Geoderma Regional 28 (2022) e00444

Contents lists available at ScienceDirect

Geoderma Regional

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

Assessment of the soil fertility status in Benin (West Africa) – Digital soil mapping using machine learning

Kpade O.L. Hounkpatin [a,*], Aymar Y. Bossa [b,c], Yacouba Yira [b,d], Mouïnou A. Igue [e], Brice A. Sinsin [b,f]

[a] Department of Soil and Environment, Swedish University of Agricultural Sciences, P.O. Box 7014, SE-75007 Uppsala, Sweden
[b] Hydro-Climate Services, BV 30051 Ouagadougou, Burkina Faso
[c] Department of Water for Agriculture and the Society, National Water Institute, University of Abomey-Calavi, Cotonou 01, P.O. Box 526, Benin
[d] Applied Science and Technology Research Institute – IRSAT/CNRST, Ouagadougou, Burkina Faso
[e] National Institute of Agricultural Research of Benin, 01 BP: 988 Cotonou, Bénin
[f] Laboratory of Applied Ecology, University of Abomey-Calavi, 01 BP 526 Cotonou, Benin

ARTICLE INFO

Keywords:
Soil fertility index
Digital soil mapping
Random forest
Multiple soil classes

ABSTRACT

A soil fertility index map (SFIm) can provide key information to decision-makers in regard to spatial planning in the context of sustainable land management. The establishment of such SFIm requires basic soil properties that can be modelled for spatial mapping. The objective of this study was to take advantage of Benin soil legacy data to produce a digital SFIm at a national level based on 8 soil properties (soil organic matter, nitrogen, pH (water), exchangeable potassium, assimilable phosphorus, sum of bases, cation exchange capacity and base saturation). Specific research aims were (1) to model and develop digital soil maps, (2) to identify the key covariates influencing soil nutrients, and (3) to build an SFIm using digital maps of the soil properties. For each soil property, modelling procedures involved the use of different covariates, including soil type, topographic, bioclimatic and spectral data, along with the comparative assessment of the cubist (CB) and quantile random forest (QRF) models. Models were evaluated not only on the basis of classical error metrics (RMSE, R^2) but also on the ability to predict local uncertainty based on the prediction interval coverage probability (PICP). The results revealed that CB performed marginally better than the QRF based on classical error metrics (R^2, RMSE) but produced the worst uncertainty with an overestimation of the local uncertainty. This suggested that the use of accuracy plots such as PICP to evaluate models can identify accuracy problems not evident with classical error metrics. The analysis revealed that the distance to the nearest stream, which was part of topographic covariates, had strong predictive ability for all the soil properties along with the bioclimatic variables. The spatial distribution of the different classes of SFIm showed a preponderance of low fertility levels with severe limitations for crop development. A limited number of high and average fertility level soils were found in the low elevation areas of southern Benin, and policy could advocate for their sole use for agricultural purposes and promote sustainable management practices.

1. Introduction

Soils deliver many ecosystem services by purifying water, reducing soil contaminants, cycling nutrients, sequestrating carbon and contributing to the production of food, fibre and fuel. These services are essential for the biosphere, especially those supporting food and agriculture as well as environmental interactions (Blume et al., 2016; Madena et al., 2012). However, the pressures of a growing population have resulted in a higher demand for soil-related products, leading to land degradation processes such as erosion, nutrient depletion, chemical contamination, acidification, and salinity. These land degradations adversely affect the soil qualities associated with these processes, resulting in the decline or loss of soil functions such as agricultural productivity (Günal et al., 2015). As a nonrenewable resource, soils ought to be preserved to ensure a sustainable future.

As in most countries in sub-Saharan Africa, Benin is experiencing rapid demographic expansion with a population growth rate of nearly 3% along with an urbanization rate of 44% (INSAE, 2017). This trend is

* Corresponding author.
E-mail address: ozias.hounkpatin@slu.se (K.O.L. Hounkpatin).

https://doi.org/10.1016/j.geodrs.2021.e00444
Received 6 July 2021; Received in revised form 1 November 2021; Accepted 3 November 2021
Available online 10 November 2021

expected to continue in the near future with increasing demand for food production and urbanization. A high rate of urbanization constitutes a threat to soil, especially with the permanent removal of the fertile topsoil horizon and its replacement by asphalt and concrete or other construction materials (Charzyński et al., 2017; Vasenev et al., 2018). In that context, decision-making in spatial planning ought to take into account that threat by keeping areas with huge potential in fulfilling their functions such as food production (Breure et al., 2012; De Groot et al., 2010; Greiner et al., 2017). This caution in the spatial planning process is necessary to fulfil sustainable development goals 2 (zero hunger) and 11 (sustainable cities and communities) of the 2030 Agenda for Sustainable Development, which focus on ensuring food security and reducing the impact of urban development on natural ecosystems (UN, 2016). To reach these goals, stakeholders involved in spatial planning need spatially explicit decision tools that could lead to optimal decisions in landscape delineation.

The assessment of single or multifunctionality of soils relies heavily on some chemical, physical and biological soil parameters (Calzolari et al., 2016). Based on some soil science studies, Greiner et al. (2017) compiled several soil properties used as inputs for assessing soil functions. For example, to assess soil function related to water infiltration and storage, approximately 7 (texture, soil organic carbon, bulk density, hydromorphic property, stone content, soil depth, horizon depth) soil properties are required. To evaluate the potential utilization and productivity of soils, all the preceding soil properties are required in addition to the pH, while assessing nutrient availability to plants considers all soil properties except hydromorphic indicators.

Various maps of basic soil properties with varying accuracy and resolution exist in different countries of Africa (Hengl et al., 2017b; Hounkpatin et al., 2018b; Minai, 2019; Silatsa et al., 2020). These maps are generally based on legacy soil data (Arrouays et al., 2017; Leenaars, 2013; Leenaars et al., 2014; Leenaars et al., 2012) coupled with the increasingly available remote sensing data using digital soil mapping (DSM) techniques (Forkuor et al., 2017; McBratney et al., 2003). The DSM relates soil point data with statistically correlated auxiliary data, and its potential has been established in many studies (Adhikari et al., 2014; Hengl et al., 2017a; Minasny and McBratney, 2016; Zeraatpisheh et al., 2019). In most cases, the application of the DSM approach has been limited for mapping single or multiple soil properties at different scales and depths for diverse landscapes and land uses. Li et al. (2018) mapped cation-exchange capacity using Bayesian modelling and proximal sensors at the field scale in Australia, while clay, sand fractions, and organic matter content were predicted for three depth layers by multivariate adaptive regression splines in Sweden by Piikki et al. (2015) at the same scale. At a catchment and regional scale, many studies have reported mapping of carbon stocks (Hounkpatin et al., 2018b; Miller et al., 2015a; Phachomphon et al., 2010), soil salinity (Taghizadeh-Mehrjardi et al., 2014), and soil electrical conductivity (Mosleh et al., 2016; Nawar et al., 2015), while at a global scale, diverse soil properties have also been considered (Hengl et al., 2017a; Ramcharan et al., 2018).

Although a substantial amount of effort has been devoted to mapping soil properties and few efforts have been devoted to mapping some soil functions at the global scale (Leenaars et al., 2018; Poggio et al., 2019), more efforts are required to translate legacy soil data into practical decision tools using DSMs. For sustainable land use and management, indicators of soil production functions such as soil SFIm can potentially be a key decision tool in spatial planning. Soil fertility refers to the soil's capacity to deliver nutrients in available forms and suitable amounts for plant growth and reproduction (Grant, 2016). The objective of this study was to exploit the soil legacy data of Benin to produce a digital SFIm at a national scale using a DSM approach. Specific research aims were (1) to model and develop digital soil maps of soil properties required for building SFIm, (2) to identify the key covariates influencing the spatial distribution of soil nutrients, and (3) to build an SFIm using digital maps of the soil properties.

2. Methodology and approach

2.1. Study area

The republic of Benin is located in West Africa, extending over an area of 112,000 km^2 between latitudes 6° and 13°N and longitudes 0° and 4°E (Fig. 1). Its terrain is mostly flat with a mean elevation of 273 m above sea level along with some undulating plains and hills. The climate distribution in Benin lies within the Sudanian, Sudano-Guinean and Guineo-Congolese zones (Hopkins, 1987). The northern part of Benin has a Sudanian-type climate with both dry and rainy seasons. However, the southern part of Benin lies in the Guineo-Congolian zone with a subequatorial climate characterized by a bimodal rainfall regime with two rainy and two dry seasons. The middle part of Benin is characterized by a transitional pattern between the southern and northern climates. Precipitation is higher in the southern part of Benin (1295 mm annually) than in the northern part (1142 mm annually) (Amoussou et al., 2016). The average minimum and maximum temperatures fluctuate between 23 and 31 °C (southern part) and between 21 and 34 °C (northern part) (data from 1960 to 2016) (Hounnou and Dedehouanou, 2018). The natural vegetation is characterized by semideciduous dense humid forests in the Guineo-Congolian zone, humid semideciduous forests, woodlands, dry dense forests and riparian forests along riverbanks or along streams in the Sudano-Guinean zone, while the landscape in the Sudanian zone is dominated by savannah land with some spots of dry clear forest and gallery forests (Neuenschwander et al., 2011). Four main soil types can be distinguished (Volkoff & Willaine, 1967) in Benin: ferruginous soils are the most widespread, occupying approximately 60% of the surface area, ferralitic soils are found in coastal sedimentary basins and inland areas on crystalline basements, vertisols in depressions and hydromorphic soils in low-altitude areas. Agricultural production is mainly rainfed and mostly characterized by small-scale farming with low input and a low level of mechanization for food crops and a reasonable input use for cotton, which represents the main cash crop. The major crop species are maize, sorghum, rice, beans,

