Food and Agriculture Organization of the United Nations

PHASE 1

Country Guidelines and Technical Specifications for

Global Soil Nutrient and Nutrient Budget Maps

GSNmap

GLOBAL SOIL PARTNERSHIP

Country guidelines and technical specifications for Global Soil Nutrient and Nutrient Budget Maps GSNmap – Phase I

Food and Agriculture Organization of the United Nations

Rome, 2022

Required citation:
FAO. 2022. *Country guidelines and technical specifications for global soil nutrient and nutrient budget maps – GSNmap: Phase 1*. Rome. https://doi.org/10.4060/cc1717en

ISBN 978-92-5-136795-7

Contents

Figures

Tables

Editors

Marcos Angelini (GSP Secretariat)

Isabel Luotto (GSP Secretariat)

Moritz Mainka (GSP Secretariat)

Christian Omuto (GSP Secretariat)

Yusuf Yigini (GSP Secretariat)

Ronald Vargas (GSP Secretariat)

Contributors and reviewers

INSII – International Network of Soil Information Institutions

ITPS – Intergovernmental Technical Panel on Soils

GSNmap WG – GSNmap Working Group

Acknowledgements

We would like to express our gratitude to Professor Gerard B. M. Heuvelink (Wageningen University, Netherlands) for his insightful discussion of the DSM approaches, as well as to Dr. Fernando García (Facultad de Ciencias Agrarias Balcarce, Universidad Nacional de Mar del Plata, Argentina) for his enthusiastic contribution on soil nutrients and laboratory methods.

Abbreviations and acronyms

CRS – coordinate reference system

DSM – digital soil mapping

GloSIS – Global Soil Information System

GLOSOLAN – Global Soil Laboratory Network

GSOCmap – Global Soil Organic Carbon Map

GSNmap – Global Soil Nutrient and Nutrient Budget Maps

GSP – Global Soil Partnership

INSII – International Network of Soil Information Institutions

ITPS – Intergovernmental Technical Panel on Soils

MAE – mean absolute prediction error

ME – mean prediction error

MEC – model efficiency coefficient

QA/QC – quality assurance/quality control

QRF – quantile regression forest

RMSE – root mean square error

SDG – sustainable development goal

SID – GSP Soil Information and Data team

SOC – soil organic carbon

SSM – sustainable soil management

Summary

This document provides guidance and technical specifications for the first phase of the GSNmap initiative which aims to generate national maps of soil nutrients and associated soil properties at 250 m resolution for agricultural lands based on a country–driven approach. On the one hand, soil nutrient maps will provide a baseline for identifying areas where their levels are critical for crop growth and will thus serve as an important decision–making tool. On the other hand, associated soil parameters such as organic carbon, pH, soil texture, bulk density, and cation exchange capacity will be mapped, which can highlight the key limits to nutrient availability. In order to obtain consistent results and to allow comparisons between countries and regions, we propose a standard methodology based on digital soil mapping techniques. General modelling procedures, data requirements and data sources are described. The final product specifications and data submission formats are also provided. This approach will require collaboration at national level between experts of GLOSOLAN and INSII. GSP will organise training sessions to support countries that require technical assistance to produce their own maps, and will facilitate the production of datasets for countries lacking the required local input data. The final product will be relevant to identify the level of nutrients and associated soil properties per regions, environments and agricultural systems, and to establish priorities for the implementation of global and national public and private policies.

Table 1: Product specifications overview of the GSNmap

Mandatory products	• Total Nitrogen 0-30 cm depth; • available Phosphorus 0-30 cm depth; • available Potassium 0-30 cm depth; • cation exchange capacity map 0-30cm depth; • soil pH map 0-30cm depth; • clay (<2 µm), silt (2-20/50 µm) and sand (50-2000 µm) fractions map, 0-30cm depth; • concentration of soil organic carbon map, 0-30cm depth ; and • bulk density map, 0-30cm depth.
Optional products	• Extractable micronutrients; • cation exchange capacity map 30-60 and 60-100 cm; • soil pH map 30-60 and 60-100 cm; • clay (<2 µm), silt (2-20/50 µm) and sand (50-2000 µm) fractions map, 30-60 and 60-100 cm; • concentration of soil organic carbon map, 30-60 and 60-100 cm; and • bulk density map 30-60 and 60-100 cm.
Depth	0-30 cm
Resolution	National level raster maps (spatial resolution of 250 m or 7.5 arc-second)
Extent	Croplands / ESA. Land Cover CCI
Projection	WGS 84 (decimal degrees geographic)
Uncertainty	Standard deviation map (raster format at 250 m or 7.5 arc-second)
Documentation	Technical report
Delivery	Online (GSP data submission tool)

Table 2: Input data requirements for the GSNmap

Input data requirements	Phase I	
Covariates	GSP google earth engine covariate repository (instructions will be provided in the GSNmap Technical Manual)	
Mandatory Soil data	Total Nitrogen (ppm) Available Phosphorus (ppm) Available Potassium (ppm) Cation exchange capacity ($cmol_c/kg$) pH Soil fractions (clay, silt and sand in g/100g) SOC (%) Bulk Density (g/cm^3)	0-30 cm
Optional soil data	Total Nitrogen (ppm) Available Phosphorus (ppm) Available Potassium (ppm) Cation exchange capacity ($cmol_c/kg$) pH Soil fractions (clay, silt and sand in g/100g) SOC (%) Bulk Density (g/cm^3)	30-60 & 60-100 cm

1 Introduction

1.1 Background and objectives

To date, a total number of around 2.3 billion people are affected by moderate and severe food insecurity (FAO *et al.*, 2022). In 2020, within the first year of the COVID-19 pandemic, an additional 320 million people became affected by food insecurity (FAO *et al.*, 2021). The current conflicts and aggravating climate change further jeopardise achieving sustainable development goal (SDG) 2 (Zero Hunger) by 2030. The situation is alarming and urgent action is needed to revert the trends and increase food security.

The current global situation requires an increase of food production while preserving natural (soil) resources, lowering greenhouse gas emissions and optimising the use of goods such as fertilisers on agricultural sites (Eisenstein, 2020). Fertiliser prices more than doubled within one year and grain prices increased by around 25 percent (Jan. 2021 - Jan. 2022) (Hebebrand and Laborde, 2022). With the start of the armed conflict in Ukraine in February 2022, this trend became more pronounced.

Growing food insecurity and rapidly increasing fertiliser prices underscore the urgent need for informed decision-making and optimised soil nutrient management. However, a large data gap exists in regards to soil nutrient stocks and soil properties that govern nutrient availability. Therefore, FAO's Global Soil Partnership (GSP) has launched the Global Soil Nutrient and Nutrient Budget map (GSNmap) initiative in an endeavour to provide harmonised and finely resolved soil nutrient data and information to stakeholders following a country-driven approach.

Up-to-date soil data on the status and spatial trends of soil nutrients and related soil attributes is key to guide policy-making to close yield gaps, and protect local natural resources. Therefore, locally-specific optimisation of soil nutrient and agricultural management are needed (Cunningham *et al.*, 2013). The soil information collected in the GSNmap thereby serves as a cornerstone in delineating priority areas for action and thereby seizes the opportunity to reduce food insecurity, close yield gaps, and reduce environmental costs arising from mismanagement of soil nutrients and especially overfertilisation.

