Validation of SoilGrids 2.0 in an Arid Region of India using In Situ Measurements

Giribabu Dandabathula, Sagar Subhashrao Salunkhe, Apurba Kumar Bera, Koushik Ghosh, Rohit Hari, Preeti Biradar, Karthik Reddy Chirala, and Mahesh Kumar Gaur

ABSTRACT

As one of the Earth’s most important natural resources, soil plays a prominent role in regulating ecosystem services, human food production systems and in facilitating a region’s sustainable development. Of late, due recognition has been given to soil sciences and soil information systems as they act as a core to achieve the targets of land degradation neutrality and help in fostering soil governance. In this regard, the availability of global soil databases paves the way for implementing successful soil information systems. Currently, harmonized world soil database from the Food and Agriculture Organization and SoilGrids from International Soil References and Information Centre serve various global soil data products in a geospatial-ready format for the scientific fraternity. In this study, SoilGrids 2.0 is validated with in situ measurements in the arid region of the Thar Desert. Soil fractions and pH at the top surface (0–5 cm) and subsurface (5–15 cm) were measured through soil sample analysis collected from the study area and compared with the values retrieved from SoilGrids 2.0 for the same location. This comparison shows that the SoilGrids 2.0 has underestimated the sand fragments up to ~28% and overestimated ~14% for silt and clay fragments. Deviation of pH in SoilGrids 2.0 was also observed with the root mean square error of one unit. However, in the comparison of soil texture classes from the field and the one given by SoilGrids 2.0, a systematic shift was found, indicating the robustness of SoilGrids prediction algorithm that can be fine-tuned by incorporating additional soil profiles (from contributing agencies) as the current snapshot of the soil database lacks dense and well-distributed soil profiles in this arid region.

Keywords: Arid Region, Digital Soil Maps, SoilGrids 2.0, Soil Texture, Soil Profile, Thar Desert.

I. INTRODUCTION

The importance of soil lies in its ability to provide ecosystem services that aid in agronomy, hydrology, climatology, and ecology; moreover, as the Earth’s most important non-renewable natural resource, soil underpins human food production system and accommodates the global biodiversity [1]. While reporting the significance of soils in the context of the United Nations Sustainable Development Goals (UN SDGs), Keesstra et al. [2] emphasized that effective soil information systems can improve the results of inter- and transdisciplinary studies on SDGs related to food security, water scarcity, climate change, biodiversity loss, and health threats. Pertinent information related to regional soils helps achieve land degradation neutrality (enshrined as target 15.3 in SDGs), which can significantly reduce a nation’s economic spending due to desertification and land degradation issues [3]. Soil governance refers to the various strategies, policies, and different decision-making by nations, states, and local governments on how the soil is utilized [4]. In this regard, soil information systems (either at the global level or the regional) act as fundamental elements to initiate soil governance and achieve the desired sustainable soil management [3], [5]-[7]. Digital soil maps and their data products act as a soul to soil information systems, which help
in soil governance and are also used as inputs to numerous environmental models.

The Harmonized World Soil Database (HWSD) is the primary source of soil information, as made available by the Food and Agriculture Organization (FAO) to the public at the global level through collaborations with various organizations [8], [9]. HWSD v1.2 can be accessed through its dedicated web portal maintained by the FAO [8]. It is a 30-arc-second raster database from which one can retrieve various soil parameters such as organic carbon, pH, available water storage capacity, soil fractions, cation exchange capacity of the soil, total exchangeable nutrients, lime and gypsum contents, sodium exchange percentage, salinity, textural class, and granulometry. Dai et al. [10], Silatsa et al. [11], and Stoorvogel et al. [12] have not only commented on the limitations of the HWSD data in terms of accuracies but also mentioned that for regional applications, it is desirable to use high-resolution soil maps. The current HWSD version with a 1-km resolution limits its use for local-level studies.

The concept of predictive soil mapping was introduced by Scull et al. [13], in which numerical or statistic models would enable a relationship between environmental variables and soil properties, which is then applied to a geographic database to create a soil map. Data from earth observation (EO) sensors, digital terrain models, and environmental covariates, along with fuzzy logic, act as an input to predictive soil mapping. However, in recent times, machine learning algorithms have been adopted for generating digital soil maps in various spatial scales and at varying depths [14], [15].