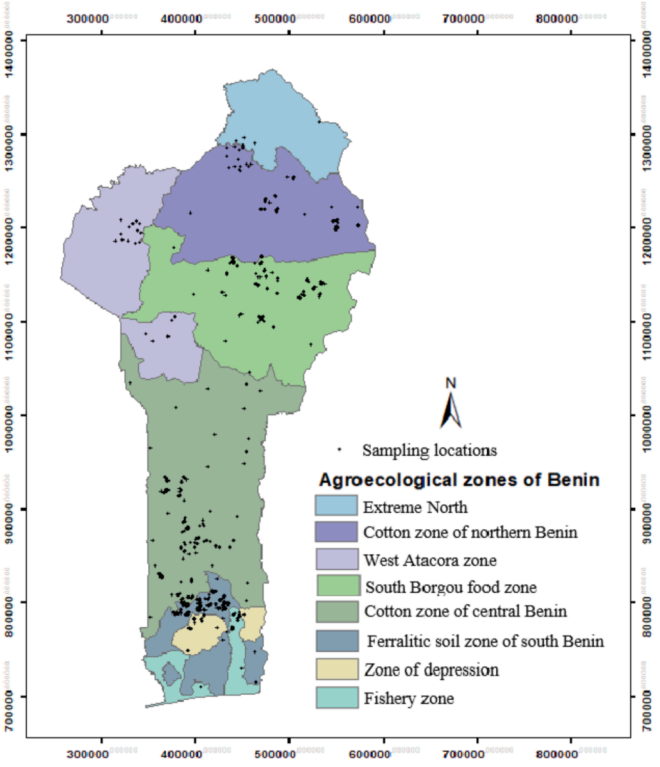

Fig. 1. Locations of the sampling sites.

cassava and yam (Dansi et al., 2012). Benin is divided into eight agro-ecological zones (Table 1) taking into consideration the different climatic zones, agro-pedological parameters and vegetation cover and cropping systems.

2.2. Soil data

Soil legacy data have been collected from different soil surveys in Benin over the past years from 2014 to 2017 (Azuka et al., 2020; Igué and Adjanohoun, 2014; Igué et al., 2018). The soil data were based on auger samplings and soil properties were considered for the topsoil (0–20 cm sample depth). The samplings were conducted in agricultural areas during different studies with random choice of targeted fields. Consequently, forestlands, savannah, and other lands were excluded from this study. Since the data were not initially collected for DSM, they present a cluster pattern, as shown by the locations of the sampling sites in Fig. 1. The results from the laboratory analysis were compiled from available reports, and the list of the available chemical and physical properties is presented in Table 2. The determination of organic matter was made according to Walkley and Black (1934), total nitrogen by the Kjeldahl method, pH (water) by using a glass electrode (1:2.5, soil/water ratio), exchangeable cation by using the method of ammonium acetate at pH 7, followed by reading at the atomic absorption spectrophotometer, the cation exchange capacity by the method of ammonium acetate 1 N pH 7 followed by a second extraction with KCl and assimilable Phosphorus by the method of Bray-1. For some of the data, soil organic carbon was obtained by dividing soil organic matter by 1.724.

2.3. Assessment of the classes of chemical fertility

The chemical fertility of the soils studied was defined based on some chemical fertility classes (Table 3). These classes were established according to Sys (1976). Although soil physical properties could affect chemical classes, they mostly represent a serious constraint in drainage situations that can impede the dynamics and absorption of nutrients. This phenomenon occurs in hydromorphic soils with poor aeration. The soils considered in Benin are well drained, and therefore it was hypothesized that physical fertility would not be a constraint on chemical fertility. The evaluation also took into account the shallow depth of the soils. Apart from these features, the classes of soil chemical fertilitywere considered the fertility levels of these soils. Five main classes were distinguished: Class I, very high level of fertility; Class II, high level of fertility; Class III, average level of fertility; Class IV, low level of fertility; and Class V, very low level of fertility (see Appendix A for full description).

Table 2
Descriptive statistics for the physical and chemical soil properties ($n = 1685$).

	min	max	mean	sd	Kurtosis	Skewness
Soil organic carbon (%)	0.00	5.53	0.88	0.49	12.11	2.34
Organic materials (%)	0.00	9.53	1.51	0.84	20.87	4.03
N (%)	0.00	1.46	0.08	0.07	215.86	12.49
pH	0.00	8.90	6.18	0.67	12.04	−1.20
Kech[1] (meq/100 g of soil)	0.00	99.00	0.34	3.41	231.13	12.15
Pass (Bray 1, ppm)	1.00	69.00	18.28	20.62	−0.03	1.3
Sum of bases (meq/100 g soil)	0.00	37.00	5.15	3.86	13.83	3.04
CEC (meq/100 g soil)	0.00	74.00	7.80	4.88	28.53	3.84
Base saturation[2] (%)*	0.00	100	62.85	18.80	−0.25	−0.16

Kech: exchangeable k, Pass: assimilable P, CEC: cation exchange capacity, min: minimum, max: maximum, sd: standard deviation.
[1] $n = 1250$.
[2] $n = 1625$.

2.4. Assessment of soil carbon stock

The determination of SOC stock (t C/ha/0–20 cm) requires data such as C content, bulk density and stone content. However, for most of the samples, bulk density and stone content were missing. Bulk density and stone content were downloaded from the ISRIC SoilGrids (Hengl et al., 2017a) portal as rasters. SoilGrids (250 m resolution) is a pixel-based map of the world built based on the spatial predictions of physical and chemical soil properties as well as soil types using ca. 150,000 soil profiles and over 158 covariates. The SOC stocks was finally computed as follows:

$$SOC_{stock} = SOC\ x\ BD\ x\ T\ x \left(1 - \frac{SC}{100}\right) \quad (1)$$

where SOC is the organic carbon content (%) of the fine earth (< 2 mm), BD is the bulk density of the fine earth oven dry (g cm^{-3}), T is the soil thickness (cm), and SC is the stone content (in volume %).

Table 1
Agro-ecological zones of Benin.

Agroecological zone	Climate	Soils	Vegetation	Main crops
Zone 1: Extreme North	Sudano-sahelian with a single rainy season (700–900 mm/year)	Ferruginous on crystalline basement, alluvial very fertile of Niger river	Shrubby savannah sparse thorn (*Acacia sieberiana*)	Millet, sorghum, beans, cotton
Zone 2: Cotton zone of northern Benin	Sudanese with a single rainy season (800–1.200 mm/year)	Tropical ferruginous on a crystalline	Shrubby trees with acacia Savannah	Cotton, yam, maize, millet
Zone 3: South Borgou food zone	Sudanese with a single rainy season (900–1300 mm/year)	Tropical ferruginous with variable characteristics	Savanna woodland shrub dominated by Butyrospermum (Shea)	Sorghum, cotton, yam, maize, cashew
Zone 4: West Atacora zone	Variation of sudano-sahelian to Sudano-guinean (1000–1300 mm/year)	Ferruginous often on deep base	Savanna wooded/shrub with Butyrospermum (Shea) and parkia (cheese monger), riparian forests	Bean, yam, groundnut, cassava
Zone 5: Cotton zone of central Benin	Sudano-guinean with two rainy seasons in the south and a rainy season in the North (1000–1200 mm/year)	Tropical Ferruginous on crystalline basement	Savannah wooded/ shrub dominated by *Danifiaohiori*	Maize, cowpeas, peanut, cotton, groundnuts
Zone 6: Ferralitic soil zone of southern Benin	Sudano-guinean with two rainy seasons in the west (600–1200 mm/year) and east (1000–1400 mm/year)	Ferrallitic, on sandy-clay sediment and Hydromorphous little humiferous with pseudo-gley, on alluvium	Dense shrubs	Maize, cassava, groundnuts
Zone 7: Zone of depression	Sudano-guinean with two rainy seasons in the west (800–1200 mm/year) and east (1000–1400 mm/year)	Vertisol with very wet soil, deep and clayey loamy soils	Dense semi-deciduous forest with large trees	Maize, cowpea, vegetable crops
Zone 8: Fishery zone	Sudano-guinean with two rainy seasons (1000–1200 mm/year)	Hydromorphous with little humification to pseudo-gley, on alluvium or Vertisol with a very humid soil surface	Grassy savannah, some mangroves,	Maize, cassava, cowpea and market gardening

Source: Laboratory of Soil Sciences, Water and Environment, (LSSEE), Center of Agricultural Research of Agonkanmey, INRAB.

Table 3
Evaluation criteria of the classes of chemical fertility.

