Science of the Total Environment 968 (2025) 178791

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

A global soil spectral grid based on space sensing

José A.M. Demattê [a,*], Rodnei Rizzo [b], Nícolas Augusto Rosin [a], Raul Roberto Poppiel [a], Jean Jesus Macedo Novais [a], Merilyn Taynara Accorsi Amorim [a], Heidy Soledad Rodriguez-Albarracín [a], Jorge Tadeu Fim Rosas [a], Bruno dos Anjos Bartsch [a], Letícia Guadagnin Vogel [a], Budiman Minasny [c], Sabine Grunwald [d], Yufeng Ge [e], Eyal Ben-Dor [f], Asa Gholizadeh [g], Cecile Gomez [h,i], Sabine Chabrillat [j,k], Nicolas Francos [c], Dian Fiantis [l], Abdelaziz Belal [m], Nikolaos Tsakiridis [n], Eleni Kalopesa [n], Salman Naimi [o], Shamsollah Ayoubi [o], Nikolaos Tziolas [d], Bhabani Sankar Das [p], George Zalidis [n], Marcio Rocha Francelino [q], Danilo Cesar de Mello [q], Najmeh Asgari Hafshejani [o], Yi Peng [r], Yuxin Ma [s], João Augusto Coblinski [t], Alexandre M.J.-C. Wadoux [h], Igor Savin [u], Brendan P. Malone [v], Konstantinos Karyotis [n], Robert Milewski [j], Emmanuelle Vaudour [w], Changkun Wang [r], Elsayed Said Mohamed Salama [m,x], Keith D. Shepherd [y]

[a] Department of Soil Science, Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, São Paulo, Brazil
[b] Center of Nuclear Energy in Agriculture (CENA), University of São Paulo, Brazil
[c] Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Australia
[d] Department of Soil, Water, and Ecosystem Sciences, University of Florida, USA
[e] Department of Biological Systems Engineering, University of Nebraska-Lincoln, Lincoln, NE 68583, USA
[f] Department of Geography, Porter School of Environmental and Earth Sciences, Faculty of Exact Science, Tel Aviv University, Tel Aviv, Israel
[g] Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Kamycka 129, Suchdol, Prague 16500, Czech Republic
[h] LISAH, Univ Montpellier, AgroParisTech, INRAE, IRD, L'Institut Agro, Montpellier, France
[i] Indo-French Cell for Water Sciences, IRD, Indian Institute of Science, Bengaluru 560012, India
[j] GFZ Helmholtz Centre for Geosciences, Telegrafenberg A17, 14473 Potsdam, Germany
[k] Leibniz University Hannover (LUH), Institute of Earth System Sciences, Soil Science Section, Herrenhaeuser Str. 2, 30419 Hannover, Germany
[l] Department of Soil Science, Faculty of Agriculture, Universitas Andalas, Kampus Unand Limau Manis, Padang 25163, Indonesia
[m] National Authority for Remote Sensing and Space Sciences, Cairo 11843, Egypt
[n] Laboratory of Remote Sensing, Spectroscopy and GIS, Aristotle University of Thessaloniki, Greece
[o] Department of Soil Science, Isfahan University of Technology, Isfahan 84156-83111, Iran
[p] Agricultural and Food Engineering Department, Indian Institute of Technology Kharagpur, West Bengal, India
[q] Department of Soils, Federal University of Viçosa, Ave. Peter Henry Rolfs s/n, 36570-900 Viçosa, Minas Gerais, Brazil
[r] State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China
[s] New South Wales Department of Climate Change, Energy, the Environment and Water, Parramatta, NSW 2150, Australia
[t] Institute of Soil Science and Plant Cultivation - State Research Institute, Czartoryskich 8, 24-100 Puławy, Poland
[u] V.V. Dokuchaev Soil Science Institute, 119017 Moscow, Russia
[v] CSIRO Agriculture and Food, Black Mountain, ACT, Australia
[w] Université Paris-Saclay, INRAE, AgroParisTech, UMR EcoSys, 91120 Palaiseau, France
[x] Department of Environmental Management, Institute of Environmental Engineering, RUDN University, 6 Miklukho-Maklaya St., Moscow 117198, Russia
[y] Innovative Solutions for Decision Agriculture (iSDA), Rothamsted Campus, West Common, Harpenden, Hertfordshire AL5 2JQ, United Kingdom

* Corresponding author.
E-mail addresses: jamdemat@usp.br (J.A.M. Demattê), narosin@usp.br (N.A. Rosin), raulpoppiel@usp.br (R.R. Poppiel), jeannovais@usp.br (J.J.M. Novais), merilyn.accorsi@usp.br (M.T.A. Amorim), hsrodrigueza@usp.br (H.S. Rodriguez-Albarracín), jorge.fimrosas@usp.br (J.T.F. Rosas), brunobartsch@usp.br (B.A. Bartsch), leticia.vogel@usp.br (L.G. Vogel), budiman.minasny@sydney.edu.au (B. Minasny), sabgru@ufl.edu (S. Grunwald), yge2@unl.edu (Y. Ge), bendor@tauex.tau.ac.il (E. Ben-Dor), gholizadeh@af.czu.cz (A. Gholizadeh), cecile.gomez@ird.fr (C. Gomez), chabri@gfz.de, chabrillat@ifbk.uni-hannover.de (S. Chabrillat), nicolas.francos@sydney.edu.au (N. Francos), tsakirin@auth.gr (N. Tsakiridis), kalopesa@auth.gr (E. Kalopesa), sn.marandi@yahoo.com (S. Naimi), ayoubi@iut.ac.ir (S. Ayoubi), ntziolas@ufl.edu (N. Tziolas), bsdas@agfe.iitkgp.ac.in (B.S. Das), zalidis@auth.gr (G. Zalidis), marcio.francelino@ufv.br (M.R. Francelino), n.asgari86@yahoo.com (N.A. Hafshejani), pengyi@issas.ac.cn (Y. Peng), yuxin.ma@environment.nsw.gov.au (Y. Ma), jcoblinski@iung.pulawy.pl (J.A. Coblinski), alexandre.wadoux@inrae.fr (A.M.J.-C. Wadoux), savin_iyu@esoil.ru (I. Savin), brendan.malone@csiro.au (B.P. Malone), kbkaryotis@auth.gr (K. Karyotis), milewski@gfz.de (R. Milewski), emmanuelle.vaudour@inrae.fr (E. Vaudour), ckwang@issas.ac.cn (C. Wang), keith.shepherd@isda-africa.com (K.D. Shepherd).

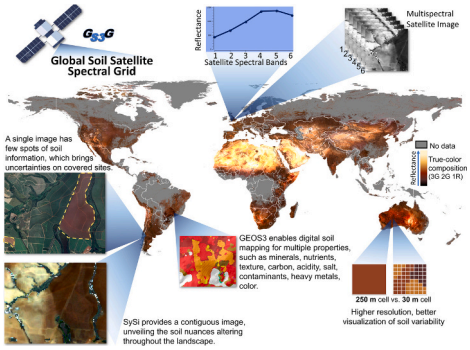
https://doi.org/10.1016/j.scitotenv.2025.178791

Available online 20 February 2025

HIGHLIGHTS

- At 800 km from the target, satellite data unveil key soil information.
- Intrinsic soil image reflectance accurately detected several soil spectral patterns.
- Satellites indicated a relative spectral similarity with terrestrial observation data.
- The GEOS3 system revealed 90 % of bare soil from agricultural areas worldwide.
- Bare soil pixels represent an unprecedented proxy at the service of society.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Christian Herrera

Keywords:
Soil reflectance spectra
Soil security
Earth observation
Digital soil mapping
Agri-environmental policy

ABSTRACT

Soils provide a range of essential ecosystem services for sustaining life, including climate regulation. Advanced technologies support the protection and restoration of this natural resource. We developed the first fine-resolution spectral grid of bare soils by processing a spatiotemporal satellite data cube spanning the globe. Landsat imagery provided a 30 m composite soil image using the Geospatial Soil Sensing System (GEOS3), which calculates the median of pixels from the 40-year time series (1984–2022). The map of the Earth's bare soil covers nearly 90 % of the world's drylands. The modeling resulted in 10 spectral patterns of soils worldwide. Results indicate that plant residue and unknown soil patterns are the main factors that affect soil reflectance. Elevation and the shortwave infrared (SWIR2) band show the highest importance, with 78 and 80 %, respectively, suggesting that spectral and geospatial proxies provide inference on soils. We showcase that spectral groups are associated with environmental factors (climate, land use and land cover, geology, landforms, and soil). These outcomes represent an unprecedented information source capable of unveiling nuances on global soil conditions. Information derived from reflectance data supports the modeling of several soil properties with applications in soil-geological surveying, smart agriculture, soil tillage optimization, erosion monitoring, soil health, and climate change studies. Our comprehensive spectrally-based soil grid can address global needs by informing stakeholders and supporting policy, mitigation planning, soil management strategy, and soil, food, and climate security interventions.

1. Introduction

Soil is essential for sustaining life on Earth and influences changes in climate, water, and food systems, as well as ecosystem services and biodiversity (Lal, 2014; Jónsson and Davíðsdóttir, 2016; Evangelista et al., 2024). Nevertheless, soil is one of the Earth's systems most threatened by human activities and requires monitoring to ensure its protection and restoration. The Food and Agriculture Organization (FAO) of the United Nations addresses ten processes that lead to global soil degradation including erosion, organic carbon loss, nutrient imbalance, acidification, contamination, waterlogging, compaction, sealing, salinization, and biodiversity loss (FAO and ITPS, 2015). These processes are exacerbated by intensive human exploitation, urban spread, unsuitable land-use practices, and climate change (Lal, 2014). Therefore, soil mapping is crucial to find ways to manage soil for agriculture and ecosystem resilience. Importantly, most of the processes mentioned above occur on the soil surface which is seen by remote sensing (RS).

This geotechnology is particularly useful for observing the soils when utilizing the visible-near-infrared (VNIR) and short-wave infrared (SWIR) spectral ranges (400–2500 nm). Optical RS allows rapid, high-resolution, and direct measurements of electromagnetic radiation (EMR) that express the chemical and physical information about the sensed object (Stenberg and Viscarra-Rossel, 2010). Viscarra Rossel et al. (2016) demonstrated relationships between various soil properties and soil spectra via EMR interaction with soil in field and laboratory settings, laying the foundation for soil spectral modeling at continental and global scales. The next step was to integrate soil spectral libraries (SSLs) in laboratory and field domains (Francos and Ben-Dor, 2022). This made it possible to fuse the spectral data from the ground and satellites, leveraging the cloud computing to predict soil attributes (Novais et al., 2025). Modeling studies enabled the creation of a global-scale soil reflectance grid at finer spatial resolutions (30 m) thereby enhancing applications such as land-use planning and precision agriculture.

Despite such progress in the soil spectral sensing domain, several issues that constrain upscaling from laboratory to field and global scale needed to be addressed. For instance, Francos and Ben-Dor (2022) proposed to identify a transfer function to scale sensing from laboratory to field and then to RS. Having overcome some challenges, a temporal imaging system was initiated, where a first attempt was performed by Demattê et al. (2016), on a municipal scale. Afterward, Gallo et al. (2018) improved the system at a regional scale. Demattê et al. (2018) implemented an automated system in the Google Earth Engine web platform (Gorelick et al., 2017). The modeling algorithm is named the Geospatial Soil Sensing System (GEOS3), which generates a product (the image), named Synthetic Soil Image (SySi). Simultaneously, a great automatized work was performed by Rogge et al. (2018) named Soil Composite Mapping Processor (SCMaP) at a country scale. Finally, Demattê et al. (2020) applied the GEOS3 to the global domain. This was tested for the Middle East (Poppiel et al., 2020), Europe (Safanelli et al., 2020) and Brazil (Safanelli et al., 2021). Rosin et al. (2023) went further

and upgraded GEOS3 to fill the gaps in areas without bare soil exposure, reaching a continuous predicted bare soil image.

Wesemael et al. (2024) performed with a similar approach, rather with different techniques to a continental scale using cloud and distributed computing, and following a two-pronged approach: one for detecting bare soils and predicting soil organic carbon (SOC) from the soil spectra and one for the permanent vegetated areas (grasslands and woodlands). One key difference between the techniques is that the GEOS system consistently combines laboratory spectra with satellite data, while Wesemael's approach relied on Sentinel-2 imagery, unlike GEOS, which used Landsat. Moreover, Wesemael's approach used satellite data from March to October which includes months where rainfall takes place in many parts of Europe. Additionally, Wesemael's method employed independent validation in specific regions across Europe, whereas GEOS performed field validations, often correlating image patterns with extensive soil mapping campaigns. Another distinction lies in the thresholding techniques for bare soil detection. GEOS applies static thresholds for the spectral indices (i.e., NDVI and NBR2), while the SCMaP employed by Wesemael uses a histogram-based method to create custom thresholds for each Sentinel-2 tile independently, enhancing regional adaptability. These differences highlight the unique contributions and scopes of each system.

