasny, B., McBratney, A.B., Biswas, A., Das, B.S., 2021. Legacy data-based national-scale digital mapping of key soil properties in India. *Geoderma* 381, 114684. <https://doi.org/10.1016/j.geoderma.2020.114684>
- Saha, S.K., Singh, B.M., 1991. Soil erosion assessment and mapping of the Algar river watershed (Uttar Pradesh) using remote sensing technique. *J. Indian Soc. Remote Sens.* 19 (2), 67–76. <https://doi.org/10.1007/BF03008122>

- Sahrawat, K.L., Wani, S.P., Pardhasaradhi, G., Murthy, K.V.S., 2010. Diagnosis of secondary and micronutrient deficiencies and their management in rainfed agroecosystems: Case study from Indian semi-arid tropics. *Commun. Soil Sci. Plant Anal.* 41, 346–360. <https://doi.org/10.1080/00103620903462340>
- 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. <https://doi.org/10.1016/j.geodrs.2017.03.003>
- Santra, P., Kumar, M., Panwar, N.R., Das, B.S., 2017b. Digital soil mapping and best management of soil resources: A brief discussion with few case studies, in: *Adaptive Soil Management: From Theory to Practices*. Springer Singapore, pp. 3–38. https://doi.org/10.1007/978-981-10-3638-5_1
- 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* 319–355. https://doi.org/10.1007/978-981-15-6864-0_9
- Sarkar, S., Roy, A. K., Martha, T. R., 2013. Soil depth estimation through soil-landscape modelling using regression kriging in a Himalayan terrain. *International Journal of Geographical Information Science* 27 (12), 2436–2454. <https://doi.org/10.1080/13658816.2013.814780>
- Sharma, J.K., Kalia, M.M., 1987. Use of aerial photographs in the study of watershed characteristics. *J. Indian Soc. Remote Sens.* 15 (2), 1–6. <https://doi.org/10.1007/BF03017778>
- Shukla, A.K., Behera, S.K., Lenka, N.K., Tiwari, P.K., Prakash, C., Malik, R.S., Sinha, N.K., Singh, V.K., Patra, A.K., Chaudhary, S.K., 2016. Spatial variability of soil micronutrients in

the intensively cultivated Trans-Gangetic Plains of India. *Soil and Tillage Research* 163, 282–289. <https://doi.org/10.1016/J.STILL.2016.07.004>

Shukla, G., Garg, R. D., Srivastava, H. S., Garg, P. K., 2018. An effective implementation and assessment of a random forest classifier as a soil spatial predictive model. *International Journal of Remote Sensing* 39 (8), 2637–2669. <https://doi.org/10.1080/01431161.2018.1430399>

Singh, S.K., Chatterji, S., 2018. Land Resource Inventory for Agricultural Land Use Planning Using Geospatial Techniques. Springer, Cham, pp. 163–183. https://doi.org/10.1007/978-3-319-78711-4_9

Sreenivas, K., Sujatha, G., Sudhir, K., Kiran, D. V., Fyze, M. A., Ravisankar, T., Dadhwal, V. K., 2014. Spatial assessment of soil organic carbon density through random forests based imputation. *Journal of the Indian Society of Remote Sensing* 42 (3), 577–587. <https://doi.org/10.1007/s12524-013-0332-x>

Sreenivas, K., Dadhwal, V.K., Kumar, S., Harsha, G.S., Mitran, T., Sujatha, G., Suresh, G.J.R., Fyze, M.A., Ravisankar, T., 2016. Digital mapping of soil organic and inorganic carbon status in India. *Geoderma* 269, 160–173. <https://doi.org/10.1016/j.geoderma.2016.02.002>

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 Regional* 22, e00294. <https://doi.org/10.1016/J.GEODRS.2020.E00294>

Tiwari, S. K., Saha, S. K., Kumar, S., 2015. Prediction modeling and mapping of soil carbon content using artificial neural network, hyperspectral satellite data and field spectroscopy. *Advances in Remote Sensing* 04 (01), 63–72. <https://doi.org/10.4236/ars.2015.41006>

United Nations, Department of Economic and Social Affairs, Population Division (2019). *World Population Prospects 2019: Highlights* (ST/ESA/SER.A/423).