Soil chemical properties	Fertility class				
	Class I	Class II	Class III	Class IV	Class V
	Very high (without limitation)	High (low limitation)	Average (moderate limitations)	Low (severe limitations)	Very low (very severe limitations)
Organic materials (%)	>2	2–1.5	1.5–1	1–0.5	<0.5
N total (%)	>0.08	0.08–0.06	0.06–0.045	0.045–0.03	<0.03
Pass (Bray 1, ppm)	>20	20–15	15–10	10–5	<5
Kech (meq/100 g of soil)	>0.4	0.4–0.3	0.3–0.2	0.2–0.1	<0.1
Sum of bases (meq/100 g soil)	>10	10–7.5	7.5–5	5–2	<2
Base saturation (%)	>60	60–50	50–30	30–15	<15
CEC (meq/100 g soil)	>25	25–15	15–10	10–5	<5
pH	5.5–6.5 6.5–7.8	5.5–6.0 6.5–7.8	5.3–5.5 7.8–8.3	5.2–5.3 8.3–8.5	<5.2 > 8.5

2.5. Explanatory covariates and feature engineering

Different covariates were considered surrogates for soil forming factors (Table 4). Topographic parameters (e.g., slope, curvature, aspect, topographic wetness index, length–slope factor) from elevation (30 m SRTM) were computed in the System for Automated Geoscientific Analyses geographic information system (SAGA GIS) following the standard approach as described by Conrad et al. (2015). Bioclimatic variables consisting of the average for the years 1970–2000 were downloaded from the WorldClim platform. Cloud-free Sentinel-2 images (Level-1C processing) were downloaded from the United States Geological Survey website (https://earthexplorer.usgs.gov/). The images were acquired from May 2016. Only six of the original spectral bands (red, blue, green, NIR, SWIR1, SWIR2) were considered along with four soil and vegetation indices that were calculated for each image. For soil types, the Benin soil map (Aholoukpè and Blavet, 2011) was considered. All the covariates were resampled to a 100 × 100 m pixel size grid.

The recursive feature elimination (RFE) implemented in the Caret package (Kuhn, 2017) was used to select the optimal set of covariates for each model. The RFE is carried out by first assessing variable importance and then iteratively eliminating the least important features (Gomes et al., 2019; Hounkpatin et al., 2018a). For each model, RFE was conducted with the full set of features; therefore, a model-specific optimal set of covariates was identified for each of the soil properties.

2.6. Modelling

2.6.1. Random Forest and quantile regression forests

Random forest (RF) is an ensemble learning algorithm that builds many decision trees and averages their predictions to either classify or estimate the value of a response variable (Breiman, 2001; Hounkpatin et al., 2018a).

RF has many features that justify its use in the present study. RF is robust to noise, parallelisable, and able to handle both categorical and numerical data without any assumption of probability distribution. RF also assesses the relative importance of the different features by permuting each covariate while keeping the remaining constant, and it computes how much the permutation reduces the accuracy by means of the OOB (data omitted from the bootstrapped samples) error. RF modelling requires tuning parameters such as the number of trees in the forest (ntree), the number of variables used to grow each tree (mtry) and the minimum number of terminal nodes (nodesize). The default value was used for the ntree (ntree = 500) to reduce the computational load, while the remaining values were set using the grid search method in the R "caret" package using tenfold cross validation with 3 repetitions. The quantile regression forest (QRF) is an extension of the RF and was established to further provide nonparametric estimates of the conditional quantiles for each node in a tree instead of returning just the mean value of the observations. The QRF was used to estimate the standard deviation related to the predictions. A more detailed description and theoretical considerations of the QRF model can be found in Meinshausen (2006). The RF and QRF models have also been used in many studies for predicting different soil properties for Africa (Flynn et al., 2019; Forkuor et al., 2017; Hengl et al., 2015).

2.6.2. Cubist model

The cubist (CB) model is a data mining technique that builds a set of regression trees and makes predictions based on linear regressions Holmes et al. (1999). Cubist partitions the training dataset by defining a set of rules, and for every feature that meets a specific rule, the associated linear model is established to predict the outcome. The neighbours and committees are to be tuned for optimal outcome, and this was carried out in the R "caret" package using tenfold cross validation with 3 repetitions. The CB model has been used in many studies for predicting different soil properties (Minasny and McBratney, 2008; Somarathna et al., 2016; Sulaeman et al., 2012; Zeraatpisheh et al., 2019).

2.7. Assessing the predictive performance of the models

The dataset of each of the soil properties was submitted to random sampling with 80% of the data used for training and 20% for independent validation. Randomly selecting different sets of training and validation sets will result in different accuracy metrics (Kuhn and Johnson, 2013). Therefore, this process was repeated 100 times to determine the distributions of the outcomes from the modelling. Consequently, both the QRF and CB models were trained 100 times with the different sets of the training dataset. Each of the models was independently validated with their corresponding 20% dataset. The model performance was assessed by comparing the averaged values of classical accuracy metrics such as R^2, Lin's concordance correlation coefficient (CCC), root mean square error (RMSE), and mean absolute error (MAE). The CCC measures both the accuracy and precision of the relationship between the observations and the predictions (Lawrence and Lin, 1989). The density plots between the actual and predicted soil properties were also analysed.

It is important to note that the classical metrics cannot provide information on the accuracy of prediction of the local uncertainty (Vaysse and Lagacherie, 2017). Therefore, the prediction interval coverage probability (PICP) was also considered, which is a measure of the total occurrences of an observed value within its related prediction interval (Malone et al., 2011; Solomatine and Shrestha, 2009; Vaysse and Lagacherie, 2017). For a given confidence level, it is expected that the same percentage of observations, equal to the related confidence level, is captured by the prediction intervals (PI). For example, one expects that 90% of the observations from the independent validation dataset will fall within its corresponding 90% PI (Kasraei et al., 2021). It is therefore the percentage of the true values (observed data) falling into a series of p-probability of (PI) limited by (1-p)/2 and (1 + p)/2 quantiles. The rule of thumb is that the lower the deviation from the accuracy line is, the

Table 4
Covariates used as soil forming factors.

Groups	Covariates	Abbr.	Res.	Description
Topography				
Morphometry	Elevation (m)	DEM	30 m	Vertical distance above sea level
	Slope (%)	Slope	30 m	Inclination of the land surface from the horizontal
	cos(Aspect)	cosAsp	30 m	North-south topographical orientation
	sin(Aspect)	sinAsp	30 m	East-west topographical orientation
	Plan curvature	CurPlan	30 m	Horizontal (contour) curvature
	Profile curvature	CurProf	30 m	Vertical rate of change of slope
	Convergence index	CI	30 m	Structure of the relief as a set of channels and ridges
	Terrain ruggedness index	TRI	30 m	Amount of elevation difference between adjacent cells
	Terrain surface convexity	TSconv	30 m	Spatial frequency of convex/concave location
	Terrain surface texture	TStex	30 m	Spatial frequency of peaks and pits
	Distance to stream network	NearDist	30 m	Distance to stream network
Hydrology	Saga wetness index	SWI	30 m	Ratio of local catchment area to slope
	Topographic wetness index	TWI	30 m	Ratio of local catchment area to slope
	LS factor	LSf	30 m	Effect of slope and slope-length on soil loss
Lightning, Visibility	Diffuse insolation	DifInso	30 m	Solar radiation that is scattered or reflected in the sky
	Direct insolation	DirInso	30 m	Solar radiation transmitted directly to the earth's surface
	Positive openness	Pos. Open	30 m	Degree of dominance of a landscape position
	Negative openness	Neg. Open	30 m	Degree of enclosure of a lanscape position
Bioclimatic variables	Annual Mean Temperature (°C)	Bio01	1 km	Annual Mean Temperature
	Mean Diurnal Range (Mean of monthly (max temp - min temp)) (°C)	Bio02	1 km	Mean Diurnal Range
	Isothermality (Bio_2/Bio_7) (* 100)	Bio03	1 km	Isothermality
	Temperature Seasonality (standard deviation *100) (°C)	Bio04	1 km	Temperature Seasonality
	Max Temperature of Warmest Month (°C)	Bio05	1 km	Max Temperature of Warmest Month
	Min Temperature of Coldest Month (°C)	Bio06	1 km	Min Temperature of Coldest Month
	Temperature Annual Range (Bio_5-Bio_6) (°C)	Bio07	1 km	Temperature Annual Range
	Mean Temperature of Wettest Quarter (°C)	Bio08	1 km	Mean Temperature of Wettest Quarter
	Mean Temperature of Driest Quarter (°C)	Bio09	1 km	Mean Temperature of Driest Quarter
	Mean Temperature of Warmest Quarter (°C)	Bio10	1 km	Mean Temperature of Warmest Quarter
	Mean Temperature of Coldest Quarter (°C)	Bio11	1 km	Mean Temperature of Coldest Quarter
	Annual Precipitation (mm)	Bio12	1 km	Annual Precipitation
	Precipitation of Wettest Month (mm)	Bio13	1 km	Precipitation of Wettest Month
	Precipitation of Driest Month (mm)	Bio14	1 km	Precipitation of Driest Month
	Precipitation Seasonality (Coefficient of Variation) (mm)	Bio15	1 km	Precipitation Seasonality (Coefficient of Variation)
	Precipitation of Wettest Quarter (mm)	Bio16	1 km	Precipitation of Wettest Quarter
	Precipitation of Driest Quarter (mm)	Bio17	1 km	Precipitation of Driest Quarter
	Precipitation of Warmest Quarter (mm)	Bio18	1 km	Precipitation of Warmest Quarter
	Precipitation of Coldest Quarter (mm)	Bio19	1 km	Precipitation of Coldest Quarter
Spectral bands and indices	Red (R)	Red	25 m	Red band of Sentinel data
	Green (G)	Green	25 m	Green band of Sentinel data
	Blue (B)	Blue	25 m	Blue band of Sentinel data
	Near infrared	NIR	25 m	Near infrared band of Sentinel data
	Shortwavelength infrared 1	SWIR 1	25 m	Shortwavelength infrared 1
	Shortwavelength infrared 2	SWIR 2	25 m	Shortwavelength infrared 2
	Hue index ((2 * R–G–B)/(G–B)	HI	25 m	Primary colours
	Brigtness index $((R^2+ G^2+ B^2)/3)^{0.5}$	BI	25 m	Average reflectance magnitude
	Colouration index (R–G)/(R + G)	ColInd	25 m	Soil colour
	Normalized Difference Vegetation Index (NIR–R)/(NIR + R)	NDVI	25 m	Health and amount of vegetation
Soil data	Soil types	TypeSoil	1:1 M	Spatial distribution of soil types

better the model (Malone et al., 2011). Consequently, for two models presenting comparable accuracy, such as R^2 and RMSE, the best model is the one having PICP values closer to the range of its corresponding confidence levels. The QRF and CB models were therefore evaluated not only on the basis of the classical metrics (R^2 and RMSE) but also on their ability to provide an accurate estimate of the associated uncertainty.