Numerous studies demonstrated RS-informed characterization of soil attributes such as organic carbon (Novais et al., 2025; Tziolas et al., 2020b; Vaudour et al., 2022), texture (Gomez et al., 2022), color (Rizzo et al., 2023), micro aggregation (Ben-Dor et al., 2023), and mineralogy (Rosin et al., 2023). Indeed, all this progress in soil spectral modeling has been made possible due to the integration of RS and extensive soil databases. The mosaics formed by RS images permit the fine-scale and contiguous mapping of soils, thereby addressing the limitations of non-contiguous approaches that rely on indirect proxies such as relief (Levin et al., 2004). Furthermore, mosaics of RS serve as proxies for soil-forming factors (McBratney et al., 2018). It is particularly useful in characterizing the spatiotemporal variations of dynamic soil-forming factors such as vegetation, biota, climate, and soil hydrology (Grunwald, 2021). Moreover, a long-term 40-year time series reveals trends and changes in the landscape, such as erosion, vegetation dynamics, and soil moisture variations, enabling the assessment of soil degradation, carbon dynamics, and restoration impacts (Mello et al., 2023). The potential digital soil mapping (DSM) usage remarks this as a significant step forward in soil science. Despite these advances, challenges remain in algorithm development and region-specific soil measurement needs.

Despite soil maps being available worldwide, most of them are represented by a coarse spatial resolution, non-contiguous, and restricted to a specific area, forming a vast puzzle with missing pieces. As soils present a huge arrangement with gradual changes, the current map techniques overgeneralize soil nuances. Problems such as gaps, obsolescence of data, and different production and validation techniques can hamper comprehensive soil monitoring (Safanelli et al., 2021). Furthermore, there are different approaches to soil mapping from classical surveys to DSM by remote sensing techniques as well as more recent DSM incorporating bare soil reflectance (Richer-de-Forges et al., 2023). In addition, the global soil cartographic information available is limited to a coarse-scale resolution, typically >1:1,000,000 (Arrouays et al., 2020; Poggio et al., 2021). Such scales also tend to disregard essential details, such as fragile soils, salt pans, or rocky outcrops. However, accurate soil information is essential for management zoning or precision agriculture (Greschuk et al., 2023). A comprehensive dataset enables effective management from a farm to a global scale.

The most detailed global-scale soil attributes mapping dataset currently available is the SoilGrid, with a spatial resolution of 250 m made with DSM techniques (Poggio et al., 2021). It is important to note that the original data used to create the maps are point-based, while DSM, which emerged in the early 2000s (McBratney et al., 2003; Lagacherie et al., 2006), mobilizes techniques to extrapolate the observations for unsampled areas. In the DSM approach, it is assumed that spatially continuous covariates represent soil-forming factors. As McBratney et al. (2018) demonstrated, this approach offers a relatively efficient, cost-effective, and environmentally conscious alternative to traditional soil mapping methods even over vegetated areas. Models such as machine learning (ML) algorithms can be used in DSM, which encompasses the spatial modeling of soil properties (Ma et al., 2019; Padarian et al., 2019; Khaledian and Miller, 2020; Wadoux et al., 2020; Grunwald, 2021). Optical RS time series have several challenges (i.e., clouds and field features such as straw, soil moisture, and rocks), which interfere with the soil spectral signature as observed in the laboratory (Ben-Dor et al., 2009).

The indicated temporal image techniques have made efforts to avoid most of these issues, and more will come with new hyperspectral satellite systems that are starting to be applicable from Earth orbit such as the current Earth Surface Mineral Dust Source Investigation (EMIT), PRecursore IperSpettrale della Missione Applicativa (PRISMA), Environmental Mapping and Analysis Program (EnMAP), Copernicus Hyperspectral Imaging Mission for the Environment (CHIME), and Surface Biology and Geology (SBG). These sensors have great spectral resolution, yet limited temporal repetition, making it difficult to perform mapping that relies on change detection. On the other hand, multispectral sensors, such as Landsat with 40 years of temporal information can depict sequential imagery of soils with and without tillage coverage. Accordingly, despite Landsat's low spectral and spatial resolutions (30–60 m) outperforms other RS in terms of its temporal resolution. Most literature shows that Landsat spectra are useful due to their strong correlation with soil properties, especially clay content (Fongaro et al., 2018) and carbon content (Padarian et al., 2022).

This work aimed to construct a global soil spectral grid (30 m spatial resolution) to identify the spatial variability of the world's soil patterns in dependence on several spatially and temporally varying environmental factors, using Landsat data. We consider the possible spatiotemporal relationships to monitor kinetic processes on the Earth's bare surface with implications for agriculture and environmental science, soil management, and climate change mitigation strategies (Ortiz-Bobea et al., 2021). To achieve this goal, we combine data from extensive SSL, wet laboratory soil analyses, and optical Earth observations.

2. Methodological procedures

Fig. 1 delineates the steps involved in constructing the global soil spectral grid. The datasets ensured broad spatial coverage representing publicly available soil data. A rigorous soil sample preprocessing screening, including drying, crushing, sieving, and VNIR-SWIR spectral measurements, was conducted to ensure laboratory consistency. The GEOS3 is a modeling concept that models satellite data in a temporal manner (Demattê et al., 2018). Each pixel created is named Temporal Soil Spectra (TESS), and the entire image is the product finally named Synthetic Soil Image (SySi). The present case depicts a Landsat time series providing a robust global representation of soil reflectance. Clustering methods, including a two-stage self-organizing map process (SOMP) and Partition Around Medoids (PAM) clustering, reduced the dimensionality of the SySi data and classified it into spectral groups for spatial interpretation. The Boruta algorithm determined the significance of environmental covariates, enabling data-driven analyses of soil-environment relationships. This approach generates a precise, scalable soil spectral grid suitable for diverse environmental and agricultural applications.

2.1. Soil samples and dataset

Data on soil attributes and laboratory spectra were obtained from two global sources, the WoSIS (Batjes et al., 2020) and the GSSL dataset (Demattê et al., 2022) (Fig. 2). The former provides data from about 96,000 georeferenced profiles (Batjes et al., 2020), while the latter GSSL

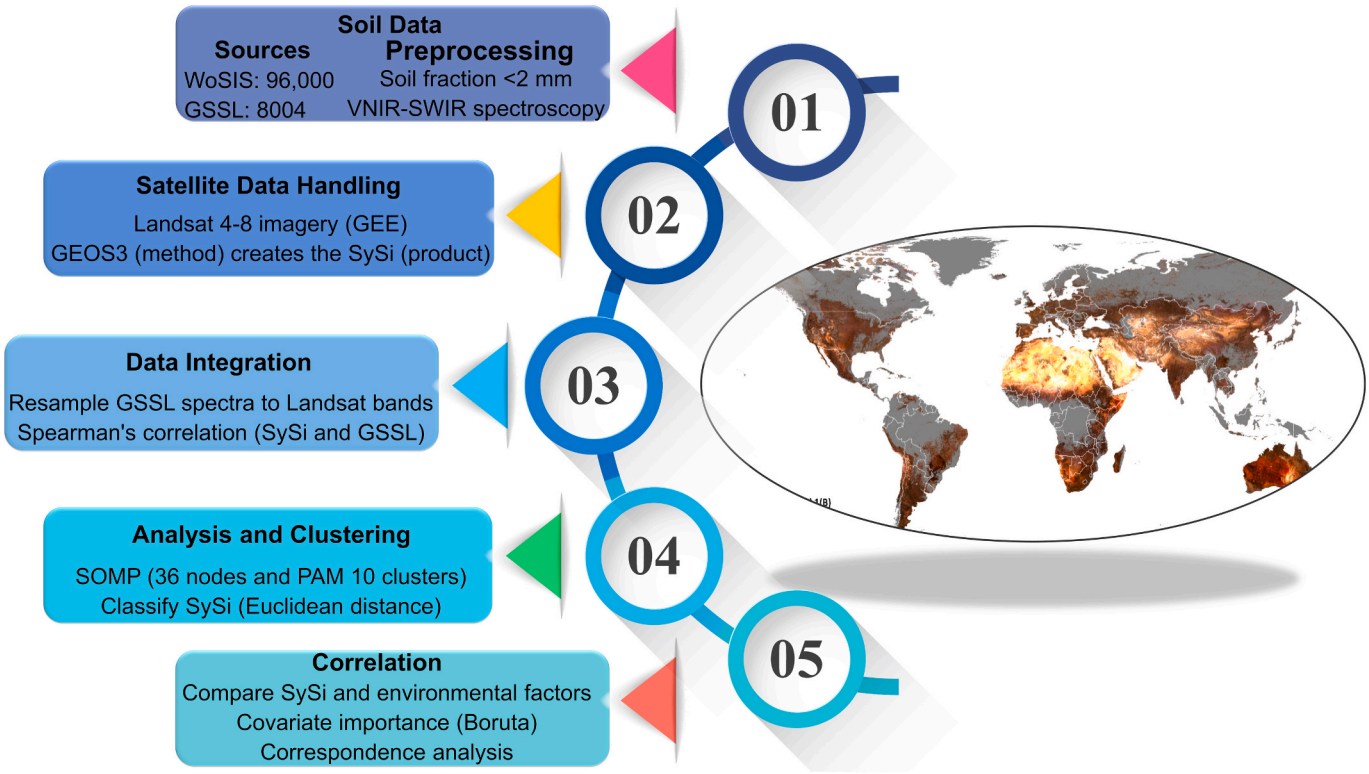


Fig. 1. Procedural steps for global bare soil grid construction. WoSIS: World Soil Information Service; GSSL: Global Soil Sampling and Site Inventory; GEE: Google Earth Engine; SySi: Synthetic Soil Image; GEOS3: Geospatial Soil Sensing System; SOMP: Self-Organizing Map Process; PAM: Partition Around Medoids.

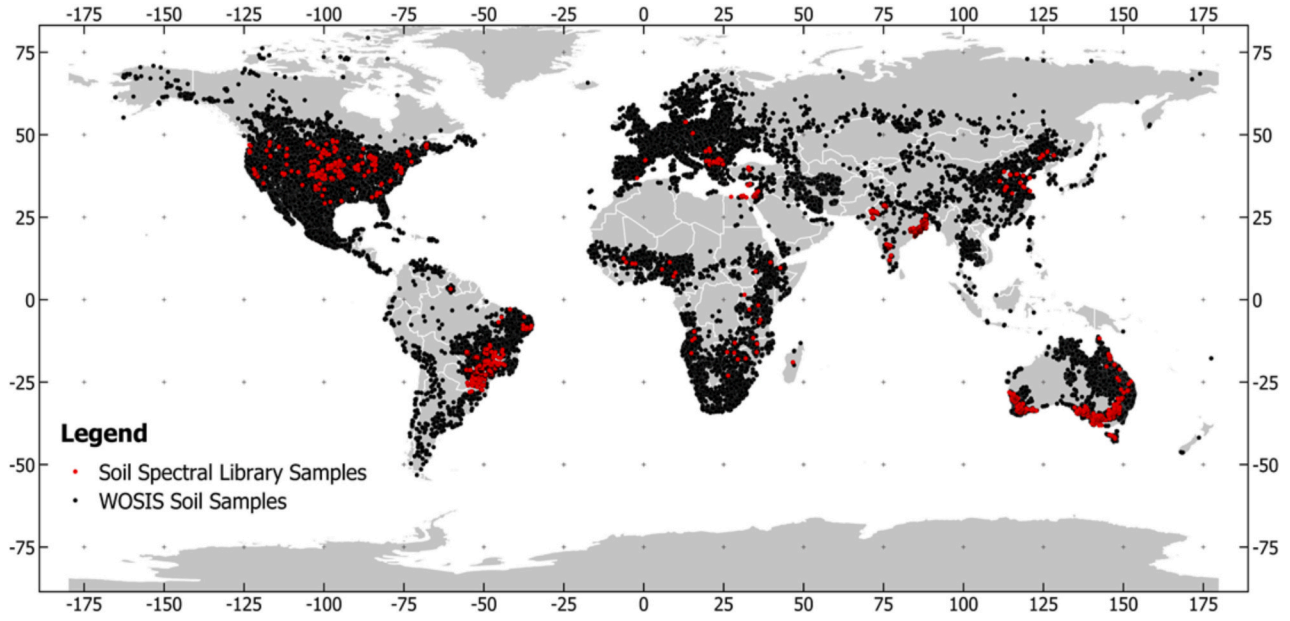


Fig. 2. Location of soil samples with data of clay, sand, silt, SOC, and VNIR-SWIR spectra measured in the laboratory belonging to the GSSL (red dots) and WoSIS (black dots) datasets.

dataset provides 8004 topsoil samples (Demattê et al., 2022). Only observations with geographic coordinate accuracy (< 0.0001 degrees, or about 10 m) were included in this case. We focused on soil attributes and topsoil samples up to 20 cm depth, including clay, sand, silt, and SOC. The number of samples for particle size fractions and SOC was 125,636 and 30,980, respectively.