Velmurugan, A., Carlos, G.G., 2009. Soil resource assessment and mapping using remote sensing and GIS. *J. Indian Soc. Remote Sens.* 37 (3), 511–525. <https://doi.org/10.1007/s12524-009-0045-3>

Vision 2050 (2015), Indian Council of Agricultural Research, New Delhi.

Wadoux, A.M.J.C., Minasny, B., McBratney, A.B., 2020a. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews* 210. <https://doi.org/10.1016/J.EARSCIREV.2020.103359>

Wadoux, A.M.J.C., Samuel-Rosa, A., Poggio, L., Mulder, V.L., 2020b. A note on knowledge discovery and machine learning in digital soil mapping. *European Journal of Soil Science* 71, 133–136. <https://doi.org/10.1111/EJSS.12909>

Xu, Y., Smith, S. E., Grunwald, S., Abd-Elrahman, A., Wani, S. P., 2017a. Incorporation of satellite remote sensing pan-sharpened imagery into digital soil prediction and mapping models to characterize soil property variability in small agricultural fields. *ISPRS Journal of Photogrammetry and Remote Sensing* 123, 1–19. <https://doi.org/10.1016/j.isprsjprs.2016.11.001>

Xu, Y., Smith, S. E., Grunwald, S., Abd-Elrahman, A., Wani, S. P., 2017b. Evaluating the effect of remote sensing image spatial resolution on soil exchangeable potassium prediction

models in smallholder farm settings. *Journal of Environmental Management* 200, 423–433.

<https://doi.org/10.1016/j.jenvman.2017.06.017>

Xu, Y., Smith, S. E., Grunwald, S., Abd-Elrahman, A., Wani, S. P., 2018a. Effects of image pansharpening on soil total nitrogen prediction models in South India. *Geoderma* 320, 52–66.

<https://doi.org/10.1016/j.geoderma.2018.01.017>

Xu, Y., Smith, S. E., Grunwald, S., Abd-Elrahman, A., Wani, S. P., Nair, V. D., 2018b. Estimating soil total nitrogen in smallholder farm settings using remote sensing spectral indices and regression kriging. *Catena* 163, 111–122.

<https://doi.org/10.1016/j.catena.2017.12.011>

Zeraatpisheh, M., Jafari, A., Bagheri Bodaghabadi, M., Ayoubi, S., Taghizadeh-Mehrjardi, R., Toomanian, N., Kerry, R., Xu, M., 2020. Conventional and digital soil mapping in Iran: Past, present, and future. *Catena* 188, 104424. <https://doi.org/10.1016/j.catena.2019.104424>

Zhang, G. L., Liu, F., Song, X. D., 2017. Recent progress and future prospect of digital soil mapping: A review. *J. Integr. Agric.* 16 (12), 2871-2885. [https://doi.org/10.1016/S2095-3119\(17\)61762-3](https://doi.org/10.1016/S2095-3119(17)61762-3)

