Geoderma 469 (2026) 117821

Contents lists available at ScienceDirect

Geoderma

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

Assessing the soil capacity to produce food and biomass worldwide

Nícolas Augusto Rosin [a,b], Alex B. McBratney [c], Raul Roberto Poppiel [a,d], Jorge Tadeu Fim Rosas [a,e], Merilyn T. Accorsi Amorim [a], Nicolas Francos [c], José A.M. Demattê [a,d,*]

[a] Department of Soil Science, Luiz de Queiroz College of Agriculture (ESALQ), University of São Paulo (USP), Piracicaba, São Paulo 13418-900, Brazil
[b] Brazilian Agricultural Research Corporation (Embrapa) –Soils, Rio de Janeiro, Rio de Janeiro 22460-000, Brazil
[c] Sydney Institute of Agriculture & School of Life & Environmental Sciences, The University of Sydney, Sydney, New South Wales 2006, Australia
[d] Center for Carbon Research in Tropical Agriculture (CCARBON), University of São Paulo, Piracicaba, São Paulo, Brazil
[e] Department of Agronomy, College of Agricultural Sciences, São Paulo Western University (UNOESTE), Raposo Tavares Hwy., Km 572, Presidente Prudente, 19067-175 São Paulo, Brazil

ARTICLE INFO

Handling Editor: B. Heung

Keywords:
Soil security
Food security
Utility functions
Digital soil mapping
Soil health

ABSTRACT

Soil security is a broad concept to address the role of soil in humankind's well-being. We aimed to map the soil capacity to produce food and biomass worldwide with a soil security assessment framework (SSAF) based on the relationship between crop yield and soil attributes (in surface and subsoil layers) across the world's ecoregions. The yield data for sugarcane, maize, rice, wheat, and soybean were transformed into a unitless utility value and used as target indicators, and clay, pH, soil organic carbon, and plant available water were used as potential indicators (PIs). The dataset was stratified by ecoregions. Utility functions were fitted between the target indicator and surface and subsoil data for each PI using generalized additive models (GAMs). GAMs were also fitted using all PIs. The final utility maps were predicted using digital soil mapping. The empirical bivariate utility functions reached an R^2 of 0.01–0.36 and showed different behavior than those expected in the literature for some cases. The behavior of the pH was closest to that expected. The octavariate models using all PIs for the surface and subsoil had better accuracies with an R^2 of 0.18–0.46. The predicted maps were related to the main crop yield for each region and enabled this information to be downscaled to 90 m and extrapolated for current non-agricultural uses. This study is the first approximation of the soil's capacity to produce food and biomass on a global scale, and limitations due to several uncertainties should be considered.

1. Introduction

The soil and its continued functionalities are existential challenges for humankind, with soil security, food security, water security, energy security, climate change mitigation, biodiversity protection, and ecosystem services delivery contributing to at least 12 of the sustainable development goals of the United Nations (Bouma and McBratney, 2013; Kopittke et al., 2022). Food security has four dimensions: availability, access, household utilization, and stability (IPC, 2021). Soil plays an important role in food production, which corresponds to the availability dimension. Soil is a non-renewable resource, and population growth and the high demand for food has increased the pressure on soil resources (Kopittke et al., 2019; Smith et al., 2016). It is necessary to preserve soil resources and intensify its use in a sustainability way to increase and guarantee service continuity for mankind.

Efforts have been made to assess the role of soil in relevant topics for mankind, and many concepts, such as soil quality, soil health, and soil security, have been developed for this purpose (Karlen et al., 2001; Koch et al., 2013; Lehmann et al., 2020). Soil security is a broad and recently developed concept, encompassing soil quality and soil health (Lehmann et al., 2020). Five dimensions (capacity, condition, capital, connectivity, and codification) and three roles are fulfilled by the soil (functions, services, and resilience to threats) (Evangelista et al., 2023; McBratney et al., 2014). Evangelista et al. (2023, 2024) proposed the Soil Security Assessment Framework (SSAF), which uses the concepts of soil security and scoring and aggregation methods, in which soil attributes can be

* Corresponding author at: Department of Soil Science, Luiz de Queiroz College of Agriculture (ESALQ), University of São Paulo (USP), Piracicaba, São Paulo 13418-900, Brazil.

E-mail addresses: nicolas.rosin@embrapa.br (N.A. Rosin), raulpoppiel@usp.br (R.R. Poppiel), jorgerosas@unoeste.br (J.T.F. Rosas), jamdemat@usp.br (J.A.M. Demattê).

https://doi.org/10.1016/j.geoderma.2026.117821
Received 19 December 2025; Received in revised form 13 April 2026; Accepted 14 April 2026
Available online 19 April 2026
0016-7061/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

used to assess soil functions, services, or threats in each soil security dimension through utility functions that relate a dataset of potential indicators (soil attributes) to a target indicator, representing the role(s) and dimension(s) to be assessed. In the present study, we focused on the soil's capacity to fulfill the functions of producing food and biomass.

The capacity dimension refers to the inherent ability of soil to perform a role for humanity and can be represented by soil attributes or properties that only change in pedological timescales (Evangelista et al., 2024; McBratney et al., 2019). The roles of producing food and biomass require the soil as support for agriculture by providing the conditions for plants to grow and supply humanity's demand for food, fiber, and fuel. The soil's capacity can be assessed by the relationship (utility function) between intrinsic soil attributes and a biomass producer variable, such as crop yield (Evangelista et al., 2023; Vogel et al., 2019). The crop yield is a target indicator for food and biomass production, and the literature has shown the relationship between soil attributes and crop yield (Greschuk et al., 2023; Sainju et al., 2022a; Sainju et al., 2022b). The subsoil layer attributes and depth that plant roots explore are less

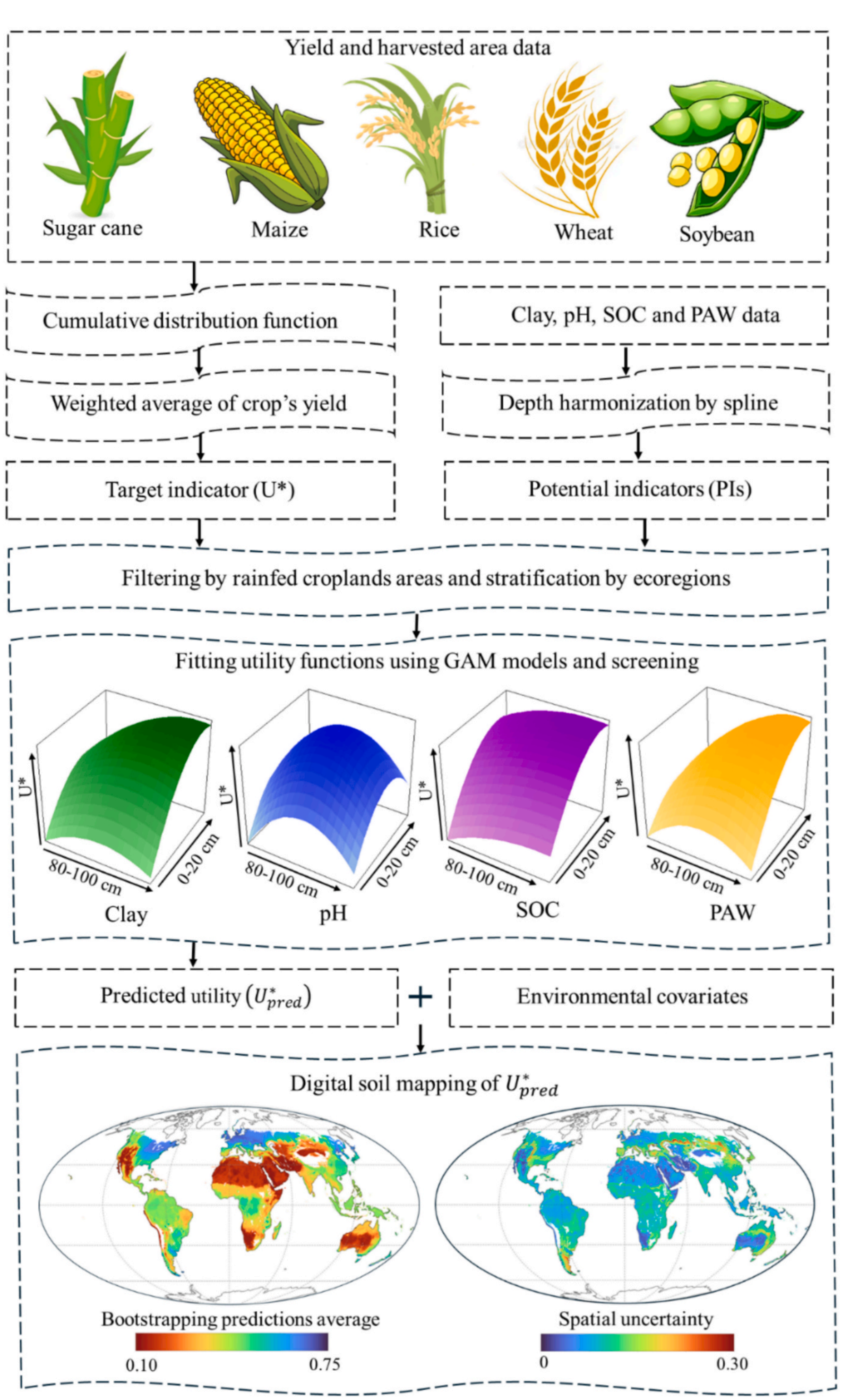

Fig. 1. Flowchart of the methodology. GAM = generalized additive model; SOC = soil organic carbon; AWC = plant available water.

studied than superficial attributes but also influence crop production (Horst-Heinen et al., 2021; de Landell et al., 2003). The subsoil layers are also important for carbon (C) pool and climate change regulation (Dubeux et al., 2024).

Crop yields worldwide are related to several factors such as climate, country income, production system, and trade (Blomqvist et al., 2020; Ray et al., 2015). Clay, pH, plant available water (PAW), and soil organic carbon (SOC) are soil attributes closely related to plant growth (He et al., 2022; Johnston, 1986; Neina, 2019; Newman, 1984) and are available worldwide (Batjes et al., 2020) or can be derived. However, even with the importance of soils in food security (Kopittke et al., 2019), no studies have related it to soil attributes on a global scale. The digital soil mapping (DSM) framework and remote sensing techniques have enabled soil security assessment in large areas (Arrouays et al., 2021), which is now possible due to higher data availability, such as soil (Batjes et al., 2020) and crop yield (Yu et al., 2020) information, and computational advances, such as the emergence of Google Earth Engine, a cloud-based tool for large scale geo-processing and modelling (Gorelick et al., 2017).

In this study, we aimed to map the soil's capacity to produce food and biomass worldwide using the SSAF based on the relationship between crop yield and soil attributes (in surface and subsoil layers) across the world's ecoregions. The five major crop yields (sugarcane, maize, rice, wheat, and soybean) were used as a reference to represent the soil's potential to support agriculture (for any crop) worldwide. We propose an adaptation of the soil security concept from a univariate to a bivariate utility function and from a bidimensional to a tridimensional utility graph, enabling the effect of a given soil attribute on the surface and subsoil to be estimated in the same function. The utility information representing the soil's capacity to provide food and biomass were assessed for bivariate utility functions fitted to each ecoregion and spatialized to the world using the DSM framework.

2. Methods and materials

SSAF (Evangelista et al., 2023) was used to assess the soil's capacity to produce food and biomass worldwide. Yield data from sugarcane, maize, rice, wheat, and soybean were transformed into unitless, aggregates and used as target indicators, and the soil attributes, namely clay, pH, SOC, and PAW, were used as potential indicators (PIs). The dataset was stratified by ecoregion, and utility functions were fitted between the target indicator and surface and subsoil data for each PI. Additionally, models were built using all PIs, with data only from the surface layer and from both layers. The final utility maps were predicted using the DSM framework. The methodology flowchart is given in Fig. 1.