$$\boldsymbol{R^2} = \mathbf{1} - \frac{\sum_{i=1}^{n}(\boldsymbol{P_i} - \boldsymbol{O_i})^2}{\sum_{i=1}^{n}(\boldsymbol{O_i} - \boldsymbol{\mu_{obs}})^2} \tag{2}$$

$$\boldsymbol{\rho_c} = \frac{2\rho\sigma_{pred}\sigma_{obs}}{\sigma_{pred}^2 + \sigma_{obs}^2 + (\mu_{pred} - \mu_{obs})^2} \tag{3}$$

$$\boldsymbol{RMSE} = \left[\frac{\mathbf{1}}{\boldsymbol{n}}\sum_{i=1}^{n}(\boldsymbol{P_i} - \boldsymbol{O_i})^2\right]^{1/2} \tag{4}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i| \quad (5)$$

where *"P"* is the predicted value, *"O"* is the observed/true value, *"μ_{obs}" and "μ_{pred}" are* the means of the observed and predicted values, respectively, "σ_{obs}^2"and "σ_{pred}^2"are the associated variances, and ρ is the correlation between the observed and predicted values.

2.8. Uncertainties and mapping the soil fertility index

The best model selected based on the previously described accuracy metrics was used to create the final maps of the different soil properties. The uncertainty assessment builds on an estimate of the sensitivity of the model to available data as well as the uncertainty of the model, as suggested by Kempen et al. (2019). To compute the sensitivity of the model to available data, a random split of 20% repeated 50 times was conducted. At each split, a model was built and used to make predictions. The standard deviation of all the predictions at the pixel level represents the sensitivity of the model to the inherent variability of the available data.

The uncertainty of the model was represented by the full conditional distribution (Vaysse and Lagacherie, 2017) of each of the soil properties using the QRF. The sensitivity of the model to available data and the uncertainty of the model were summed to determine the overall uncertainty, as considered in Yigini et al. (2018). The total uncertainty was then expressed as a percentage by taking the ratio between the overall uncertainty and the mean prediction of the corresponding soil property. Finally, the soil fertility was established by pixelwise classifying the final maps of the soil chemical properties according to the descriptions and the indications in Table 3. The SFIm was evaluated based on expert knowledge on soil fertility in Benin and previous scientific publications.

3. Results

3.1. Modelling performance of the cubist and random forest models

The validation statistics of the CB and QRF models for the different soil properties are presented in Table 5. For most of the soil properties, CB exhibited slightly higher R^2 value than QRF. For Pass, OM, SOC, and SOC stock R^2 values were mostly above 50% for CB, with R^2 values of 71%, 52%, 0.53% and 50%, respectively. The corresponding R^2 values for the QRF models for Pass, OM, SOC, and SOC stock R^2 were 72%, 47%, 47% and 41%, respectively. For pH, CEC, and SumBas had R^2 values between 30% and 40% with the CB predictions, while the QRF presented values between 31% and 37%. Except for Pass, the CB also presented the lowest records for the observed RMSE from the validations. The results show a high CCC for assimilable Pass (83%/84%, CB/QRF), SOC (71%/67%, CB/QRF), OM (71%/68%, CB/QRF), SOC stock (71%/61%, CB/QRF), and CEC (60%/65%, CB/QRF), while exchangeable K (Kech) and N recorded the lowest values of 15% (CB, QRF) and 14%/16% (CB/QRF), respectively. The latter two soil properties also recorded the lowest R^2. The biases were low, with values less than or equal to 1 regardless of the models.

The density distribution between the actual and predicted soil properties is presented in Fig. 2 for the CB and QRF models. Both models generally present a similar trend for the density distributions, which show an underestimation of the lower and higher values, while values centred around the mean were overestimated. Fig. 3 shows the empirical coverage frequencies at the corresponding confidence levels. The PICP plots show that the empirical coverage frequencies were more comparable to the corresponding nominal coverage probabilities for the QRF than for the CB. Between the 90 and 25% confidence levels, it can be seen with the CB that the observed proportions within the interval demonstrate a pattern of increasing deviation from the expected result for all the soil properties. In all these cases, the CB overestimated uncertainties regardless of the soil properties. However, the QRF also displayed some overestimations of uncertainty for the N, Pass, CEC, SumBass and BS but at a lower magnitude than the CB.

3.2. Variable importance

Fig. 4 presents the variable importance of the different soil properties ranked by the CB and QRF models. Here, only the top 5 variables for each of the soil properties were presented, although some have even lower counts of variables after recursive feature elimination. For both the CB and QRF models, the nearest distance to the stream network (NearDist) was the most influential variable in most cases for the different soil properties, followed by the bioclimatic variables. NearDist was associated with a negative correlation with all the soil properties except for assimilable P (Pass). This suggests that higher values of nutrients were located closer to streams compared to areas that are far away from water bodies. The NearDist actually indirectly implies the influence of topography, which regulates water and sediment redistribution.

Table 5
Performance of the cubist and quantile random forest models for predicting soil properties.

		R^2	CCC	RMSE	bias	MAE
Cubist	pH	0.395 (±0.022)	0.572 (±0.018)	0.474 (±0.009)	0.005 (±0.013)	0.331 (±0.006)
	OM (%)	0.527 (±0.040)	0.719 (±0.023)	0.548 (±0.034)	−0.015 (±0.028)	0.312 (±0.014)
	SOC (%)	0.532 (±0.041)	0.718 (±0.025)	0.318 (±0.022)	−0.001 (±0.014)	0.181 (±0.008)
	SOC stock (t C/ha)	0.507 (±0.037)	0.696 (±0.023)	8.990 (±0.513)	0.058 (±0.417)	5.135 (±0.225)
	N (%)	0.092 (±0.014)	0.168 (±0.012)	0.083 (±0.001)	−0.007 (±0.001)	0.024 (±0.001)
	Pass (Bray1)	0.710 (±0.037) (	0.834 (±0.019)	10.588 (±0.679)	– 0.490 (±0.435)	6.18 (±0.34)
	Kech (meq/100 g soil)	0.053 (±0.028)	0.158 (±0.039)	0.238 (±0.022)	−0.019 (±0.007)	0.089 (±0.005)
	CEC (meq/100 g soil)	0.386 (±0.046)	0.607 (±0.035)	3.597 (±0.205)	−0.059 (±0.105)	1.966 (±0.066)
	SumBas (meq/100 g soil)	0.297 (±0.049)	0.506 (±0.045)	3.257 (±0.145)	−0.004 (±0.101)	1.747 (±0.065)
	BS (%)	0.339 (±0.029)	0.520 (±0.022)	14.650 (±0.353)	−0.528 (±0.561)	10.745 (±0.256)
Quantile Random Forest	pH	0.371 (± 0.023)	0.582 (± 0.018)	0.490 (± 0.011)	0.008 (± 0.014)	0.347 (± 0.008)
	OM (%)	0.479 (± 0.035)	0.681 (± 0.026)	0.584 (± 0.025)	0.054 (± 0.023)	0.359 (± 0.012)
	SOC (%)	0.470 (± 0.032)	0.675 (± 0.022)	0.343 (± 0.015)	0.033 (± 0.013)	0.208 (± 0.006)
	SOC stock (t C/ha)	0.410 (± 0.024)	0.610 (± 0.019)	9.740 (± 0.241)	1.020 (± 0.317)	6.570 (± 0.169)
	N (%)	0.085 (± 0.016)	0.143 (± 0.015)	0.083 (± 0.001)	−0.006 (± 0.001)	0.025 (± 0.000)
	Pass (Bray1)	0.729 (± 0.023)	0.843 (± 0.013)	10.239 (± 0.447)	−0.642 (± 0.330)	6.044 (± 0.249)
	Kech (meq/100 g soil)	0.051 (± 0.012)	0.152 (± 0.012)	0.230 (± 0.004)	0.003 (± 0.007)	0.095 (± 0.004)
	CEC (meq/100 g soil)	0.350 (± 0.030)	0.550 (± 0.026)	3.610 (± 0.105)	−0.150 (± 0.103)	1.980 (± 0.053)
	SumBas (meq/100 g soil)	0.310 (± 0.043)	0.550 (± 0.038)	2.930 (± 0.153)	0.270 (± 0.102)	1.760 (± 0.066)
	BS (%)	0.285 (± 0.029)	0.505 (± 0.026)	15.599 (± 2.250)	−0.815 (± 0.523)	11.212 (± 0.283)

OM: organic materials, SOC: soil organic carbon, Kech: exchangeable K, Pass: assimilable P, CEC: cation exchange capacity, SumBas: sum of bases, CCC: Lin's concordance correlation coefficient, MAE: mean absolute error, RMSE: root mean square error.