For standard quality conditions of soil spectral data, the soil samples from GSSL were air- or oven-dried, crushed, and sieved to a size <2 mm. Sample holders were at least 6 cm in diameter and 1 cm deep, except when measurements were made with a fiber optic. The spectrometer has a spectral resolution of at least 10 nm in the VNIR-SWIR region (between 400 and 2500 nm). The illumination source was a halogen lamp, and a spectralon© white reference optimized and calibrated the sensor. Spectra were recorded as diffuse reflectance varying from 0 to 1 (more details refer to Demattê et al., 2022).

2.2. Bare surface composite image production

We analyzed Tier 1 Collection 1 surface reflectance data processed by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) and Surface Reflectance Code (LaSRC) algorithms (USGS, 2020a, 2020b) from the Landsat 4 Thematic Mapper (TM), Landsat 5 TM, Landsat 7 Enhanced Thematic Mapper Plus (ETM+), and Landsat 8 Operational Land Imager (OLI) sensors. These datasets are accessible through the Google Earth Engine (GEE) cloud platform (Gorelick et al., 2017). Given the different spectral configurations of the OLI sensors, such as spectral resolution, Landsat 8 imagery was transformed to ETM+ following a harmonization process (Roy et al., 2016).

The GEOS3 was used to identify bare surfaces in each available Landsat satellite image from 1984 to 2022. These images have a temporal resolution of 16 days and a spatial resolution of 30 m. The bare surfaces were labeled based on several indices: the greenness index (GREEN), the normalized difference vegetation index (NDVI), the normalized burn ratio index (NBR), and the normalized burn ratio index 2nd version (NBR2). The term 'bare surface' was assigned when the following conditions were met: GREEN <0.65, $-0.05 <$ NDVI <0.25, NBR > -0.23, and $-0.05 <$ NBR2 < 0.15. Clouds, shadows, and other non-bare ground pixels were removed using the quality assessment band (QA band). Each pixel in the QA band has a decimal value indicating surface, atmospheric, and sensor conditions that affect the overall usefulness of the data (USGS, 2020a, 2020b). The band is in unsigned 16-bit format and provides information on various pixel conditions, including water, cloud, cloud shadow, and snow.

We created a mask to eliminate pixels not marked as "clear" to ensure that only land surface reflectance was retained. The mask was then created and applied to the entire satellite image collection, preserving only bare land surface reflectance data, thereby eliminating the influence of clouds or snow. It is noteworthy that the GEOS3 enables the assessment of spatiotemporal variability in two forms: Firstly, by analyzing all the time series (i.e., from 1984 to 2024), and secondly, by periods, which may be annually, monthly, or even daily, depending on the sensor revisit. The resulting product was a space-time cube of masked images containing bare surface pixels, characterized by a 30 m resolution and 16 days of revisit.

The images within the space-time cube were then aggregated, using the sparse spatiotemporal occurrences, into a single composite product representing the median surface reflectance value named SySi. SySi captures natural abiotic surfaces, such as bare soil, sand, and rock outcrops, where vegetation negligibly affects the reflected signal (Demattê et al., 2020). Masked pixels had no data on the SySi and did not impact processing; these areas are typically associated with natural vegetation fragments or soil surfaces under frequent succession or crop residues.

2.3. Comparison between SySi and GSSL

The initial assessment involved comparing data from SySi with the GSSL spectral library. Given their different spectral resolutions, the spectra from the GSSL were resampled to align with the six reflectance bands from Landsat sensors, namely, blue (450–520 nm), green (520–600 nm), red (630 nm - 690 nm), NIR (770–90 nm), SWIR1 (1550–1750 nm), and SWIR2 (2080–2350 nm) (Demattê et al., 2018). Afterward, Spearman's correlation coefficients served as comparison parameters for the two spectra types, where $|r| \geq 0.7$: Strong relationship; $0.4 \leq |r| < 0.7$: Moderate relationship; and $|r| < 0.4$: Weak relationship, generally considered poor for meaningful predictions.

2.3.1. Definition of SySi spectral groups

A quantitative analysis of spectra patterns was conducted at various locations worldwide to assess the consistency of SySi. This process involved a clustering process of SySi's pixels, followed by detailed interpretation. The spectral groups (SGs) were defined using six image bands through a SOMP, an unsupervised neural network capable of reducing the dimensionality of large datasets (Kohonen, 2001). This algorithm attenuates the impact of data complexity or potential outliers. Initially, SOMP was applied solely to SySi reflectance data at the WoSIS sampling locations.

SOMP used the spectra data ("input space") to generate a lower-dimensional representation ("map space"). The map space is divided into a specific number of components (also known as nodes), resulting in a hexagonal or rectangular grid. In this study, we defined a hexagonal grid with 36 nodes, which allows for representing the variability of the data and drastically reducing the dimensionality of the dataset. Every node in the map space is associated with a weight vector, representing the node's position in the input space. While the nodes in the map space remain fixed, the algorithm moves the weight vectors toward the input data, aiming to minimize the Euclidean distance without compromising the topology from the map space.

During the process, SOMP performs several successive iterations until the weight vectors represent, as closely as possible, the input patterns closer to the nodes in the two-dimensional map. In the second stage of the process, the spatial analysis occurred using a classical clustering algorithm called the PAM (Kaufman and Rousseeuw, 1990). The SOMP vectors are considered as local averages of the input data. The final number of clusters was defined according to the methodology proposed by Zappia and Oshlack (2018), resulting in 10 clusters. The two-stage method reduces sensitivity to outliers and noise while lowering computational costs.

Finally, SGs were extrapolated to the entire data SySi by classifying each pixel according to the similarity to the closest medoids. The similarity between medoids and pixels was calculated using the Euclidean distance metric. Euclidean distance is particularly suitable for this task due to its simplicity, intuitive interpretation, and computational efficiency.

2.3.2. Relationship between SySi spectral groups and environmental factors

We compared various environmental covariates previously used in global DSM (Hengl et al., 2017) with the six synthetic reflectance bands from SySi. This comparison enabled us to assess associations between 13 relief attributes (i.e., Aspect, Eastness, Elevation, Gaussian Curvature, Hillshade, Horizontal Curvature, Maximal Curvature, Mean Curvature, Minimal Curvature, Northness, Shape Index, Slope and Vertical Curvature) and the SySi bands (blue, green, red, NIR, SWIR1 and SWIR2). The importance of each covariate was calculated using the Boruta algorithm (Kursa and Rudnicki, 2010). During this process, covariates were ranked by considering five specific soil attributes: clay, sand, silt, carbonates, and soil organic carbon. We selected these attributes because they are spectrally active soil constituents frequently used as physicochemical and biological indicators for soil quality assessments (Liu et al., 2020).

Additionally, once the SGs were defined, they were compared to environmental factors directly related to soils and their formation. The factors considered in our analysis included global temperature-moisture domains, relief, land use for the year 2015 (Sayre et al., 2020), world lithology (Hartmann and Moosdorf, 2012), soil classes (FAO-UNESCO, 1974), and topsoil textural classes derived from WoSIS samples. Associations between factors and SySi were investigated using correspondence analysis (CA) (Lele et al., 2007). This analysis enables the reduction of data dimensions into orthogonal components, which explain most of the variation. Continuous and categorical data can be managed with CA in a contingency table (Viscarra Rossel et al., 2016). In our case, a contingency table with units (spectral group) and variables (environmental factor) was computed and later converted into a smaller set of dimensions (principal coordinates). Those dimensions were visualized by plotting the scores for SGs alongside each environmental factor in ordination graphs.

2.3.3. Boruta algorithm to define SySi importance in soil mapping

The Boruta algorithm provides a means to a robust evaluation of SySi's importance in DSM processes. The method relies on the random

forests (RF) classifier to rank the most important covariates for predicting a specific soil attribute (Kursa and Rudnicki, 2010). Initially, the algorithm extends the covariate space by adding randomly permuted existing covariates (pCs), which eliminates their correlation with the response variable (soil attribute). Later, RF predicts a specific soil attribute and calculates the Z-scores (a covariate importance indicator).

The maximum Z-score (MaxZ) among the pCs is defined and actual covariates that scored better than MaxZ are defined as "important" and retained. Attributes with undetermined importance are submitted to a two-sided test of equality with the MaxZ. Attributes with an importance significantly lower than MaxZ were defined as 'unimportant' and permanently removed from the system. Finally, all pCs were removed and the procedure was repeated. The algorithm was performed recursively until importance was assigned for all the attributes, or the algorithm had reached the previously set limit of the RF runs.

3. Results and discussion

3.1. Data consistency

Addressing the need for more detailed soil information, this research "uncovered" 90 % of Earth's bare soils under agricultural or natural land use and land cover at a resolution of 30 m. This product not only offers finer resolution than those previously available ($\geq$ 250 m) but also demonstrates high consistency considering early attempts (i.e., Demattê et al., 2018, 2020). This approach can be considered a method of sustainable survey, as it minimizes environmental impacts, costs, and the time required for execution, reducing analyses needed for large-scale surveys (McBratney et al., 2018). None of the other studies that produced images of bare surfaces at global scales had reached this scope and detail level (Poggio et al., 2021). The EMR is regionally dependent, and the algorithm evaluates each case throughout the entire globe. Furthermore, only a limited number of these studies evaluated the reliability of RS data (cf. Safanelli et al., 2020; Tziolas et al., 2020a, 2020b).

On data consistency, a rigorous control process for soil sampling and laboratory measurements showed correlation values between observed and predicted data. Accuracy concerning field observation data varies from moderate to strong relationships (Huete and Escadafal, 1991). The values ranged from 0.5 to 0.9. As an example of this consistency, spectra over the Syrian territory exhibited higher correlation values than those reported in studies with a local focus. This endeavor is noteworthy, considering the constraints imposed by the scale of our analysis and the utilization of a global dataset derived from multiple sources (Poggio et al., 2021).

Table 1 presents the spectra obtained by SySi in comparison with those from the Global Soil Spectral Library (GSSL), which yielded significant results ($p < 0.0001$). Spearman's correlations (r) ranged from 0.67 to 0.84, with higher values observed for the blue, green, and SWIR spectral bands, key ranges for soil properties modeling (Demattê et al., 2020). This observation indicates that considering the whole studied area of bare soil, the filtering of the GEOS3 system found relative consistency between the ground and space sensing data.

Table 1
Spearman correlation values between satellite data and laboratory spectra (n = 104,004).

Simulated bands (SySi)	Landsat bands					
	Blue	Green	Red	NIR	SWIR 1	SWIR 2
Blue	0.84	0.83	0.69	0.71	0.64	0.51
Green	0.82	0.82	0.71	0. 72	0.65	0.51
Red	0.71	0.74	0.69	0.70	0.63	0.53
NIR	0.75	0.78	0.73	0.75	0.69	0.59
SWIR 1	0.77	0.80	0.76	0.79	0.77	0.66
SWIR 2	0.74	0.77	0.74	0.77	0.75	0.67

3.2. Relation between SySi's spectral behavior and environmental factors

The SySi successfully captured the Earth's surface reflectance patterns intrinsic to environmental variability (Fig. 3a), enabling the observation of spectral characteristics at various locations worldwide (Fig. 3b). Despite focusing on only six bands from Landsat sensors, it represents the wavelengths in the VNIR-SWIR spectral ranges and a practical evaluation was achieved by analyzing spectral shapes and reflectance patterns using SySi. Ten spectral groups (SG) represented the diversity of the world's soils. Subsequently, these groups were compared with environmental factors directly influencing soil formation and characteristics.