2.1. Soil, yield, and support data obtaining

A georeferenced legacy soil database was compiled with data manly from the World Soil Information Service (WoSIS) (Batjes et al., 2020). This dataset contained information on the clay, silt, sand, SOC, and pH in water for most locations, achieving reasonable coverage of the world's agricultural areas. PAW was calculated by a flux-based method (de Melo et al., 2023), using clay, silt, and sand data. The dataset was harmonized into five 20-cm layers from the surface to 1-m depth by spline interpolation (Bishop et al., 1999; Malone et al., 2009). This study utilized clay, pH, SOC, and PAW data for the 0–20 (representing the surface layer with more anthropogenic influence) and 80–100 cm layers (representing the pedogenic horizon with less anthropogenic influence), totaling 251,903 observation points. More information about the dataset and soil analyses can be found in Batjes et. al. (2020), and the map showing the sample density is given in Supplementary Material 1.

The yield data for the top five consumed crops in the world (sugarcane, maize, rice, wheat, and soybean) were obtained from Yu et al. (2020). These five crops were chosen because they occupy large areas, and together, they are present in most parts of world, providing enough data for a worldwide study. This dataset consisted of ≈10-km resolution raster files with global yield and harvested area data for each pixel representing 2010. The files were obtained using a spatial production allocation model, a downscaling method based on average crop data from 2009 to 2011, encompassing the closest years to areas where data from this period were missing (Yu et al., 2020). Yu et al. (2020) classified the data in four faming systems: 1) irrigated farming using a high input level, such as modern varieties, fertilizers, and advanced management techniques; 2) rainfed high-input farming oriented to market, using high-yield varieties, machinery, and optimum nutrient application and phytosanitary control; 3) rainfed low-input farming oriented for personal consumption using traditional varieties and manual labor with low nutrient application and chemical phytosanitary control; and 4) subsistence farming system, consisting of areas in which cropland and suitable areas do not exist but farmland is still present, with production primarily for personal consumption. For this study, we used only the yield and harvested area data of rainfed crops, consisting of high-input (2), low-input (3), and subsistence (4) farming systems. The maps showing the yield of these five crops is provided in Supplementary Material 1.

We also obtained raster layers with land use and land cover (LULC) for 2010 (Potapov et al., 2022) with a resolution of 30 m, presence/absence of high irrigation (>2000 ha in 86 km^2), and a resolution of ≈9 km (Nagaraj et al., 2021) and an ecoregion map (Dinerstein et al., 2017) that splits the world terrestrial area into 14 ecoregions (Fig. 2).

The sugarcane, maize, rice, wheat, and soybean yield and harvested area data and LULC, irrigation, and ecoregion information were sampled from the harmonized soil dataset.

2.2. Scoring yield data

The yield data for each crop were scored between 0 and 1 using an empirical cumulative distribution function (eCDF) with the *ecdf* package in R (R Core Team, 2024). eCDF is a step function that increases by $1/n$ in each observation, where n is the number of observations. It is constant between 0 and 1 (Dekking, 2005). The eCDF function for dataset $x = (x_1, x_2, \cdots, x_n)$ can be described as follows:

$$\widehat{F}_n(x) = \frac{1}{n}\sum\nolimits_{i=1}^{n} I(x_i \le x) \tag{1}$$

where $\widehat{F}_n(x)$ is the eCDF function of x; n is the number of observations in x; and $I(x_i \le x)$ is the indicator function, that is 1 if $x_i \le x$ and 0, otherwise. The input for eCDF (x dataset) is the yield for each crop, and the output is the transformed utility of that crop (U).

After scoring, the utilities of sugarcane (U_{sg}), maize (U_m), rice (U_r), wheat (U_w), and soybean (U_{sb}) were aggregated using the weighted average, resulting in the average utility (U^*). The weight (W) for each crop was calculated for all points as follows:

$$W = \frac{A}{A_{sg} + A_m + A_r + A_w + A_{sb}} \tag{2}$$

where A_{sg} is the sugarcane area; A_m is the maize area; A_r is the rice area; A_w is the wheat area; and A_{sb} is the soybean area. A is the area of each crop, and the formula above was applied with the A of all crops in nominators (A_{sg}, A_m, A_r, A_w, or A_{sb}) to obtain the W for all crops (W_{sg}, W_m, W_r, W_w, and W_{sb}). Missing A values (no data for a specific crop in a specific point) resulted in $W = 0$, and this crop was then ignored in the next equation. U^* was calculated as follows:

$$U^* = \frac{(U_{sg} \times W_{sg}) + (U_m \times W_m) + (U_r \times W_r) + (U_w \times W_w) + (U_{sb} \times W_{sb})}{W_{sg} + W_m + W_r + W_w + W_{sb}} \tag{3}$$

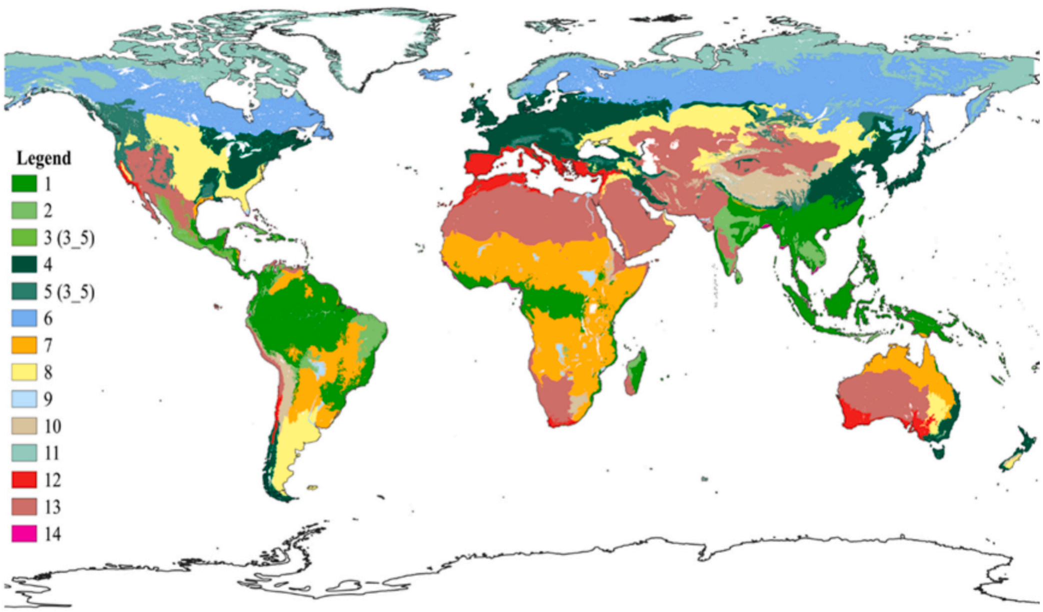

Fig. 2. World's ecoregions from Dinerstein et al. (2017) and available at "http://ecoregions.appspot.com/". 1 = Tropical and subtropical moist broadleaf forests; 2 = Tropical and subtropical dry broadleaf forests; 3 = Tropical and subtropical coniferous forests; 4 = Temperate broadleaf and mixed forests; 5 = Temperate coniferous forests; 7 = Tropical and subtropical grasslands, savannas, and shrublands; 8 = Temperate grasslands, savannas, and shrublands; 12 = Mediterranean forests, woodlands, and scrub; 13 = Deserts and xeric shrublands; and 3_5 = Tropical, subtropical, and temperate coniferous forests.

2.3. Utility modelling

The points were stratified by ecoregion, reducing the variability of environmental factors and enhancing the influence of soil attributes in the behavior utility functions. The points were filtered according to LULC, retaining only the points from rainfed croplands in the fitting dataset. The models were fitted using a generalized additive model (GAM), that is based in sum of smooth functions and is useful to access nonlinear effects of the variables (Hastie and Tibshirani, 1986), implemented in R *mgcv* package (R Core Team, 2024; Wood, 2017).

Four bivariate utility functions, namely clay, pH, SOC, and PAW, were fitted, one for each PI, using data from the surface (0–20 cm) and subsoil (80–100 cm) for each soil attribute as an independent variable and U^{*} as the dependent variable. The GAM with an interactive term can be described as follows:

$$U^{*}\ s(PI_{0-20cm}) + s(PI_{80-100cm}) + s(PI_{0-20cm}, PI_{80-100cm}) \tag{4}$$

where *s* is the smoothing function. Exploratory analyses were performed for all ecoregions, but ecoregions without enough points to fit the GAMs or that did not reach the significance threshold of $p = 0.05$ in the F-test for at least one PI, indicating a lack of statistical significance in modelling, were left out of the next step. For visualization, surface plots with level curves were built using the *ggplot2* package in R (R Core Team, 2024; Wickham, 2016), where the U^{*}_{pred} values for the GAM fitting dataset were also screened.

To obtain more accurate results, two more GAMs were fitted for each ecoregion using all PIs. One model was built using only data from the surface, as described in Eq. (5); the other model was built using data from the surface and subsoil layers of all PIs, as described in Eq. (6).

$$U^{*}\ s(Clay_{0-20cm}) + s(pH_{0-20cm}) + s(PAW_{0-20cm}) + s(SOC_{0-20cm}) \tag{5}$$

$$\begin{aligned} U^{*}\ s(Clay_{0-20cm}) &+ s(Clay_{80-100cm}) + s(pH_{0-20cm}) + s(pH_{80-100cm}) \\ &+ s(PAW_{0-20cm}) + s(PAW_{80-2100cm}) + s(SOC_{0-20cm}) \\ &+ s(SOC_{80-100cm}) \end{aligned} \tag{6}$$

The GAMs for each ecoregion fitted using non-irrigated cropland areas were used to predict the utility $\left(U^{*}_{pred}\right)$ for all points of the ecoregion, independent of landcover. The U^{*}_{pred} for all ecoregions was compiled in a unique dataset, resulting in six datasets for all ecoregions with U^{*}_{pred} from clay, pH, PAW, SOC, all PIs using data from the surface layer, and all PIs using data from both layers.

2.4. Digital soil mapping

The U^{*}_{pred} derived from the PIs (clay, pH, SOC, and PAW) and the other two models were mapped at a 90-m resolution for the world (except ecoregions without a utility function) using a geospatial mapping pipeline (van den Hoogen et al., 2021) implemented in Python using Google Colaboratory Interface and Google Earth Engine (GEE) (Gorelick et al., 2017). The mapping procedure was based on the DSM framework, in which environmental covariates representing SCORPAN model factors were used to perform the spatial predictions (McBratney et al., 2003).

A set of proxies (covariates) representing the soil variations were obtained in GEE: 1) A soil–vegetation image, representing the soil (directly), organisms (directly), age (indirectly), and parent material (indirectly) in which the bare soil pixels were flagged by vegetation indices from Landsat time series images (1984–2022) according to Demattê et al. (2018) and the locations without bare soil were filled with a vegetation mosaic; 2) the annual temperature and precipitation represented the climate factor (directly) (Karger et al., 2017); and 3) the elevation, slope, and multiscale topographic position index represented the relief factor (directly) (Safanelli et al., 2020; Theobald et al., 2015; Yamazaki et al., 2017); the Euclidean distance to rivers represented the spatial position (directly) (Behrens et al., 2018). The covariates were obtained or resampled to a 90-m resolution.

The pipeline from Hoogen et al. (2021) utilizes the Random Forest machine learning model (Breiman, 2001), which is composed of the following steps: 1) sampling the environmental covariates for the U^{*}_{pred} georeferenced dataset; 2) hyperparameter tuning for the Random Forest model (testing 27 combinations); 3) 10-fold cross-validation and calculation of performance metrics: coefficient of determination (R^2), root mean square error (RMSE), and ratio of the interquartile performance (RPIQ); 4) selection of the best model by the lowest RMSE; 5) spatial prediction by bootstrapping (Efron and Tibshirani, 1993) with 100 repetitions; 6) generation of 90-m maps for each U^{*}_{pred} by calculating the mean and uncertainty maps ($90\% pedictioninterval = 95Q - 5Q$, where Q is the prediction quartile).