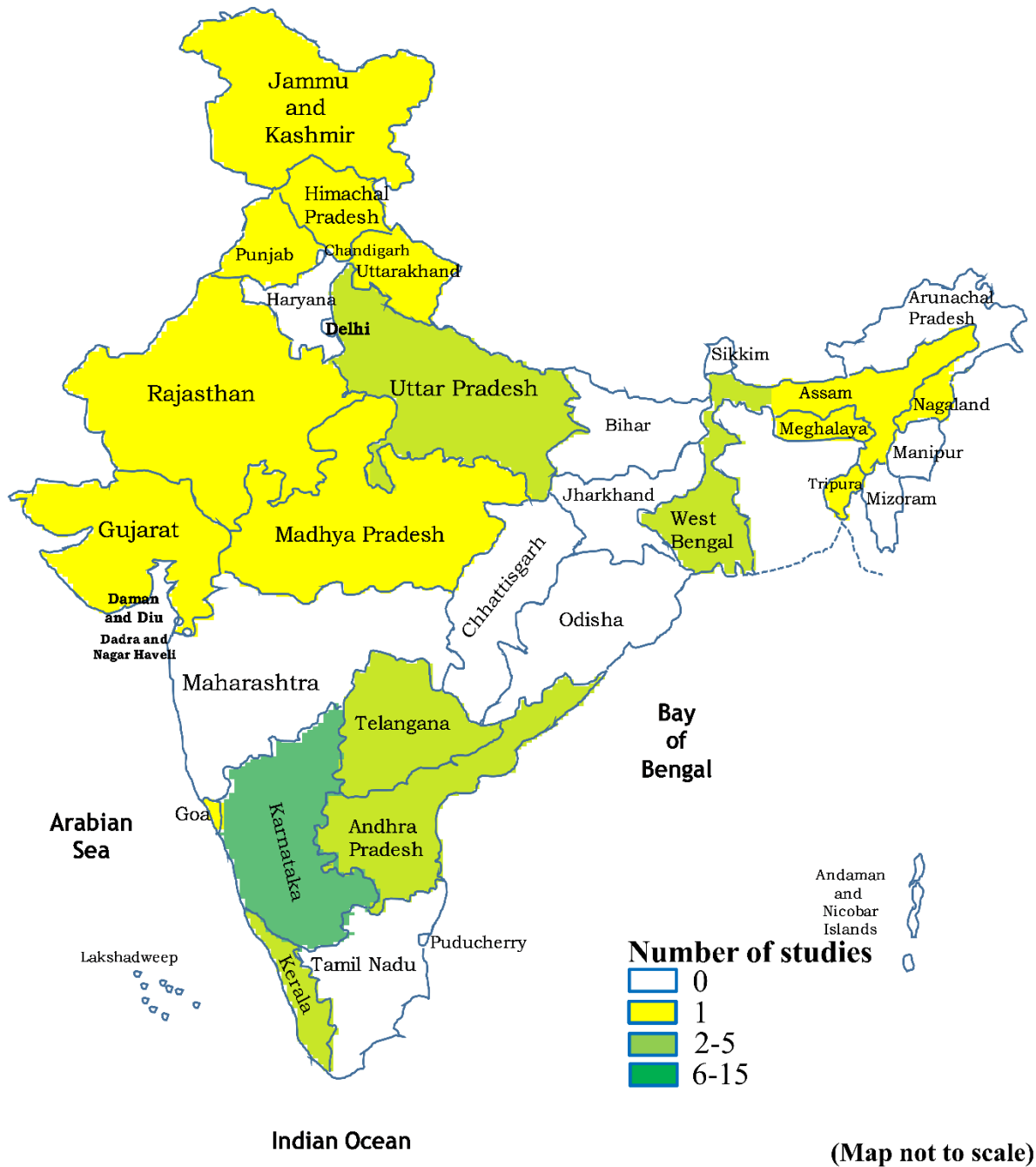


Fig. 1. State and Union Territory wise geographic distribution of DSM activities in India

(Note: This figure does not include DSM activities performed at a national or global extent)