Hengl et al. [16] are credited for conceptualizing and generating the SoilGrids database at the 250-m spatial resolution that contains layers representing physical (percentage of sand-silt-clay, coarse fragments, and bulk density) as well as chemical properties (soil organic carbon, pH, cation exchange capacity, and nitrogen) of soil at seven depths (0, 5, 15, 30, 60, 100, and 200 cm). Predictive soil mapping was incorporated in SoilGrids using ~150,000 soil profiles for training the non-linear machine learning models along with numerous EO-based soil covariates. The initial version of SoilGrids (version 1.0) at 1 km resolution could not help predict the model for the regions with arid, semi-arid, deserts, and sand dunes. However, in the recent version (i.e., 2.0), these regions are given special treatment based on assertive pseudo-observations [16]. This version of SoilGrids is improved by using ~240,000 soil profiles worldwide, along with 400 global environmental covariates describing vegetation, terrain morphology, climate, geology, and hydrology [17]. These datasets were curated by the International Soil Reference and Information Centre (ISRIC) under World Soil Information Service (WoSIS) and are hosted at https://soilgrids.org. This data can be downloaded from https://files.isric.org/soilgrids/latest/ or https://maps.isric.org under the Open Database License (ODbL). Recent technical updates, accuracy/uncertainty issues, and database applicability/use cases are described in [17], [18], [19], [20].

An accuracy assessment of SoilGrids was done by [16] for Tasmania and Australia using the Soil Survey Geographic Dataset (SSURGO) developed by the National Cooperative Soil Survey and Soil and Landscape Grid of Australia, respectively. The correlation coefficients for the SoilGrids of these two regions with the respective source data were 0.79 and 0.71, respectively. Liang et al. [21] generated national-level soil organic carbon maps at six depth intervals for China using parameters from SoilGrids such as soil organic content (SOC), bulk density, sand, and silt as environmental covariates in their model to re-compute the SOC. The experiment done by [21] found that integrating SoilGrids data products into SOC modeling can be a good strategy for improving model performance at the national scale. [22] used the SoilGrids dataset to generate a 3D soil hydraulic database for Europe; when the resultant dataset was validated with 1500 samples, it was observed that the performance of soil hydraulic predictions had improved up to 21% over the models, which does not include the SoilGrids dataset as input.

This research aims to validate the physical soil parameters, such as texture fragments (percentage of sand, silt, and clay), and chemical properties, such as pH, as predicted by the SoilGrids version 2.0 for the part of the Thar Desert. This validation is intended for surface and subsurface soils at two depth ranges, 0–5 cm and 5–15 cm, respectively.

II. MATERIALS AND METHODS

A. Study Area

The study area for this research is the arid region of the great Indian Thar Desert, located in the Rajasthan state of India. To validate the SoilGrids database, samples were collected from four districts of Rajasthan, namely Jaisalmer, Bikaner, Nagaur, and Jodhpur. These districts comprise segments of the desert environment with the annual rainfall varying from 100 to 400 mm; importantly, the total rainy days per annum in these districts range from 20 to 45, with the mean maximum/minimum temperature ranging between 42–44 °C and 22–25 °C during summers and between 22–25 °C and 4–8 °C during winters for day and night respectively [23]. Fig. 1 shows the location map of the study area. While estimating soil water content for this study area, in which the soil types of this region were mentioned as aridisols, entisols, and inceptisols by Santra et al. [24].

B. Sample Sites

A total of 24 soil samples, each with ~2 kg weight, were collected from the surface (at ~5 cm depth) and subsurface (between 5 and 15 cm depth) from the study area. The distribution of these sample sites is shown in Fig. 1c. The samples were randomly selected but ensured to be well distributed with a rationale that the sample site is having a minimum of 5 ha of homogeneous signature with respect to soil texture variation as appeared from the satellite imagery; for this, Sentinel-2 A/B optical data are used for inference. While selecting sampling locations, it was made sure that these sites did not have any litter deposition or weed growth. Locations with excessive moisture content were also avoided. The samples were chosen from the landscape that contained sparse vegetation. Areas with sand dunes were avoided due to their potential to participate in the aeolian process. Species such as Calligonum polygonoides, Aerva persica, Citrullus colocynthis, Prosopis juliflora, Ziziphus mauritiana, Prosopis cineraria, and Fagonia bruguieri were predominantly found in this landscape.
Fig. 1. Map of the study area: a) Map of India highlighting the state of Rajasthan; b) Map of Rajasthan highlighting the districts from where samples were collected; c) Distribution of the sampled sites in the study area.