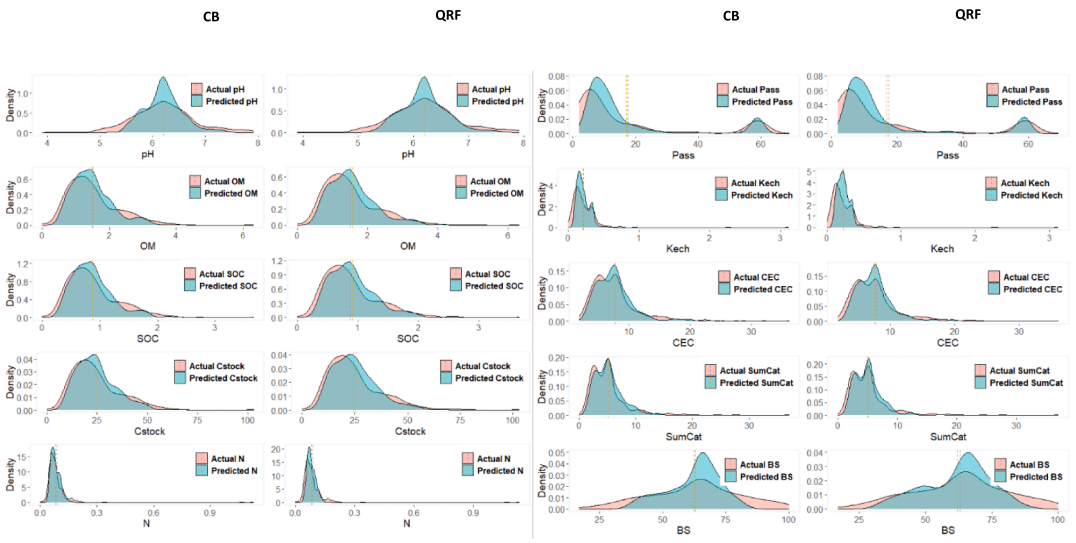

Fig. 2. Density plots between actual and predicted soil properties from the cubist and random forest models. CB: cubist model, QRF: random forest model, OM: organic matter, SOC: soil organic carbon, N: nitrogen, Pass: assimilable P, Kech: exchangeable K, CEC: cation exchange capacity, SumBas: sum of bases, BS: base saturation.

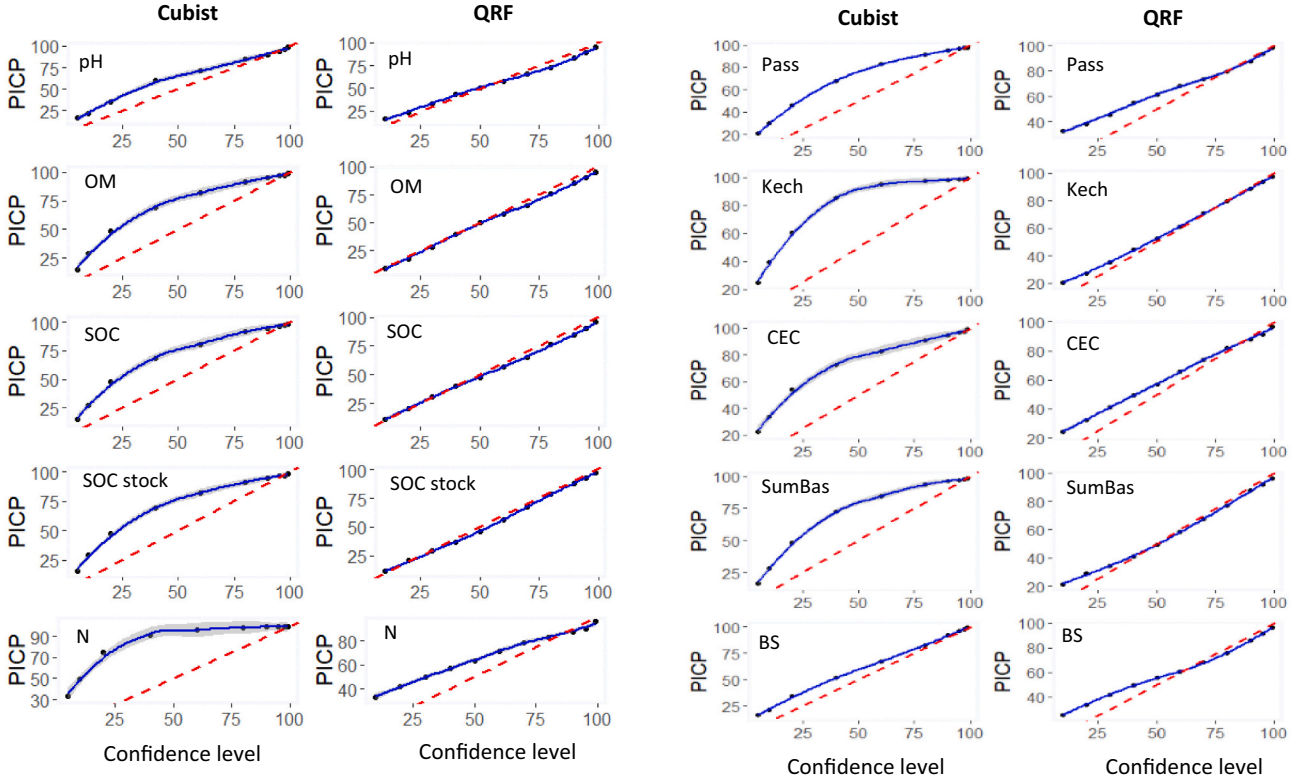

Fig. 3. Plot of the prediction interval coverage probability (PICP) and confidence level based on validation of the cubist model and quantile regression forest. OM: organic matter, SOC: soil organic carbon, N: nitrogen, Pass: assimilable P, Kech: exchangeable K, CEC: cation exchange capacity, SumBas: sum of bases, BS: base saturation.

The bioclimatic covariates ranked next to the NearDist, taking pre-eminence over the remaining covariates. For these variables, the QRF model reported more precipitation-based indices (PBI) (17 times) than temperature-based indices (TBI) (6 times), while the opposite was noted for the CB models (19 times PBI versus 14 TBI). The top ranked variables among the PBIs are the precipitation of the warmest quarter (PWQ, Bio_18), precipitation of the wettest month (Bio_13), and precipitation of the coldest quarter (Bio_19), while the mean temperature of the coldest quarter (mTCQ, Bio_11) and mean temperature of the driest quarter (Bio_09) are the most reported among the TBIs. A significant negative correlation was also observed between the PWQ and the different soil properties.

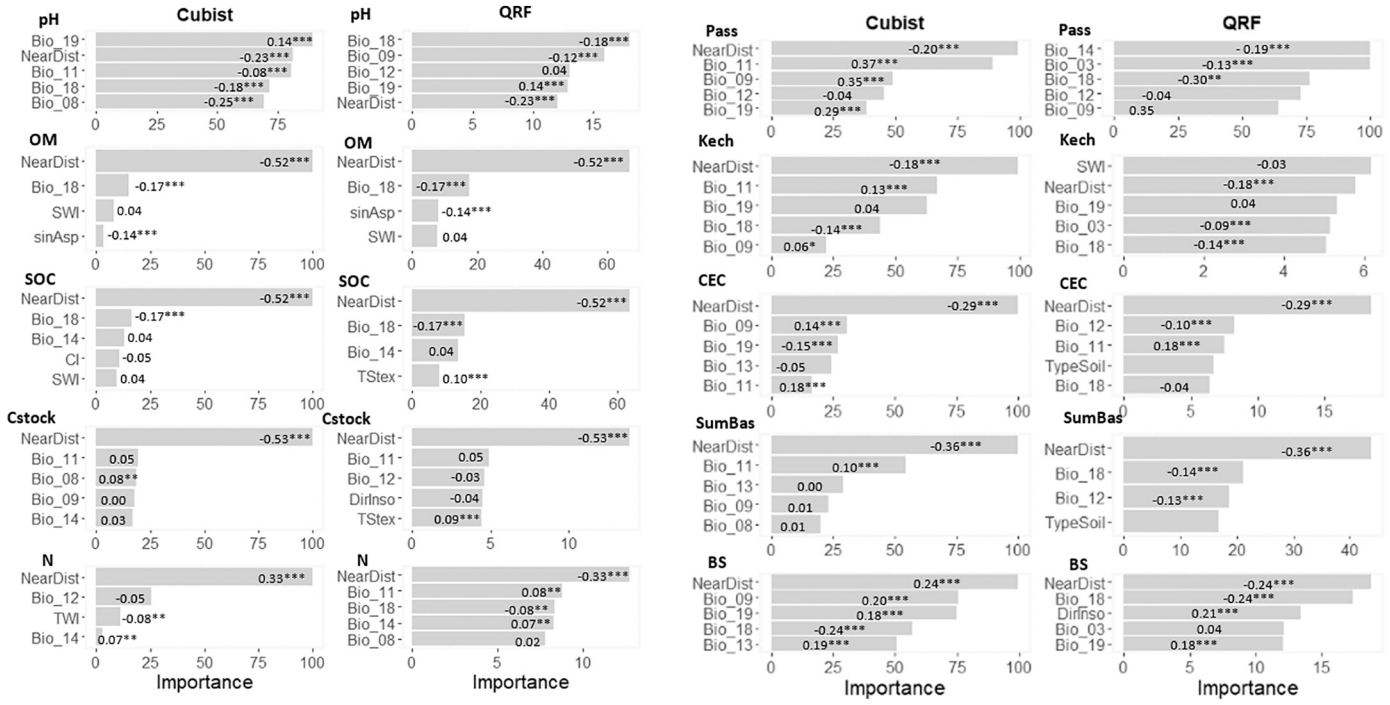

Fig. 4. Variable importance of the cubist and quantile regression forest. Models. OM: organic matter, SOC: soil organic carbon, N: nitrogen, Pass: assimilable P, Kech: exchangeable K, CEC: cation exchange capacity, SumBas: sum of base, BS: base saturation.