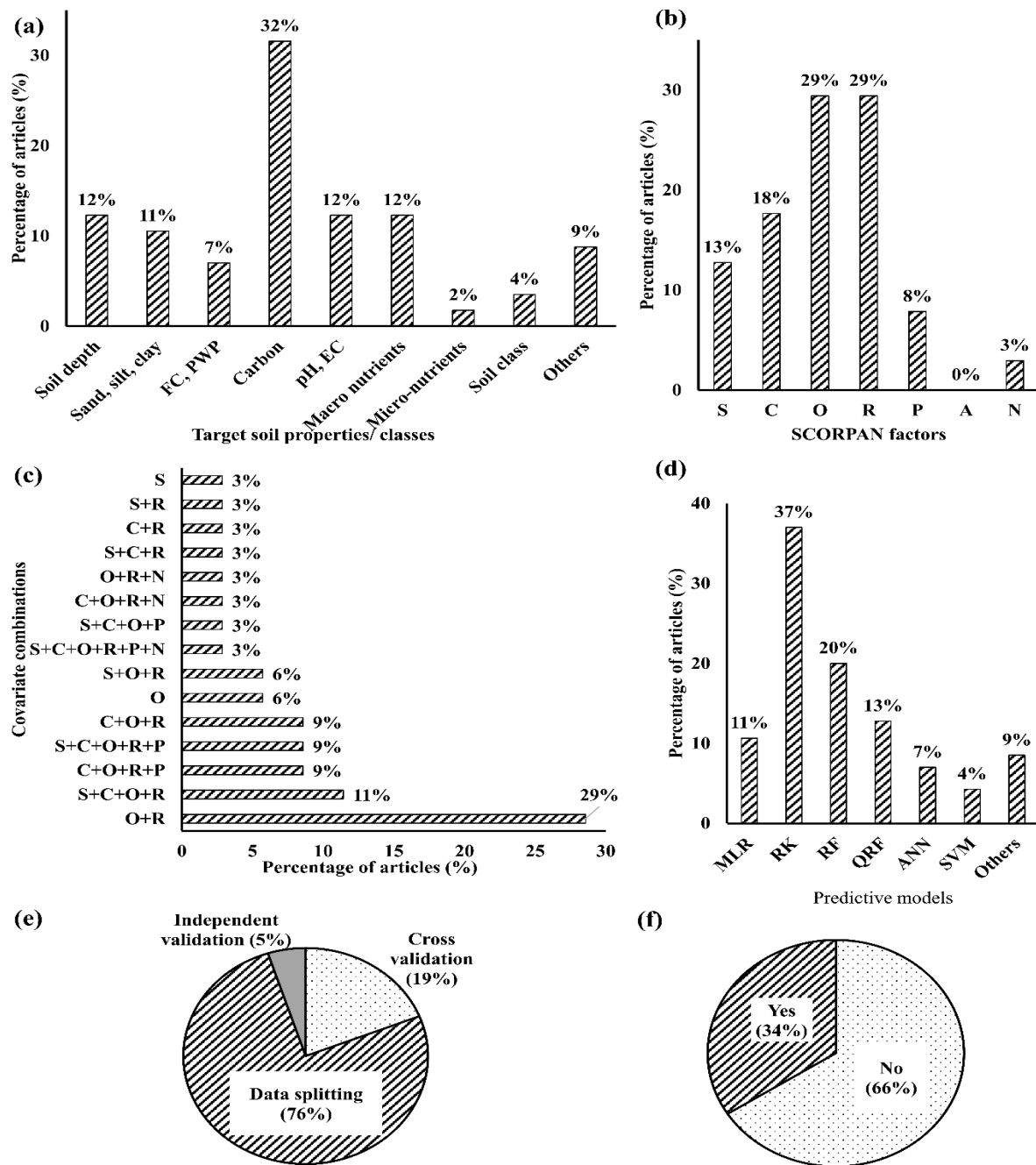


Fig. 2. Summary of key outputs (a) frequency of soil properties or classes predicted (others include CEC, coarse fragments, soil colour, and soil albedo); (b) frequency of use of environmental covariates; (c) frequency of use of covariate combinations; (d) frequency of use of predictive models (Note: others include models which are used in less than two articles); (e) model evaluation approaches; (f) mapping of uncertainty.

Table 1. An overview of DSM approaches in India

Sl. No.	Reference	Year	Location (state)	Geographical area (km ²)	Mapping extent ^a	SD (number of samples per km ²)	Method of sampling ^b	Maximum depth of sampling (m)	Target variables ^c	Environmental covariates	Prediction models (Best model in comparison studies underlined) ^d	Validation approach (uncertainty maps prepared or not in parenthesis)
1	Krishnan et al., 2007	2007	Goa	88,000	Regional	0.0040	L	0.3	SOC stock	S+C+O+R+P	MART	Data splitting (No)
2	Kuriakose et al., 2009	2009	Kerala	9.5	Local	29.0	R	3	Soil depth	O+R	RK (MLR+kriging)	Data splitting (Yes)
3	Sarkar et al., 2013	2013	West Bengal	330.35	Local	0.4000	R	4	Soil depth	O+R	MLR, <u>RK (MLR+kriging)</u>	Data splitting (Yes)
4	Sreenivas et al., 2014	2014	Andhra Pradesh and Karnataka	4,67,000	Regional	0.0010	StR	0.3	SOC density	S+C+O+R	RF	Data splitting (No)
5	Tiwari et al., 2015	2015	Uttar Pradesh	189	Local	0.3000	R	0.1	SOC	S	ANN	Data splitting (No)
6	Sreenivas et al., 2016	2016	Whole country	3,287,000	National	0.0004	StR	1	SOC density, SIC density	S+C+O+P	RF	Data splitting (No)
7	Santra et al., 2017a	2017	Rajasthan and Gujarat	3,20,000	Regional	0.0003	L	2	Sand content	S+C+R	RK (RF+kriging)	Cross validation (Yes)
8	Dharumarajan et al., 2017	2017	Andhra Pradesh	248	Local	0.5000	R	0.3	pH, EC, SOC	S+O+R	RF	Data splitting (No)
9	Xu et al., 2017a	2017	Telangana	4.65	Local	54.8	StR	0.15	TN, Kex	O+R	RF	Data splitting (No)
10	Xu et al., 2017b	2017	Telangana & Karnataka	10	Local	25.5	StR	0.15	Kex	O+R	RK (MLR+kriging)	Data splitting (No)
11	Mondal et al., 2017	2017	Madhya Pradesh	20,558	Regional	0.0102	StR	0.15	SOC	S+O+R	RK (MLR+kriging)	Data splitting (No)