1.2 Global Soil Partnership

The Global Soil Partnership (GSP) was established in December 2012 as a mechanism to develop a strong interactive partnership and to enhance collaboration and generate synergies between all stakeholders to raise awareness and protect the world's soil resources. From land users to policymakers, one of the main objectives of GSP is to improve governance and promote sustainable management of soils. Since its creation, GSP has become an important partnership platform where global soil issues are discussed and addressed by multiple stakeholders at different levels.

The mandate of GSP is to improve governance of the planet's limited soil resources in order to guarantee productive agricultural soils for a food-secure world. In addition, it supports other

essential soil ecosystem services in accordance with the sovereign right of each Member State over its natural resources. In order to achieve its mandate, GSP addresses six thematic action areas to be implemented in collaboration with its regional soil partnerships (Figure 1).

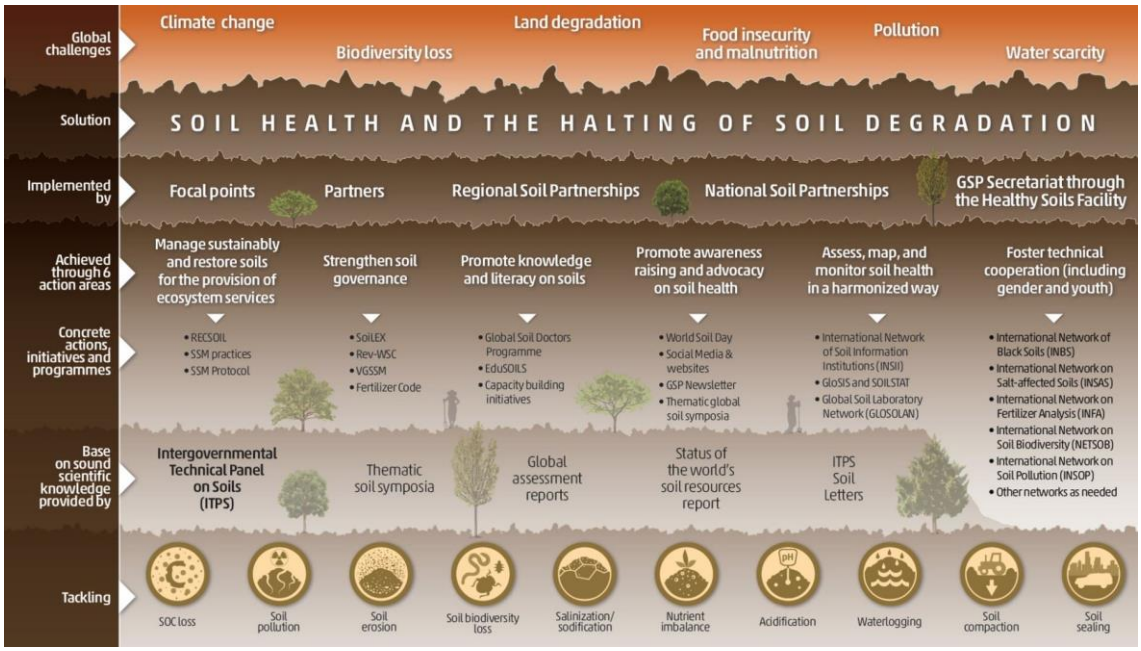

Figure 1: GSP action framework

The area of work on Soil Information and Data (SID) of the GSP builds an enduring and authoritative global system (GloSIS) to monitor and forecast the condition of the Earth's soil resources and produce map products at the global level. The secretariat is working with the international network of soil data providers (INSII - International Network of Soil Information Institutions) to implement data related activities.

1.3 Country-driven approach and tasks

The GSNmap initiative will be jointly implemented by the International Network of Soil Information Institutions (INSII) and the GSP Secretariat. The process will be country-driven, involving and supporting all Member States in developing their national GSNmap data products. The GSNmap products will be developed following a two phase approach:

- Phase I: development of soil nutrient and associated soil property maps.
- Phase II: quantification, analysis, projections of nutrient budgets for agricultural land use systems at national, regional and global scale.

These guidelines only concern GSNmap Phase I, while the guidelines for the GSNmap Phase II will be published in the fourth quarter of 2022.
Depending on national data availability and technical capacities, ad-hoc solutions will be developed by the GSNmap WG to support countries during the national GSNmap production and/or harmonisation phase. Where possible, GSP Secretariat will use publicly available data to gap-fill the areas which are not covered by the national submissions unless the country requests to be left blank on the GSNmap products.

1.4 Cooperation with the Global Soil Laboratory Network (GLOSOLAN)

GSNmap data products demand recent, accurate and dense ground data measurements of relevant soil nutrients and associated soil properties. Considering the importance of ground data also being held by non-INSII institutions, especially by the laboratories within countries, INSII members are encouraged to liaise with the national laboratories which are members of the GSP's Global Soil Laboratory Network (GLOSOLAN) to access the required data inputs (refer to Section 3). GSP Secretariat will share the list of GLOSOLAN laboratories with INSII members and facilitate the process.

2 Digital soil mapping of soil nutrients and associated soil attributes

Digital soil mapping (DSM) is a methodological framework to create soil attribute maps on the basis of quantitative relationships between spatial soil databases and environmental covariates. The quantitative relations can be modelled by different statistical approaches, most of them considered machine learning techniques. Environmental covariates are spatially explicit proxies of soil-forming factors that are employed as predictors of the geographical distribution of soil properties. The methodology has evolved from the theories of soil genesis developed by Vasil Dokuchaev in his work "*The Russian Chernozems*" (1883), which later were formalised by Jenny (1941) with the equation of the soil-forming factors. The conceptual equation of soil-forming factors has been updated by McBratney *et al.* (2003) as follows:

$$S = f(s,c,o,r,p,a,n) \quad (1)$$

where S is the soil classes or attributes (to be modelled) as a function of "s" as other soil properties, "c" as climatic properties, "o" as organisms, including land cover and human activity, "r" as terrain attributes, "p" as parent material, "a" as soil age, and "n" as the geographic position.

Digital soil mapping has been used to produce maps of soil nutrients. For instance, Hengl *et al.* (2017) predicted 15 soil nutrients at a 250 m resolution in Africa, using a random forest model (Breiman, 2001), topsoil nutrient observations at point locations and a set of spatially-explicit environmental covariates. In 2021, Hengl *et al.* applied the same modelling approach to estimate total phosphorus in semi-natural soils at the global scale.

In this document, we present three DSM frameworks to map soil nutrients and associated soil properties. One approach for soil observations with latitude and longitude data (point-support) (Figure 3), another approach for soil observations with administrative unit information as the only geographical reference (area-support) (Figure 4), and a blend of both methods when both area- and point-support are available (Figure 5).

2.1 Soil data

In the first phase of this initiative, INSII members will produce maps of the following soil attributes (see Table 3).