SGs 1 and 2 were the most common groups (Fig. 3b), primarily associated with hilly croplands and grasslands in tropical and temperate regions. The parent material consists of unconsolidated sediments, metamorphic rocks, or, in some cases, evaporites. These factors are predominantly found in medium-textured soils (in SG1). Furthermore, SG 1 and 2 spectra showed similar characteristics, with medium reflectance intensity and convex shape. The higher reflectance intensity of SG1 is attributed to the increased presence of coarse particles and the soils' lower soil organic carbon content (SOC), distinguishing it from SG2 (Demattê et al., 2020). This finding can be used to distinguish soil, i.e.: by texture (Fongaro et al., 2018; Gomez et al., 2022), SOC (Padarian et al., 2022), color (Rizzo et al., 2023), or floodplains (Mello et al., 2023).

Joint observations of Fig. 3a and b show that SG3 prevailed in plain pastures and croplands, highlighted in western China and western Asia. Similarly, SGs 4 and 5 predominated in natural landscapes with complex relief. SGs 3, 4, and 5 exhibited the lowest reflectance intensities, with shapes varying from convex (SG3) to flat or concave (SG5) caused by the soil physicochemical conditions. These three SGs also exhibited the highest SOC among all SGs, which reduced reflectance in the VNIR region, particularly in the visible region, as Stoner and Baumgardner (1981) observed. Interestingly, SG3 was associated with soils containing argic horizons (i.e., Phaeozems, and Planosols), resulting in lower soil surface clay content due to remotion processes (Buol et al., 2011).

As shown in Fig. 3a, SG4 corresponded to soil classes with dark surface horizons because the curve presents low albedo in the visible region. As an example of these soil types, we identified the Chernozems, Vertisols, or Rendzinas, with blackish colors due to organic matter accumulation or moisture (Liu et al., 2020). The map of soil groups (Fig. 3b) showed a higher occurrence of SG4 in some regions of Russia, Ukraine, and Kazakhstan, which is consistent with the occurrence of the soil as mentioned earlier in classes of the FAO-UNESCO (1974) world soil map (see Supplemental Figs. S1 and S2). Nevertheless, the low reflectance of SG 5 (Fig. 3a) is influenced by parent material, specifically volcanic rocks rich in mafic minerals. These minerals absorb more electromagnetic energy, producing a distinctive pattern in the soil spectra (Stoner and Baumgardner, 1981). Moreover, the fine and very fine soil texture of the locations with SG5 (Fig. 3b) also contributed to the predominance of these reflectance patterns.

In contrast, SGs 6 to 10 have been observed mainly in tropical, subtropical, or temperate deserts, where soils are mainly unconsolidated with higher quartz or salt concentrations reflecting most light that affects them. These characteristics and the low SOC content are the main reasons for the highest reflectance (Howari et al., 2002; Stenberg and Viscarra-Rossel, 2010). SG 6 was observed sparsely throughout Asia, while SG 7 was predominantly found in the Namibia Desert (southwestern Africa) and the Thar Desert (India). These two groups were primarily associated with Yermosols, soils common in semiarid to arid hot climates.

SG 8 corresponded to Arenosols and Solonchaks, sandy and salic soils, observed in a semi-desert area south of the Sahara, the Simpson Desert (Australia), or temperate deserts such as the Taklamakan in China. Desert areas with salt-affected soils (e.g., Solonchaks) were also observed in SG 9, highlighting the potential of remote sensing products

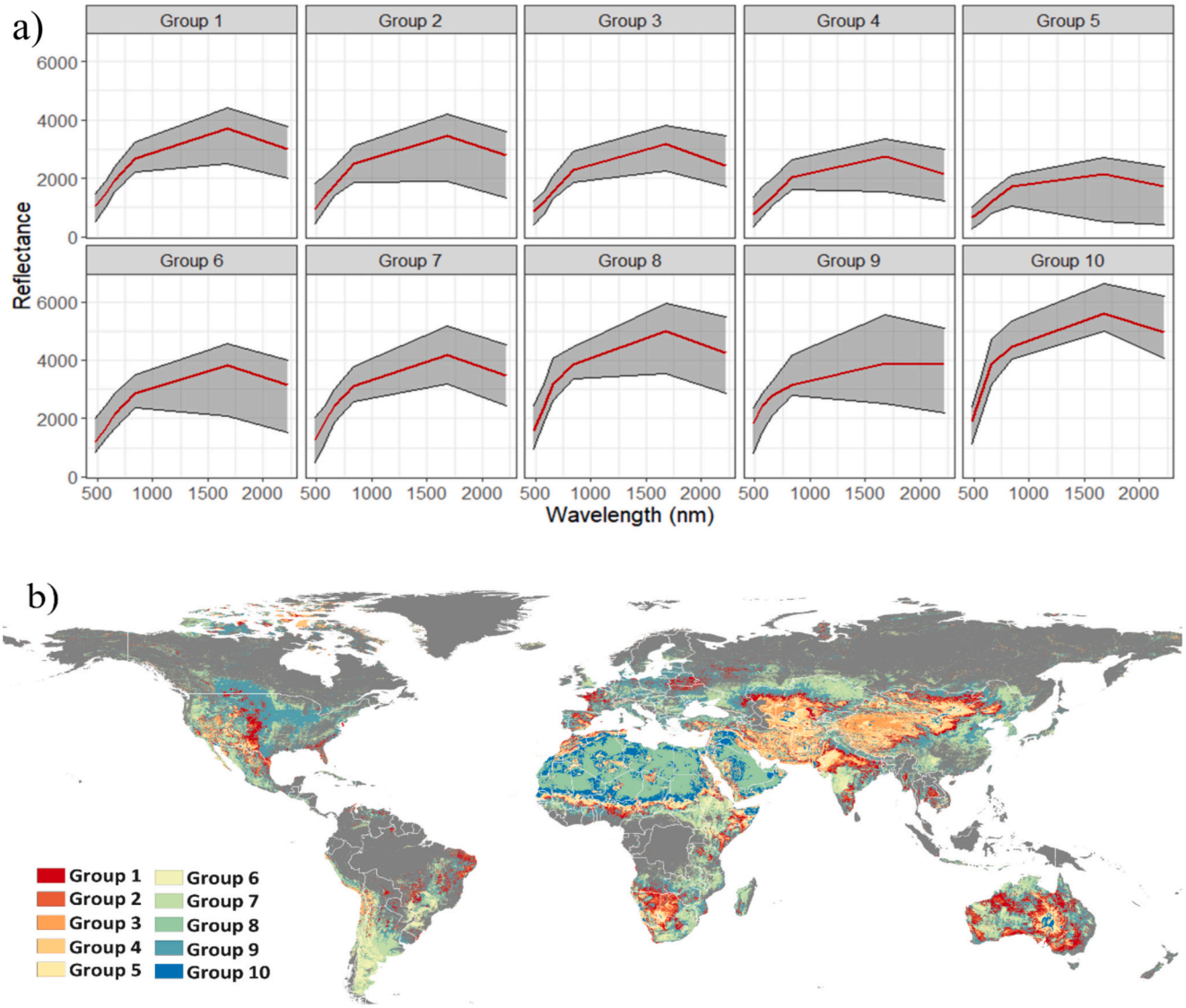


Fig. 3. Soil spectral patterns worldwide. a) Mean spectra for the ten soil spectral groups with minimum and maximum values and b) map of the ten soil spectral groups (SG).

such as SySi in identifying and mapping such soils (Naimi et al., 2021; Wang et al., 2021). Finally, SG 10 was the dominant group in the Sahara and is strongly associated with Yermosols (FAO-UNESCO, 1974). Considering the characteristics of such soils and the regional climatic conditions (tropical desert), we expected a higher intensity of this SG due to its very convex shape. Middle-Western and boundary of the Sahara, with sandy soil prevalence, are the most representative regions with the SG 10 (Fig. 3b).

3.3. SySi relevance for Earth topsoil modeling

The findings indicate a strong correlation between SySi and topsoil attributes or soil classes suggesting that it is a valuable tool for soil characterization, mapping, and classification (Demattê et al., 2020) although only the soil surface was evaluated. It has been demonstrated that bare surface images can enhance DSM, particularly at local and regional scales (Ng et al., 2023; Urbina-Salazar et al., 2023; Rizzo et al., 2023). Consequently, we compared the environmental covariates previously utilized in the global DSM (Hengl et al., 2017; Poggio et al., 2021) and the six reflectance bands from the SySi. Similarities in soil patterns can be observed in both grids (Fig. 4a and b). The potential of these variables to support global mapping, such as the mapping of topsoil color, was discussed following Rizzo et al. (2023).

Spectral patterns are present at specific positions due to combinations of specific environmental factors such as climate, land use, and geology. The long-term mean of precipitation and temperature, representing climatic covariates, were weakly to moderately relevant to infer soil properties (Fig. 5a). We expected a more robust correlation, especially considering that different soil moisture regimes affect the leaching of base cations, enriching the surface layers of soil with pigments such as oxide. This climatic factor is the leading cause of horizontal and vertical variations in soil color (Liu et al., 2020). Many factors influence the comparison with climate zones, but an important consideration is the low accuracy of the climate maps. Soil spectral data can change drastically within a small area, thus, higher spatial resolution covariates could enhance inference on spatially explicit soil properties and soil forming processes.

Among the LULC covariates (Fig. 5b), Forest, Cropland, Shrubland and Grassland achieved the highest importance in the model, as these LULC types exert a strong influence on reflectance (Demattê et al., 2020). We found that only the plain areas were moderately relevant among relief data (Fig. 5c). Topographic landscape covariates are

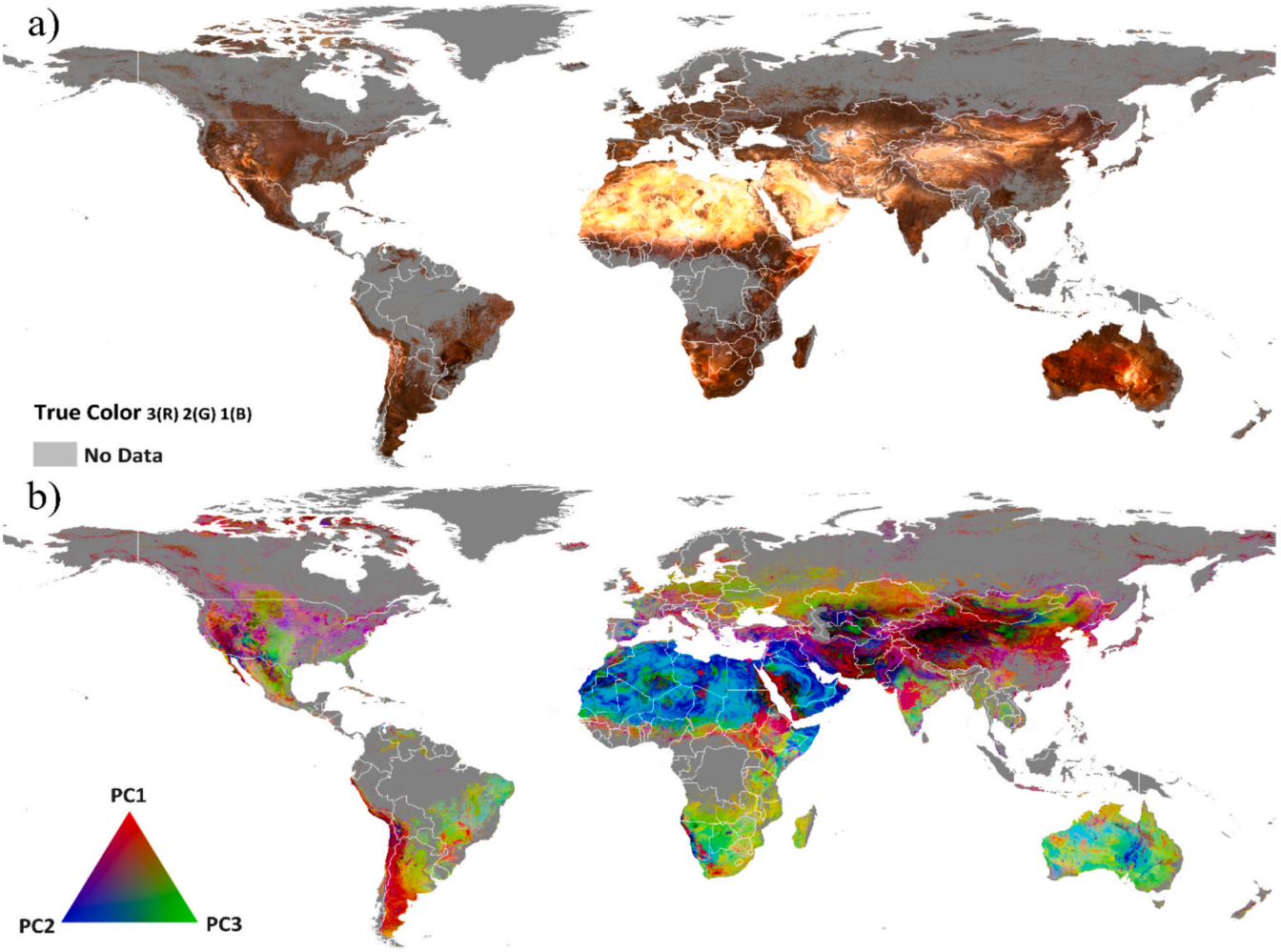


Fig. 4. Distribution of spectral patterns and statistics over Earth's bare surfaces. (a) True-color image of bare surfaces (Landsat ETM+ band 1 = blue, band 2 = green, and band 3 = red). (b) Red, green, and blue (RGB) image depicting pixel-based spatialization of the three principal components (PC1, PC2, PC3) on the bare surfaces.

related to water dynamics and indicate erosion intensity, redistribution of soil material, particle sorting processes, and thermal regimes. Some of the terrain attributes are critical covariates (e.g., Topographic Wetness Index, Multiresolution Index of Valley Bottom Flatness, and LS factor) in DSM (McBratney et al., 2018). However, these secondary topographic data were not used in our study due to their computational demand which makes the processing difficult, even on cloud computing platforms (Safanelli et al., 2020). Fig. 5c shows the relation of SG. SG1 predominates arid (i.e., Middle East) and semiarid areas (i.e., northeastern Brazil) while SG4 prevails in mountainous landscapes (i.e., north and southwestern Asia).