The averages of the clay, pH, SOC, and PAW U^{*}_{pred} maps were calculated, generating a unique map derived from all bivariate utility functions, and models using all attributes with data from the surface layer and using all attributes with data from both layers were generated from a unique model.

3. Results

3.1. Scoring yield data to utility

After sampling the global dataset, 131,438 points for wheat, 117,270 points for maize, 93,118 for soybean, 31,800 for rice, and 25,119 for sugarcane were obtained. The yields of sugarcane, maize, rice, wheat, and soybean and the utilities scored by the eCDF are presented in Fig. 3. Sugarcane had the most sampled points with high yields, varying from 0 ($U_{sg} = 0$) to 120 ton ha^{-1} ($U_{sg} = 1$) and a median close to 75 ton ha^{-1} ($U_{sg} = 0.50$) (Fig. 3a). Maize, rice, and wheat presented the most sampled points with low yields. The maize yield varied from 0 to 16 ton ha^{-1}, with a median of 5 ton ha^{-1}. The rice yield varied from 0 to 8 ton ha^{-1} with a median of 2.5 ton ha^{-1}, and wheat yield varied from 0 to 10 ton ha^{-1} with a median of 3 ton ha^{-1} (Fig. 3bcd). Soybean presented more uniform yields worldwide, varying from 0 to 5 ton ha^{-1} with a median value close to 2.3 ton ha^{-1} (Fig. 3d). The weighted mean utilities of all crop (U*) distributions are shown in Fig. 3f.

3.2. Soil and yield data stratified by ecoregions

Ecoregions 6. Boreal forests and taiga, 9. Flooded grasslands and savannas, 10. Montane grasslands and shrublands, and 14. Mangroves did not have data for rainfed croplands or reach the number of samples necessary to fit the GAM or degree of significance determined in

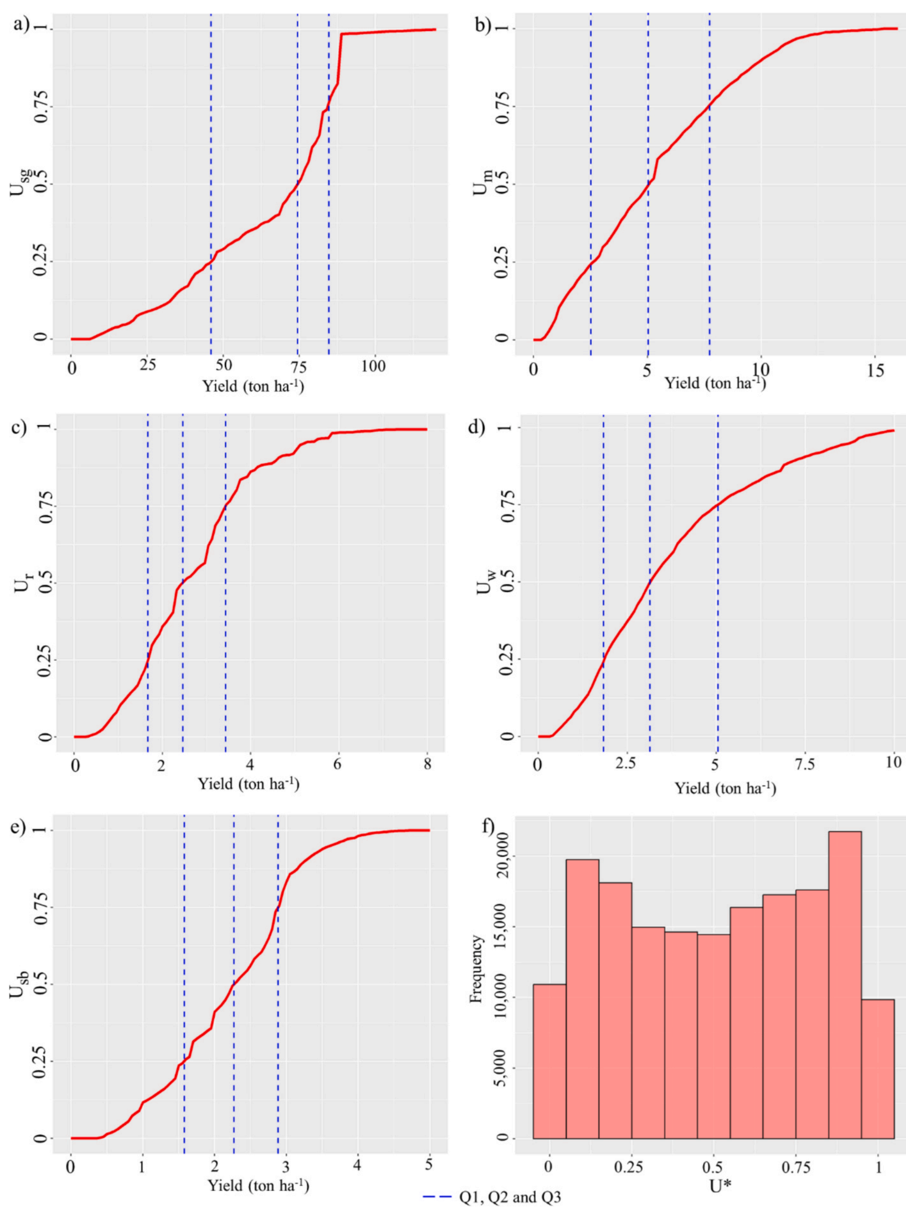

Fig. 3. Empirical cumulative distribution function plots for sugarcane (a), maize (b), rice (c), wheat (d), and soybean (e), and histogram of the weighted mean of all crop utilities (U*). U_{sc} = sugarcane utility; U_m = maize utility; U_r = rice utility; U_w = wheat utility; U_w = soybean utility; Q = quartile.

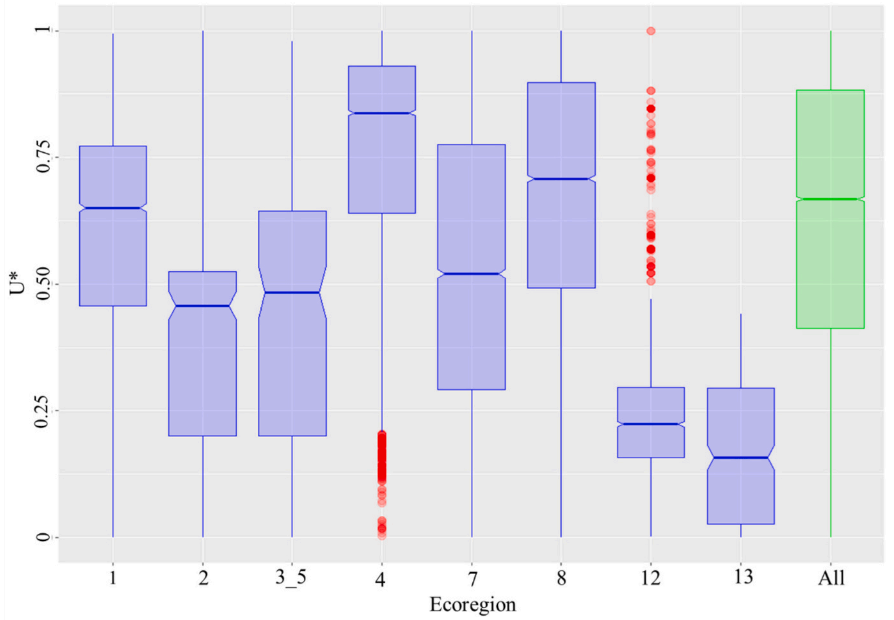

Fig. 4. Distribution of the utility (U*) of rainfed croplands for all points and ecoregions. 1 = Tropical and subtropical moist broadleaf forests; 2 = Tropical and subtropical dry broadleaf forests; 3_5 = Tropical, subtropical, and temperate coniferous forests; 4 = Temperate broadleaf and mixed forests; 7 = Tropical and subtropical grasslands, savannas, and shrublands; 8 = Temperate grasslands, savannas, and shrublands; 12 = Mediterranean forests, woodlands, and scrub; and 13 = Deserts and xeric shrublands.

preliminary tests; thus, they were removed from subsequent steps. The GAMs and mapping were only performed for the following ecoregions: 1. Tropical and subtropical moist broadleaf forests, 2. Tropical and subtropical dry broadleaf forests, 3. Tropical and subtropical coniferous forests, 4. Temperate broadleaf and mixed forests, 5. Temperate coniferous forests, 7. Tropical and subtropical grasslands, savannas, and shrublands, 8. Temperate grasslands, savannas, and shrublands, 12. Mediterranean forests, woodlands, and scrub, and 13. Deserts and xeric shrublands. The samples of ecoregions 3 and 5 were placed together in the same GAM and collectively called Tropical, subtropical, and temperate coniferous forests (3_5).

A total of 31,991 points were obtained for rainfed crops for these ecoregions, with at least enough data for one of the PIs: clay, pH, PAW, and SOC. The filtered points showed a high U* value, with a median of 0.66 (Fig. 4) compared to the dataset of all samples with U* (Fig. 3f). Ecoregions 1, 4, 7, and 8 contained most of the world's agricultural areas (Fig. 2 and Supplementary Material 1) and a large U* (Fig. 4). Ecoregions 4 and 8 had a higher U* with median values of 0.84 and 0.70, followed by ecoregions 1 and 7 with median values of 0.65 and 0.53. Ecoregions 3_5 and 2 had U* median values of 0.48 and 0.45, respectively. The lowest U* values were found for ecoregions 12 and 13 with median values of 0.22 and 0.16, respectively. The number of samples per ecoregion and per PI are given in Table 1.

3.3. GAM model performance

The accuracy of GAMs is shown in Tables 1 and 2, together with the number of samples used to fit each model in each ecoregion. To model each PI individually, the R^2 ranged from 0.01 to 0.30 for clay, from 0.06 to 0.22 for pH, from 0.01 to 0.36 to PAW, and from 0.06 to 0.30 for SOC. The R^2 for modelling all PIs for the superficial layer ranged from 0.04 to 0.38, and modelling all PIs for both layers ranged from 0.18 to 0.46. Generally, the best fitting model was found for all attributes from both layers, followed by all attributes for the superficial layer, and the worst accuracies were found for the individually fitted PIs.

The models were used to reach U^*_{pred} for all points (rainfed agriculture or not), resulting in 105,235 data points for clay, 102,286 for pH, 82,315 for PAW, and 77,204 for SOC. The models using all attributes resulted in less points for further mapping with 86,754 using the models for all attributes with data from superficial layer and 49,499 for the model using data from both layers.

3.4. Utility function behavior

The Tropical and subtropical moist broadleaf forest ecoregion presented high U^*_{pred} values when the clay content for both layers was higher than 200 g kg^{-1} (Fig. 5a), presenting two peaks for low and high clay values. Lower U^*_{pred} values were found for points with a low clay content in the superficial layer and a high content in the subsoil layer. Tropical and subtropical dry broadleaf forests presented two U^*_{pred} peaks when the clay content for both layers was intermediary and high (Fig. 5b). Tropical, subtropical, and temperate coniferous forests showed higher U^*_{pred} values for sandy soils with an influence from the subsoil layer (Fig. 5c). Temperate broadleaf and mixed forests presented higher U^*_{pred} values for sandy soils with more influence from the subsoil (Fig. 5d). Tropical and subtropical grasslands, savannas, and shrublands presented higher U^*_{pred} values when the clay content for both layers was between 200 and 750 g kg^{-1} (Fig. 5e). Lower U^*_{pred} values were found for points with a low clay content in the superficial layer and a high content in the subsoil layer. Temperate grasslands, savannas, and shrublands presented two U^*_{pred} peaks for a moderate clay content in both layers (Fig. 5f). Generally, these ecoregions presented U^*_{pred} values for more sandy soils in the subsoil layer. Mediterranean forests, woodlands, and scrub presented higher U^*_{pred} values for an intermediate clay content in the surface layer and a low content in the subsoil layer (Fig. 5g). The points with a high clay content in the subsoil layer and a lower clay content in the superficial layer presented a very low U^*_{pred}. Deserts and

Table 1
Accuracy of generalized additive models (GAMs) for clay, pH, plant available water (PAW), and soil organic carbon (SOC) utility functions.