Fig. 2. Typical locations of sample sites. Sample’s locations are overlaid on Sentinel 2A imagery. a) Sample site at 27.062788°N, 71.557659°E near Bhadriya village, Jaisalmer district; b) Sample site at 27.397835°N, 70.473327°E near Ramgarh village, Jaisalmer district; c) Sample site at 28.242899°N, 72.876350°E near Poogal village, Bikaner district; d) Sample site at 26.260903°N, 70.329479°E near Khayala Matt, Jaisalmer district.
Dried in the laboratory before sieving them through a 2-mm sieve. The sieving process ensured the removal of boulders, small pebbles, root particles, and litter from the samples. The international pipette method determined the sand, silt, and clay percentages. Based on different sedimentation velocities of particles with different diameters, the pipette method is considered one of the standard methods to determine the distribution of individual grain size fractions [25]. Organic carbon estimated for all soil samples using the method endorsed by [26] yielded a value of <0.09%; due to this negligible value, organic carbon was not validated in this study even though SoilGrids 2.0 contains soil organic carbon as a layer. We used a pH meter to estimate the soil pH of the samples.

Of the 24 surface soil samples, 22 had sand fragments in the range of ~80% to ~86%; thus, their soil texture was determined as loamy sand; the percentage of clay in these samples varied between ~6% and ~10%. Earlier studies have confirmed the predominance of loamy sand texture for surface soils in the current study area [27], [28], [29]. The pH in the topsoil for all samples was in the range of 8.55 to 9.32.

Similarly, out of the 23 collected samples for subsurface soils, 22 had sand fragments in the range of ~77% to ~87%, of which the texture of 15 samples was loamy sand, and the remaining seven were sandy loam. The texture of one sample was sandy clay loam. For the one intended sample, bedrock was found at a depth of 6 cm and could not be collected. In the subsurface soils of all the samples, pH was in the range of 8.2 to 9.8. Table I shows results from the laboratory analysis for the soil samples and the location (latitude, longitude), percentage of soil fragments (i.e., sand, silt, and clay), and pH at the surface (0–5 cm) and subsurface (5–15 cm) along with their corresponding soil textural class as per the United States Department of Agriculture soil texture classes.

The data in Table I represent the location of the samples, soil fragments at the surface (0–5 cm), and subsurface (5–15 cm), along with their pH values.

### Table I: Results from the Laboratory Analysis of Soil Samples Collected for Validating the SoilGrids 2.0 Database

| Sample Id. | Location (latitude, longitude) | Soil fragments (%), texture class and pH at surface | Soil fragments (%), texture class and pH at subsurface |
|------------|--------------------------------|-----------------------------------------------------|-------------------------------------------------------|
|            |                                | Sand|Silt|Clay|Texture|pH  | Sand|Silt|Clay|Texture|pH |
| 1          | 26.460N, 72.41080E             | 83.15 | 7.91 | 8.94 | LS | 8.98 | 82.83 | 8.05 | 9.11 | LS | 8.44 |
| 2          | 27.100S, 71.84358E             | 85.63 | 7.32 | 7.05 | LS | 8.84 | 79.09 | 9.66 | 11.26 | SL | 8.54 |
| 3          | 27.062N, 71.55700E             | 84.42 | 7.15 | 8.44 | LS | 8.78 | 82.03 | 8.33 | 9.63 | LS | 8.62 |
| 4          | 26.989N, 71.27500E             | 81.83 | 8.07 | 10.10 | LS | 9.07 | 82.96 | 7.74 | 9.30 | LS | 8.87 |
| 5          | 26.7136N, 70.78195E            | 82.12 | 8.87 | 9.02 | LS | 8.67 | 82.38 | 8.52 | 9.10 | LS | 8.80 |
| 6          | 26.5030N, 70.59525E            | 82.59 | 8.07 | 9.34 | LS | 8.84 | 83.76 | 7.79 | 8.44 | LS | 8.53 |
| 7          | 26.2609N, 70.32948E            | 80.86 | 8.45 | 10.68 | LS | 8.94 | 78.72 | 10.00 | 11.28 | SL | 9.94 |
| 8          | 26.7207N, 70.59396E            | 80.79 | 8.54 | 10.67 | LS | 8.97 | 77.83 | 11.06 | 11.10 | SL | 8.85 |
| 9          | 26.8520N, 70.37035E            | 82.96 | 7.80 | 9.24 | LS | 9.09 | 81.85 | 8.44 | 9.70 | LS | 9.10 |
| 10         | 27.3970N, 70.10081E            | 84.05 | 7.26 | 8.70 | LS | 9.04 | 78.13 | 10.47 | 11.40 | SL | 8.90 |
| 11         | 27.4063N, 70.10398E            | 83.03 | 8.62 | 8.36 | LS | 9.09 | 86.30 | 7.56 | 6.14 | LS | 9.10 |
| 12         | 27.3970N, 70.47333E            | 81.28 | 10.18 | 8.54 | LS | 9.04 | 81.90 | 9.28 | 8.83 | LS | 8.23 |
|   | Latitude         | Longitude        | Sand (g/kg) | Silt (g/kg) | Clay (g/kg) | pH  | Soil Type |
|---|-----------------|------------------|-------------|-------------|-------------|-----|-----------|
| 12| 27.3979N,       | 70.4733E        | 81.28       | 10.18       | 8.54        | LS  | 9.04      | 81.90 | 9.28 | 8.83 | LS  | 8.23 |
| 13| 27.3154N,       | 70.2584E        | 80.91       | 10.61       | 8.47        | LS  | 8.93      | 82.13 | 7.66 | 10.21 | LS  | 8.73 |
| 14| 27.6070N,       | 70.8279E        | 80.85       | 10.35       | 8.79        | LS  | 8.97      | 82.08 | 8.17 | 9.75  | LS  | 9.02 |
| 15| 27.4218N,       | 70.6679E        | 64.78       | 15.69       | 19.52       | SL  | 8.78      | 79.79 | 9.99 | 11.22 | SL  | 8.89 |
| 16| 26.8719N,       | 70.7307E        | 63.37       | 15.16       | 21.47       | SCL | 8.55      | Bed Rock |
| 17| 27.7158N,       | 72.6132E        | 81.15       | 9.04        | 9.81        | LS  | 8.99      | 78.95 | 8.64 | 12.41 | SL  | 9.80 |
| 18| 28.2428N,       | 72.8756E        | 86.94       | 6.10        | 6.96        | LS  | 9.09      | 84.39 | 6.98 | 8.63  | LS  | 8.90 |
| 19| 28.7009N,       | 72.5773E        | 86.54       | 6.57        | 6.90        | LS  | 9.18      | 86.98 | 5.89 | 7.12  | LS  | 9.10 |
| 20| 28.3368N,       | 72.5018E        | 86.94       | 6.47        | 6.58        | LS  | 9.20      | 83.82 | 7.40 | 8.79  | LS  | 9.08 |
| 21| 28.1500N,       | 72.28039E       | 84.96       | 6.99        | 8.06        | LS  | 9.32      | 79.46 | 10.49 | 10.05 | SL  | 9.20 |
| 22| 28.2106N,       | 73.2982E        | 80.84       | 9.39        | 9.78        | LS  | 8.79      | 81.83 | 8.40 | 9.77  | LS  | 9.08 |
| 23| 27.9445N,       | 73.3995E        | 86.97       | 6.23        | 6.79        | LS  | 9.08      | 81.32 | 8.96 | 9.73  | LS  | 9.26 |
| 24| 27.0425N,       | 73.58113E       | 81.14       | 9.32        | 9.55        | LS  | 9.07      | 55.77 | 19.14 | 25.08 | SCL | 9.72 |