Table 3: Overview of the laboratory methods that are used to measure the soil properties mapped in the GSNmap

Soil attribute	Unit	Laboratory method
Total Nitrogen	ppm	Dumas dry combustion method (FAO, 2021a) or Kjeldahl method (FAO, 2021b)
Available Phosphorus	ppm	Bray I and II, Mehlich I, Olsen (FAO, 2021c; FAO, 2021d; FAO, 2021e)
Available Potassium	ppm	Mehlich III (Mehlich, 1984)
Cation exchange capacity	$cmol_c/kg$	Ammonium acetate (Schollenberger and Simon, 1945)
pH	-	Soil pH in H_2O, KCl, $CaCO_2$ (FAO, 2021f)
Soil fractions (clay, silt, sand)	g/100g	Hydrometer (e.g. Bouyoucos, 1962)
SOC	%	Dumas dry combustion, Walkley-Black, Tyurin spectrophotometric (FAO,2019a; FAO,2019b; FAO, 2021g)
Bulk density	g/cm^3	Overview of methods provided by Blake (1965)
Nutrients (Ca, S, Mg, Fe, B, Cl, Mn, Zn, Cu, Mo, Ni, Si)	ppm	DTPA extraction method (FAO, 2022), Mehlich III (Mehlich, 1984), aqua regia extraction (Berrow and Stein, 1983)

Figure 2 shows the alternative methodologies to be followed, depending on whether the soil observations have corresponding coordinates or just a reference to an administrative unit. When data have XY coordinates, the DSM protocol for point support (Section 2.2) must be followed. Instead, if data do not have XY coordinates but information on the administrative unit, Section 2.3 is the alternative method (area–support). When data is mixed, the option DSM for point–support and area–support (Section 2.4) should be followed. In this case, the area–support data will be used for generating an environmental covariate, while the point–support will be used to produce the final map. Finally, if no data is available, gap–filling with publicly available layers will be used.

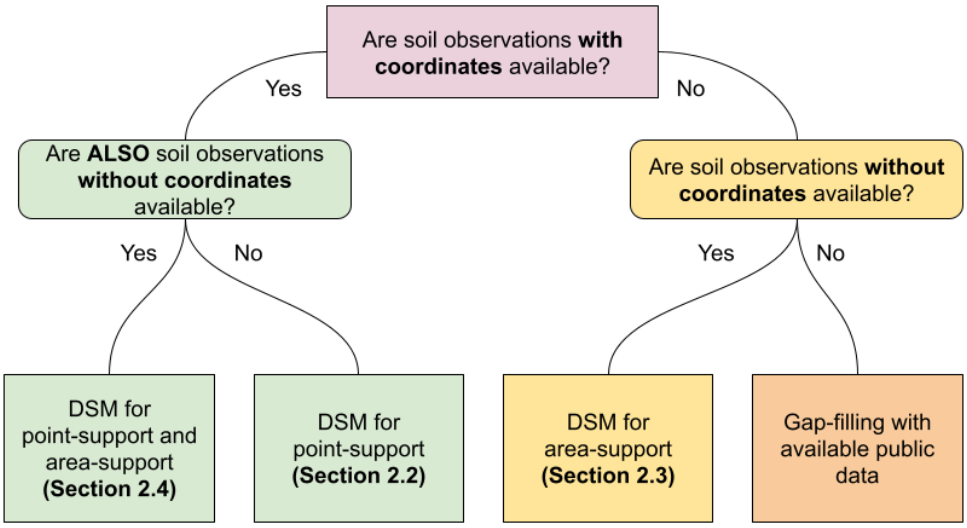

Figure 2: Flowchart for defining the methodology

2.2 DSM approach for point–support data

We present here a DSM framework to map soil nutrients and associated soil properties based on point–support data (with XY coordinates) (Figure 3). The steps in Figure 3 will also be implemented if both point- and area-support data are available.

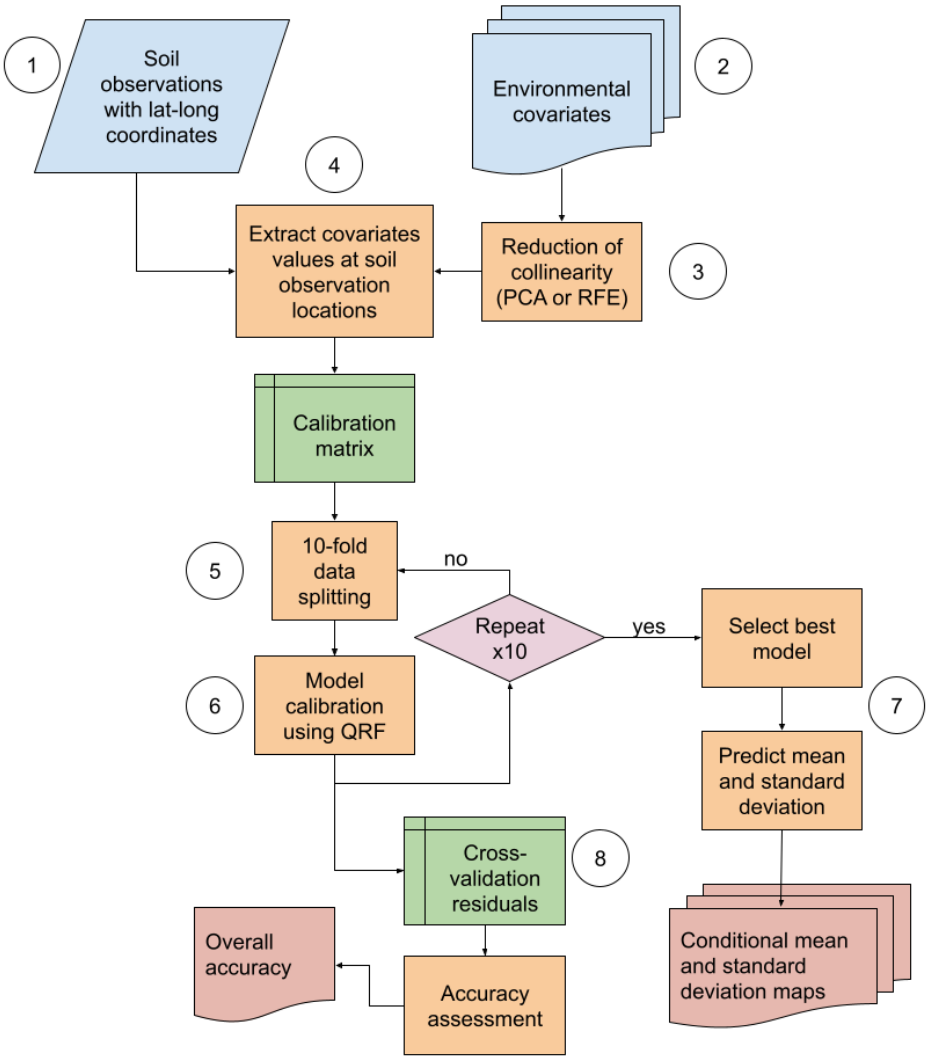

Figure 3: DSM approach for soil observations with latitude and longitude data (point-support). Circles represent steps described in the main text

2.2.1 Step 1: prepare point data

Soil data consist of measurement at a specific geographical location, time and soil depth. Therefore, it is necessary to arrange the data following the format shown in Table 4.

Table 4: Format example of a soil dataset

Profile ID	Horizon ID	Lat	Long	Year	Top	Bottom	Soil property	Value	Lab method
1	1_1	12.123456	1.123456	2018	0	20	SOC	3.4	W&B
1	1_2	12.123456	1.123456	2018	20	40	SOC	2.1	W&B
2	2_1	23.123456	2.123456	2019	0	30	SOC	2.9	W&B

Profile ID = unique profile identifier; Horizon ID = unique layer identifier; Lat = latitude in decimal degrees; Long = longitude in decimal degrees; Year = sampling year; Top = upper limit of the layer in cm; Bottom = lower limit of the layer in cm; Soil property = name of the soil property; Value = numerical value of the measure; Lab method = name of the laboratory protocol used for measuring the soil property.