Ecoregion	Clay			pH			PAW			SOC		
	n	a-R^2	RMSE	n	a-R^2	RMSE	n	a-R^2	RMSE	n	a-R^2	RMSE
1. Tropical and subtropical moist broadleaf forests	3,682	0.20	0.19	3,701	0.12	0.20	3,599	0.19	0.19	3,241	0.16	0.19
2. Tropical and subtropical dry broadleaf forests	298	0.19	0.17	307	0.18	0.18	282	0.17	0.18	258	0.10	0.18
3–5. Tropical, subtropical, and temperate coniferous forests	154	0.03	0.26	149	0.02	0.27	152	0.06	0.26	125	0.13	0.25
4. Temperate broadleaf and mixed forests	8,058	0.11	0.19	6,681	0.09	0.21	5,851	0.06	0.19	4,464	0.06	0.21
7. Tropical and subtropical grasslands, savannas, and shrublands	6,660	0.30	0.23	6,756	0.22	0.25	6,141	0.31	0.23	5,465	0.30	0.23
8. Temperate grasslands, savannas, and shrublands	8,077	0.08	0.22	7,874	0.20	0.21	7,791	0.07	0.28	6,098	0.08	0.22
12. Mediterranean forests, woodlands, and scrub	1,164	0.25	0.16	1,112	0.06	0.12	397	0.36	0.18	1,304	0.13	0.16
13. Deserts and xeric shrublands	258	0.01	0.13	258	0.07	0.12	256	0.01	0.13	184	0.27	0.11

n = number of samples; a-R^2 = adjusted coefficient of determination; RMSE = root mean square error.

Table 2
Accuracy of generalized additive models (GAMs) for clay, pH, plant available water (PAW), and soil organic carbon (SOC) utility functions.

Ecoregion	0–20 layer			Both layers		
	n	a-R^2	RMSE	n	a-R^2	RMSE
1. Tropical and subtropical moist broadleaf forests	3,228	0.20	0.19	3,194	0.28	0.18
2. Tropical and subtropical dry broadleaf forests	280	0.23	0.18	235	0.32	0.16
3–5. Tropical, subtropical, and temperate coniferous forests	159	0.11	0.25	98	0.28	0.22
4. Temperate broadleaf and mixed forests	8.727	0.14	0.20	2,614	0.18	0.18
7. Tropical and subtropical grasslands, savannas, and shrublands	5,595	0.38	0.23	4,878	0.44	0.20
8. Temperate grasslands, savannas, and shrublands	6,409	0.18	0.21	4,828	0.44	0.19
12. Mediterranean forests, woodlands, and scrub	1,852	0.04	0.23	127	0.42	0.14
13. Deserts and xeric shrublands	286	0.28	0.11	160	0.46	0.09

n = number of samples; a-R^2 = adjusted coefficient of determination; RMSE = root mean square error.

xeric shrublands presented generally lower U^*_{pred} values, which were higher for clayey soil with more influence in the surface layer (Fig. 5h). The peaks of U^*_{pred} for clay are shown in Table 3.

Tropical and subtropical moist broadleaf forest ecoregions presented the highest U^*_{pred} values for a pH lower than 7 in the subsoil layer (Fig. 6a). In the surface layer, high U^*_{pred} values were found with a high pH. Tropical and subtropical dry broadleaf forests presented high U^*_{pred} values with a high pH in the superficial layer and a low pH in the subsoil layer (Fig. 6b). Tropical, subtropical, and temperate coniferous forests presented higher U^*_{pred} values for points with a high pH in the superficial layer and a lower pH in the subsoil layer (Fig. 6c). Temperate broadleaf and mixed forests presented a higher U^*_{pred} when the pH was high in the surface layer and lower in the subsoil layer (Fig. 6d). Tropical and subtropical grasslands, savannas, and shrublands presented a peak in U^*_{pred} for slightly acidic pH values in both layers (Fig. 6e). This ecoregion presented a low U^*_{pred} for a pH higher than 7 for both layers. Temperate grasslands, savannas, and shrublands presented a peak in U^*_{pred} for slightly acidic pH values in both layers (Fig. 6f). This ecoregion presented a low U^*_{pred} for pH values higher than 8 in the surface layer. Mediterranean forests, woodlands, and scrub and Deserts and xeric shrublands presented a peak in U^*_{pred} for slightly basic pH values in both layers (Fig. 6g). Deserts and xeric shrublands presented generally lower U^*_{pred} values with higher values for a low pH in the superficial layer (Fig. 6h). The peaks of U^*_{pred} are shown in Table 3.

For PAW, most ecoregions showed a behavior similar to clay (Figs. 5 and 7). The tropical and subtropical moist broadleaf forest ecoregion presented high U^*_{pred} values for intermediary and high PAW values in both layers (Fig. 7a). Tropical and subtropical dry broadleaf forests presented a high U^*_{pred} for high and relatively high PAW values in both layers (Fig. 7b). Tropical, subtropical, and temperate coniferous forests presented higher U^*_{pred} values for points with higher PAW values in the superficial layer and lower PAW values in the subsoil layer (Fig. 7c). Temperate broadleaf and mixed forests presented a peak in U^*_{pred} with intermediary PAW values in the superficial layer and lower PAW values in the subsoil layer (Fig. 7d). For most points, high or relativity high U^*_{pred} values were found. Tropical and subtropical grasslands, savannas, and shrublands presented higher U^*_{pred} values when PAW values in both layers were between 0.10 and 0.25 m^{-1} m^{-1}, with 2 peaks (Fig. 7e). Temperate grasslands, savannas, and shrublands presented a peak in U^*_{pred} for PAW intermediary values in both layers, as U^*_{pred} was kept high for this range in the superficial layer even with a low PAW in the subsoil layer (Fig. 7f). Mediterranean forests, woodlands, and scrub and Deserts and xeric shrublands presented higher U^*_{pred} values for intermediary PAW values in the superficial layer and low PAW values in the subsoil layer (Fig. 7g). Deserts and xeric shrublands presented generally lower U^*_{pred} values, with a higher U^*_{pred} for high PAW values in the superficial layer and low PAW values in the subsoil layer (Fig. 7h). The peaks of U^*_{pred} are given in Table 3.

Most points for all ecoregions showed lower SOC values (Fig. 8). The Tropical and subtropical moist broadleaf forest ecoregion presented high U^*_{pred} values in two main peaks with an SOC from 5 to 10 g kg^{-1} and from 15 to 20 g kg^{-1} in the superficial layer and 5 g kg^{-1} and from 10 to 15 g kg^{-1} in the subsoil layer (Fig. 8a). Tropical and subtropical dry broadleaf forests presented high U^*_{pred} values for an SOC content higher than the 10 g kg^{-1} in the subsoil layer (Fig. 8b). Tropical, subtropical, and temperate coniferous forests presented higher U^*_{pred} values for points with higher SOC values in both layers, with a higher influence in the superficial layer (Fig. 8c). Temperate broadleaf and mixed forests presented high U^*_{pred} values with low SOC values (close to 0 g kg^{-1}) in the subsoil layer, a SOC value higher than 10 g kg^{-1} in the superficial layer, and two peaks for high SOC values in the subsoil (Fig. 8d). Tropical and subtropical grasslands, savannas, and shrublands presented higher U^*_{pred} values when SOC was higher than 5 g kg^{-1} in the superficial layer, showing four peaks (Fig. 8e). Temperate grasslands, savannas, and shrublands presented four peaks for SOC values higher than 5 g kg^{-1} in

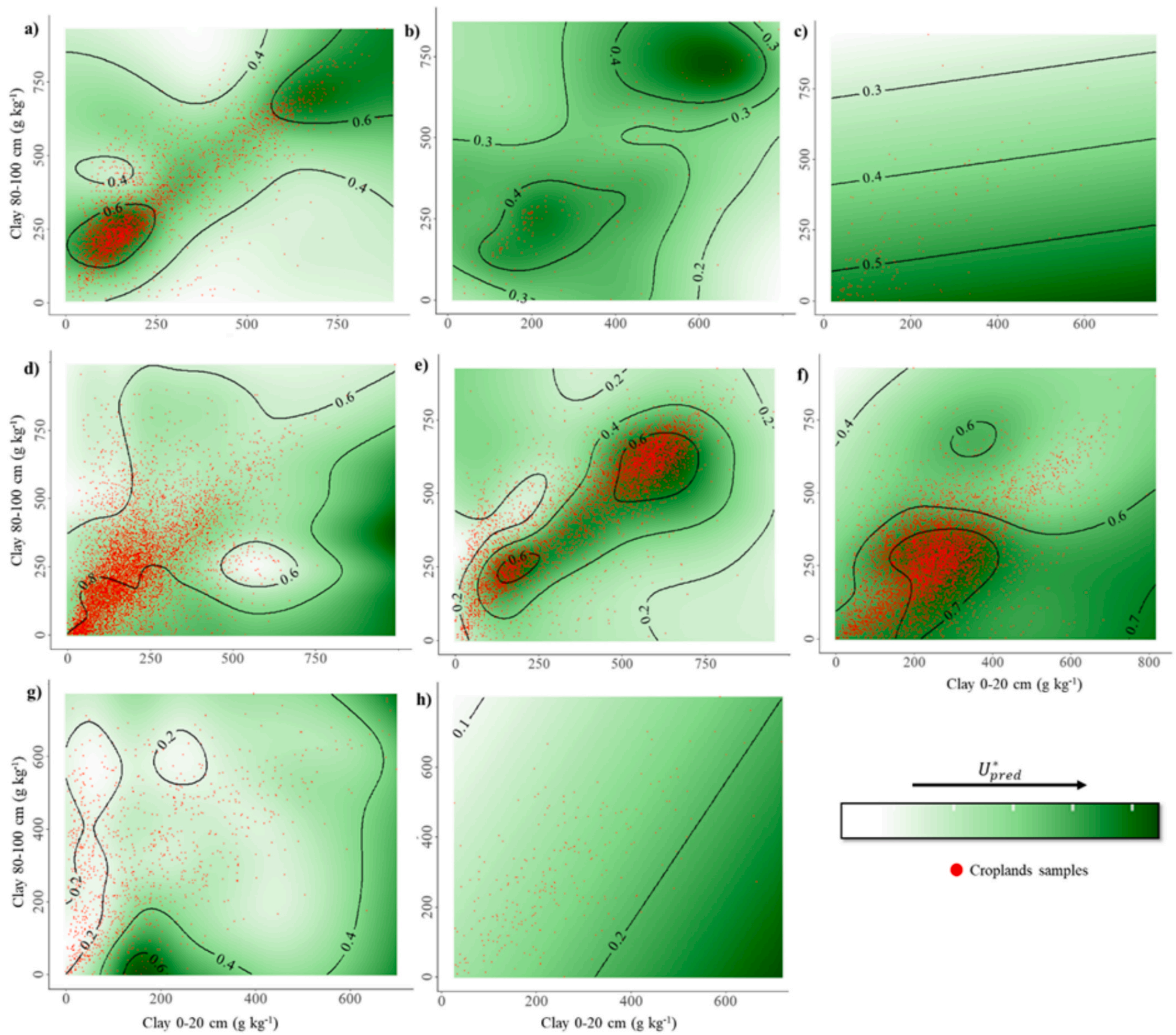

Fig. 5. Utility function plots between the clay content (surface and subsoil) and predicted utility (U^*_{pred}) for Tropical and subtropical moist broadleaf forests (a), Tropical and subtropical dry broadleaf forests (b), Tropical, subtropical, and temperate coniferous forests (c), Temperate broadleaf and mixed forests (d), Tropical and subtropical grasslands, savannas, and shrublands (e), Temperate grasslands, savannas, and shrublands (f), Mediterranean forests, woodlands, and scrub (g), and Deserts and xeric shrublands (h).

the superficial layer. Considering where most points were concentrated, there was a peak for U^*_{pred} with SOC values from 15 to 35 g kg^{-1}. Mediterranean forests, woodlands, and scrub presented higher U^*_{pred} values for SOC values from 5 to 15 g kg^{-1} in the superficial layer and from 5 to 10 g kg^{-1} in the subsoil layer (Fig. 8g). Deserts and xeric shrublands presented higher U^*_{pred} values, when SOC values were high in the superficial layer and low in the subsoil layer (Fig. 8h). The specific peaks for U^*_{pred} are shown in Table 3.