LS: Loamy sand; SCL: Sandy clay loam; SL: Sandy loam

---

D. Soil Fragments (% of sand, silt, and clay) and pH from SoilGrids 2.0

ISRIC provides flexible methods to download the layers of the SoilGrids 2.0 database [32]. The download methods include web map service, web coverage service, Google Earth Engine, and WebDav (where complete or a part of global maps can be downloaded in virtual raster format). In this research, layers of the mean content of sand, silt, clay, and pH for two depths (0–5 cm and 5–15 cm) were downloaded from the web portal https://files.isric.org/soilgrids/latest/ using QGIS software. Once the data were downloaded from this web portal, values of mean content (g/kg) for sand, silt, clay, and pH were retrieved at the same geo-locations as sample sites (shown in Table II). Fig. 4 shows the map depicting sand content (g/kg) from SoilGrids 2.0 for the extent of the study area overlaid with the sample locations.

For comparison with samples obtained from the field, the mean content (g/kg) for sand, silt, and clay was converted to percentage values. The retrieved values of soil fragments (in percentage), their uncertainties (as given in the corresponding uncertainty layers), and pH from the SoilGrids 2.0 database are presented in Table II for all locations from where the ground samples were collected.

---

Fig. 4. Sand content (g/kg) from SoilGrids 2.0 for the extent of the study area overlaid with the sample locations where in situ observations were made.
Out of 24 records of the predictions made by SoilGrids 2.0, 21 and 20 predictions fall in sandy clay loam texture class on the surface and subsurface, respectively. The range of percentage sand content in the top soils is ~46–61%; for subsurface soils, the range is ~44–62%. Values in Table II include soil fragments at the surface (0–5 cm) and subsurface (5–10 cm), texture class, and pH, along with the uncertainties for soil fragments prediction.