Soil data usually require a pre-processing step to solve common issues such as, arranging the data format, fixing the consistency of the soil horizon depth, detecting unusual soil property measurements, among other issues.

Once the original dataset is clean and consistent, data harmonisation is needed to produce synthetic horizons (such as 0–30 cm layer), as well as to make compatible measurements from different lab methods. Horizon harmonisation will be done with the mass preserving spline function (Bishop *et al.*, 1999, Malone *et al.*, 2009) fitted to each individual soil profile, which requires more than a layer per profile. In the cases of single–layer samples, which is common in sampling for nutrient determination, a pedotransfer function locally calibrated should be applied. Pedotransfer functions will be also required to harmonise the laboratory methods. Experts from GLOSOLAN will provide advice in this regard.

2.2.2 Step 2: prepare environmental covariates

The *SCORPAN* equation (Eq. 1) refers to the soil–forming factors that determine the spatial variation of soils. However, these factors cannot be measured directly. Instead, proxies of these soil forming factors are used. One essential characteristic of the environmental covariates is that they are spatially explicit, covering the whole study area. Table 5 shows a summary of the environmental covariates that can be implemented under the DSM framework.

Apart from the environmental covariates mentioned in Table 5, other types of maps could also be included, such as Global Surface Water Mapping Layers and Water Soil Erosion from the Joint Research Centre (JRC).

Since environmental covariates will be available at different resolutions and coordinate reference systems (CRS), they have to be harmonised at a common resolution and CRS. The target resolution in GSNmap is 250 m x 250 m, therefore, all covariates will be aggregated (from higher to lower resolution) or disaggregated (from lower to higher resolution) to this resolution. This process involves a raster resampling method, which is usually implemented by a bilinear approach for continuous covariates, and by the nearest–neighbour approach for categorical covariates.

Table 5: Summary of soil–forming factors and their potential proxies (environmental covariates)

Factor	Environmental covariate	Freely available source
Soils	Legacy soil maps of different scales, soil property maps produced with an independent dataset	Global layers https://gitlab.com/openlandmap/global-layers SoilGrids https://soilgrids.org/
Climate	Climatic data such as monthly/yearly/seasonal temporal mean and standard deviation of precipitation, temperatures (min, max, mean, etc.), evapotranspiration, radiation, snow occurrence, aridity index, etc.	Chelsa climate https://chelsa-climate.org/
Organism	Vegetation temporal and spatial patterns are the main proxies of the effect of living organisms. They can be characterised by remote sensing data from optical sensors such as vegetation indices (NDVI, EVI, SAVI), visual bands, NIR, SWIR, TIR bands from, as well as other band ratios. Land cover and land cover change maps are also included in this category.	Landsat mission; MODIS mission; Sentinel 2 mission; ESA global land cover; Dynamic World; https://code.earthengine.google.com/
Relief	Terrain attributes derived from digital elevation models including elevation, slope, terrain curvatures, channel network base level, vertical distance to channel network, terrain wetness index, etc.	Multi–Error–Removed Improved–Terrain DEM (MERIT DEM): http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM/
Parent material and age	Geological maps might be used to derive surface parent material data, including their age, but these are the least available type of data	

Note that the target resolution of GSNmap has been set at 250 m, which can be considered a moderate resolution for a global layer. However, those countries that require a higher resolution are free to develop higher resolution maps and aggregate the resulting maps to the target resolution of GSNmap for submission.

2.2.3 Step 3: reduce collinearity in environmental covariates

Multicollinearity is usually present in remote sensing data and terrain attributes. While this was an issue for multiple linear regression models, current models such as random forest can deal with high dimensionality. However, the main reasons to reduce the number of environmental covariates are that a model with fewer predictors can be interpreted more easily, thus extracting new knowledge, redundant information increasing the computational demand, and improve prediction results (Behrens *et al.*, 2014).

Covariate selection can be done by supervised or unsupervised methods (Behrens el al., 2010). Supervised methods work on the basis of prediction results, hence they are based on a given dataset. For instance, recursive feature elimination (RFE) in caret R package (Kuhn, 2008) provides a tool for selecting covariates according to their predicting contribution. Instead, unsupervised methods are used to reduce the dimensionality of the dataset by removing redundant information without taking into account a particular target variable. Principal component analysis is one of the most widely used for this purpose, however, it does not ensure that specific discriminant features are kept within the main factors (Behrens *et al.*, 2014). Another drawback of this technique is that model interpretation can be reduced when using factors instead of the original covariates.

In this initiative, both methods will be tested for the specific country case. The final decision on what of them will be used will be made by the national experts.

2.2.4 Step 4: merging soil data and environmental covariates

A calibration dataset consists of soil observations and a matrix of predictors, where each row is a soil observation paired with the values of the corresponding covariates for the given spatial location. Some common issues and solution when merging soil observations and covariates are:

- Mismatch of coordinate reference system (CRS): it requires to convert the CRS of point data to the raster or polygon covariate CRS.
- Categorical covariates: some covariates may be categorical, such as land use/cover, legacy soil maps or geological maps. A common problem in this case is that some classes may not be sampled with any soil observation, causing an error when using the layer for prediction, since the model cannot predict over a class that was not part of the model calibration step. Also, because of the cross–validation procedure, it is advised to have, at least, three soil samples per class for the same reason.

2.2.5 Step 5: setting up repeated *k*–fold cross validation

Cross validation is one of the most used methods in DSM for assessing the overall accuracy of the resulting maps (Step 8, Figure 3). Since this is implemented along with the model calibration step, we explain the process at this stage.

Cross validation consists of randomly splitting the input data into a training set and a testing set. However, a unique testing dataset can bias the overall accuracy. Therefore, *k*–fold cross validation randomly splits the data into *k* parts, using 1/*k* part of it for testing and *k*-1/*k* part for training the model. In order to make the final model more robust in terms of parameter

estimations, we include repetitions of this process. The final approach is called repeated *k*–fold cross–validation, where *k* will be equal to ten in this process. A graphical representation of the 10-fold cross validation is shown in Figure 4. Note that green balls represent the samples belonging to the testing set and yellow balls are samples of the training set. Each row is a splitting step of the 10-folds, while each block (repetitions) represent the repetition step.