3.5. Utility maps

U^*_{pred} showed a pattern for the most cultivated crop in each region (Supplementary Material 1, Figs. 4 and 9). We obtained three U^*_{pred} maps for the world; the first was obtained from the average of all attributes. The highest U^*_{pred} values were in northeastern North America, northern Europe, and northeastern Asia, covering parts of ecoregions 4 and 8 (Fig. 9a). For the other two maps, high U^*_{pred} values were also found in those regions and in southeastern South America and central Africa, partially covering ecoregions 1, 4, 7, and 8 (Fig. 9bc). The lowest U^*_{pred} values were found in northern and southern Africa, the Middle East, central Asia, western North America, and central Australia for all maps in ecoregion 13. The other ecoregions (2, 3_5, and 12) showed intermediary U^*_{pred} values in all maps. The maps with models using all attributes (Fig. 9bc) showed more variations inside the ecoregions than in the average map (Fig. 9a). The main difference between the map using all attributes with data from the superficial layer and the map with all attributes from both layers was in ecoregion 13, showing North America and northern Europe.

The R^2 of spatial prediction using the attributes individually ranged from 0.66 to 0.69 and a mean uncertainty from 0.11 to 0.13. The maps using all attributes in the superficial layer had an R^2 of 0.68 and

Table 3
Specific peaks for the predicted utility (U^*_{pred}) of each utility function.

Ecoregion/Layer	Clay (g kg^{-1})		pH		PAW (m^{-1}m^{-1})		SOC (g kg^{-1})	
	Superficial	Subsoil	Superficial	Subsoil	Superficial	Subsoil	Superficial	Subsoil
1. Tropical and subtropical moist broadleaf forests	150–250 >650	150–250 >650	6.5–8	3.5–6	0.1–0.15 >0.25	0.1–0.15 >0.25	5–10 15–20	5 10–15
2. Tropical and subtropical dry broadleaf forests	200–400 450–700	200–400 600–800	>8 <5	>6.5 <5.5	0.25	0.25	0–30	>10
3–5. Tropical, subtropical, and temperate coniferous forests	–	< 250	>7	<6	>0.15	<0.05	>40	>15
4. Temperate broadleaf and mixed forests	–	< 200	6–7	6–7	0.1–0.2	0	>10	0 >20
7. Tropical and subtropical grasslands, savannas, and shrublands	250 500–750	250 500–750	5.5–7	5.5–6.5	0.15 0.20–0.25	0.15 0.20–0.25	5–20 30–40	>5
8. Temperate grasslands, savannas, and shrublands	150–400	50–300	5.5–7	5.5–7	0.10–0.20	0.10–0.20	5–35 45–70	<20
12. Mediterranean forests, woodlands, and scrub	100–300	0	7.5–8.5	6–8	0.10–0.15	0	5–15	5–10
13. Deserts and xeric shrublands	>600	–	<6.5	–	>0.15	<0.10	>10	<15

PAW = plant available water; SOC = soil organic carbon.

uncertainty mean of 0.11, but using all attributes in both layers resulted in values of 0.64 and 0.13, respectively. The U^*_{pred} maps for each attribute individually and the uncertainty maps are given in Supplementary Material 2.

4. Discussion

We used bivariate utility functions, a novel adaptation of the SSAF approach proposed by Evangelista et al. (2023), to establish the relationships of soil attributes (clay, pH, PAW, and SOC) in surface and subsoil layers with scored yield data used as the target indicator for the soil's capacity to produce food and biomass worldwide. The obtained indices are generic and can be used to infer the soil's capacity to support most crops of interest, not just the crops used in this study. We also assessed how the soil attributes drive food and biomass production across the world's ecoregions and their spatial distribution by high resolution mapping. This approach enabled the evaluation of the availability dimension (food production) of Food Security.

4.1. Crop yield worldwide

Most crops have high yields in the temperate climates of North America and Europe with the exception of sugarcane, presenting high yields in tropical conditions (Figs. 2 and 3, Supplementary Material 1). Blomqvist et al. (2020) desegregated the three drivers of the world's yields: cropping intensity, geographic distribution of croplands, and crop composition. The first is related to farm practices, and the latter two to diets and trade. Ray et al. (2015) reported the influence of climate on yield, explaining up to 39% of the variability and cited the influence of specific local factors on high spatial variability yields, such as the availability and use of agronomic inputs, pest and pathogen infestations, soil management, distribution of varied crop maturity types, socioeconomic conditions, and political or social issues. Soil is the basis for food security, but there are no studies showing that relationship on a global scale, especially for crop intensification (Kopittke et al., 2019).

These conditions collectively resulted in different production systems and management practices that influence the yield worldwide. For example, for maize, the rainfed, low input, and subsistence systems are in South America and sub-Saharan Africa, and high input, rainfed systems are in China, North America, Europe, and other regions, as reflected by yield (Supplementary Material 1) (Yu et al., 2020). In this context, the use of scored yields of five crops aggregated by the weighted average of their area as target indicators for SSAF reducing the variation across the globe and enhancing the effects of soil attributes in utility functions.

4.2. GAM accuracy

The fitting of bivariate utility functions using the soil attributes individually enabled determination of the influence of each attribute in the utility, as described in Section 4.3, but some attributes showed very low accuracies (Table 1). Several factors could have impacted the accuracy of the GAMs, such as low resolution of yield data raster, error in soil analyses or imprecise geographical coordinates, and uncontrolled non-soil factors that influence the crop yield, as explained previously. The low accuracy in utility function propagated uncertainties in the final maps. Aiming to partially overcome this issue, two models with all attributes were fitted, resulting in more accurate utility functions. The model for all attributes using data from superficial layers showed better results than the utility functions with soil attributes used individually; however, the better model was reached using all attributes in the surface and subsoil layers (Table 2). The model with all attributes in both layers resulted in less points for prediction. In multivariate modelling, the combination of independent variables related to the dependent variable improves the model's accuracy. Ng et al. (2024) also found high accuracy by combining several attributes in a deep learning model than a single utility function for each PI. The high accuracy for the model using the two layers showed the importance of subsoil in crop productivity as demonstrated by de Landell et al. (2003) and Horst-Heinen et al. (2021).

4.3. Utility function behavior

In SSAF, the behavior of a scoring function for a given soil attribute to perform a role was classified as: "higher is better", "less is better", "optimum", "binary", or "categorical" (Evangelista et al., 2023). The expected behavior was "more is better" for clay, PAW, and SOC and "optimum" for pH (Cherubin et al., 2016; Volchko et al., 2014). We adapted this concept from bidimensional to tridimensional utility graphs, and the expected behavior for the bivariate utility functions is illustrated in Fig. 1. This novel approach enables the utility graph to be modelled and screened using data from two soil layers, as in this study, or from two soil attributes using only one utility function. The subsoil layer attributes, which also impact the crop productivity (Horst-Heinen et al., 2021; de Landell et al., 2003), have rarely been explored.

Soil quality and health studies often use an ordination method, such as principal component analysis to assign contributing weights to soil attributes (Andrews et al., 2002; Cherubin et al., 2016; Greschuk et al., 2023; Li et al., 2024). In these cases, the behavior of the predicted utility always follows the expected pattern (theorical or previously fitted), as only the weight of each attribute changes. The results obtained in this study contrast with the expected behavior for several cases (Figs. 4, 5, 6,

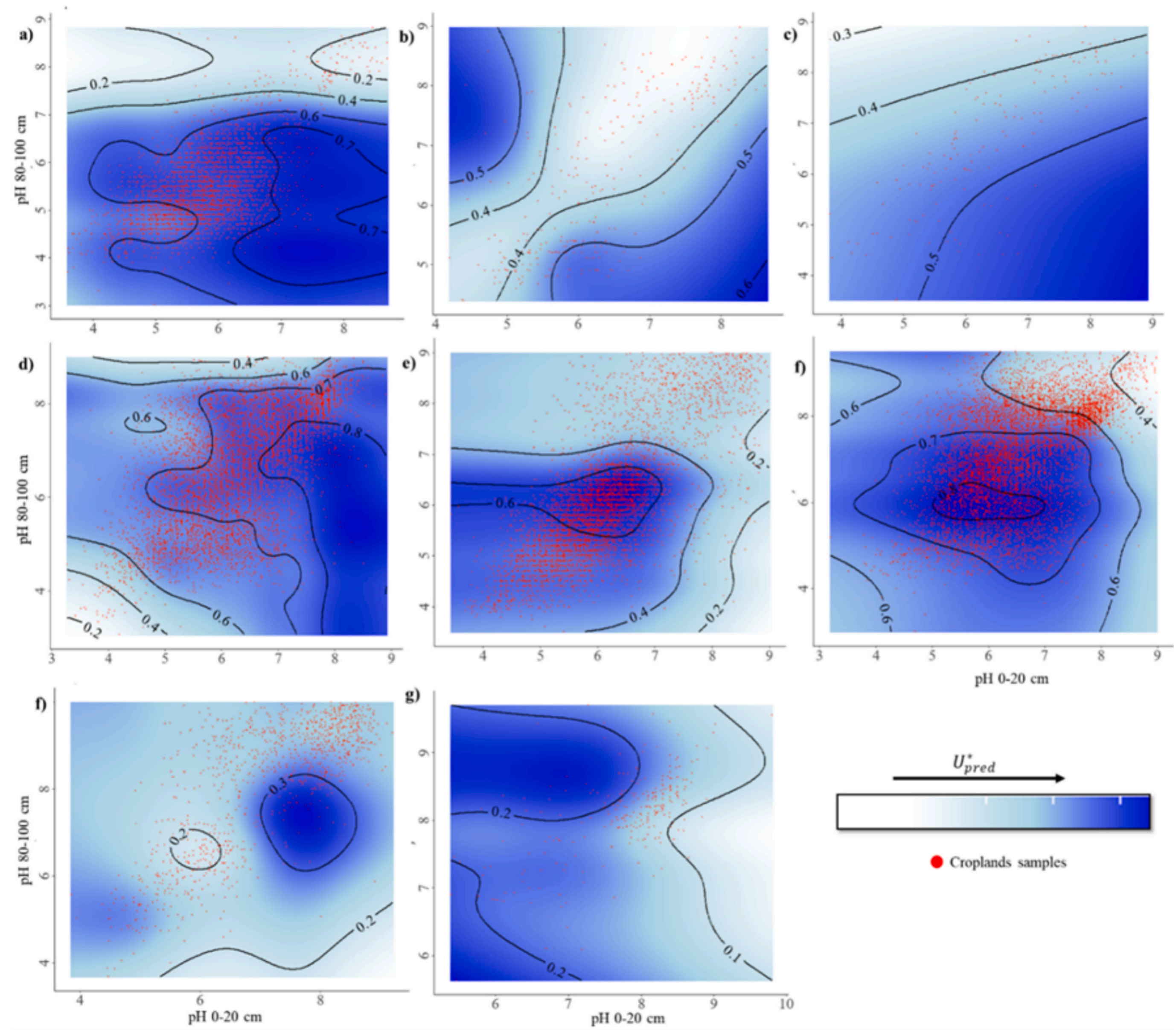

Fig. 6. Utility function plots between pH (surface and subsoil) and predicted utility (U^{*}_{pred}) for Tropical and subtropical moist broadleaf forests (a), Tropical and subtropical dry broadleaf forests (b), Tropical, subtropical, and temperate coniferous forests (c), Temperate broadleaf and mixed forests (d), Tropical and subtropical grasslands, savannas, and shrublands (e), Temperate grasslands, savannas, and shrublands (f), Mediterranean forests, woodlands, and scrub (g), and Deserts and xeric shrublands (h).

and 7; Table 2). This occurred because the utility function was fitted using real yield data, which permits the expression of the empirical relationship between the soil indicators and the target indicator (utility). The use of regression to fit the utility function eliminates redundancies in a dataset, reduces the number of PIs, and shows relatively high accuracy in a multiple PI deep learning modelling (Ng et al., 2024). However, it enables the use of external factors not related to soil influences in utility function fitting, such as climate, production system, farm management, and economic questions. The issues regarding yield data worldwide are well known (Ray et al., 2012, 2015). Ng et al. (2024) also found unexpected patterns for some soil attributes using SSAF to assess soil functions: habitat for biodiversity and nutrient regulation and storage.