### E. Methodology

Percentages of soil fragments, pH, and soil texture classes from the field observations and SoilGrids 2.0 were compared as part of the validation procedure. The comparison has been made for both the top surface (0–5 cm) and subsurface (5–15 cm). Fig. 5 shows the methodology used in validating the SoilGrids 2.0 database using in situ observations. Root Mean Square Error (RMSE) and Mean Bias Error (MBE) were chosen to validate the performance of SoilGrids 2.0 in predicting soil fragments and pH with in situ measurements.

RMSE is frequently used to measure the difference between predicted values and actual values of the environment. RMSE aggregates all individual differences called residuals and gives a single measure of predictive power. The formula for RMSE is given in (1). MBE helps estimate the average bias in observations and actual values. Its formula is given in (2).

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_i - O_i)^2} \tag{1}
\]

\[
MBE = \frac{1}{n} \sum_{i=1}^{n} (P_i - O_i) \tag{2}
\]

where \(P_i\) is the predicted value (in our case, values from SoilGrids 2.0) and \(O_i\) is the observation value (in our case, values from the in-situ data).

The next section discusses the results obtained after computing RMSE and MBE.
III. RESULTS

RMSE and MBE have been used in research such as [31], [32] for the performance validation of digital soil maps with observed values of soil properties from in situ sampling data. Table III shows the results after comparing the data from SoilGrids 2.0 with in situ measurements.

Similarly, Table IV shows the quantified results from the computation of RMSE and MBE.

Table IV shows a difference of ~28% and ~27% in estimations of sand fragments by SoilGrids 2.0 in terms of RMSE. Fig. 6 illustrates the bias in estimating the soil fragments by SoilGrids 2.0 with respect to in situ measurements, whereas it is evident from Fig. 6a, 6b, and 6c that there is an underestimation of the sand fraction by a magnitude of ~28 and overestimation of both silt and clay fractions by the magnitude of ~14. Predictions for sand, silt, and clay in the subsurface are also along similar lines. The prediction of pH in the study area, as per the SoilGrids 2.0 standard, has an RMSE of ~1 at both surface and subsurface levels (underestimated by 1 unit).

IV. DISCUSSION

Fig. 7 shows soil texture classes from in situ samples and SoilGrids 2.0 for surface and subsurface soils. As seen in Fig. 7a, the majority of the samples from surface soils analyzed by in situ measurements have the texture class of loamy sand, whereas the texture class for the soils derived from the soil fractions of SoilGrids 2.0 at the same location as in situ measurement is majorly classified as sandy clay loam. As discussed earlier, this difference can be attributed to the underestimation of sand fragments and the overestimation of silt and clay fragments by SoilGrids 2.0. Similarly, for subsurface soils, the texture class identified from the in situ observations is primarily that of loamy sand and sandy loam, whereas SoilGrids 2.0 predicts these as sandy clay loam.

| TABLE III: RESULTS FROM THE COMPARISON OF SOIL FRAGMENTS, pH AND TEXTURE CLASSES FROM SOILGRIDS 2.0 AND IN SITU MEASUREMENTS |
|---------------------------------------------------------------|
| Quality measure                                              | Value at surface soils (0–5 cm) (n = 24) | Value at subsurface soils (5–15 cm) (n = 23) |
|---------------------------------------------------------------|
| Range of sand fraction (%)                                    | SoilGrids 2.0 | In situ | SoilGrids 2.0 | In situ |
| Range of clay fraction (%)                                    | ~46–61%      | ~80–86% | ~44–62%      | ~77–87% |
| Dominant soil texture class                                   | SCL          | LS      | SCL          | LS and SL |
| Range of pH                                                   | 7.8–9.1      | 8.5–9.32 | 7.8–9.1      | 8.4–9.7  |

LS: Loamy sand; SCL: Sandy clay loam; L: Loam

| TABLE IV: RESULTS FROM THE COMPUTATION OF RMSE AND MBE FOR THE VALUES FROM SOIL FRAGMENTS AND pH FROM SOILGRIDS 2.0 AND IN SITU MEASUREMENTS |
|---------------------------------------------------------------|
| Quality measure                                              | Value at surface (0–5 cm) (n = 24) | Value at subsurface (5–15 cm) (n = 23) |
|---------------------------------------------------------------|
| RMSE for soil fractions and pH                                | Sand | Silt | Clay | pH | Sand | Silt | Clay | pH |
| RMSE for soil fractions and pH                                | 28.73 | 14.14 | 18.23 | 1.02 | 27.88 | 13.28 | 14.97 | 1.09 |
| MBE for soil fractions and pH                                 | –27.32 | 14.14 | 14.59 | –1 | –26.65 | 12.10 | 14.49 | –1 |

Fig. 6. Mean bias error in SoilGrids 2.0 prediction for sand, silt and clay and the in situ measurements.
However, an important observation is that for both surface and subsurface soils, a systematic shift is observed in the prediction of texture classes by SoilGrids 2.0; the direction of texture class shift is oriented uniformly, that is, along the north-east direction as evident from Fig. 7a and 7b.