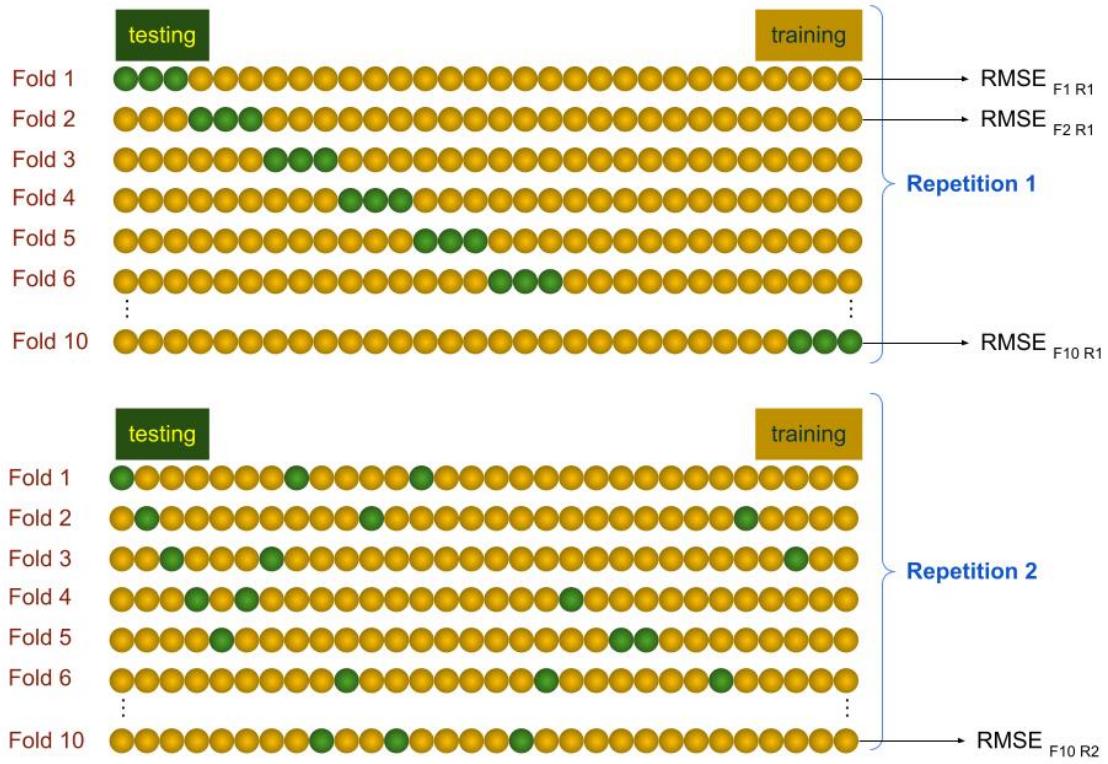

Figure 4: Schematic representation of the repeated cross–validation process

Step 5 in Figure 4 represents the repeated cross–validation, but note that after each single splitting step (the rows in Figure 4) the training data go to model calibration, which will be explained in Step 6 (next section), and the testing data will be used with the calibrated model to produce the residuals (Step 8, Section 2.2.8).

Repeated cross validation has been nicely implemented in the caret R package (Kuhn, 2008), along with several calibration methods.

2.2.6 Step 6: model calibration

The model calibration step involves the use of a statistical model to find the relations between soil observations and environmental covariates. One of the most widely used models in DSM is random forest (Breiman, 2001). Random forest is considered a machine learning method which belongs to the decision–tree type of model. Random forest creates an ensemble of trees using a random selection of covariate. The prediction of a single tree is made based on the observed samples mean in the leaf. The random forest prediction is made by taking the

average of the predictions of the single trees. The size of the number of covariates at each tree (*mtry*) can be fine–tuned before calibrating the model.

Quantile regression forests (QRF; Meinshausen, 2006) are a generalisation of the random forest models, capable of not only predicting the conditional mean, but also the conditional probability density function. This feature allows one to estimate the standard deviation of the prediction, as well as the likelihood of the target variable falling below a given threshold. In a context where a minimum level of a soil nutrient concentration may be decisive for improving the crop yield, this feature can play an important role for the GSNmap initiative.

Model calibration will be implemented using the caret package (Kuhn, 2008). While we suggest to use QRF, caret provides a large set of models (https://topepo.github.io/caret/available- models.html#) that might perform better in specific cases. In this regard, it is up to the user to implement a different model, ensuring the product specifications (Section Product Specifications).

2.2.7 Step 7: predicting soil attributes mean and standard deviation

After calibrating the model, caret will select the best set of parameters and will fit the model using the whole dataset. Then, the final model can be used to predict the target soil properties. The process uses the model and the values of the covariates at target locations. This is generally done by using the same input covariates as a multilayer raster format, ensuring that the names of the layers are the same as the covariates in the calibration dataset. In this step we will predict the conditional mean and conditional standard deviation at each raster cell.

2.2.8 Step 8: overall accuracy assessment

Accuracy assessment is an essential step in digital soil mapping. One aspect of the accuracy assessment has been done in Step 7 by predicting the standard deviation of the prediction, which shows the spatial pattern of the uncertainty. Another aspect of the uncertainty is the estimation of the overall accuracy to measure the model performance. This will be measured using the model residuals generated by caret during the repeated cross validation step.

The residuals produced by caret consist of tabular data with observed and predicted values of the target soil property. They can be used to estimate different accuracy statistics. Wadoux *et al.* (2022) have reviewed and evaluated many of them. While they concluded that there is not a single accuracy statistic that can explain all aspect of map quality, they recommended the following:

- mean prediction error (ME), that estimates the prediction bias;
- mean absolute prediction error (MAE) and root mean squared prediction error (RMSE) to estimate the magnitude of the errors; and
- model efficiency coefficient (MEC) (Janssen and Heuberger, 1995) as an estimator of the proportion of variance explained by the model.

While solar diagrams (Wadoux, *et al.* 2022) are desired, we propose to produce a scatterplot of the observed vs predicted values maintaining the same range and scale for the X and Y axes.

Finally, note that accuracy assessment has been discussed in Wadoux *et al.* (2021), since the spatial distribution of soil samples might constrain the validity of the accuracy statistics. This is especially true in cases where the spatial distribution of observations is clustered. The authors recommended creating a kriging map of residuals before using them for assessing the map quality.

2.3 DSM approach for area–support data

The area–support methodology will be required in the case that soil observations (soil nutrients and/or soil properties) only have an administrative unit as geographical reference. This means that the locations of the samples are uncertain, although constrained to a specific area. The smaller the administrative unit, the smaller the location's uncertainty.

When should this process be applied?

The method is meant to be applied in cases where the density of samples in the administrative unit is around one sample per square kilometre. When this is the only source of data (no point data), it could also be applied with lower density, but note that the uncertainty will increase.

The process in this section differs from the previous one in that we need to assign coordinates to each sample and consecutively follow a similar approach as done in Section 2.2. This process is repeated several times to generate a large number of predictions which is later aggregated by the mean of the predictions.

Figure 5 shows the workflow for soil data with area–support instead of geographical coordinates. The workflow can be divided into the following steps, some of which have been explained in Section 2.2.

2.3.1 Step 1: assign coordinates to soil samples

Table 6 shows an example of the type of data needed for this process. It is required that the name or code of the administrative unit in the table is associated with a polygon map with the same name or code, so the samples can be associated with a specific area in the map.

Table 6: Format example of a soil dataset without latitude and longitude

Profile ID	Horizon ID	District	Year	Top	Bottom	Soil property	Value	Lab method
1	1_1	Gers	2018	0	20	SOC	3.4	W&B
1	1_2	Gers	2018	20	40	SOC	2.1	W&B
2	2_1	Hérault	2019	0	30	SOC	2.9	W&B

Profile ID = unique profile identifier; Horizon ID = unique layer identifier; District = district name or code (administrative unit); Year = sampling year; Top = upper limit of the layer in cm; Bottom = lower limit of the layer

in cm; Soil property = name of the soil property; Value = numerical value of the measure; Lab method = name of the laboratory protocol used for measuring the soil property.