For clay, some ecoregions showed an unexpected inverse pattern, such as high U^{*}_{pred} for sandy and loam soils, while other ecoregions showed peaks in U^{*}_{pred} for loam and clay (Fig. 4 and Table 2). Several agricultural areas with high yields (Yu et al., 2020) (Supplementary material 1) were observed in soils with variable soil textures, including areas with sandy soils (Huang and Hartemink, 2020; Poggio et al., 2021). Sandy soils can be managed to improve the crop yield, ensuring food security (Musei et al., 2024). Despite soil management, it is also possible that a specific cultivar can be adapted to these conditions (Reighard and Newall, 1993; Zhou and Gwata, 2016). For PAW, U^{*}_{pred} followed the same pattern as clay because this property was derived from clay plus the sand and silt contents (de Melo et al., 2023) (Fig. 6 and Table 2). The pH showed utility graphs closer to the expected pattern compared to other attributes with some biomes presenting a higher U^{*}_{pred} for a pH ranging from 5.5 to 8, mainly in the 0–20 cm layer. The 80–100 cm layer generally showed a high U^{*}_{pred} for pH values slightly lower than those in the 0–20 cm layer (Fig. 5 and Table 2). The pH in the 0–20 cm layer shows great anthropic influence and is closely related to the chemical conditions for plant growth (Neina, 2019). The

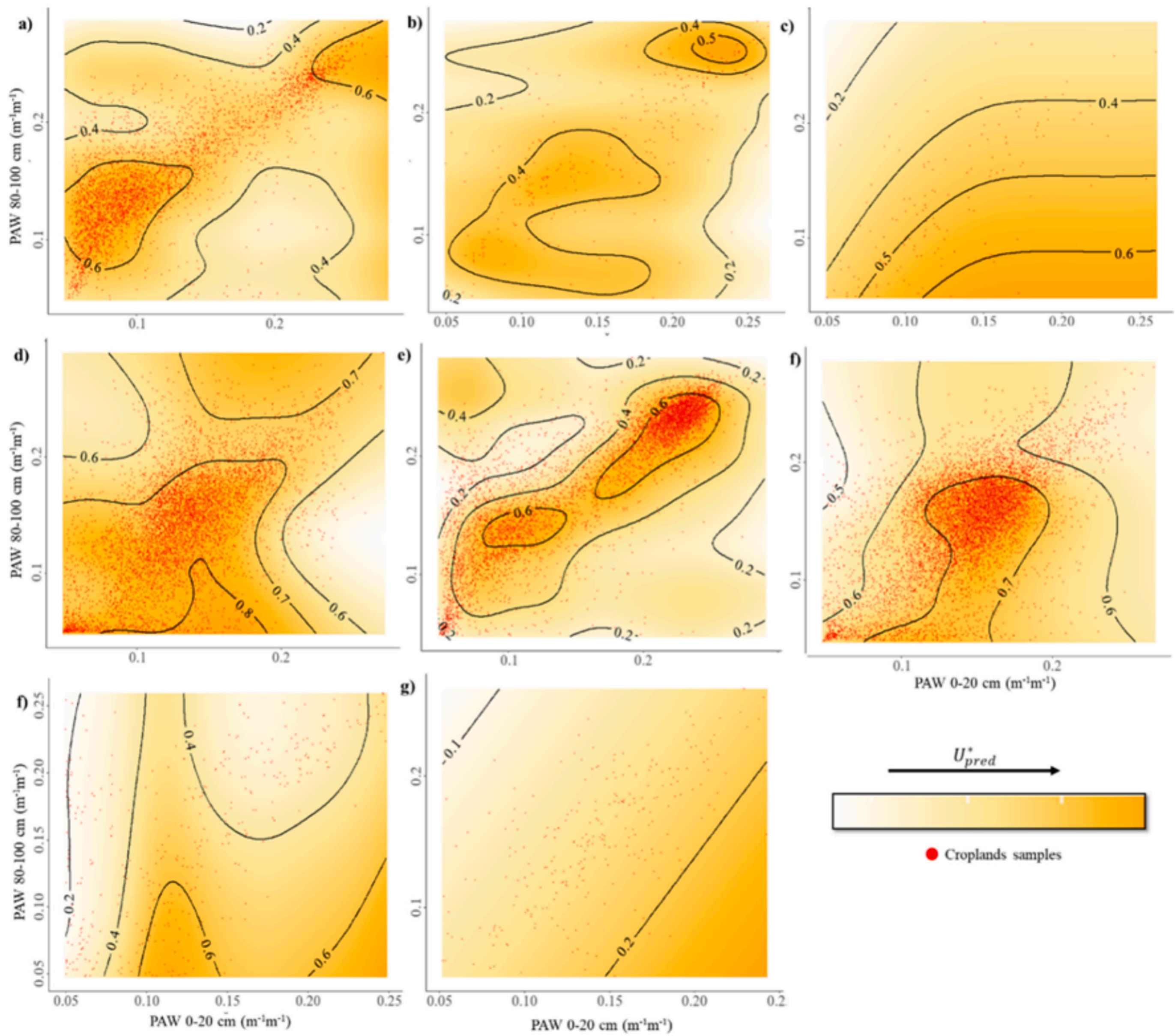

Fig. 7. Utility function plots between plant available water (PAW) (surface and subsoil) and predicted utility (U^{*}_{pred}) for Tropical and subtropical moist broadleaf forests (a), Tropical and subtropical dry broadleaf forests (b), Tropical, subtropical, and temperate coniferous forests (c), Temperate broadleaf and mixed forests (d), Tropical and subtropical grasslands, savannas, and shrublands (e), Temperate grasslands, savannas, and shrublands (f), Mediterranean forests, woodlands, and scrub (g), and Deserts and xeric shrublands (h).

pH in the 80–100 cm layer is more related to soil genesis processes with less anthropogenic influence. Globally, the water balance drives the soil pH in dry climates, having more alkaline soils, and wet climates have more acid soil (Slessarev et al., 2016). A low pH in the subsoil creates an inadequate environment for root growth by impacting water and nutrient absorption. For SOC, an unexpected behavior was observed in some of the ecoregions, as a high U^{*}_{pred} was associated with a low SOC (Fig. 7 and Table 2). Croplands normally lose SOC compared to native forests and grasslands (Morais et al., 2019), and high yields were found in areas with agriculture for a long time (Ray et al., 2012; Yu et al., 2020), as they likely experienced SOC degradation. External factors, without a relationship to the soil, explain the unexpected behavior of some utility functions.

4.4. Digital utility maps

As the first effort to evaluate the soil's capacity to produce food and biomass worldwide, there were some expected limitations due to the cascade of uncertainty from the input data for spatial predictions. The first map consisted of an average U^{*}_{pred} map for each attribute and could be considered the worst map because of its low accuracy in some utility functions. The high U^{*}_{pred} values for this map are mainly in areas of traditional agriculture with high yields, with few extrapolations (Fig. 9a and Supplementary Material 1), as high U^{*}_{pred} values were only found in ecoregions 4 and 8. The second and third maps come from more accurate GAMs (Table 2), presenting more variations in the ecoregions with a

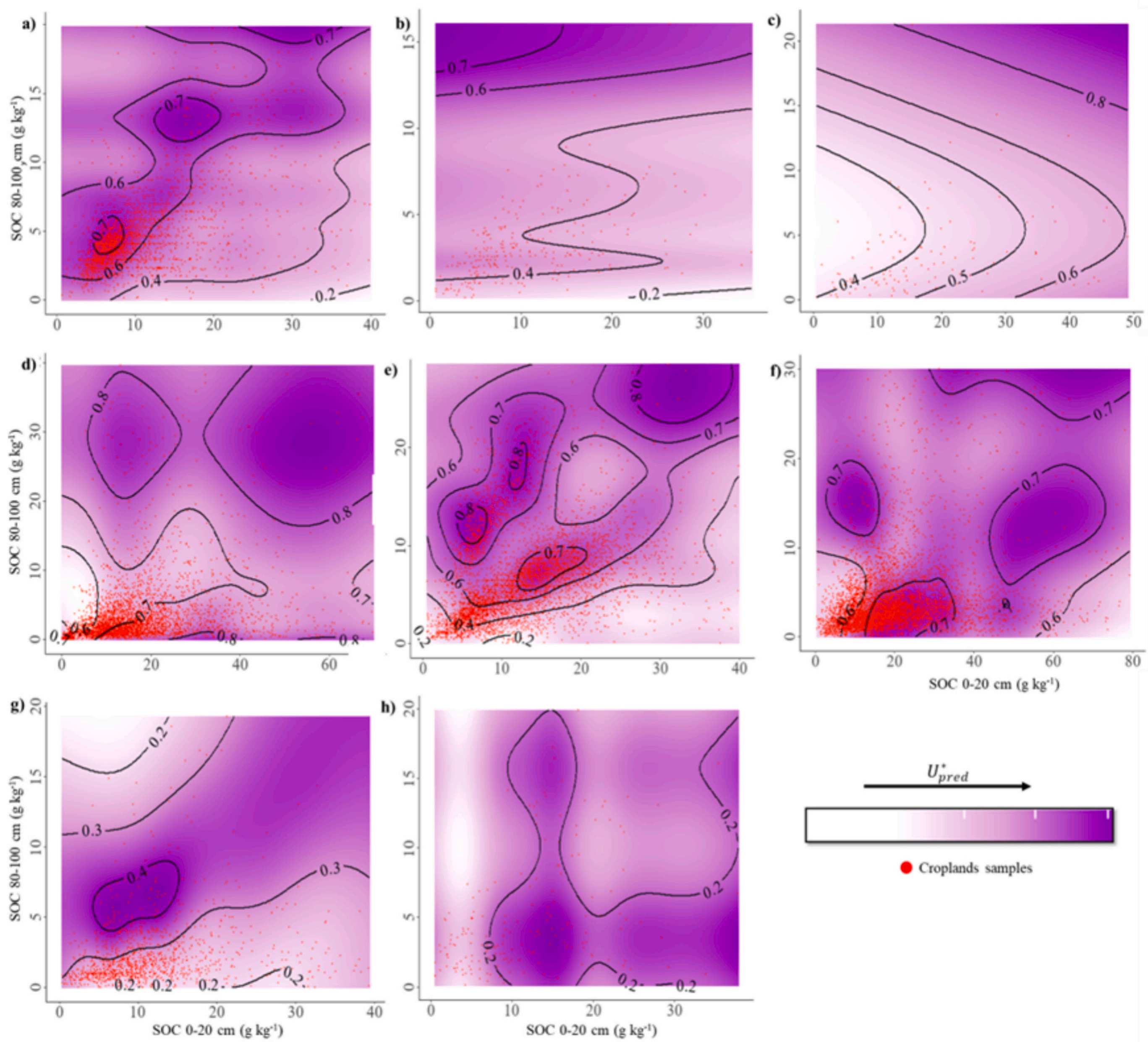

Fig. 8. Utility function plots between plant available water (PAW) (surface and subsoil) and predicted utility (U^*_{pred}) for Tropical and subtropical moist broadleaf forests (a), Tropical and subtropical dry broadleaf forests (b), Tropical, subtropical, and temperate coniferous forests (c), Temperate broadleaf and mixed forests (d), Tropical and subtropical grasslands, savannas, and shrublands (e), Temperate grasslands, savannas, and shrublands (f), Mediterranean forests, woodlands, and scrub (g), and Deserts and xeric shrublands (h).

high U^*_{pred} for ecoregions 1, 4, 7, and 8 and more extrapolation for non-agricultural areas with the same climate and soil conditions (Fig. 9bc). The third map had less points than the second map, showing higher uncertainty. Thus, the second map had the best accuracy in spatial prediction, and the third map showed more accuracy in GAMs (Table 2). The limited availability of subsoil information worldwide (Batjes et al., 2020) was also a factor. More effort could be made to obtain data for subsoil layers.