Even though the magnitude of error in estimating the sand fragment in this study area by SoilGrids 2.0 is in the order of ~28%, the uncertainty values flagged for these pixel locations are in the range only of 1%–2% (refer to column 7 in Table II). However, for the same locations, the overestimation of silt and clay fragments has a magnitude bias of ~14%, and the uncertainty of silt and clay is in the range of ~5%–8% (refer to columns 8 and 9 in Table II). Thus, the prediction algorithm included in SoilGrids 2.0 can be corrected by reducing the percentage of silt and clay fragments and increasing the sand fragment component for this study area.

During the SoilGrids 2.0 database generation, numerous influential parameters, such as soil profile data (density and distribution), environmental covariates, and variables associated with model tuning and model fitting, are analyzed. Importantly, the lack of well-distributed soil profile data among the above parameters leads to gap areas and uncertainties in model prediction [17].

While describing the details related to soil profile data associated with SoilGrids 2.0 database generation, [18] presented the spatial distribution of soil profiles for the latest snapshot, in which the total distribution at the country level for India includes only ~200 profiles (refer to Fig. 2 and Appendix C in [18]).

Fig. 7. Soil texture classes derived for soil fractions as predicted by SoilGrids 2.0 and from the in-situ samples for surface and subsurface soils. a) The texture class according to SoilGrids 2.0 for surface soils: sandy clay loam; in situ: sandy loam; b) The texture class according to SoilGrids 2.0 for subsurface soils: sandy clay loam; in situ: loamy sand/sandy loam.

Fig. 8. Density of soil profiles in the Indian subcontinent for SoilGrids 2.0 database generation. a) Low density of soil profiles for the Indian subcontinent; b) A single soil profile governs the prediction algorithm for arid regions covering Jaisalmer and Bikaner districts.
The distribution and density of soil profiles for India, as well as for the study area of this research, are reproduced in Fig. 8. From Fig. 8b, it is evident that there is only one soil profile in the soil prediction algorithm for significant parts of Jaisalmer and Bikaner districts.

The ISRIC–WoSIS system has the provision for contributing and sharing geo-tagged soil profiles from the data providers like soil survey organizations, research institutes, and individual experts [17]. All the data shared by contributors are first stored ‘as is’ in the ISRIC data repository. However, these data subsequently undergo quality assessment and standardizing mechanisms as determined by the WoSIS standard operating procedure (available at [30]). The accumulated soil profiles will play an immense role in refining the accuracy of the SoilGrids 2.0 database. A significant number of soil profiles are already shared for countries in North America, Western Europe, southern parts of Africa, eastern regions of Asia-Pacific, and Australia for the SoilGrids 2.0 database. From Fig. 8, it is evident that there exists a significant gap in soil profiles for the Indian subcontinent. Thus, contributions from soil data providers for soil profiles are needed for this subcontinent to strengthen the prediction algorithm of SoilGrids 2.0. Specifically for the study area used in this research, there is a need for up-to-date knowledge related to soils so the risk of desertification could be reduced. In this context, databases such as SoilGrids 2.0, which is currently available at 250 m, will not only be helpful in various environmental studies but also pave the way to identify the zones of higher desertification risk.

Given the very low distribution of soil profiles feed for predicting the properties for the arid region, as in this study, the SoilGrids 2.0 database has shown relatively better accuracy with a systematic bias. However, for improving the absolute accuracy of the SoilGrids 2.0 database in arid regions like those specified in this research, more soil profiles, as well as fine-tuning of the model fit of the predictive algorithm, are required.

V. CONCLUSION

Pertinent information about soils is essential for various applications related to Earth sciences. The availability of soil databases such as HWSD, with 1 km resolution, and SoilGrids 2.0, with 250 m resolution, has paved the way for understanding the dynamics of soil at the global level. In this study, we validated the physical and chemical properties of soils from the SoilGrids 2.0 database in the Thar Desert area; for this, soil fragments and pH were validated with in situ measurements at surface and subsurface levels. Results show that the sand percentage predicted by SoilGrids 2.0 for the study area is underestimated by ~28% and overestimated the silt and clay fragments by ~14%. The prediction of pH in this study area by SoilGrids 2.0 has the RMSE of ~1, which indicates the need of fine-tuning the model for chemical properties as well. However, the comparison of soil texture classes from SoilGrids 2.0 with those from in situ samples shows a systematic bias indicating that the flaws in the prediction algorithms can be identified and corrected for the arid regions comparatively accurately. Towards this, the first observation is that there is a lack of enough soil profiles for the study area (or the Indian subcontinent at large). SoilGrids 2.0 portal facilitates data providers to upload/contribute soil profiles that will fine-tune the model. In this regard, for this region, a partnership between local expert institutions and ISRIC needs to be improved, which in turn would benefit the scientific fraternity in conducting studies that need accurate soil information.