12	Mitran et al., 2018	2018	Telangana and Karnataka	3,03,000	Regional	0.0010	R	0.3	Total carbon stock	S+C+O+R+P	<u>RK</u> (GWR+kriging), RK (MLR+kriging)	Data splitting (Yes)
13	Shukla et al., 2018	2018	Uttar Pradesh	12,421	Regional	0.4000	R	2	Soil class	S+C+O+R+P+N	RF, CDT, CEB	Data splitting (No)
14	Hinge et al., 2018	2018	Assam, Meghalaya, Nagaland, and Tripura	1,17,800	Regional	0.0020	L	0.5	SOC	C+O+R+N	RK (RF+kriging)	Data splitting and independent validation (No)
15	Xu et al., 2018a	2018	Karnataka	40	Local	6.5	StR	0.15	TN	O+R+N	RF	Data splitting (No)
16	Kalambukattu et al., 2018	2018	Himachal Pradesh	10	Local	7.5	Gr	0.15	SOC and N	O+R	ANN	Data splitting and independent validation (No)
17	Okonkwo et al., 2018	2018	Punjab	50,362	Regional	0.0007	L	0.2	EC and pH	S+C+O+R	ANN	Data splitting (No)
18	Xu et al., 2018b	2018	Telangana	10	Local	25.5	StR	0.15	TN	O+R	RK (MLR+kriging)	Data splitting (No)
19	Gomez et al., 2019	2019	Karnataka	84	Local	1.5	StR	0.15	Textural class	O	SVM	Data splitting (Yes)
20	Mitran et al., 2019	2019	Karnataka	4,656	Local	0.0100	R	0.15	Sand, silt, clay content	O+R	MLR & <u>RK</u> (MLR+kriging)	Data splitting (Yes)
21	Dharumarajan et al., 2019	2019	Karnataka (Location-1)	55.08	Local	16.0	Gr	0.2	pH, EC, SOC	O+R	QRF	Cross validation (Yes)
22	Bhunia et al., 2019a	2019	West Bengal	323.6	Local	0.2	R	0.2	SOC	O	MLR	Data splitting (No)
23	Bhunia et al., 2019b	2019	West Bengal	NA	NA		R	0.15	SOC	O	MLR	Data splitting (No)
24	Chhipa et al., 2019	2019	Uttarakhand	0.4	Plot	240.0	R	0.3	pH, EC, TOC	S+R	RK (MLR+kriging)	Cross validation (No)

25	Bangroo et al., 2020	2020	Jammu and Kashmir	261	Local	0.4000	R	0.2	SOC, TN	O+R	RK (MLR+kriging)	Data splitting (No)
26	Dharumarajan and Hegde, 2020	2020	Andhra Pradesh	1,62,000	Regional	0.0140	L	2	Soil texture	C+O+R	RF	Data splitting (No)
27	Dharumarajan et al., 2020a	2020	Karnataka	1,91,800	Regional	0.0300	L	2	Soil depth	C+O+R	QRF, RK (RF+kriging)	Data splitting (No)
28	Dharumarajan et al., 2020b	2020	Karnataka	3,655	Local	0.0200	R	2	pH, SOC, clay, sand, sand, silt, CEC, FC, PWP	O+R	QRF	Cross validation (Yes)
29	Reddy et al., 2021	2020	Whole country	3,287,000	National	0.0005	L	2	Sand, clay, pH, SOC	C+R	3D-RK (MLR+kriging)	Data splitting (Yes)
30	Dharumarajan et al., 2021a	2021	Kerala and Karnataka	56,763	Regional	0.0030	R	2	SOC stock	S+C+O+R	QRF	Cross validation (Yes)
31	Dharumarajan et al., 2021b	2021	Karnataka	1,91,976	Regional	0.0279	L		FC, PWP	C+O+R+P	RK (RF+kriging)	Data splitting (No)
32	Dharumarajan et al., 2021c	2021	Andhra Pradesh	1,62,000	Regional	0.0140	L	2	Soil depth, FC, PWP	C+O+R+P	MLR, SVM, Cubist, <u>RF</u>	Data splitting (Yes)
33	Dharumarajan et al., 2021d	2021	Karnataka	97.8	Local	1.9325	R & Gr	1.77	Soil depth, coarse fragments, soil colour, pH, EC, SOC, P, K, S, Fe, Mn, Cu, Zn, B	C+O+R+P	RF	Cross validation (No)
34	Lalitha et al., 2021	2021	Andhra Pradesh	160,205	Regional	0.0178	L	2	Soil depth	C+O+R	QRF	Data splitting (Yes)