In this step, the latitude will be assigned to each sample in a random manner within the administrative unit to which it belongs. If the samples only belong to a single land use, such as croplands, a cropland mask can be used to reduce the probable area within the administrative unit.

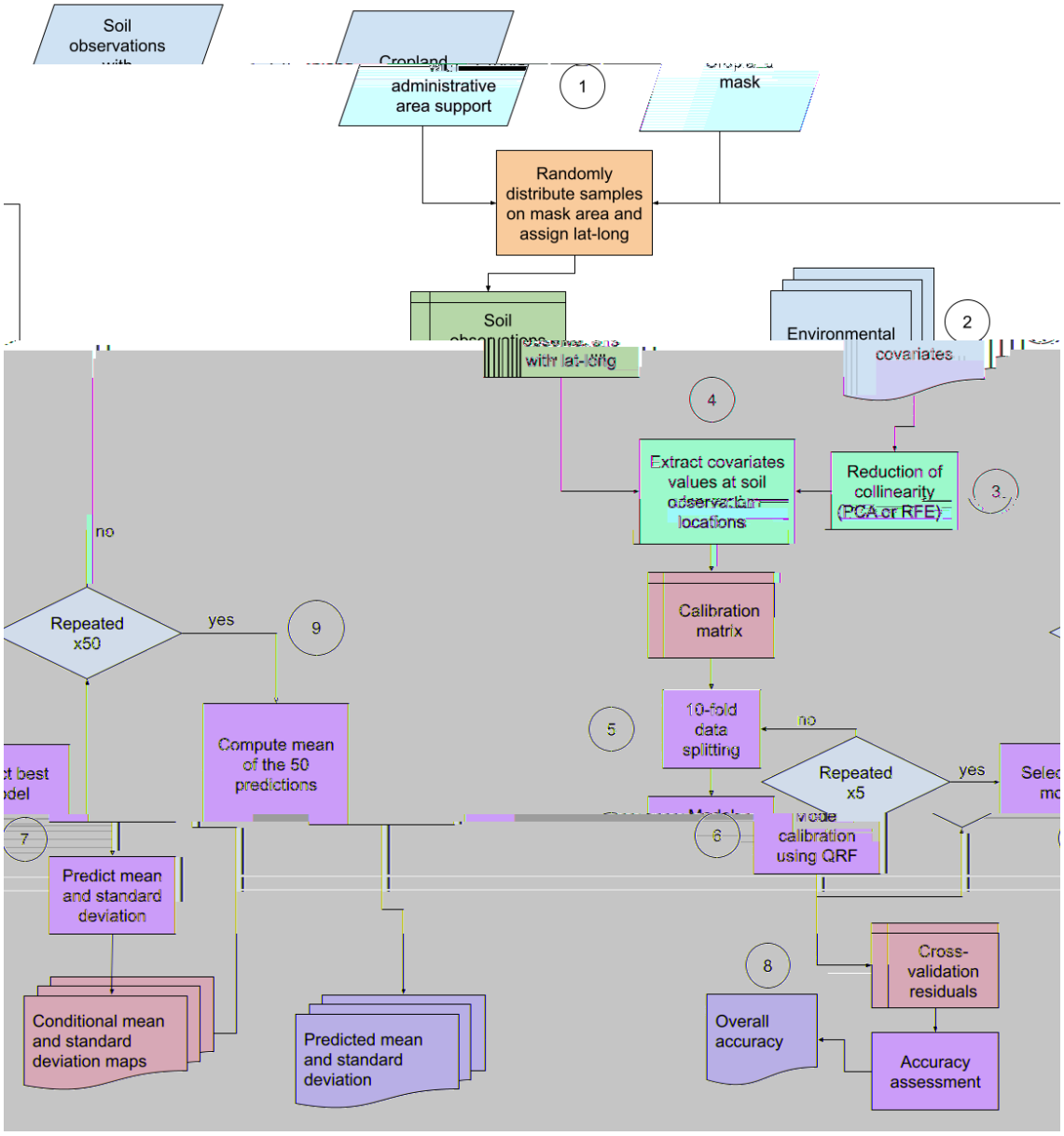

Figure 5: DSM approach for soil data with administrative unit information as the only geographical reference (area–support). Circles represent steps described in the main text

2.3.2 Steps 2 to 8: business as usual

The steps two to eight are the same as Section 2.2.2 to Section 2.2.8, respectively. The only difference is that the repetitions of 10-fold cross–validation (Step 5) can be reduced to half or less to reduce the computational demand of the whole process.

2.3.3 Steps 9: process repetition and final predictions

The steps of Sections 2.3.1 and 2.3.2 will be repeated a large number of times according to the computational capacity. Ideally, it should be repeated more than 50 times. This is because we target to estimate the mean of all predictions to get a stable mean for each pixel of the study area. The final maps will be estimated by the means of the conditional mean and conditional standard deviation.

2.4 DSM approach for point- and area–support data

This approach is a blend of the two previous methods. In essence, the soil data with area–support is used to produce a layer following the methodology of Section 2.3 which is subsequently used as a covariate applying the methodology of Section 2.2 (Figure 6). It is expected that the contribution of the area–support data as a covariate have a great impact in the final map. Then, the method is worth applying in the case that the sample size of the area–support data is considerably larger than the point–support data.

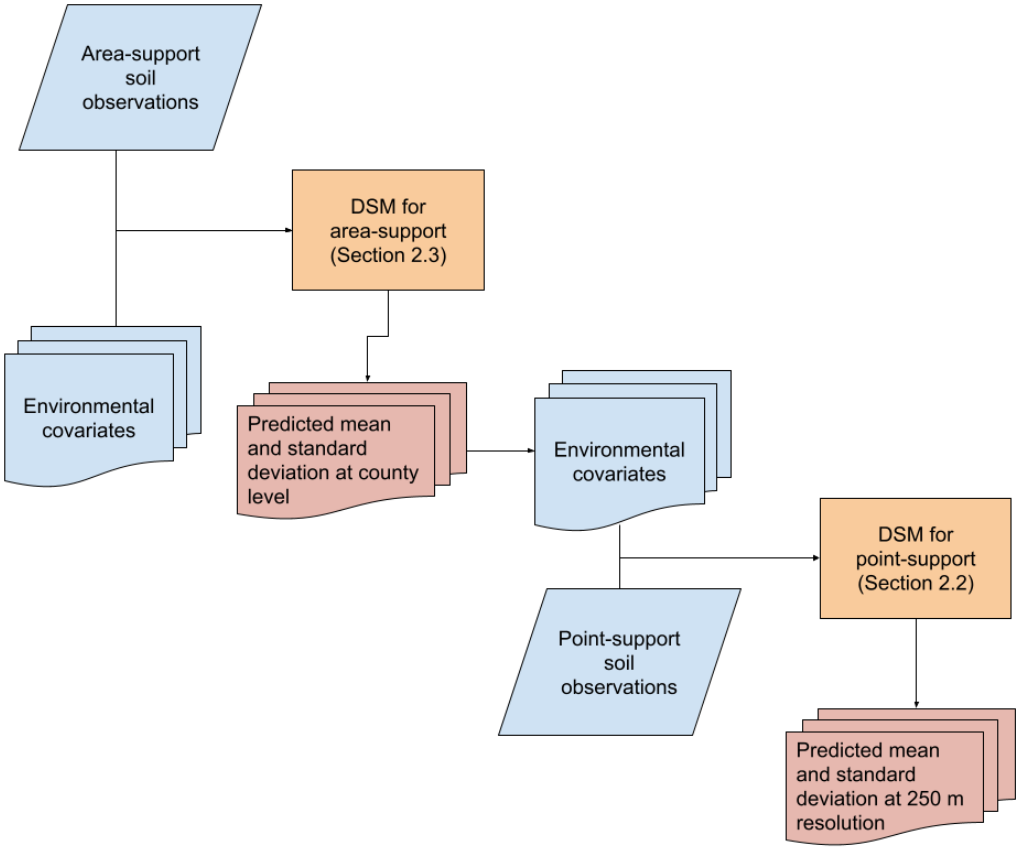

Figure 6: Flowchart for using both area– and point–support for mapping soil nutrients