ACKNOWLEDGMENT

Authors thank ISRIC for providing open access to various data products that are used in this research. Foremost, authors are indebted to Director, National Remote Sensing Centre, ISRO, Hyderabad, India and express sincere gratitude for encouraging the research activities and providing necessary infrastructure to this research. Authors would like to express sincere gratitude to Chief General Manager, Regional Centres, NRSC for permitting this work. Authors would like to express special thanks of gratitude to the Mr. Janaki Ram Suressh of NRSC, Hyderabad, Mr. Sushil Rehpaide of RRSC-Central, Nagpur and officials of RRSC-West for providing technical insights in this research.

DATA AVAILABILITY

All the data supporting in this study are publicly available and mentioned in manuscript.

CONFLICT OF INTEREST

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Smith P. Managing the global land resource. Proceedings of the Royal Society B: Biological Sciences, 2018; 285(1874), 20172798. http://dx.doi.org/10.1098/rspb.2017.2798.
[2] Keesstra SD, Bouma J, Wallinga J, Tittonell P, Smith P, Cerdà A, Montanarella L. et al. The significance of soils and soil science towards realization of the United Nations Sustainable Development Goals. Soil, 2016; 2(2), 111–128. http://dx.doi.org/10.5194/soil-2-111-2016.
[3] Erdogan HE, Havlicek E, Dazzi C, Montanarella L, Van Liedekerke M, Vrščaj B, Krasilnikov P, et al. Soil conservation and sustainable development goals (SDGs) achievement in Europe and Central Asia: Which role for the European soil partnership? International Soil and Water Conservation Research, 2021; 9(3), 360–369. https://doi.org/10.1016/j.iswcr.2021.02.003.
[4] FAO. People & soil – Global soil governance: Status and future perspectives. Geneva Eurosoil, 2021. [Internet] Retrieved from: www.fao.org/global-soil-partnership/areas-of-work/soil-governance/en/.
[5] Hannam I. Soil governance and land degradation neutrality. Soil Security, 2022; 6, 100030. https://doi.org/10.1016/j.soisec.2021.100030.
[6] Montanarella L. Agricultural policy: Govern our soils. Nature, 2015; 528(7580), 32–33. https://doi.org/10.1038/528032a.
[7] Velayutham M, Mishra BB. Soil and future. In: B. Mishra (Ed.) The soils of India. World Soils Book Series. Springer; 2020. https://doi.org/10.1007/978–3–030–31082–0_14.
[8] Fischer G, Nachtergaele F, Prieler S, Van Velthuizen HT, Verelst L, Wiberg D. Global agro-ecological zones assessment for agriculture (GAEZ 2008). IAASA and FAO. 2008.
[9] Nachtergaele F, Van Velthuizen H, Verelst L, Batjes N, Dijkshoorn K, Van Engelen V, Fischer G, et al. The harmonized world soil database. Proceedings of the 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane, Australia; 1–6 August 2010; pp. 34–37.
[10] Dai Y, Shangguan W, Wei N, Xin Q, Yuan H, Zhang S, Liu S, et al. A review of the global soil property maps for Earth system models. Soil, 2019; 3(2), 137–158. https://doi.org/10.5194/soil-3-137-2019.
[11] Silatza FB, Yemefack M, Tabi FO, Heuvelink GB, Leenaars JG. Assessing countrywide soil organic carbon stock using hybrid machine
learning modelling and legacy soil data in Cameroon. *Geoderma*, 2020; 367, 114260. https://doi.org/10.1016/j.geoderma.2020.114260.

[12] Stoorvogel JJ, Bakkenes M, Temme AJ, Batjes NH, ten Brink BJ. SW-world: A global soil map for environmental modelling. *Land Degradation & Development*, 2017; 28(1), 22–33. https://doi.org/10.1002/ldr.2656.

[13] Scull P, Franklin J, Chadwick OA, McArthur D. Predictive soil mapping: A review. *Progress in Physical Geography*, 2003; 27(2), 171–197. https://doi.org/10.1191/0309133303pp366ra.

[14] Sweta K, Dharamrajjan S, Supintha A, Lalitha M, Vasundhara R, Kalasselli B, Hegde R, et al. (2022). Transforming soil paradigms with machine learning. In *Data science in agriculture and natural resource management*. Springer, 2022: 243-265.