3.3. Spatial distribution of the soil properties

The QRF model was used to map each soil property across the study area, and the associated uncertainty (%) was computed (Fig. 5). The QRF was preferred over the CB model because of its better PICP.

Soil pH varied from slightly acidic (4.6–6) in southern and northwestern Benin to weakly acidic (6–7) in central to northern Benin along with some alkaline spots (> 7). The spatial patterns of OM, SOC and SOC stocks were similar across the study area. Obviously, OM and SOC were linked by the multiplicative factor (1.724), while the SOC stock had a relatively weak correlation ($r = 0.31$) with both properties (SI 1.). The southern and lower central parts of Benin have higher spot values of OM, SOC and SOC stocks in the landscape.

Southern Benin was characterized by low nitrogent content in general (< 0.15%), low (< 20 ppm) assimilable P (Pass) and some low spots of exchangeable potassium (< 0.2–0.3 meq/100 g of soil). Central benin generally presents even lower nitrotren content (< 0.1%), moderate spots of Pass values (20–30 ppm) and low Kech values (< 0.23 meq/100 g of soil). Finally, Northern Benin displays generally also low nitrogen content (< 0.15%), although slightly higher values were located along the Niger bassin (0.15–0.25%), moderate Pass (< 10–20 ppm) level along with low Kech values (< 0.23 meq/100 g of soil). However, for most of these soil properties, slightly higher values can be found in some locations of extreme northern and western northern Benin.

The distribution of CEC shows low potential of available nutrient supply, with the major parts of Benin presenting lower values (< 10 meq/100 g of soil), although Southern Benin displays relatively higher spot values (10–15 meq/100 g of soil). A similar trend was observed in general for the sum of bases (SumBas) with values <5 meq/100 g all over Benin soils, while relatively higher spots (5.5–10 meq/100 g of soil) are observed in the southern and upper parts of Northern Benin. The base saturation map indicated that bases form the major component of the CEC by occupying approximately 50 to 70% of the exchange sites of the soil's adsorption complex in most soils. This was consistent with the pH distribution, which showed no trend of strongly acidic soils.

The uncertainties related to the maps of the different soil properties showed reasonable levels of accuracy, with most varying between 40% and 50% for the values of the ratio between the sum of the model and data uncertainty over the mean maps of the respective soil properties. It appears that spatial uncertainties were higher in locations where sampling data were not available (Fig. 1), such as the extreme North, lower Southern and northwestern areas of Benin. It was obvious therefore that as more sampling data become available in these areas, the predictions will gain spatial accuracy.

3.4. Soil fertility index map

Fig. 6 presents the SFIm of Benin based on the pixelwise distribution of the soil properties in relation to the rules indicated in Table 3. It appears that most of the agricultural lands of Benin belong to the low fertility class, suggesting that these soils present severe limitations for production. In fact, based on the average values of the soil properties, half of the agroecological zones (zones 2–6) of Benin (Table 6) have soils with low fertility (Class IV) characterized by severe limitations. The limitations in the low fertility class areas refer to situations that cause yield reduction or the implementation of cropping techniques that adversely affect the profitability of the production system. Agroecological zone 1 (Class III on average) and zones 6–7 (Classes II and III on average) presented high and average levels of fertility, as displayed by the SFIm (Fig. 6, Table 6). They also contain, on average, relatively higher soil carbon stocks than the remainder.

4. Discussion

This section mainly focuses on discussing the performance of the QRF and CB models in relation to the classical error metrics and PICP, the ranking in relation to variable importance, the spatial distribution of the soil properties and the resulting SFIm. Finally, the limitations and implications of the study are also presented.

4.1. Performance of the QRF and CB models

The findings of this study showed that the prediction of the different soil properties results in various ranges of classical error metrics. The CB

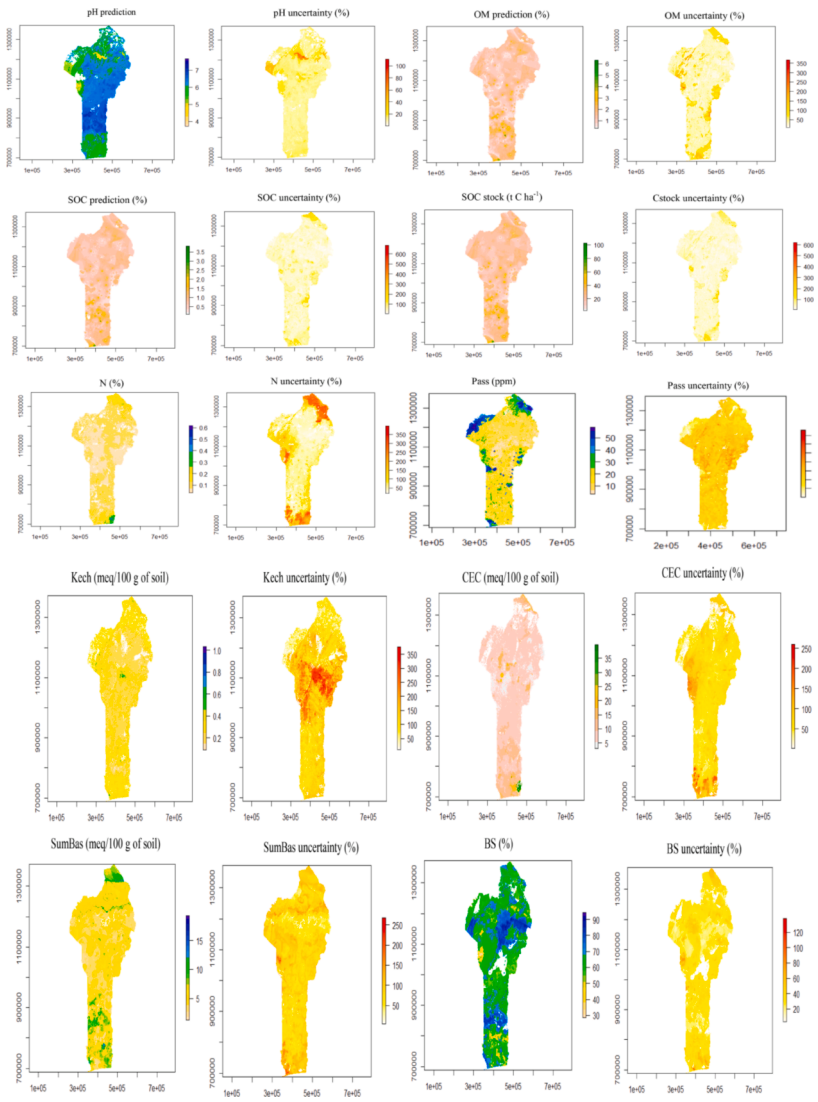

Fig. 5. Prediction and uncertainty maps of the soil properties by QRF. OM: organic materials, SOC: soil organic carbon, Kech: exchangeable K, Pass: assimilable P, CEC: cation exchange capacity, SumBas: sum of base, BS: base saturation.

models presented generally higher R^2 and lower RMSE values than the QRF models. Compared to other studies, the results vary based on the models used, the predicted soil properties and the location of the studies. For example, Hengl et al. (2017b) recorded a higher R2 for N (66%) and extractable K (64%) for the topsoil (0–30 cm) of Sub-Saharan Africa while recording a lower R^2 for extractable P (12%) with the QRF model. Compared to the present study, mapping macronutrients in agricultural (depth 0–15 cm) lands in Iran with the cubist model also resulted in a lower validation R^2 of 13% (CCC = 24%) for N but in a higher, although relatively still low, R^2 of 17% (CCC = 24%) for P (Shahbazi et al., 2019). Dharumarajan et al. (2017) obtained much lower R^2 values of 0.23 (CCC = 0.38) and 0.30 (CCC = 0.37) for SOC and pH, respectively, with the QRF in the semiarid tropics of South India at 30 cm depth. The cubist model yielded an R^2 of 66% after validation of CEC for the top 30 cm using a large regional dataset over Europe (Padarian et al., 2019). The QRF validation R^2 for the SOC stock for the present study is two times higher than in the finding of Hounkpatin et al. (2018b) for the topsoil (0–30 cm) of the Dano catchment in Burkina Faso. Since CB and QRF are data driven, disparities between current findings and previous studies might be attributed to the interplay of inherent climate, soil type and local agricultural practices.