Generally, the U^*_{pred} distribution was related to the yield maps of the main crop in each region (Fig. 9 and Supplementary Material 1). However, the use of SSAF enabled downscaling 90 m (farm level) and extrapolation of the utility information to non-agricultural areas. For example, the U^*_{pred} of South America's high yield regions was extrapolated for areas in the same ecoregion and with the same soil conditions in Africa (Fig. 9bc). By identifying two non-agricultural regions in the same ecoregion with the same vegetation, the Amazon and the Congo rainforests, differences were observed in U^*_{pred}, with the Congo rainforest presenting high values (Fig. 9bc). This can be explained by differences in climate, topography, and soil type. The Congo rainforest had a higher altitude and low precipitation and showed a predominance of Rhodic

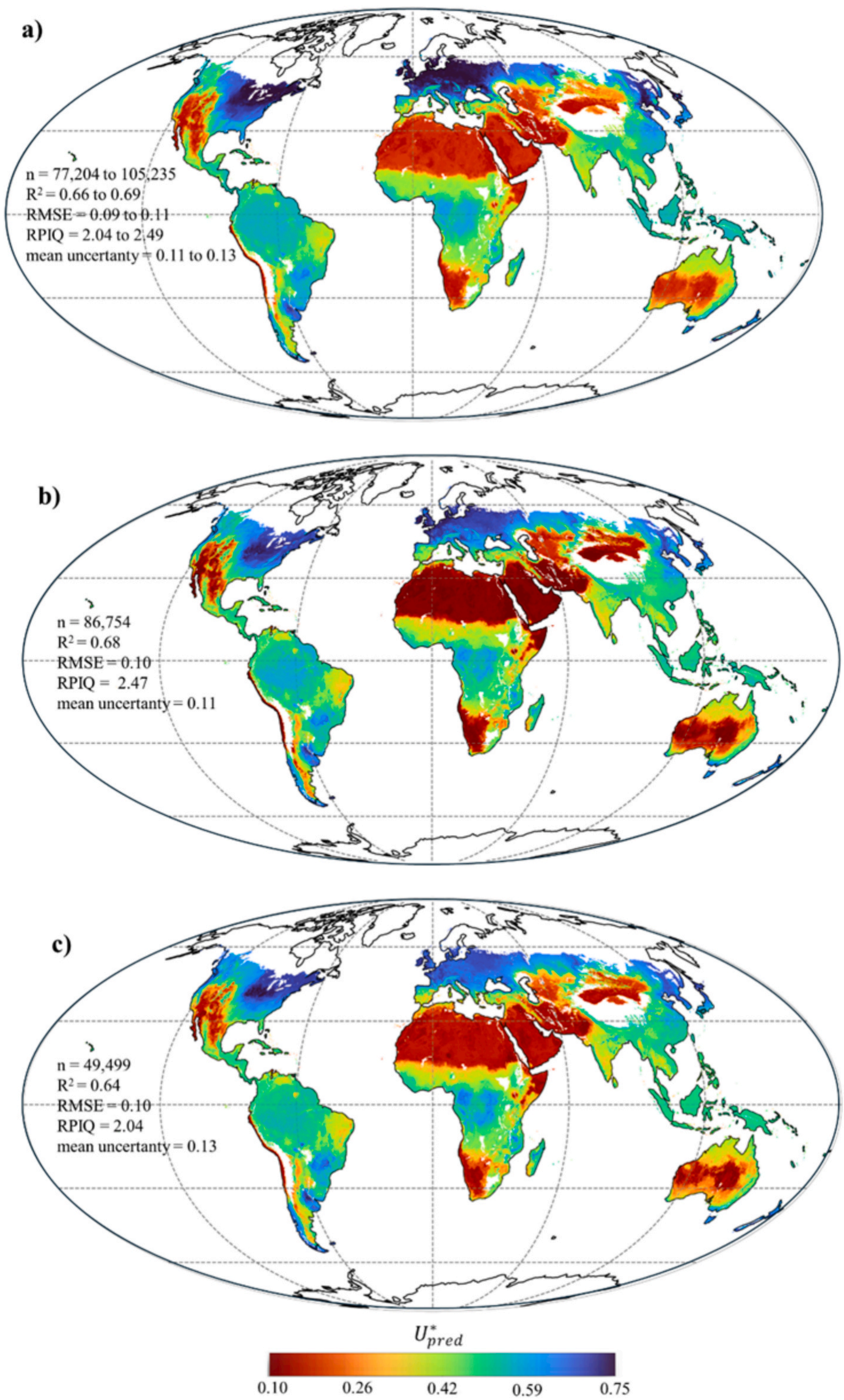

Fig. 9. Global maps for the predicted utility (U^*_{pred}) of clay, pH, plant available water (PAW), soil organic carbon (SOC): map average (a), model for all attributes in the superficial layer (b), and model for all attributes in both layers (c). n = number of samples; R^2 = coefficient of determination; RMSE = root mean square error; RPIQ = ratio of the interquartile performance.

Ferralsols, while Xanthic Ferralsols, a more leachate soil class, were predominant in the Amazon basin (FAO-UNESCO, 1974; Karger et al., 2017; Yamazaki et al., 2017).

Francos et al. (2025) generated soil capacity and condition maps for Americas' Horse Latitudes using SSAF and the Soil Health Index (SHI) method (developed by Poppiel et al. (2025)). Their soil capacity map covered all of these regions, enabling a general comparison with our results. SHI methodology utilizes a set of soil attributes and an ordination strategy (Poppiel et al. 2025), without fitting empirical utility functions that consider a target indicator (i.e., crop yield). There is a general similarity between our results (Fig. 9) and the SHI results from Francos et al. (2025), as shown by the occurrence of moderate capacity in most of South America and a match in regions with low capacity (northern Mexico, northeastern Brazil, and the Pacific coast of South America). However, there are also some considerable differences; the Paraná basin showed a high capacity in our results and a moderate–low capacity in the results of Francos et al. (2025). In the Andes Mountain ranges, Central America, and South Brazil, the opposite was observed. This difference occurred because the SHI is a fully soil-based index (Poppiel et al., 2025), while we used soil attributes and crop yield data, meaning that crop production also influenced the map pattern.

The increasing global food demand impacts land requirements and puts pressure on soil resource (Kastner et al., 2012; Kopittke et al., 2019; Pozza and Field, 2020). The high increases in yield occur in regions with recently established agriculture or historically low yields, while traditional agricultural showed stagnation or only moderate increases (Ray et al., 2012). The map of the world's soil capacity to produce food and biomass could be useful to determine areas where crops can be intensified or expanded, reducing the environmental impact. In this context, our product just assessed the soil component, which is only one factor to be considered in farm planning, and unrestrained cropland expansion is a factor for environmental degradation.

5. Conclusions

The soil's capacity to produce food and fuel were assessed worldwide within the Soil Security Assessment Framework (SSAF), providing an understanding of the availability dimension of Food Security. We scored and aggregated the yields of the five most consumed crops (sugarcane, maize, rice, wheat, and soybean) as utility. The obtained index was generic and can be applied/extended to other crops of interest not just these five. Bivariate utility functions using clay, pH, plant available water (PAW) and soil organic carbon (SOC) in the surface and subsoil layers were fitted by generalized additive models (GAMs) and presented an R^2 ranging from 0.01 to 0.36. The empirical bivariate utility functions relating the utility to soil attributes showed an unexpected behavior for some cases. The pH showed a behavior similar to the literature. The model using all of these soil attributes in the surface and subsoil had better accuracy, presenting an R^2 ranging from 0.18 to 0.46. However, a model using all soil attributes but only in the surface layer resulted in a map with less uncertainty due to high data availability. The predicted utility maps were related to the yield maps of the main crop for each region, but the use of SSAF enabled downscaling to 90 m (farm level) and extrapolation of the utility information for non-agricultural areas. However, it is the first approximation of the soil's capacity to produce food and biomass on a global scale, showing limitations due to input data uncertainties and uncontrolled factors. The obtained map of the soil's capacity to produce food and biomass can be used in sustainable agricultural intensification to promote Food Security.

CRediT authorship contribution statement

Nícolas Augusto Rosin: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Conceptualization. **Alex B. McBratney:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Conceptualization. **Raul Roberto Poppiel:** Writing – review & editing. **Jorge Tadeu Fim Rosas:** Writing – review & editing. **Merilyn T. Accorsi Amorim:** Writing – review & editing. **Nicolas Francos:** Writing – review & editing. **José A.M. Demattê:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Funding

This work was supported by the São Paulo Research Foundation (FAPESP) [grant numbers: 2023/11467–9; 2021/10063–6; 2021/05129–8; 2021/10573–4] and the Australian Research Council Laureate Fellowship (FL210100054) on Soil Security entitled "A calculable approach to securing Australia's soils".

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

The authors are grateful to the Department of Soil Science of ESALQ-USP, to the USP project nº 22.1.09345.01.2 and to the members of Geotechnologies in the Soil Science Group (GEOCIS) (https://esalq-geocis.wixsite.com/english).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2026.117821.

Data availability

The authors do not have permission to share data.

References

Andrews, S.S., Karlen, D.L., Mitchell, J.P., 2002. A comparison of soil quality indexing methods for vegetable production systems in Northern California. Agr. Ecosyst. Environ. 90 (1), 25–45. https://doi.org/10.1016/S0167-8809(01)00174-8.

Arrouays, D., Mulder, V.L., Richer-de-Forges, A.C., 2021. Soil mapping, digital soil mapping and soil monitoring over large areas and the dimensions of soil security – a review. Soil Secur. 5, 100018. https://doi.org/10.1016/j.soisec.2021.100018.

Batjes, N.H., Ribeiro, E., van Oostrum, A., 2020. Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). Earth Syst. Sci. Data 12 (1), 299–320. https://doi.org/10.5194/essd-12-299-2020.

Behrens, T., Schmidt, K., Viscarra Rossel, R.A., Gries, P., Scholten, T., MacMillan, R.A., 2018. Spatial modelling with Euclidean distance fields and machine learning. Eur. J. Soil Sci. 69 (5), 757–770. https://doi.org/10.1111/ejss.12687.

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. https://doi.org/10.1016/S0016-7061(99)00003-8.

Blomqvist, L., Yates, L., Brook, B.W., 2020. Drivers of increasing global crop production: a decomposition analysis. Environ. Res. Lett. 15 (9), 0940b6. https://doi.org/10.1088/1748-9326/ab9e9c.

Bouma, J., McBratney, A., 2013. Framing soils as an actor when dealing with wicked environmental problems. Geoderma 200–201, 130–139. https://doi.org/10.1016/j.geoderma.2013.02.011.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Cherubin, M.R., Karlen, D.L., Cerri, C.E.P., Franco, A.L.C., Tormena, C.A., Davies, C.A., Cerri, C.C., 2016. Soil quality indexing strategies for evaluating sugarcane expansion in Brazil. PLoS One 11 (3), e0150860. https://doi.org/10.1371/journal.pone.0150860.

de Melo, M.L.A., Inforsato, L., Pinheiro, E.A.R., de Jong van Lier, Q., 2023. Plant available water predicted by a flux-based approach. Geoderma 429, 116253. https://doi.org/10.1016/j.geoderma.2022.116253.

Dekking, M., 2005. A modern introduction to probability and statistics: understanding why and how. Springer, London.

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.

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N.D., Wikramanayake, E., et al., 2017. An ecoregion-based approach to protecting half the terrestrial realm. Bioscience 67 (6), 534–545. https://doi.org/10.1093/biosci/bix014.