[15] Wadoux AMJC, Minasny B, McBratney AB. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, 2020; 210, 103359. https://doi.org/10.1016/j.earscirev.2020.103359.

[16] Heng T, Mendes de Jesus J, Heuvelink GBM, Gonzalez MR, Kilibarda M, Blagotić A, Shangguan W, et al. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, 2017; 12(2), e0169748. https://doi.org/10.1371/journal.pone.0169748.

[17] Poggio L, De Sousa LM, Batjes NH, Heuvelink G, Kempen B, Ribeiro E, Rossiter D. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *Soil*, 2021; 7(1), 217–240. https://doi.org/10.5194/soil-7-217-2021.

[18] Batjes NH, Ribeiro E, Van Oostrom A. Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). *Earth System Science Data*, 2020; 12(1), 299–320. https://doi.org/10.5194/essd-12–299-2020.

[19] Batjes NH, Ribeiro E, Van Oostrom A, Leenaars J, Heng T, Mendes de Jesus J. WoSIS: Providing standardised soil profile data for the world. *Earth System Science Data*, 2017; 9(1), 1–14. https://doi.org/10.5194/essd-9–1–2017.

[20] Rossiter DG, Poggio L, Beaudette D, Libohova Z. How well does predictive soil mapping represent soil geography? An investigation from the USA [Preprint]. *Soil Discussions*, 2021: 1–35. https://doi.org/10.5194/soil-2021–80.

[21] Liang Z, Chen S, Yang Y, Zhou Y, Shi Z. High-resolution three-dimensional mapping of soil organic carbon in China: Effects of SoilGrids products on national modeling. *Science of the Total Environment*, 2019; 685, 480–489. https://doi.org/10.1016/j.scitotenv.2019.05.332.

[22] Tóth B, Weynants M, Pázmány L, Hengl T. 3D soil hydraulic database of Europe at 250 m resolution. *Hydrological Processes*, 2017; 31(14), 2662–2666. https://doi.org/10.1002/hyp.11203.

[23] Kumar M, Goossens E, Goossens R. Assessment of sand dune change detection in Rajasthan (Thar) Desert, India. *International Journal of Remote Sensing*, 1993; 14(9), 1689–1703. https://doi.org/10.1080/01431169308935995.

[24] Santra P, Kumar M, Kumawat RN, Painuli DK, Hati KM, Heuvelink GBM, Batjes NH. Pedotransfer functions to estimate soil water content at field capacity and permanent wilting point in hot arid western India. *Journal of Earth System Science*, 2018; 127(3), 1–16.

[25] Šinkovičová M, Igaz D, Kondrllová E, Jarošová M. Soil particle size analysis by laser diffractometry: Result comparison with pipette method. In IOP Conference Series: *Materials Science and Engineering*, 2017; 245(7), 072025. IOP Publishing. doi:10.1088/1757–899X/245/7/072025.

[26] IS2720-Part22. Indian standard methods of tests for soils – Determination of organic matter. Bureau of Indian Standards; 1972.

[27] Gupta RK. (1968). Anthropogenic influences on the vegetation of Western Rajasthan. *Vegetatio Acta Geobot*, 2008; 16(1), 79–94. https://doi.org/10.1007/BF00261358.

[28] Rathore VS, Singh JP, Bhardwaj S, Nathawat NS, Kumar M, Roy, M. M. Potential of native shrubs Haloxylon salicornicum and Calligonum polygonoides for restoration of degraded lands in arid western Rajasthan, India. *Environmental Management*, 2015; 55(1), 205–216. https://doi.org/10.1007/s00267–014–0372–1.

[29] Singh KC, Rao AS. A short note on water use and production potential of Lasiurus scindicus in the Thar desert of Rajasthan, India. *Journal of Arid Environments*, 1996; 33(2), 261–262. https://doi.org/10.1016/j.jarl.1996.0062.

[30] SoilGrids. *Global Gridded Soil Information*. (n.d.). www.isric.org. https://www.isric.org/explore/soilgrids.

[31] Bargaaoui YE, Walter C, Michot D, Saby NP, Vincent S, Lemercier B. Validation of digital maps derived from spatial disaggregation of legacy soil maps. *Geoderma*, 2019; 356: 113907. https://doi.org/10.1016/j.geoderma.2019.113907.

[32] Kmocho A, Kanal A, Astover A, Kull A, Virro H, Helm A, Pärtel M, et al. *EstSoil-EH*: A high-resolution eco-hydrological modelling parameters dataset for Estonia. *Earth System Science Data*, 2021; 13(1), 83–97. https://doi.org/10.5194/essd-13–83–2021.