Based on the PICP plots (Fig. 3), the empirical coverage frequencies from the QRF were relatively more comparable to the corresponding PICP compared to the CB. The overestimation of the local uncertainty observed for CB was higher than the pattern observed with the QRF. This trend came in contrast to the fact that the CB generally presented a slightly higher R^2 and lower RMSE than the QRF. The RMSE and R^2 are used as goodness of fit criteria to evaluate the quality of the predictions, while the PICP is considered for the determination of the efficacy of the uncertainty estimates. The trend recorded in this study aligns with the results from Vaysse and Lagacherie (2017), who used regression kriging (RK) and QRF for mapping pH, SOC and clay. In their modelling, although the QRF recorded slightly lower accuracy (high RMSE, lower R^2) in predicting pH, clay and SOC compared to RK, its performance showed a higher accuracy regarding the uncertainty predictions. Likewise, Szatmári and Pásztor (2019) reported in their study that regression kriging produced contradictory results by giving on the one hand the best spatial predictions according to some classical error metrics (RMSE, mean error) while on the other hand underestimating the uncertainty according to the related accuracy plot. The authors attributed this trend to the kriging variance reflecting only the position of unsampled locations in geographical space without taking into account their position in feature space. The conclusion from the preceding reports is that classical error metrics could be misleading and accuracy plots can help visualize the effectiveness of the prediction of local uncertainties. To our knowledge, no study has compared the accuracy based on PICP between

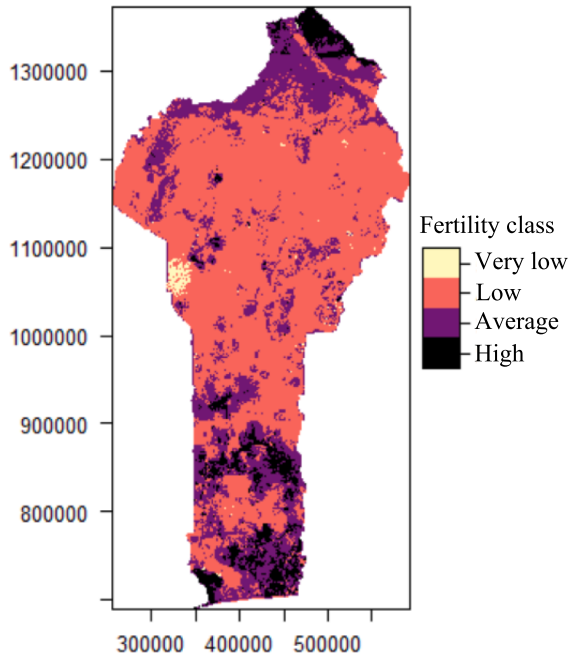

Fig. 6. Fertility class distribution.

QRF and CB. Although the reasons for this current trend require further investigation, it might be related to the fact that unlike QRF, CB uses a linear fitting after partitioning the data with a given condition.

Both the CB and QRF models underestimated the lower and higher values of the different soil properties, while values centred around the mean were overestimated. This confirms earlier findings that machine learning models such as cubist and random forest tend to underestimate high values and overestimate low values (Čeh et al., 2018; Horning, 2010; Hu et al., 2020; Ma et al., 2017). Although this typical behaviour is due to predictions resulting from average values for both models, it should also be noted that lower and higher values are underrepresented with most of the values centred around the means in the training set.

4.2. *Covariates influencing the spatial distribution of the soil properties*

The soil properties were found to be primarily influenced by the nearest distance to the stream network followed by the bioclimatic covariates (Fig. 4). The preeminence of bioclimatic variables alongside the NearDist is an indicator of the interaction between climate variables and topography leading to hydrology and toposequential sequences of soil properties at agricultural sites. Under tropical rains, water and sediments tend to move and accumulate in lower slope areas in reaction to landscape gradients in gravitational potential energy. These topography-driven erosional processes lead to the transport of elements from upper slope areas, which are associated with lower nutrient concentrations, to lower slope areas, which are generally characterized by higher levels of nutrients (Kumhálová et al., 2011; Wubie and Assen, 2020). These low slopes are areas of deeper soils with high clay content. This is in line with existing studies that also found higher organic matter (OM), SOC, SOC stock, and soil nutrients closer to streams (Abate and Kibret, 2016; Cincotta et al., 2019; Holleran et al., 2015).

The key covariates among the bioclimatic variables (Fig. 4) were precipitation of warmest quarter (PWQ, Bio_18), precipitation of wettest month (Bio_13), precipitation of coldest quarter (Bio_19) while the mean temperature of coldest quarter (mTCQ, Bio_11) and mean temperature of driest quarter (Bio_09). These indices translate to either water availability or water stress conditions. Northern Benin, with its semiarid climate, will most likely suffer from water-stress conditions compared to the subhumid climate of southern Benin. With climate change leading to increased temperature, the water supply to crops acts as a limiting factor of crop growth and yield (Challinor et al., 2007). Climatic variables such as annual temperature average, range and seasonality and annual precipitation and seasonality values were the most important variables for predicting soil chemical attributes (pH, CEC, base saturation, Al) in a Brazilian catchment (Poppiel et al., 2019). A study carried out in Kenya recorded the annual mean temperature, annual precipitation and the mean temperature of the wettest quarter as the key factors that affect the distribution of maize (Kogo et al., 2019). Temperature affects the germination, growth and development of thermophilic C4 crops such as maize, millet, and sorghum, which are cultivated in the study area.

Since the PWQ index provides the total precipitation during the warmest three months of the year water, it might act as a limiting factor for plant growth and yield because soil water dynamics play a crucial role in the spatial distribution of SOC and soil nutrients. The significant negative correlation of the PWQ with the different soil properties might be related to a higher level of decomposition under the high temperature of the warmest month and subsequent increase in SOC and nutrient elements in the soil. High temperature in the warmest quarter will translate into higher evapotranspiration in combination with a reduction in erosional processes due to the lack of water. Consequently, SOC, SOC stock and soil nutrients might tend to relatively accumulate on site under a situation of low water availability.

4.3. *Spatial distribution of the soil properties and resulting soil fertility index map*

Considering the spatial distribution of the soil properties (Fig. 5), pH was mostly acidic in Benin, and this trend is in line with some studies carried out in Benin (Amonmide et al., 2019; Igue et al., 2013) and West Africa (Hien et al., 2006; Saiz et al., 2012). High values of OM, SOC and SOC stocks were found in the southern and lower central parts of Benin (Fig. 5). These areas correspond to the Oueme River Basin (ORB), which is subject to higher rainfall and more intensive agricultural production than northern Benin, which is drier with lower agricultural intensity (Sonneveld et al., 2012). Additionally, the ORB also benefits from dense stream networks, which suggest that processes of deposition of sediments in lower and flat areas are a major part of soil carbon distribution.

The results also showed that most of the agricultural lands of Benin belong to the low fertility class, presenting severe limitations for cropping (Fig. 7). The main characteristics of soils in Benin, as in many subsaharian countries, are their low organic matter content and their high mineralization rate. Most often than not, farmers burn crop residues in Benin, while high temperature leads to a faster decomposition of organic matter, resulting in poor croplands (Kouelo et al., 2016). Intensive and unsuitable use of tillage equipment at the farm level associated with the burning or removal of crop residues and the reduction of fallow periods result in the decline of soil organic matter content and the destruction of their structure (Dossou-Yovo et al., 2016; Igué et al., 2018). Approximately 60% of the tropical ferruginous soils that made up the total land area of Benin (Table 1) are affected by this phenomenon (Igué et al., 2018), and further degradation could occur in the absence of sustainable management practices.

Studies conducted in southern, central and northern Benin reported suboptimal distributions of the N, P, and K elements and CEC for crop production (Igue et al., 2013; Igue et al., 2016; Saïdou et al., 2018). The high decay rate of organic matter results in soils with poor nitrogen, which becomes a limiting factor in most agricultural soils. Even some soils located in converging locations (e.g., hydromorphic soils) that receive materials from relatively higher elevation areas are intensively cultivated to the extent of sometimes having even lower levels of fertility compared to the surroundings (Hounkpatin et al., 2018b; Igue et al., 2016). For phosphorus, the type and nature of the dominant clay, which is mostly kaolinite, explains its low level in both Benin and many countries in subsaharian Africa (Koné et al., 2010; Saïdou et al., 2018). This makes an additional supply of these macronutrients as fertilizers

Table 6
Descriptive statistics of the mapping values of the soil properties and associated fertility class for each agroecological zone.