Dubeux, J.C.B., Lira Junior, M.de.A., Simili, F.F., Bretas, I.L., Trumpp, K.R., Bizzuti, B.E., et al., 2024. Deep soil organic carbon: a review. CABI Reviews. https://doi.org/10.1079/cabireviews.2024.0024.

Efron, B., Tibshirani, R.J., 1993. An introduction to the bootstrap. Chapman & Hall, Londres.

Evangelista, S.J., Field, D.J., McBratney, A.B., Minasny, B., Ng, W., Padarian, J., et al., 2023. A proposal for the assessment of soil security: soil functions, soil services and threats to soil. Soil Secur. 10, 100086. https://doi.org/10.1016/j.soisec.2023.100086.

Evangelista, S.J., Field, D.J., McBratney, A.B., Minasny, B., Ng, W., Padarian, J., et al., 2024. Soil security—strategizing a sustainable future for soil (pp. 1–70). https://doi.org/10.1016/bs.agron.2023.10.001.

FAO-UNESCO, 1974. Soil map of the world at 1:5000000 scale.

Francos, N., Ng, W., Evangelista, S., Styc, Q., Sharififar, A., Pachon, J., McBratney, A., 2025. Assessing soil condition and capacity using the global pedogenon map within the americas' horse latitudes. Catena 259, 109372. https://doi.org/10.1016/j.catena.2025.109372.

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 (1), 14103. https://doi.org/10.1038/s41598-023-39981-y.

Hastie, T., Tibshirani, R., 1986. Generalized additive models. Stat. Sci. 1 (3), 297–310.

He, D., Oliver, Y., Rab, A., Fisher, P., Armstrong, R., Kitching, M., Wang, E., 2022. Plant available water capacity (PAWC) of soils predicted from crop yields better reflects within-field soil physicochemical variations. Geoderma 422, 115958. https://doi.org/10.1016/j.geoderma.2022.115958.

Horst-Heinen, T.Z., Dalmolin, R.S.D., ten Caten, A., Moura-Bueno, J.M., Grunwald, S., de Pedron, F., et al., 2021. Soil depth prediction by digital soil mapping and its impact in pine forestry productivity in South Brazil. For. Ecol. Manage. 488, 118983. https://doi.org/10.1016/j.foreco.2021.118983.

Huang, J., Hartemink, A.E., 2020. Soil and environmental issues in sandy soils. Earth Sci. Rev. 208, 103295. https://doi.org/10.1016/j.earscirev.2020.103295.

IPC, 2021. Technical Manual Version 3.1: Evidence and Standards for Better Food Security and Nutrition Decisions. Rome: FAO . https://www.ipcinfo.org/fileadmin/user_upload/ipcinfo/manual/IPC_Technical_Manual_3_Final.pdf. Accessed 8 September 2025.

van den Hoogen, J., Robmann, N., Routh, D., Lauber, T., van Tiel, N., Danylo, O., Crowther, T.W., 2021. A geospatial mapping pipeline for ecologists. bio Rxiv.

Johnston, A.E., 1986. Soil organic matter, effects on soils and crops. Soil Use Manag. 2 (3), 97–105. https://doi.org/10.1111/j.1475-2743.1986.tb00690.x.

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., et al., 2017. Climatologies at high resolution for the earth's land surface areas. Sci. Data 4 (1), 170122. https://doi.org/10.1038/sdata.2017.122.

Karlen, D.L., Andrews, S.S., Doran, J.W., 2001. Soil quality: current concepts and applications (pp. 1–40). https://doi.org/10.1016/S0065-2113(01)74029-1.

Kastner, T., Rivas, M.J.I., Koch, W., Nonhebel, S., 2012. Global changes in diets and the consequences for land requirements for food. Proc. Natl. Acad. Sci. 109 (18), 6868–6872. https://doi.org/10.1073/pnas.1117054109.

Koch, A., McBratney, A., Adams, M., Field, D., Hill, R., Crawford, J., et al., 2013. Soil security: solving the global soil crisis. Global Pol. 4 (4), 434–441. https://doi.org/10.1111/1758-5899.12096.

Kopittke, P.M., Berhe, A.A., Carrillo, Y., Cavagnaro, T.R., Chen, D., Chen, Q.-L., et al., 2022. Ensuring planetary survival: the centrality of organic carbon in balancing the multifunctional nature of soils. Crit. Rev. Environ. Sci. Technol. 52 (23), 4308–4324. https://doi.org/10.1080/10643389.2021.2024484.

Kopittke, P.M., Menzies, N.W., Wang, P., McKenna, B.A., Lombi, E., 2019. Soil and the intensification of agriculture for global food security. Environ. Int. 132, 105078. https://doi.org/10.1016/j.envint.2019.105078.

Landell, M.G.de.A., Prado, H.do, Vasconcelos, A.C.M.de, Perecin, D., Rossetto, R., Bidoia, M.A.P., et al., 2003. Oxisol subsurface chemical attributes related to sugarcane productivity. Scientia Agricola, 60(4), 741–745. https://doi.org/10.1590/S0103-90162003000400020.

Lehmann, J., Bossio, D.A., Kögel-Knabner, I., Rillig, M.C., 2020. The concept and future prospects of soil health. Nat. Rev. Earth Environ. 1 (10), 544–553. https://doi.org/10.1038/s43017-020-0080-8.

Li, Y., Zhao, H., Liu, J., Chaonan, C., Yuxuan, G., 2024. A framework for selecting and assessing soil quality indicators for sustainable soil management in waste dumps. Sci. Rep. 14 (1), 8491. https://doi.org/10.1038/s41598-024-58930-x.

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. https://doi.org/10.1016/J.GEODERMA.2009.10.007.

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

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117 (1–2), 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.

McBratney, A.B., Field, D., Morgan, C.L.S., Huang, J., 2019. On soil capability, capacity, and condition. Sustainability 11 (12), 3350. https://doi.org/10.3390/su11123350.

Morais, T.G., Teixeira, R.F.M., Domingos, T., 2019. Detailed global modelling of soil organic carbon in cropland, grassland and forest soils. PLoS One 14 (9), e0222604. https://doi.org/10.1371/journal.pone.0222604.

Musei, S.K., Kuyah, S., Nyawira, S., Ng'ang'a, S.K., Karugu, W.N., Smucker, A., Nkurunziza, L., 2024. Sandy soil reclamation technologies to improve crop productivity and soil health: a review. Front. Soil Sci. 4. https://doi.org/10.3389/fsoil.2024.1345895.

Nagaraj, D., Proust, E., Todeschini, A., Rulli, M.C., D'Odorico, P., 2021. A new dataset of global irrigation areas from 2001 to 2015. Adv. Water Resour. 152, 103910. https://doi.org/10.1016/j.advwatres.2021.103910.

Neina, D., 2019. The role of soil pH in plant nutrition and soil remediation. Appl. Environ. Soil Sci. 2019, 1–9. https://doi.org/10.1155/2019/5794869.

Newman, A.C.D., 1984. The significance of clays in agriculture and soils. Philos. Trans. Roy. Soc. London. Ser. A, Math. Phys. Sci. 311 (1517), 375–389. https://doi.org/10.1098/rsta.1984.0035.

Ng, W., Evangelista, S.J., Padarian, J., Pachon, J., O'Donoghue, T., Xue, P., et al., 2024. Estimating surrogates, utility graphs and indicator sets for soil capacity and security assessments using legacy data. Soil Res. 62 (2). https://doi.org/10.1071/SR23138.

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

Poppiel, R.R., Cherubin, M.R., Novais, J.J.M., Demattê, J.A.M., 2025. Soil health in Latin America and the Caribbean. Commun. Earth Environ. 6 (1), 141. https://doi.org/10.1038/s43247-025-02021-w.

Potapov, P., Hansen, M.C., Pickens, A., Hernandez-Serna, A., Tyukavina, A., Turubanova, S., et al., 2022. The global 2000-2020 land cover and land use change dataset derived from the landsat archive: first results. Front. Remote Sens. 3. https://doi.org/10.3389/frsen.2022.856903.

Pozza, L.E., Field, D.J., 2020. The science of soil security and food security. Soil Secur. 1, 100002. https://doi.org/10.1016/j.soisec.2020.100002.

R Core Team, 2024. R: a language and environment for statistical computing. https://www.R-project.org/. Accessed 17 June 2024.

Ray, D.K., Gerber, J.S., MacDonald, G.K., West, P.C., 2015. Climate variation explains a third of global crop yield variability. Nat. Commun. 6 (1), 5989. https://doi.org/10.1038/ncomms6989.

Ray, D.K., Ramankutty, N., Mueller, N.D., West, P.C., Foley, J.A., 2012. Recent patterns of crop yield growth and stagnation. Nat. Commun. 3 (1), 1293. https://doi.org/10.1038/ncomms2296.

Reighard, G.L., Newall Jr, W.C., 1993. Orchard middles and rootstock cultivar influence peach tree growth and fruiting on deep sandy soils. Acta Horticulturae 349, 159–162. https://doi.org/10.17660/ActaHortic.1993.349.23.

Safanelli, J.L., Poppiel, R., Ruiz, L., Bonfatti, B., Mello, F., Rizzo, R., Demattê, J., 2020. Terrain analysis in google earth engine: a method adapted for high-performance global-scale analysis. ISPRS Int. J. Geo-Inform. 9 (6), 400. https://doi.org/10.3390/ijgi9060400.

Sainju, U.M., Liptzin, D., Dangi, S.M., 2022a. Enzyme activities as soil health indicators in relation to soil characteristics and crop production. Agrosyst. Geosci. Environ. 5 (3). https://doi.org/10.1002/agg2.20297.

Sainju, U.M., Liptzin, D., Jabro, J.D., 2022b. Relating soil physical properties to other soil properties and crop yields. Scientific Reports 12 (1), 22025. https://doi.org/10.1038/s41598-022-26619-8.

Slessarev, E.W., Lin, Y., Bingham, N.L., Johnson, J.E., Dai, Y., Schimel, J.P., Chadwick, O. A., 2016. Water balance creates a threshold in soil pH at the global scale. Nature 540 (7634), 567–569. https://doi.org/10.1038/nature20139.

Smith, P., House, J.I., Bustamante, M., Sobocká, J., Harper, R., Pan, G., et al., 2016. Global change pressures on soils from land use and management. Global Change Biol. 22 (3), 1008–1028. https://doi.org/10.1111/gcb.13068.

Theobald, D.M., Harrison-Atlas, D., Monahan, W.B., Albano, C.M., 2015. Ecologically-relevant maps of landforms and physiographic diversity for climate adaptation planning. PLOS ONE 10 (12), e0143619. https://doi.org/10.1371/journal.pone.0143619.

Vogel, H.-J., Eberhardt, E., Franko, U., Lang, B., Ließ, M., Weller, U., et al., 2019. Quantitative evaluation of soil functions: potential and state. Front. Environ. Sci. 7. https://doi.org/10.3389/fenvs.2019.00164.

Volchko, Y., Norrman, J., Rosèn, L., Norberg, T., 2014. A minimum data set for evaluating the ecological soil functions in remediation projects. J. Soils Sediments 14 (11), 1850–1860. https://doi.org/10.1007/s11368-014-0939-8.

Wickham, H., 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag, New York.

Wood, S.N., 2017. Generalized additive models: an introduction with R, (2a.). Chapman and Hall/CRC, New York.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J.C., et al., 2017. A high-accuracy map of global terrain elevations. Geophys. Res. Lett. 44 (11), 5844–5853. https://doi.org/10.1002/2017GL072874.

Yu, Q., You, L., Wood-Sichra, U., Ru, Y., Joglekar, A.K.B., Fritz, S., et al., 2020. A cultivated planet in 2010 – part 2: the global gridded agricultural-production maps. Earth System Sci. Data 12 (4), 3545–3572. https://doi.org/10.5194/essd-12-3545-2020.

Zhou, M.M., Gwata, E.T., 2016. Quantifying sugarcane cultivar genetic gains in the midlands region of South Africa. Agron. J. 108 (1), 342–348. https://doi.org/10.2134/agronj2015.0141.