35	Lagacherie et al., 2022	2022	Karnataka	84	Local	2.5833	R	3	Clay content, albedo, soil horizon thickness, FC, PWP, rock fragments	S+C+O+R	RK (RF+kriging), QRF	Cross validation (No)
----	-------------------------	------	-----------	----	-------	--------	---	---	---	---------	----------------------	-----------------------

^a Plot: 0-1 km²; local: > 1 km²-10⁴ km²; regional: >10⁴-10⁷ km²; national: 3,287,590 km²

^b R: random, StR: stratified random, Gr: grid, L: legacy soil data

^c SOC: soil organic carbon, SIC: soil inorganic carbon, EC: electrical conductivity, TN: total nitrogen, K_{ex}: exchangeable potassium, N: available nitrogen, CEC: cation exchange capacity, FC: field capacity, PWP: permanent wilting point, P: phosphorus, K: potassium, S: sulphur, Fe: iron, Mn: manganese, Cu: copper, Zn: zinc, B: boron

^d MART: multiple additive regression tree, RK: regression kriging, MLR: multiple linear regression, RF: random forest, ANN: artificial neural network, GWR: geographically weighted regression, CDT: CART (classification and regression tree) decision tree, CEB: CART ensemble bagger, SVM: support vector machine, QRF: quantile regression forest

Table 2. Sources of the environmental covariates (other than field observations) used for DSM studies in India

	Environmental covariate	Sources of environmental covariates	References
S	Soil order, sub-order, great group, soil type, depth, texture, drainage, gravelliness, physiography	Legacy soil maps from NBSS&LUP	Krishnan et al., 2007; Sreenivas et al., 2014, 2016; Santra et al., 2017a; Mitran et al., 2018; Okonkwo et al., 2018; Dharumarajan et al., 2017; 2021a, 2021b
	Landform, soil order map at 1:50,000 cartographic scale	The Karnataka State Remote Sensing Applications Centre	Lagacherie et al., 2022
C	Temperature, precipitation, temperature seasonality, precipitation seasonality, mean diurnal range of temperature, annual range of temperature, isothermality, precipitation of the driest month, precipitation of the wettest month	Worldclim data	Sreenivas et al., 2014, 2016; Santra et al., 2017a; Mitran et al., 2018; Okonkwo et al., 2018; Reddy et al., 2021; Dharumarajan and Hegde, 2020; Dharumarajan et al., 2020a, 2021a, 2021b, 2021c;