3 Product specifications

3.1 Mandatory products

The GSNmap consists of two phases and requires the mandatory submission of certain gridded data products. The items must be submitted as raster files (GeoTiff) at a resolution of 250 m for the soil depth of 0–30 cm. The acceptable age of the samples underlying each map is specified in parentheses. In Phase 1, the following products are mandatory:

- total Nitrogen (sampling period: preferably 2017–2022);
- available Phosphorus (sampling period: preferably 2017–2022);
- available Potassium (sampling period: preferably 2017–2022);
- cation exchange capacity (sampling period: 1980–2022);
- soil pH (sampling period: 1980–2022);
- soil texture (sampling period: no constraint);
    - clay (< 2 µm)
    - silt (2–20/50 µm)
    - sand (50–2000 µm)
- soil organic carbon map (sampling period: preferably 2000–2022); and
- bulk density (sampling period: no constraint).

3.2 Optional datasets

Country members are encouraged to deliver the following products and supplementary data in Phase 1:

- cation exchange capacity at 30–60 and 60–100 cm depth (sampling period: 1980–2022);
- soil pH at 30–60 and 60–100 cm depth (sampling period: 1980–2022);
- soil texture (as specified in 4.1) at 30–60 and 60–100 cm depth (sampling period: no constraint);
- bulk density at 30–60 and 60–100 cm depth (sampling period: no constraint); and
- other nutrients at 0–30 cm depth (sampling period: no constraint)
    - Calcium (Ca), Sulphur (S), Magnesium (Mg)
    - Iron (Fe), Boron (B), Chlorine (Cl), Manganese (Mn), Zinc (Zn), Copper (Cu), Molybdenum (Mo), Nickel (Ni).

3.3 Spatial entity

3.3.1 Horizontal and vertical resolution

The mandatory product of both phases of the GSNmap will cover a soil depth of 0–30 cm. Additionally, countries are highly encouraged to provide maps for 30–60 and 60–100 cm depth.

The maps shall be produced at regular fixed horizontal dimensions of 7.5 arc–seconds grid (approximately only 250 x 250 m) at the equator. A generic, empty, global 7.5 arc–second

grid will be prepared and shared with all participating countries. Countries will be expected to deliver their datasets using these standard grids.

3.3.2 Spatial reference

World Geodetic System 1984 (WGS84) geographic (latitude/longitude) projection will be used for all submitted maps. The final GSNmap layers will also be delivered at this coordinate reference system.

3.3.3 Extent

The GSNmap layers will be developed only for croplands.

3.3.4 Excluded areas

Data providers are expected to deliver a continuous surface for their predictions for soils under croplands. Data providers should not attempt to mask out the excluded areas from the grid (e.g. water surfaces, urban areas). The GSP Secretariat will mask excluded areas using standard spatialized layers. Values in the excluded grid cells will be identified as no data (NA) in the final global product.

3.3.5 Uncertainty assessment

The uncertainties will be calculated and salong with the GSNmap layers.

3.3.6 Data submission

File naming conventions and directory structure:

The GSP Secretariat will provide an online data submission facility. Deliverables can be uploaded as individual files or as compressed archives of files (.zip, .rar, 7z). Structure is as follows:

Phase 1

Mandatory maps:

|_ Total Nitrogen map ([ISO3CountryCode]_GSNmap_Ntot_Map030.tiff)

|_ Available Phosphorus map ([ISO3CountryCode]_GSNmap_Pav_Map030.tiff)

|_ Available Potassium map ([ISO3CountryCode]_GSNmap_Ktot_Map030.tiff)

|_Cation Exchange Capacity map ([ISO3CountryCode]_GSNmap_CEC_Map030.tiff)

|_Soil pH map ([ISO3CountryCode]_GSNmap_pH_Map030.tiff)

|_Soil Clay map ([ISO3CountryCode]_GSNmap_Clay_Map030.tiff)

|_Soil Silt map ([ISO3CountryCode]_GSNmap_Silt_Map030.tiff)

|_Soil Sand map ([ISO3CountryCode]_GSNmap_Sand_Map030.tiff)

|_Soil Organic Carbon map ([ISO3CountryCode]_GSNmap_SOC_Map030.tiff)

|_Bulk density map ([ISO3CountryCode]_GSNmap_BD_Map030.tiff)

Uncertainty maps:

|_ All maps except for the soil texture class map come with upper and lower uncertainty maps that are denominated as follows:

[ISO3CountryCode]_GSNmap_[Product name]_UncertaintyMap030.tiff

Documents:

|_ Report ([ISO3CountryCode]_Report. pdf)

|_ FactSheet_([ISO3CountryCode]_Report. pdf)

Optional data products:

In case maps for 30–60 or 60–100 cm depth are submitted, they should be named as following:

[ISO3CountryCode]_GSNmap_[Product name]_Map3060.tiff

[ISO3CountryCode]_GSNmap_[Product name]_Map60100.tiff

File formats:

GIS files shall be delivered in GeoTIFF format. GeoTIFF is a standard .tif or image file format that includes additional spatial (georeferencing) information embedded in the .tif file as tags. These are called embedded tags, tif tags. These tags include raster metadata such as spatial extent, coordinate reference system, resolution, no data values.

4 Quality assurance/quality control

Each country will be responsible for carrying out basic Quality Assurance/Quality Control (QA/QC) of all data before providing it to the GSP Secretariat. Quality Assurance can be described as the process of preventing errors from entering the datasets; while Quality Control can be described as the process of identifying and correcting existing errors in the datasets.

All datasets should be checked for:

- spatial errors (extent, projection);
- units;
- completeness of data;
- consistency with data shown in any accompanying documents (such as reports or drawings);
- compliance with the Data Standards described in this document; and
- consistency of reported validation results with the provided data.

Final QA/QC for the global datasets will be facilitated by the GSP Secretariat through its technical networks (INSII, GSNmap WG, and the Intergovernmental Technical Panel on Soils (ITPS) will give final clearance to the global dataset prior to public release).

5 Data policy

The final global dataset will be distributed under the endorsed GSP Data Policy (http://www.fao.org/3/a-bs975e.pdf). As suggested in the GSP Data Policy, a Creative Commons licence will be assigned to the global dataset. Data providers will retain the ownership of national datasets.

References

Behrens, T., Zhu, A.X., Schmidt, K. & Scholten, T. 2010. Multi–scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma*, 155(3–4): 175–185.

Behrens, T., Schmidt, K., Ramirez–Lopez, L., Gallant, J., Zhu, A.X. & Scholten, T. 2014. Hyper–scale digital soil mapping and soil formation analysis. *Geoderma*, 213: 578–588.