Lithology showed a strong relationship with soil spectral behavior. Basaltic rock areas were close to SG5 (Fig. 5d) as are Ferralsols (Fig. 5f). Indeed, these patterns are consistent with field observation, such as the typical occurrence of Ferralsols on Basalt in southwestern Brazil (Gallo et al., 2018). In summary, all this information is essential for decision-making regarding managing or protecting specific areas.

3.4. Remote sensing informed global analysis of soil variation and diversity

Using a surface soil proxy is essential in quantifying various soil attributes as Demattê et al. (2020) have already observed. This approach is well supported by >50 years of laboratory research focused on establishing relationships between soil attributes and reflectance, mainly conducted in laboratory settings (Knox et al., 2015; Viscarra Rossel et al., 2015; Clingensmith and Grunwald, 2022; Ben-Dor and Demattê, 2024). Indeed, the statistics of each variable of the matrix (Fig. 6) from our study show the significant importance of soil properties and spectral data on a global scale. Elevation and SWIR2 band show the highest importance with 78 and 80 %, respectively, followed by spectral data, while the lowest values were found for SOC and calcium carbonate. The combination of elevation and spectral data achieved a higher R^2 than using elevation alone, which is in agreement with the findings of Fongaro et al. (2018).

The evidence presented here suggests that SySi bands provide critical information to describe the spatial variability of topsoil color, which is consistent with findings by Rizzo et al. (2023). This statement is justified because bare surface images capture real-time data about soils through reflectance. While SySi bands effectively capture topsoil color variability, they have several limitations: weak correlations with specific soil properties, inadequate representation of certain soil groups, spatially varying accuracy due to environmental factors, uncertainties arising from the sensor or atmospheric conditions, and the generalization of temporal aspects by harmonizing the pixels. These issues should be considered to improve the reliability of soil map products.

3.5. Insights for societal issues

The majority of DSM studies do not use SySi information and therefore may face difficulties in capturing adequate soil variability, which is a challenge for accurate pedological delineation and mapping.

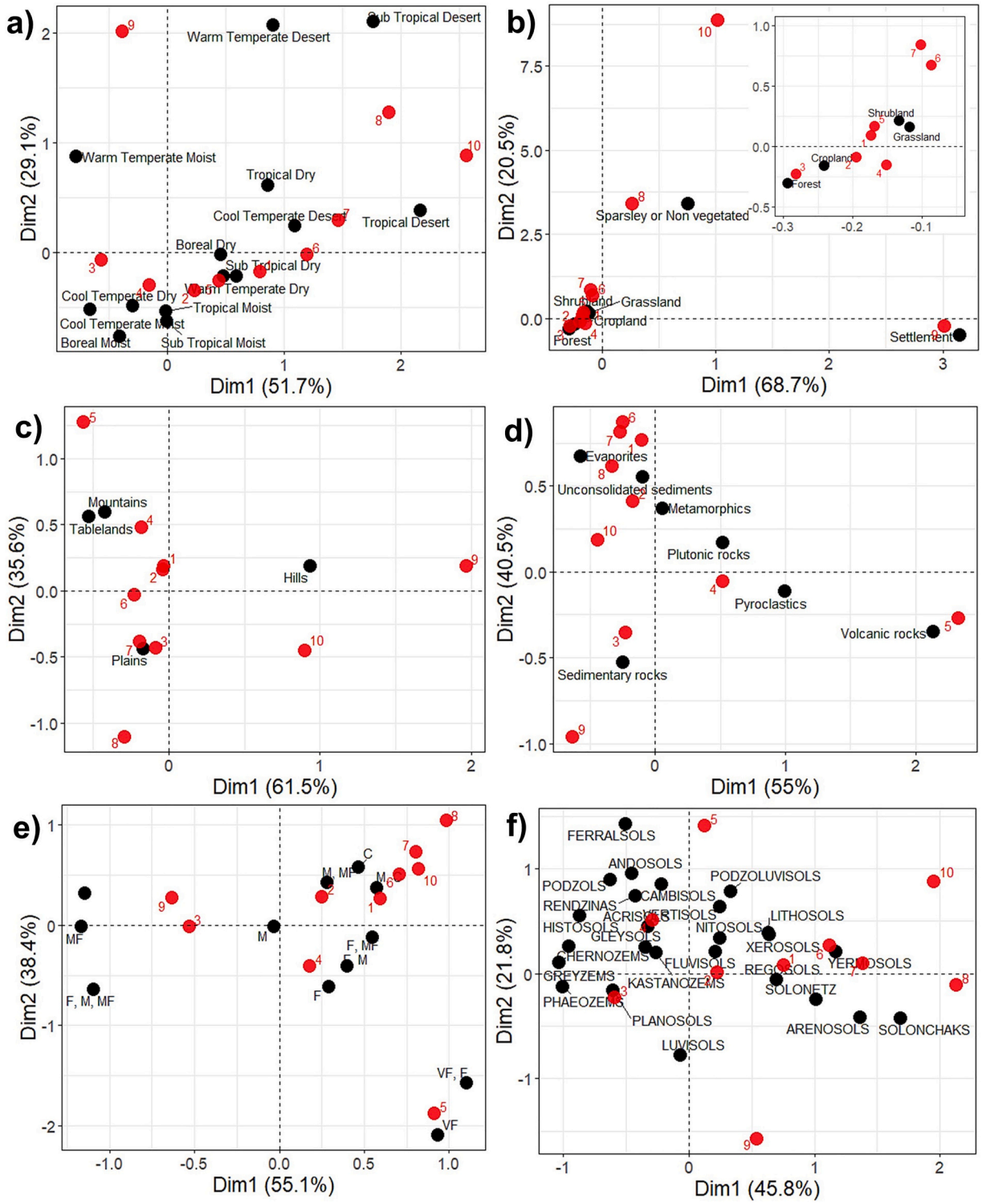


Fig. 5. Comparison of the position of spectral groups (red circles) to a) climate zones of the world (Sayre et al., 2020), b) land use and land cover (Sayre et al., 2020), c) landforms distribution (Sayre et al., 2020), d) Earth's geological groups (Hartmann and Moosdorf, 2012), e) soil textural classes (Batjes et al., 2017), and f) soil classes (FAO-UNESCO, 1974) (a, b, c, d, e, in black circles).

For example, users cannot observe soil variation beneath the soil surface (Fig. 7a). In Fig. 7b, a comparison is shown between pedological delineation without SySi (cyan line) and with SySi (brown line). SySi provides critical important background information that makes the result different and affects several subsequent decisions, such as soil management, fertilization, and conservation. Fig. 7c highlights a classical soil survey delineation (yellow line) based on the site (borehole) plot and relief. Applying the SySi (Fig. 7d), the areas with different spectra (related to other soil properties) are highlighted and affect new locations to insert the sites, causing an utterly new delineation, as Safanelli et al. (2021) observed in a similar situation.

Both conventional (Hartemink et al., 2013; Negasa et al., 2017) and digital (Chen et al., 2022; Ma et al., 2019) soil sampling methods may fail to capture the full range of soil variation if they do not consider soil

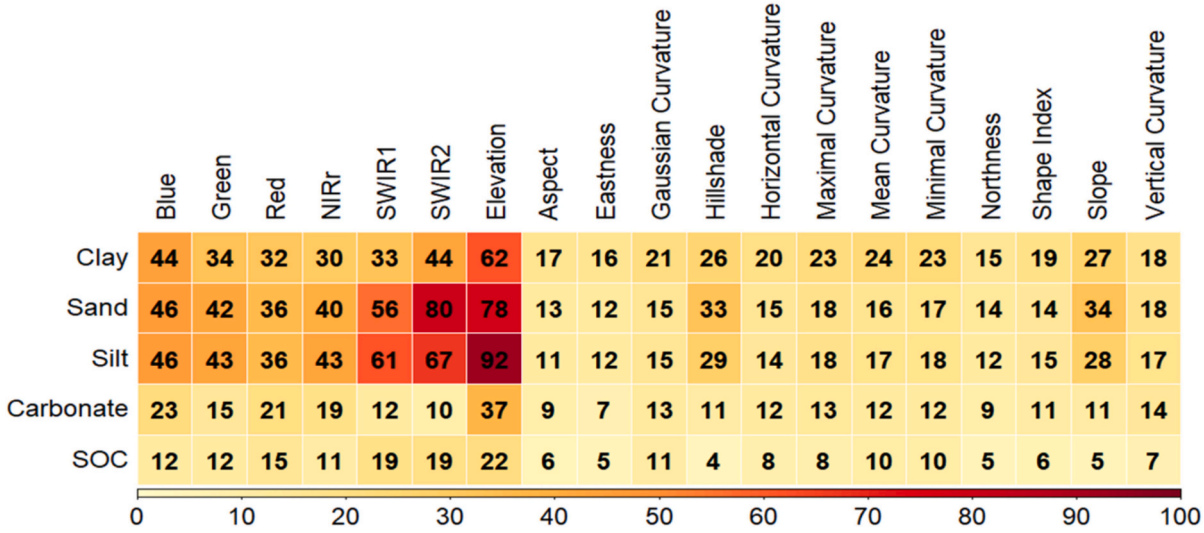


Fig. 6. Contribution of SySi bands and terrain covariates with soil attributes modeling (sand, silt, clay, carbonate, and SOC). Values in percentage (%).