Mean annual temperature, mean annual rainfall, number of dry months	Bioclimate map	Krishnan et al., 2007
Land surface temperature	MODIS data	Mitran et al., 2018
Precipitation, temperature	Indian Meteorological Department (IMD)	Hinge et al., 2018
Monthly precipitation and temperature	Climate change knowledge portal of the World Bank Group and the open government data platform of India	Shukla et al., 2018
Rainfall, air humidity, wind velocity, air temperature, precipitation, and global radiation	Automatic weather station	Lagacherie et al., 2022
O NDVI, EVI, LULC, NPP	MODIS	Sreenivas et al., 2014, 2016; Dharumarajan and Hegde, 2020; Dharumarajan et al., 2017, 2019, 2020b, 2021a, 2021b, 2021c; Shukla et al., 2018; Mitran et al., 2018; Hinge et al., 2018; Okonkwo et al., 2018; Lalitha et al., 2021
LULC	LULC map prepared by NRSC	Sreenivas et al., 2014, 2016; Mitran et al., 2018; Hinge et al., 2018
Band values, band ratios, NDVI, GSI	Landsat 8	Kalambukattu et al., 2018; Shukla et al., 2018; Mitran et al., 2019; Bangroo et al., 2020; Dharumarajan and Hegde, 2020; Dharumarajan et al., 2020a, 2021a, 2021b, 2021c; Lalitha et al., 2021
Band values	Sentinel 2	Dharumarajan et al., 2019, 2020b; Lagacherie et al., 2022
Reflectance, IMI	Landsat 8, SAR data (RISAT-1, Sentinel-1A), ALOS DEM	Shukla et al., 2018
BSI, spectral reflectance, band ratio, vegetation indices	WorldView-2, Pleiades-1A, Landsat-8	Xu et al., 2018a
Land cover	Forest Map of South India	Krishnan et al., 2007
Land use	ASTER image	Kuriakose et al., 2009
LULC	Cartosat-1 and Resourcesat – 1 LISS-IV	Sarkar et al., 2013
Command area	Irrigation atlas of India	Sreenivas et al., 2016
Agro-ecological regions	Agro-ecological regions of India map	Sreenivas et al., 2016
Band values and NDVI	Resourcesat-2	Dharumarajan et al., 2017; Lagacherie et al., 2022
Landform	World Soil Survey archive and catalogue	Okonkwo et al., 2018
LULC	GlobeLand30	Okonkwo et al., 2018

R	Elevation, slope gradient, slope aspect, slope length, slope angle, plan curvature, profile curvature, TPI, TWI, TRI, L-S factor, MrRTF, MrVBF, flow accumulation, hill shade, channel network base level, convergence index, cross-sectional curvature, longitudinal curvature, relative slope position, valley depth, vertical distance to channel network	SRTM DEM	Krishnan et al., 2007; Dharumarajan and Hegde, 2020; Dharumarajan et al., 2017, 2020a, 2020b, 2021a, 2021b, 2021c; Santra et al., 2017a; Mitran et al., 2018; Hinge et al., 2018; Reddy et al., 2021; Lalitha et al., 2021
	Elevation, aspect, flow accumulation, flow direction, slope	ASTER DEM	Xu et al., 2018a
	Slope, elevation, aspect, CTI, slope, curvature	ALOS DEM	Shukla et al., 2018
	Elevation, slope, aspect, general curvature, distance from the stream, flow direction, plan curvature, profile curvature, northeastness, northwestness, TPI, TWI, CTI, SPI, STI, MrVBF, MrRTF	Cartosat DEM	Sarkar et al., 2013; Okonkwo et al., 2018; Kalambukattu et al., 2018; Chhipa et al., 2019; Mitran et al., 2019; Dharumarajan et al., 2019; Bangroo et al., 2020; Lagacherie et al., 2022
P	Geology (rock type)	Geological Survey of India	Krishnan et al., 2007; Sreenivas et al., 2016; Mitran et al., 2018
	Lithology and geomorphology	The Karnataka State Remote Sensing Applications Centre	Dharumarajan et al., 2020a, 2021b
	Lithology	Andhra Pradesh State Remote Sensing Applications Centre	Dharumarajan et al., 2021c

(DEM: digital elevation model, NDVI: normalized difference vegetation index, EVI: enhanced vegetation index, LULC: land use land cover, GSI: grain size index, NPP: net primary productivity, IMI: integral moisture index, BSI: bare soil index, TPI: topographic position index, TWI: terrain wetness index, TRI: terrain ruggedness index, STI: sediment transport index, MrRTF: multi-resolution ridge top flatness, MrVBF: multi-resolution index of valley bottom flatness, CTI: compound topographic index, SPI: stream power index)

Highlights

- Indian soil scientists have paid more attention to digital soil mapping recently.
- Two out of 35 articles conducted national extent digital soil mapping in India.
- Environmental covariates have been used most frequently.
- Regression kriging has been the most frequently used predictive model.



DSPACE

<https://dspace.org/>

Identifying opportunities to improve digital soil mapping in India: a systematic review

Dash, Prava Kiran; Panigrahi, Niranjana; Mishra, Antaryami

2021-12-27

Elsevier

<https://doi.org/10.1016/j.geodrs.2021.e00478>

Dash PK, Panigrahi N, Mishra A. (2022) Identifying opportunities to improve digital soil mapping in India: a systematic review. Geoderma Regional, Volume 28, March 2022, Article number e00478