Berrow, M. L. and Stein, W. M. 1983. Extraction of metals from soils and sewage sludges by refluxing with aqua regia. *Analyst*, 108(1283): 277–285.

Bishop, T. F. A., McBratney, A. B. & Laslett, G. M. 1999. Modelling soil attribute depth functions with equal–area quadratic smoothing splines. *Geoderma*, 91(1–2): 27–45.

Blake, G. R. 1965. Bulk density. In: C.A. Black, eds. *Methods of Soil Analysis: Part 1 Physical and Mineralogical Properties, Including Statistics of Measurement and Sampling*, pp. 374–390. Madison, USA, American Society of Agronomy.

Bouyoucos, G. J. 1962. Hydrometer method improved for making particle size analyses of soils. *Agronomy journal*, 54(5): 464–465.

Breiman, L. 2001. Random forests. *Machine Learning*, 45: 5–32.

Cunningham, S. A., Attwood, S. J., Bawa, K. S., Benton, T. G., Broadhurst, L. M., Didham, R. K. McIntyre, S. *et al.* 2013. To close the yield–gap while saving biodiversity will require multiple locally relevant strategies. *Agriculture, Ecosystems & Environment*, 173: 20–27.

Dokuchaev, V. V. 1883. *The Russian chernozem report to the free economic society*. St. Petersburg, Imperial University of St. Petersburg.

Eisenstein, M. 2020. Natural solutions for agricultural productivity. *Nature*, 588: S58–S59.

FAO. 2019a. *Standard operating procedure for soil total carbon – Dumas dry combustion method*. Rome, FAO. Available at: https://www.fao.org/3/ca7781en/ca7781en.pdf

FAO. 2019b. *Standard operating procedure for soil organic carbon – Walkley–Black method. Titration and colorimetric method*. Rome, FAO. Available at: https://www.fao.org/3/ca7471en/ca7471en.pdf

FAO. 2021a. *Standard operating procedure for soil total nitrogen – Dumas dry combustion method*. Rome, FAO. Available at: https://www.fao.org/3/cb3646en/cb3646en.pdf

FAO. 2021b. *Standard operating procedure for soil nitrogen – Kjeldahl method*. Rome, FAO. Available at: https://www.fao.org/3/cb3642en/cb3642en.pdf

FAO. 2021c. *Standard operating procedure for soil available phosphorus, Bray I and Bray II method*. Rome, FAO. Available at: https://www.fao.org/3/cb3460en/cb3460en.pdf

FAO. 2021d. *Standard operating Procedure for soil available phosphorus – Mehlich I method*. Rome, FAO. Available at: https://www.fao.org/3/cb5427en/cb5427en.pdf

FAO. 2021e. *Standard operating procedure for soil available phosphorus – Olsen method*. Rome, FAO. Available at: https://www.fao.org/3/cb3644en/cb3644en.pdf

FAO. 2021f. *Standard operating procedure for soil pH determination*. Rome, FAO. Available at: https://www.fao.org/3/cb3637en/cb3637en.pdf

FAO. 2021g. *Standard operating procedure for soil organic carbon: Tyurin spectrophotometric method*. Rome, FAO. Available at: https://www.fao.org/3/cb4757en/cb4757en.pdf

FAO, IFAD, UNICEF, WFP & WHO. 2021. *The State of Food Security and Nutrition in the World 2021. Transforming food systems for food security, improved nutrition and affordable healthy diets for all*. Rome, FAO. https://doi.org/10.4060/cb4474en

FAO. 2022. *Standard operating procedure for soil available micronutrients (Cu, Fe, Mn, Zn) and heavy metals (Ni, Pb, Cd), DTPA extraction method*. Rome, FAO. Available at: https://www.fao.org/3/cc0048en/cc0048en.pdf

FAO, IFAD, UNICEF, WFP & WHO. 2022. *The State of Food Security and Nutrition in the World 2022. Repurposing food and agricultural policies to make healthy diets more affordable*. Rome, FAO. https://doi.org/10.4060/cc0639en

Hebebrand, C. & Laborde, D. 2022. *High fertiliser prices contribute to rising food security concerns*. URL (last cited on 16 August, 2022): https://www.ifpri.org/blog/high-fertilizer-prices-contribute-rising-global-food-security-concerns

Hengl, T., Leenaars, J. G., Shepherd, K. D., Walsh, M. G., Heuvelink, G., Mamo, T., Tilahun, H. *et al.* 2017. Soil nutrient maps of Sub–Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. *Nutrient Cycling in Agroecosystems*, 109(1): 77–102.

Hengl, T., Miller, M. A., Križan, J., Shepherd, K. D., Sila, A., Kilibarda, M., Antonijević, O. *et al.* 2021. African soil properties and nutrients mapped at 30 m spatial resolution using two–scale ensemble machine learning. *Scientific Reports*, 11(1): 1–18.

Janssen, P. H. M. & Heuberger, P. S. C. 1995. Calibration of process–oriented models. *Ecological Modelling*, 83(1–2): 55–66.

Jenny, H. 1941. *Factors of Soil Formation: A System of Quantitative Pedology.* New York, Dover Publications.

Kuhn, M. 2008. Building predictive models in R using the caret package. *Journal of Statistical Software*, 28: 1–26.

Kuhn, M. 2014. caret: Classification and Regression Training. Comprehensive R Archive Network.

Malone, B. P., McBratney, A. B., Minasny, B. & Laslett, G. M. 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma*, 154(1–2): 138–152.

McBratney, A. B., Santos, M. M. & Minasny, B. 2003. On digital soil mapping. *Geoderma*, 117(1–2): 3–52.

Mehlich, A. 1984. Mehlich 3 soil test extractant: A modification of Mehlich 2 extractant. *Communications in soil science and plant analysis,* 15(12): 1409–1416.

Meinshausen, N. 2006. Quantile regression forests. *Journal of Machine Learning Research*, 7: 983–999.

Schollenberger, C. J. & Simon, R. H. 1945. Determination of exchange capacity and exchangeable bases in soil—ammonium acetate method. *Soil science*, 59(1): 13–24.

UN Department of Economic and Social Affairs, Population Division. 2019. *World Population Prospects 2019: Highlights*. URL: https://population.un.org/wpp/publications/files/wpp2019_highlights.pdf (last access: 12.08.2022)

Wadoux, A.M.C., Heuvelink, G.B., De Bruin, S. & Brus, D.J., 2021. Spatial cross–validation is not the right way to evaluate map accuracy. *Ecological Modelling*, 457: 109692.

Wadoux, A.M.C., Walvoort, D.J. & Brus, D.J. 2022. An integrated approach for the evaluation of quantitative soil maps through Taylor and solar diagrams. *Geoderma*, 405: 115332.

GLOBAL SOIL PARTNERSHIP

The Global Soil Partnership (GSP) is a globally recognized mechanism established in 2012. Our mission is to position soils in the Global Agenda through collective action. Our key objectives are to promote Sustainable Soil Management (SSM) and improve soil governance to guarantee healthy and productive soils, and support the provision of essential ecosystem services towards food security and improved nutrition, climate change adaptation and mitigation, and sustainable development.

Australian Government
Department of Agriculture, Water and the Environment

AFACI
Asian Food & Agriculture Cooperation Initiative

Ministry of Finance of the Russian Federation

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Rural Development Administration