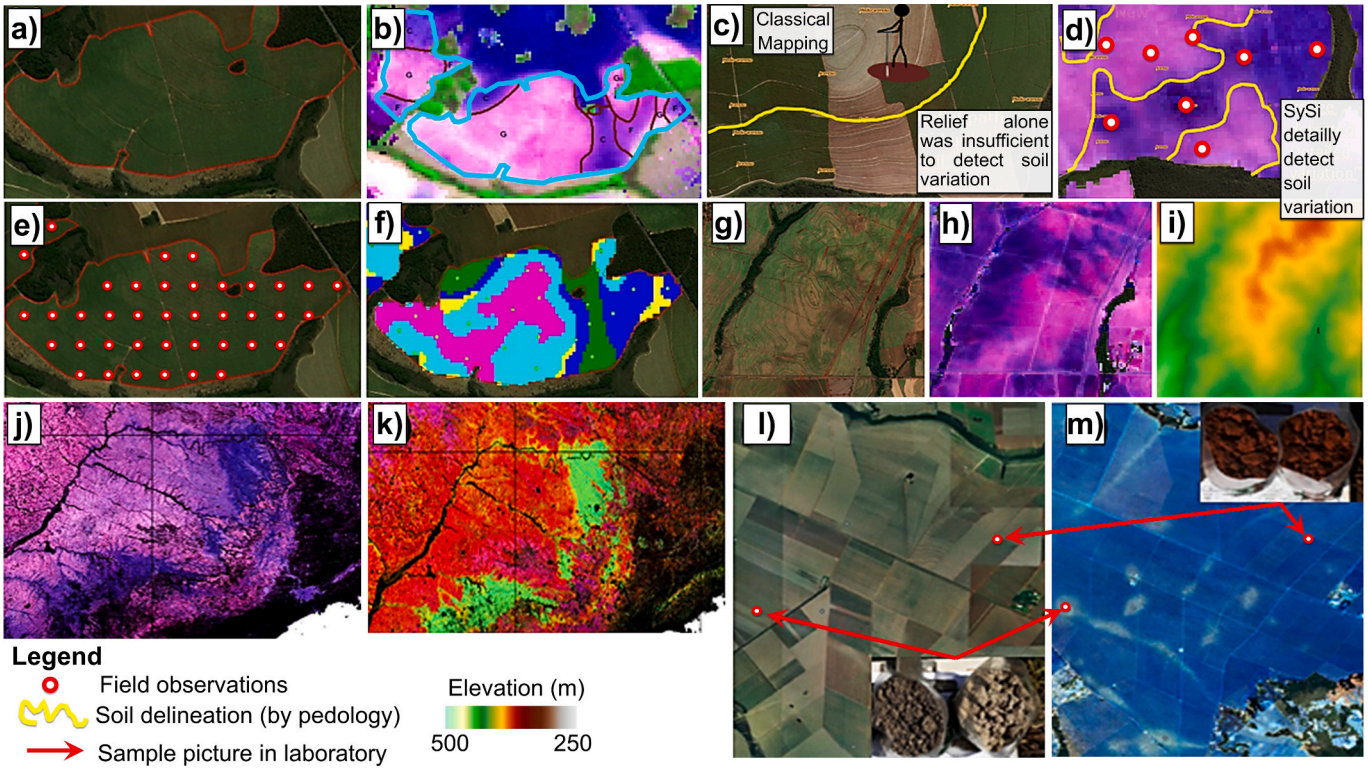


Fig. 7. Scale is not presented due to high variability ranging from farm to state scale. The examples are from Brazil. a) A Google image in true color composition showing sugarcane crop in Goiás state, Brazil, b) SySi 5R4G3B composition with traditional soil survey (using only relief) in blue lines, and new soil type surveyed in brown lines (using relief plus SySI), c) conventional soil survey delineation (yellow line) at a site; d) the same site with new delineation using SySi and the locations to soil sampling, e) precision agriculture grid sampling, f) same site with new soil site allocation using SySi, g) a Google Earth image with natural color covered by sugarcane cultivation, h) the same site with SySi, highlighting soil variations, i) the same area with elevation model, j) SySi image of São Paulo, and Párana States, Brazil, in 5R4G3B color composition, k) soil potential productivity map(green indicates a high potential and red indicates a low potential) (Greschuk et al., 2023), l) Google Earth image of a cropland area, and m) same area with SySi 5R4G3B composition (Mello et al., 2023), the red arrows go from the laboratory sample pictures to the sampling places.

as a continuous feature across the landscape. Conventional surveying typically relies on manual sampling and observation, with surveyors collecting soil samples based on predefined locations, often using relief-focused approaches. In contrast, digital surveying incorporates technology, spatial data, and modeling techniques, such as RS and GIS, to map soil properties more continuously across a landscape. Traditional surveyors typically use relief to delineate soil classes on the landscape; however, when bare soil information is added, multiple underlying variations are revealed, leading to significant changes in delineation. In summary, when positioning soil samples, SySi data could be considered, when available, to ensure a more accurate representation. For instance, incorporating SySi data in stratified random sampling designs is a

valuable asset to improve the placement of soil sampling locations in future studies.

Precision agriculture and soil management zoning are strategies for collecting soil samples at specific locations to guide nutrient management with optimized fertilizer applications. Optimizing soil management relies on accurately mapping soil variability controlled by various soil-environmental factors (Fig. 7e, f). In the first case, if a farmer makes a regular grid, it is not possible to observe the differences. However, with SySi the sampling allocation can be targeted to areas of soil variation, which subsequently improves the accuracy of the precision farming system. Fig. 7g, h, and i compare a standard image, an image with SySi, and an image with only relief (DEM). Soil variation does not necessarily correspond to relief variation; thus, a pedologist can only interpret the delineation using this traditional information. Indeed, pedologists frequently delineate the soil transitions based on landforms, with or without the support of aerial photographs or altitude maps. However, SySi provides new information to be integrated into this process. It captures direct measurements of the topsoil, representing the diversity of multiple soil properties. Thus, while the relief may appear uniform, spectra reveal changes that would otherwise go unnoticed. Pedologists likely lose accuracy in their delineations of soils without incorporating soil spectral information.

Spectral reflectance can be linked to particle size distribution, carbon content, carbon stock, pH, chemistry elements, clay content and mineralogy (Nocita et al., 2015; Novais et al., 2025). Therefore, it is reasonable to expect that spectral data could not only help improve DSM and agricultural productivity but also be applied to assess soil health. Fig. 7j shows areas of bare soil in four Brazilian states (Paraná, São Paulo, Minas Gerais, and Mato Grosso do Sul), where lighter colors represent soils formed from sandstone, and dark purple indicates basalt. Fig. 7k shows a map of soil potential productivity (Greschuk et al., 2023), where green features represent higher potential productivity, while red areas represent lower potential productivity. These latter areas are characterized by high clay, iron, and carbon content. Notably, areas of high productivity correlate with regions shown in dark blue in the SySi images.

Fig. 7l and m shows that SySi images reveal sandy areas (patches) with poor drainage within a generally clayey site, which agrees with Mello et al. (2023). In addition, significant relationships between the bare color composition images and soil types could be observed worldwide. Supplemental notes 1 and 2 detail the performance of spectral clustering for Asia and America in Figs. S1 and S2. We observe in many cases that areas indicated as sandy soil were clayey. This behavior occurs because straw plays a critical goal and sometimes cannot be depicted by the algorithm (Demattê et al., 2018).

4. Challenges and limitations of the approach

Despite the promising results, there is still room for improvement in the bare soil spectra. We observed inconsistencies in areas with high straw covering the soil surface. We also observed the importance of laboratory spectral libraries to train the model and validate the data from a satellite. The more robust and representative SSL the better the model at the satellite level. Thus, obtaining these inconsistencies in the final product is plausible, while human experts must be involved in the interpretation of the spectra. We strongly encourage efforts to improve algorithms on the detection of bare soil, for instance, identifying soils free of vegetation but remaining wet. On the one hand, upcoming hyperspectral images (i.e., EMIT, PRISMA, EnMAP, CHIME, and SBG) may partially address this issue as they are relatively recent in terms of spatial coverage and are not designed to capture the entire globe (except the Terrestrial Ecosystem and Airborne Observatory with Global Ecosystem Resolution - TANAGER-1). On the other hand, globally available Landsat and Sentinel data, for instance, are valuable because they can image larger areas with bare soil, particularly as no-till agricultural practices become more common. The new hyperspectral data will gain in quality but lose coverage over larger spatial areas. Therefore, we should focus on both spatial resolution and spectral range in the future.

Working with RS and open (public) data poses several issues and limitations. Traditionally, soil carbon analysis has been performed by Walkley and Black's method, despite the combustion method being considered the most accurate (Drover and Manner, 1975). Although, currently an extensive dataset is unavailable. Thus, we have to work with what we have until soil carbon monitoring adopts state-of-the-art soil analytics and new-generation global databases (Novais et al., 2025). Another point is that wet soil laboratories can have differences in their analysis, which will also impact the sensor analysis (Paiva et al., 2022). Importantly, the spectral reflectance is dependent on the soil analysis. Open data, even when fused with private data, comes from different soil laboratories which add uncertainties. The timing of the soil analysis also matters, mostly for carbon. A carbon analysis from 1980 in a given site, can be different in the same site 20 years later. Thus, carbon analysis made at different times also interferes with the relationship of spectral modeling.

The RS technique also has issues. In the laboratory, there are fewer uncertainties in measuring the reflectance and harmonization of SSL can be done today using the IEEE P4005 WG strategy (Ben-Dor et al., 2024). Orbital sensor harmonization and calibration is still an issue that needs more attention when compared to laboratory settings (Ben-Dor et al., 2024). The sensors inside the Landsat system, and any other platform, also degrade over time which may impact the spectral result if routine radiometric calibration is not well performed. The spectral resolution is another issue for multispectral sensors since it is not able to detect the spectral features of straw or any other residue in the field. Other issues could be cloud interference, soil moisture variations, and aggregation (Ben-Dor et al., 2009). Many of these issues were overcome by automated systems that can choose the best images and low cloud coverage, evaluate the best pure pixel for soil, and create the median of the spectra to diminish the error.

An overlap between the dates of RS data acquisition and the soil sampling can be achieved in projects that use legacy and continuous sampling when new data is being collected. The use of historical soil data, such as many of the soil sampling points in the WoSIS dataset, typically leaves a significant time gap between the date of sampling and the date of RS data acquisition. This would be expected to introduce additional uncertainty, especially if the work is using bare soil composite images to quantify soil properties such as SOC. This element is known to change under agricultural land management practices. The same is true for other work using remotely sensed data stacks for prediction. For example, Sothe et al. (2022) produced a SOC map of Canada using soil sample data from 1977 to 1995 and RS data from 2000 to 2019. Does this work reflect SOC nowadays? It is more likely a map of SOC from the median age of the data, as we have also stated in our manuscript. Significant losses of SOC in agricultural landscapes over the last 40 years make this temporal gap a notable challenge for future research. Despite this problem, the technique still provides useful insights into SOC, with less uncertainty in some regions and more uncertainty in others. Therefore, SOC content assessments must be approached with caution.

In summary, spatial resolution matters to capture a mixture of soil types in a given field. The spectral resolution limitation may not mimic the absorption bands of many chromophores, and thus, underestimate the real soil condition. The mismatch between soil sample support (i.e., a small soil sample collected over an area of about 20 cm^2, considering the augered borehole) and spatial resolution of the sensor (e.g., Landsat with 30 m spatial resolution) may add ambiguities in the soil-spectral modeling. It is necessary to consider that electromagnetic energy comes from the sun for 149,600,000 km at lightspeed until it reaches the Earth's surface. The EMR crosses the atmosphere, reaches the soil and reflects and goes back to the sensor for >800 km from the target. In this way, it suffers interference from aerosols such as clouds and smog. After

addressing all these challenges and issues, we achieved a Spearman's correlation between soil laboratory and satellite spectra between 0.51 and 0.84, with 80 % of the correlations over 0.7 (Table 1). This outcome demonstrates coherence between bare soil information and field data.

5. Conclusions

We successfully developed Earth's topsoil reflectance spectra at a 30 m spatial resolution. The GEOS3 processing is the system and the SySi the product that effectively covers agricultural and natural areas, providing societal insights. The Landsat time series spanned over 38 years of imaging, and its overlapping data has demonstrated a significant similarity to field truth spectra. Therefore, the system has "uncovered" soil in over 90 % of the world's agricultural areas. Considering that the spectral signature was captured by a sensor 800 km above the ground with a spatial resolution of 30 m, while the soil sample was collected in the field at 0.1 × 0.1 m, this outcome is an impressive advancement in soil science. We also obtained ten different spectral groups regarding the shape and intensity of specific soil properties.

The spectral data, even being limited, showed remarkable variation, reflecting each specific soil type's unique and complex nature. As a rich information source, the spectra can be integrated into pedo-transfer functions to estimate several soil properties, such as sand, silt, clay, cation exchange capacity, mineral composition, and nutrients, which showed a significant correlation with SySi. Elevation and SWIR 1 and 2 were the variables that explained the higher importance in correlation with these attributes. GEOS3 also presented a contiguous map that improves soil synoptic visualization, decreasing errors in mapping. This may pave the way for improved soil monitoring and facilitate an understanding of soil variation and the operationalization of smart agriculture.

The global SySi provides a valuable tool for monitoring and analyzing soil properties across diverse landscapes. These findings can impact disciplines such as DSM, soil surveying, precision agriculture, soil productivity, land use planning, tillage, conservation management, and irrigation. This product can guide the future allocation of soil sampling sites, thereby reducing the costs of human labor, soil analysis, and field campaigns. They can also be helpful for scientists, consultants, farmers, and policymakers and complement climate monitoring efforts. In the future, as higher spectral, spatial and temporal resolution data become available, this paper can serve as a proof of concept to inform future global studies mapping topsoil properties in a novel way. Despite this, we advocate that there is still space to explore create new algoritms for bare soil detection.

CRediT authorship contribution statement

José A.M. Demattê: Writing – review & editing, Writing – original draft, Supervision, Resources, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Rodnei Rizzo:** Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Nícolas Augusto Rosin:** Software, Methodology. **Raul Roberto Poppiel:** Software, Methodology, Formal analysis, Conceptualization. **Jean Jesus Macedo Novais:** Writing – review & editing, Visualization, Validation, Investigation, Formal analysis. **Merilyn Taynara Accorsi Amorim:** Software, Methodology. **Heidy Soledad Rodriguez-Albarracín:** Software, Methodology. **Jorge Tadeu Fim Rosas:** Software, Methodology. **Bruno dos Anjos Bartsch:** Software, Methodology. **Letícia Guadagnin Vogel:** Software, Methodology. **Budiman Minasny:** Writing – review & editing, Supervision, Formal analysis. **Sabine Grunwald:** Writing – review & editing, Data curation. **Yufeng Ge:** Writing – review & editing, Data curation. **Eyal Ben-Dor:** Writing – review & editing, Supervision, Formal analysis. **Asa Gholizadeh:** Writing – review & editing, Data curation. **Cecile Gomez:** Writing – review & editing, Data curation. **Sabine Chabrillat:** Writing – review & editing, Data curation. **Nicolas Francos:** Writing – review & editing, Data curation. **Dian Fiantis:** Writing – review & editing, Data curation. **Abdelaziz Belal:** Writing – review & editing, Data curation. **Nikolaos Tsakiridis:** Writing – review & editing, Data curation. **Eleni Kalopesa:** Writing – review & editing, Data curation. **Salman Naimi:** Writing – review & editing, Data curation. **Shamsollah Ayoubi:** Writing – review & editing, Data curation. **Nikolaos Tziolas:** Writing – review & editing, Data curation. **Bhabani Sankar Das:** Writing – review & editing, Data curation. **George Zalidis:** Writing – review & editing, Data curation. **Marcio Rocha Francelino:** Writing – review & editing, Data curation. **Danilo Cesar de Mello:** Writing – review & editing. **Najmeh Asgari Hafshejani:** Writing – review & editing, Data curation. **Yi Peng:** Writing – review & editing, Data curation. **Yuxin Ma:** Writing – review & editing, Conceptualization. **João Augusto Coblinski:** Writing – review & editing, Data curation. **Alexandre M.J.-C. Wadoux:** Writing – review & editing, Conceptualization. **Igor Savin:** Writing – review & editing, Data curation. **Brendan P. Malone:** Writing – review & editing, Data curation. **Konstantinos Karyotis:** Writing – review & editing, Data curation. **Robert Milewski:** Writing – review & editing, Data curation. **Emmanuelle Vaudour:** Writing – review & editing, Data curation. **Changkun Wang:** Writing – review & editing, Data curation. **Elsayed Said Mohamed Salama:** Writing – review & editing, Data curation. **Keith D. Shepherd:** Writing – review & editing, Investigation, Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

First Author ackowledge the CNPq for a researcher scholarship. The authors would like to express their gratitude to the São Paulo Research Foundation (FAPESP) for the financial support provided through grant number 2021/05129-8. The authors also wish to thank researchers and institutions in other countries for providing data. We thank the GeoCiS research group https://esalqgeocis.wixsite.com/english for technical support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2025.178791.

Data availability

The authors do not have permission to share data.

References

Arrouays, D., McBratney, A., Bouma, J., Libohova, Z., Richer-de-Forges, A.C., Morgan, C. L.S., Roudier, P., Poggio, L., Mulder, V.L., 2020. Impressions of digital soil maps: the good, the not so-good, and making them ever better. Geoderma Reg. 20, e00255. https://doi.org/10.1016/j.geodrs.2020.e0025z.

Batjes, N.H., Ribeiro, E., van Oostrum, A., Leenaars, J., Hengl, T., Mendes de Jesus, J., 2017. WoSIS: providing standardized soil profile data for the world. Earth Syst Sci Data 9, 1–14. https://doi.org/10.5194/essd-9-1-2017.

Batjes, N.H., Ribeiro, E., Van Oostrum, A., 2020. Standardized soil profile data to support global mapping and modelling (WoSIS snapshot 2019). Earth Syst Sci Data 12, 299–320. https://doi.org/10.5194/essd-12-299-2020.

Ben-Dor, E. and Demattê, J.A.M. Remote sensing of soil in the optical domains. (2024). In: Remote Sensing Handbook, Second edition, Ed. Prasad S. Thenkabail, Taylor and Francis Inc Press, pg. 643–703, DOI:https://doi.org/10.1201/9781003442127-21. (November 2024. Chapter 16).

Ben-Dor, E., Chabrillat, S., Demattê, J.A.M., Taylor, G.R., Hill, J., Whiting, M.L., Sommer, S., 2009. Using imaging spectroscopy to study soil properties. Remote Sens. Environ. 113, S38–S55. https://doi.org/10.1016/j.rse.2008.09.019.

Ben-Dor, E., Granot, A., Wallach, R., Francos, N., Heller Pearlstein, D., Efrati, B., Borůvka, L., Gholizadeh, A., Schmid, T., 2023. Exploitation of the SoilPRO® (SP) apparatus to measure soil surface reflectance in the field: five case studies. Geoderma 438, 116636. https://doi.org/10.1016/j.geoderma.2023.116636.

Ben-Dor, E., Karyotis, K., Chabrillat, S., Schmid, T., 2024. Standardization and protocol development for soil spectral reflectance in laboratory and field settings: an overview of IEEE SA P4005 activity. In: IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 268–270. https://doi.org/10.1109/IGARSS53475.2024.10641689.

Buol, S.W., Southard, R.J., Graham, R.C., McDaniel, P.A., 2011. Soil genesis and classification: sixth edition. In: Soil Genesis and Classification: Sixth Edition. https://doi.org/10.1002/9780470960622.

Chen, S., Arrouays, D., Mulder, V.L., Poggio, L., Minasny, B., Roudier, P., Walter, C., 2022. Digital mapping of GlobalSoilMap soil properties at a broad scale: a review. Geoderma 409, 115567. https://doi.org/10.1016/j.geoderma.2021.115567.

Clingensmith, C.M., Grunwald, S., 2022. Predicting soil properties and interpreting Vis-NIR models from across continental United States. Sensors Journal 22 (Article 3187), 1–17. https://doi.org/10.3390/s22093187.

Demattê, J.A.M., Alves, M.R., Terra, F.S., Bosquilia, R.W.D., Fongaro, C.T., Barros, P.P.S., 2016. Is it possible to classify topsoil texture using a sensor located 800 km away from the surface? The Brazilian Journal of Soil Science 40, e0150335. https://doi.org/10.1590/18069657rbcs20150335.

Demattê, J.A.M., Fongaro, C.T., Rizzo, R., Safanelli, J.L., 2018. Geospatial Soil Sensing System (GEOS3): a powerful data mining procedure to retrieve soil spectral reflectance from satellite images. Remote Sens. Environ. 212, 161–175. https://doi.org/10.1016/j.rse.2018.04.047.

Demattê, J.A.M., Safanelli, J.L., Poppiel, R.R., Rizzo, R., Quiñonez Silvero, N.E., Mendes, W. de S., Bonfatti, B.R., Dotto, A.C., Urbina Salazar, D.F., Mello, F.A. de O., Paiva, A.F. da S., Souza, A.B., Santos, N.V. dos, Nascimento, C.M., de Mello, D.C., Bellinaso, H., Neto, L.G., Amorim, M.T.A., Resende, M.E.B. de, Vieira, J. da S., Queiroz, L.G. de, Gallo, B.C., Sayão, V.M., Lisboa, C.J. da S., 2020. Bare Earth's surface spectra as a proxy for soil resource monitoring. Sci. Rep. 10, 4461. https://doi.org/10.1038/s41598-020-61258-w.

Demattê, J.A.M., Paiva, A.F.d.S., Poppiel, R.R., Rosin, N.A., Ruiz, L.F.C., Mello, F.A.d.O., Minasny, B., Grunwald, S., Ge, Y., Ben-Dor, E., Gholizadeh, A., Gomez, C., Chabrillat, S., Francos, N., Ayoubi, S., Fiantis, D., Biney, J.K.M., Wang, C., Belal, A., Silvero, N.E.Q., 2022. The Brazilian Soil Spectral Service (BraSpecS): a user-friendly system for global soil spectra communication. Remote Sens. 14 (3), 740. https://doi.org/10.3390/rs14030740.

Drover, D.P., Manner, H., 1975. A comparison between the walkley-black and a dry combustion method for determining organic carbon in a humic brown soil, Papua New Guinea. Communica- tions in Soil Science and Plant Analysis 6 (5), 495–500. https://doi.org/10.1080/00103627509366586.

Evangelista, S.J., Field, D.J., McBratney, A.B., Minasny, B., Ng, W., Padarian, J., Dobarco, M.R., Wadoux, A.M.J.-C., 2024. Soil security-zstrategizing a sustainable future for soil. Adv. Agron. 183. https://doi.org/10.1016/bs.agron.2023.10.00.

FAO, 1974. Soil Map of the World. Volume I, Legend. Unesco, Paris. https://data.apps.fao.org/map/catalog/srv/eng/catalog.search#/metadata/cc45a270-88fd-11da-a88f-000d939bc5d8.

FAO and ITPS, 2015. Status of the World's Soil Resources (SWSR) – Main Report. Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils, Rome, Italy. http://www.fao.org/3/a-i5199e.pdf (650 p.).

Fongaro, C.T., Demattê, J.A.M., Rizzo, R., Lucas Safanelli, J., Mendes, W.D.S., Dotto, A. C., Vicente, L.E., Franceschini, M.H.D., Ustin, S.L., 2018. Improvement of clay and sand quantification based on a novel approach with a focus on multispectral satellite images. Remote Sens. 10 (10), 1555. https://doi.org/10.3390/rs10101555.

Francos, N., Ben-Dor, E., 2022. A transfer function to predict soil surface reflectance from laboratory soil spectral libraries. Geoderma 405, 115432. https://doi.org/10.1016/j.geoderma.2021.115432.

Gallo, B.C., Demattê, J.A.M., Rizzo, R., Safanelli, J.L., Mendes, W.S., Lepsch, I.F., Sato, M. V., Romero, D.J., Lacerda, M.P.C., 2018. Multi-temporal satellite images on topsoil attribute quantification and the relationship with soil classes and geology. Remote Sens. 10, 1571. https://doi.org/10.3390/rs10101571.

Gomez, C., Vaudour, E., Féret, J.-B., de Boissieu, F., Dharumarajan, S., 2022. Topsoil clay content mapping in croplands from Sentinel-2 data: influence of atmospheric correction methods across a season time series. Geoderma 423, 115959. https://doi.org/10.1016/j.geoderma.2022.115959.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Greschuk, L.T., Demattê, J.A.M., Silvero, N.E.Q., Rosin, N.A., 2023. A soil productivity system reveals most Brazilian agricultural lands are below their maximum potential. Sci. Rep. 13, 14103. https://doi.org/10.1038/s41598-023-39981-y.

Grunwald, S., 2021. Grand challenges in pedometrics-AI research. Frontiers in Soil Science-Pedometrics 1, 714323. https://doi.org/10.3389/fsoil.2021.714323 (1–8).

Hartemink, A.E., Krasilnikov, P., Bockheim, J.G., 2013. Soil maps of the world. Geoderma 207–208, 256–267. https://doi.org/10.1016/j.geoderma.2013.05.003.

Hartmann, J., Moosdorf, N., 2012. The new global lithological map database GLiM: a representation of rock properties at the Earth surface. Geochem. Geophys. Geosyst. 13, 12004. https://doi.org/10.1029/2012GC004370.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., 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 (2), e0169748. https://doi.org/10.1371/journal.pone.0169748.

Howari, F.M., Goodell, P.C., Miyamoto, S., 2002. Spectral properties of salt crusts formed on saline soils. J. Environ. Qual. 31, 1453–1461. https://doi.org/10.2134/jeq2002.1453.

Huete, A.R., Escadafal, R., 1991. Assessment of biophysical soil properties through spectral decomposition techniques. Remote Sens. Environ. 35, 149–159. https://doi.org/10.1016/0034-4257(91)90008-T.

Jónsson, J.Ö.G., Davíðsdóttir, B., 2016. Classification and valuation of soil ecosystem services. Agric. Syst. 145, 24–38. https://doi.org/10.1016/j.agsy.2016.02.010.

Kaufman, L., Rousseeuw, P.J., 1990. Finding Groups in Data. https://doi.org/10.1002/9780470316801.

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.

Knox, N.M., Grunwald, S., McDowell, M.L., Bruland, G.L., Myers, D.B., Harris, W.G., 2015. Modelling soil carbon fractions with VNIR and MIR spectroscopy. Geoderma 239-240, 229–239. https://doi.org/10.1016/j.geoderma.2014.10.019.

Kohonen, T., 2001. Self-organizing maps, 30, 502. https://doi.org/10.1007/978-3-642-56927-2.

Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the Boruta package. J. Stat. Softw. 36 (11), 1–13. https://doi.org/10.18637/jss.v036.i11.

Digital soil mapping: an introductory perspective. In: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), 2006. Developments in Soil Science, 1st ed. Elsevier, Amsterdam; Boston, ISBN 978-0-444-52958-9.

Lal, R., 2014. Soil conservation and ecosystem services. International Soil and Water Conservation Research 2 (3), 36–47. https://doi.org/10.1016/S2095-6339(15)30021-6.

Lele, S.R., Dennis, B., Lutscher, F., 2007. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. Ecol. Lett. 10, 551–563. https://doi.org/10.1111/j.1461-0248.2007.01047.x.

Levin, N., Ben-Dor, E., Karnieli, A., 2004. Topographic information of sand dunes as extracted from shading effects using Landsat images. Remote Sens. Environ. 90, 190–209. https://doi.org/10.1016/j.rse.2003.12.008.

Liu, F., Rossiter, D.G., Zhang, G.L., Li, D.C., 2020. A soil colour map of China. Geoderma 379, 114556.

Ma, Y., Minasny, B., Malone, B.P., Mcbratney, A.B., 2019. Pedology and digital soil mapping (DSM). Eur. J. Soil Sci. 70 (2), 216–235. https://doi.org/10.1111/ejss.12790.

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.

McBratney, A.B., Minasny, B., Stockmann, U. (Eds.), 2018. Pedometrics, 1st ed. vol. XIII. Springer, p. 720. https://doi.org/10.1007/978-3-319-63439-5.

Mello, F.A.O., Demattê, J.A.M., Bellinaso, H., et al., 2023. Remote sensing imagery detects hydromorphic soils hidden under agriculture system. Sci. Rep. 13, 10897. https://doi.org/10.1038/s41598-023-36219-9.

Naimi, S., Ayoubi, S., Zeraatpisheh, M., Dematte, J.A.M., 2021. Ground observations and environmental covariates integration for mapping of soil salinity: a machine learning-based approach. Remote Sens. 13, 4825. https://doi.org/10.3390/rs13234825.

Negasa, T., Ketema, H., Legesse, A., Sisay, M., Temesgen, H., 2017. Variation in soil properties under different land use types managed by smallholder farmers along the toposequence in southern Ethiopia. Geoderma 290, 40–50. https://doi.org/10.1016/j.geoderma.2016.11.021.

Ng, W., Minasny, B., McBratney, A., de Caritat, P., Wilford, J., 2023. Digital soil mapping of lithium in Australia. Earth System Science Data Discussions 2023, 1–25. https://doi.org/10.5194/essd-15-2465-2023.

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Ben-Dor, E., Brown, D.J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J.A.M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Wetterlind, J., 2015. Soil spectroscopy: an alternative to wet chemistry for soil monitoring. Adv. Agron. 132, 139–159. https://doi.org/10.1016/bs.agron.2015.02.002.

Novais, J.J.M., Melo, B.M.D., Junior, A.F.N., Lima, R.H.C., Souza, R.E., Melo, V.F., Amaral, E.F., Tziolas, N., Demattê, J.A.M., 2025. Online analysis of Amazon's soils through reflectance spectroscopy and cloud computing can support policies and the sustainable development. J. Environ. Manage. 375, 124155. https://doi.org/10.1016/j.jenvman.2025.124155.

Ortiz-Bobea, A., Ault, T.R., Carrillo, C.M., Chambers, R.G., Lobell, D.B., 2021. Anthropogenic climate change has slowed global agricultural productivity growth. Nat. Clim. Chang. 11 (4), 306–312. https://doi.org/10.1038/s41558-021-01000-1.

Padarian, J., Minasny, B., McBratney, A.B., 2019. Using deep learning for digital soil mapping. Soil 5 (1), 79–89. https://doi.org/10.5194/soil-5-79-2019.

Padarian, J., Stockmann, U., Minasny, B., McBratney, A.B., 2022. Monitoring changes in global soil organic carbon stocks from space. Remote Sens. Environ. 281, 113260. https://doi.org/10.1016/j.rse.2022.113260.

Paiva, A.F.S., Poppiel, R.R., Rosin, N.A., Greschuk, L.T., Rosas, J.T.F., Demattê, J.A.M., 2022. The Brazilian Program of soil analysis via spectroscopy (ProBASE): combining spectroscopy and wet laboratories to understand new technologies. Geoderma 421 (1), 115905. https://doi.org/10.1016/j.geoderma.2022.115905.

Poggio, L., De Sousa, L.M., Batjes, N.H., Heuvelink, G.B., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil 7 (1), 217–240. https://doi.org/10.5194/soil-7-217-2021.

Poppiel, R.R., Lacerda, M.P.C., Rizzo, R., Safanelli, J.L., Bonfatti, B.R., Silvero, N.E.Q., Demattê, J.A.M., 2020. Soil color and mineralogy mapping using proximal and remote sensing in Midwest Brazil. Remote Sens. 12 (7), 1197. https://doi.org/10.3390/rs12071197.

Richer-de-Forges, A.C., Chen, Q., Baghdadi, N., Chen, S., Gomez, C., Jacquemoud, S., Martelet, G., Mulder, V.L., Urbina-Salazar, D., Vaudour, E., Weiss, M., Wigneron, J.-

P., Arrouays, D., 2023. Remote sensing data for digital soil mapping in French research—a review. Remote Sens. 15, 3070. https://doi.org/10.3390/rs15123070.
Rizzo, R., Wadoux, A.M.J.-C., Demattê, J.A.M., Minasny, B., Barrón, V., Ben-Dor, E., Francos, N., Savin, I., Poppiel, R., Silvero, N.E.Q., Terra, F. da S., Rosin, N.A., Rosas, J.T.F., Greschuk, L.T., Ballester, M.V.R., Rico Gómez, A.M., Belllinaso, H., Safanelli, J.L., Chabrillat, S., Fiorio, P.R., Salama, E.S.M., 2023. Remote sensing of the Earth's soil color in space and time. Remote Sens. Environ. 299, 113845. https://doi.org/10.1016/j.rse.2023.113845.
Rogge, D., Bauer, A., Zeidler, J., Mueller, A., Esch, T., Heiden, U., 2018. Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014). Remote Sens. Environ. 205, 1–17. https://doi.org/10.1016/j.rse.2017.11.004.
Rosin, N.A., Demattê, J.A.M., Poppiel, R.R., Silvero, N.E.Q., Rodriguez-Albarracin, H.S., Rosas, J.T.F., Greschuk, L.T., Bellinaso, H., Minasny, B., Gomez, C., Marques Júnior, J., Fernandes, K., 2023. Mapping Brazilian soil mineralogy using proximal and remote sensing data. Geoderma 432, 116413. https://doi.org/10.1016/j.geoderma.2023.116413.
Roy, D.P., Kovalskyy, V., Zhang, H.K., Vermote, E.F., Yan, L., Kumar, S.S., Egorov, A., 2016. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. Remote Sens. Environ. 185, 57–70. https://doi.org/10.1016/j.rse.2015.12.024.
Safanelli, J.L., Chabrillat, S., Ben-Dor, E., Demattê, J.A.M., 2020. Multispectral models from bare soil composites for mapping topsoil properties over Europe. Remote Sens. 12, 1369. https://doi.org/10.3390/rs12091369.
Safanelli, J.L., et al., 2021. Fine-scale soil mapping with Earth Observation data: a multiple geographic level comparison. Rev. Bras. Ciênc. Solo 45, e0210080. https://doi.org/10.36783/18069657rbcs20210080.
Sayre, R., Karagulle, D., Frye, C., Boucher, T., Wolff, N.H., Breyer, S., Wright, D., Martin, M., Butler, K., Van Graafeiland, K., Touval, J., Sotomayor, L., McGowan, J., Game, E.T., Possingham, H., 2020. An assessment of the representation of ecosystems in global protected areas using new maps of World Climate Regions and World Ecosystems. Global Ecology and Conservation 21, e00860. https://doi.org/10.1016/j.gecco.2019.e00860.
Sothe, C., Gonsamo, A., Arabian, J., Kurz, W.A., Finkelstein, S.A., Snider, J., 2022. Large soil carbon storage in terrestrial ecosystems of Canada. Glob. Biogeochem. Cycles 36, e2021GB007213. https://doi.org/10.1029/2021GB007213.
Stenberg, B., Viscarra-Rossel, R.A., 2010. Diffuse reflectance spectroscopy for high-resolution soil sensing. Proximal Soil Sensing 29–47. https://doi.org/10.1007/978-90-481-8859-8_3.
Stoner, E.R., Baumgardner, M.F., 1981. Characteristic variations in reflectance of surface soils. Soil Sci. Soc. Am. J. 45, 1161–1165. https://doi.org/10.2136/sssaj1981.03615995004500060031x.
Tziolas, N., Tsakiridis, N., Ben-Dor, E., Theocharis, J., Zalidis, G., 2020a. Employing a multi-input deep convolutional neural network to derive soil clay content from a synergy of multi-temporal optical and radar imagery data. Remote Sens. 12, 1389. https://doi.org/10.3390/rs12091389.
Tziolas, N., Tsakiridis, N., Ogen, Y., Kalopesa, E., Ben-Dor, E., Theocharis, J., Zalidis, G., 2020b. An integrated methodology using open soil spectral libraries and Earth Observation data for soil organic carbon estimations in support of soil-related SDGs. Remote Sens. Environ. 244, 111793. https://doi.org/10.1016/j.rse.2020.111793.
U.S. Geological Survey, 2020a. Landsat 4-7 Collection 1 (C1) Surface Reflectance (LEDAPS) product guide. URL. https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/atoms/files/LSDS-1370_L4-7_C1-SurfaceReflectance-LEDAPS_ProductGuide-v3.pdf (WWW Document) (accessed September 1, 2021).
U.S. Geological Survey, 2020b. Landsat 8 Collection 1 (C1) Land Surface Reflectance Code (LaSRC) product guide URL https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fspublic/atoms/files/LSDS-1368_L8_C1-LandSurfaceReflectanceCode-LASRC_ProductGuide-v3.pdf (WWW Document) (accessed September 1, 2021).
Urbina-Salazar, D., Vaudour, E., Richer-de-Forges, A.C., Chen, S., Martelet, G., Baghdadi, N., Arrouays, D., 2023. Sentinel-2 and Sentinel-1 bare soil temporal mosaics of 6-year periods for soil organic carbon content mapping in Central France. Remote Sens. 15, 2410. https://doi.org/10.3390/rs15092410.
Vaudour, E., Gholizadeh, A., Castaldi, F., Saberioon, M., Borůvka, L., Urbina-Salazar, D., Fouad, Y., Arrouays, D., Richer-de-Forges, A.C., Biney, J., Wetterlind, J., Van Wesemael, B., 2022. Satellite imagery to map topsoil organic carbon content over cultivated areas: an overview. Remote Sens. 14, 2917. https://doi.org/10.3390/rs14122917.
Viscarra Rossel, R.A., Chen, C., Grundy, M.J., Searle, R., Clifford, D., Campbell, P.H., 2015. The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. Soil Research 53 (8), 845–864. https://doi.org/10.1071/SR14366.
Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B.G., Bartholomeus, H.M., Bayer, A.D., Bernoux, M., Böttcher, K., Brodský, L., Du, C.W., Chappell, A., Fouad, Y., Ji, W., 2016. A global spectral library to characterize the world's soil. Earth Sci. Rev. 155, 198–230. https://doi.org/10.1016/j.earscirev.2016.01.012.
Wadoux, A.M.C., Minasny, B., McBratney, A.B., 2020. Machine learning for digital soil mapping: applications, challenges and suggested solutions. Earth Sci. Rev. 210, 103359.
Wang, F., Yang, S., Wei, Y., Shi, Q., Ding, J., 2021. Characterizing soil salinity at multiple depth using electromagnetic induction and remote sensing data with random forests: a case study in Tarim River Basin of southern Xinjiang. China. Science of The Total Environment 754, 142030. https://doi.org/10.1016/j.scitotenv.2020.142030.
Wesemael, B.V., Abdelbaki, A., Ben-Dor, E., Chabrillat, S., d'Angelo, P., Demattê, J.A.M., Genova, G., Gholizadeh, A., Heiden, U., Karlshoefer, P., Milewski, R., Poggio, L., Sabetizade, M., Sanz, A., Schwind, P., Tsakiridis, N., Tziolas, N., Yagüe, J., Žížala, D., 2024. Geoderma 452, 117113. https://doi.org/10.1016/j.geoderma.2024.117113.
Zappia, L., Oshlack, A., 2018. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. Gigascience 7, 1–9. https://doi.org/10.1093/gigascience/giy083.