		Zone* 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zone 7	Zone 8
pH	min	5.50	3.72	3.43	4.62	4.74	4.78	5.12	4.74
	max	6.59	7.11	7.37	7.12	7.72	7.22	6.97	7.00
	mean	6.11	6.05	6.33	5.88	6.47	5.93	5.93	6.02
	sd	0.14	0.42	0.17	0.43	0.34	0.21	0.23	0.16
OM (%)	min	0.37	0.36	0.33	0.36	0.24	0.42	0.43	0.45
	max	3.35	3.51	4.17	3.47	7.00	6.75	5.21	4.58
	mean	1.43	1.23	1.24	1.21	1.45	1.53	1.74	1.53
	sd	0.41	0.36	0.37	0.40	0.64	0.63	0.68	0.34
SOC (%)	min	0.18	0.14	0.14	0.13	0.09	0.09	0.09	0.14
	max	1.87	2.18	2.34	2.00	4.09	3.99	2.99	2.92
	mean	0.81	0.70	0.71	0.67	0.83	0.88	1.03	0.90
	sd	0.24	0.20	0.21	0.20	0.38	0.39	0.46	0.20
SOC stock (t C/ha)	min	3.94	4.07	3.83	3.83	2.42	2.47	2.53	3.91
	max	54.66	65.41	61.49	61.70	112.95	109.16	81.40	80.37
	mean	23.40	19.79	20.04	18.84	23.32	24.29	28.35	24.62
	sd	7.25	5.89	6.28	6.05	10.66	10.63	12.78	5.46
N (%)	min	0.06	0.03	0.04	0.04	0.03	0.04	0.03	0.05
	max	0.42	0.20	0.15	0.17	0.28	0.60	0.83	0.27
	mean	0.13	0.08	0.07	0.08	0.08	0.13	0.14	0.10
	sd	0.04	0.02	0.01	0.02	0.02	0.08	0.08	0.02
Pass (Bray1)	min	7.74	3.40	3.01	3.94	3.69	3.42	4.99	3.96
	max	57.36	48.09	58.35	59.00	58.78	58.42	56.52	56.52
	mean	31.56	10.65	10.66	25.20	16.69	14.42	19.26	14.32
	sd	12.40	5.12	6.96	17.73	11.19	9.19	13.17	6.13
Kech (meq/100 g soil)	min	0.17	0.11	0.09	0.12	0.09	0.07	0.07	0.10
	max	0.73	1.86	2.40	0.64	1.05	0.89	1.29	0.63
	mean	0.33	0.23	0.23	0.27	0.24	0.27	0.31	0.28
	sd	0.04	0.04	0.06	0.06	0.04	0.06	0.06	0.05
CEC (meq/100 g soil)	min	5.77	3.80	2.93	2.99	2.81	3.09	4.99	5.50
	max	22.39	17.62	23.61	25.68	23.92	38.66	45.50	34.31
	mean	10.37	6.85	6.94	7.04	7.25	12.63	12.57	10.13
	sd	2.87	1.40	3.05	2.79	2.05	7.30	4.40	2.48
SumBas (meq/100 g soil)	min	3.97	1.76	1.40	1.32	1.39	1.42	1.58	2.08
	max	15.21	19.56	17.57	12.37	20.82	16.68	17.74	15.51
	mean	7.01	4.78	4.51	4.47	5.29	5.66	6.45	6.79
	sd	1.96	1.30	1.04	1.13	1.85	2.05	2.14	1.85
BS (%)	min	45.67	32.95	32.03	28.98	28.69	26.74	28.77	30.96
	max	85.30	93.12	94.43	89.96	91.27	91.38	83.15	93.58
	mean	63.77	66.34	69.49	64.17	64.43	60.94	64.27	58.02
	sd	5.05	8.42	8.16	8.06	8.78	9.55	9.38	6.64
Zonal fertility class for OM		III	III	III	III	III	II	II	II
Zonal fertility class for N		I	II	II	II	II	I	I	I
Zonal fertility class for Pass		I	III	III	I	II	III	II	III
Zonal fertility class for Kech		III	IV	IV	III	IV	III	III	III
Zonal fertility class for SumBas		III	IV	IV	IV	III	III	III	III
Zonal fertility class for BS		I	I	I	I	I	I	I	II
Zonal fertility class for CEC		III	IV	IV	IV	IV	III	III	III
Zonal fertility class for pH		I	I	I	I	I	I	I	I
Overall zonal fertility class		III	IV	IV	IV	IV	IV	II	III

OM: organic materials, SOC: soil organic carbon, Kech: exchangeable K, Pass: assimilable P, CEC: cation exchange capacity, SumBas: sum of bases min: minimum, max: maximum, sd: standard deviation, *: see definition of zone in Table 1.

necessary for a higher productivity of the croplands in Benin.

Agroecological zone 1 (Class III on average) and zones 6–7 (Class II on average) were found to have high and average levels of fertility (Fig. 6, Table 6). These areas are generally located at low elevations and are characterized by alluvial plains along with hydromorphic soils in valleys, bas-fond and vertisols (Table 1). These soils present fewer limitations regarding organic matter content and nitrogen (Igue et al., 2013). However, in the context of increasing population rates and climate change, intensified agriculture in wetlands such as inland valleys is most likely to further fragilize wetland ecosystems with subsequent consequences on greenhouse Gaz emissions in the absence of proper monitoring and sustainable means of production.

4.4. Limitation and implication of the study

Obviously, the uncertainties related to the prediction of the soil properties will translate into the resulting SFIm, especially for areas of low sampling coverage. Although creating a full error budget was beyond the scope of the present study, the uncertainty maps covered to some extent the sensitivity of the model to available data along with the uncertainty of the model (Yigini et al., 2018). For some of the soil properties, such as N and Kech, model accuracy was limited while being reasonable for the remainder, although the predictions can still be improved. The limitation in accuracy can be related to the low correlation between the covariates and the soil properties (SI 1), differences in environmental conditions between the training and validation datasets, and inherent measuring errors related to the training and validation sets (Bonfatti et al., 2016; Nelson et al., 2011).

The clustering pattern in the distribution of the spatial samplings prevents full coverage of the feature space, resulting in higher prediction uncertainty for locations that have never been learned during the modelling stage by the models. Consequently, a future sampling design that covers the full covariate feature space is necessary to improve the prediction accuracy. A multiscale approach can also be considered, as the spatial distribution of many soil properties is subject to factors operating at different spatial and temporal scales (Behrens et al., 2014;

Miller et al., 2015b; Wilson et al., 2017). In addition, the SOC stocks were computed based on the bulk density and stone content maps of the soil grid platform. Consequently, the current values reported in this study should be updated with more local data. There is therefore a local need to systematically take into account density and stone content in further sampling or develop pedo transfer functions in that regard.

The results of this study are a representation of the current situation based on the available soil data and remote sensing variables. In that regard, they are valuable as baselines for studies focusing on potential yield estimation, crop growth simulation modelling, long-term dynamics of soil nutrients and SOC stocks. In addition, only toposil (0–20 cm) data were considered for establishing the SFIm, limiting its use for fertile management practices, which cannot be advised on the basis of only surface data. Consequently, more data was available, and one should consider the whole soil profile. Additionally, as boclimatic variables were critical in predicting the different soil properties, future studies could investigate the impact of different climate scenarios on soil fertility to assess the magnitude and direction of change in future crop productivity at the end of the century.

5. Conclusion

The main outcomes of this research were:

1. It is feasible to produce a spatial soil fertility index map by considering the combination of the values of different soil properties, such as soil organic carbon matter, nitrogen, pH (water), exchangeable potassium, assimilable phosphorus, sum of bases, cation exchange capacity and base saturation. This can be done by using a DSM approach that requires georeferenced soil data, different covariates (soil type, topographic, bioclimatic, spectral data, etc.) and models such as cubist and quantile random forest.
2. Evaluating a model prediction accuracy based only on classical error metrics (RMSE, R^2) might be misleading, as contrasting results can appear when compared to their corresponding prediction interval coverage probability plot. Although the cubist models generally recorded the lowest RMSE and highest R^2, the local uncertainties of the different soil properties were overestimated. However, there is a need for further research to investigate why better performance based on classical error metrics does not necessarily align with the prediction of local uncertainty.
3. The identification of the key covariates influencing the spatial distribution of soil nutrients reavealed that the nearest distance to the stream network has strong predictive ability for all the soil properties along with the bioclimatic variables.
4. The predicted maps of the soil properties showed high spatial variability with high uncertainties in areas of low sampling density. The resulting SFIm revealed that most of the soils in Benin have low fertility. In addition, a limited number of soils with high and average fertility levels are located in the Oueme River Bassin, and policy could advocate for their use in agriculture and promote sustainable management practices.

Declaration of Competing Interest

None.

Acknowledgement

The authors express their gratitude to the Hydro-Climate Services (www.hcs-in.org) for initiating the stuty and its support for data treatments and analysis. We also thank all the reviewers and the editor for their valuable comments on the paper.

Appendix A

Description of the five chemical fertility classes considered in the study:

(1) Class I, very high level of fertility: soils presenting no limitation, optimal condition for crop growth and yield.
(2) Class II, high level of fertility: soils presenting slight limitations, referring to situations that could slightly reduce yields without, however, demanding special cultivation techniques. Soils are of class II fertility when the characteristics have no more than 3 moderate limitations, possibly associated with low limitations.
(3) Class III, average level of fertility: soils presenting moderate limitations, referring to situations that cause a greater decrease in yields or the implementation of special cultural techniques. These limitations do not adversely affect the profitability of the production system. They are characterized by more than 3 moderate limitations associated with only one severe limitation.
(4) Class IV, low level of fertility: soils presenting severe limitations, referring to situations that cause yield reduction or the implementation of cropping techniques that adversely affect the profitability of the production system. Soils belonging to this class present more than one indicator of severe limitations.
(5) Class V, very low level of fertility: soils presenting very severe limitations, referring to situations that do not allow the use of the land for production. This class not only has severe limitations but is also associated with more than one indicator of very severe limitations.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geodrs.2021.e00444.

References

Abate, N., Kibret, K., 2016. Effects of land use, soil depth and topography on soil physicochemical properties along the toposequence at the Wadla Delanta Massif. Northcentral Highlands Ethiopia Environ. Pollut. 5 (2).
Adhikari, K., Hartemink, A.E., Minasny, B., Kheir, R.B., Greve, M.B., Greve, M.H., 2014. Digital mapping of soil organic carbon contents and stocks in Denmark. PLoS One 9 (8), e105519.
Aholoukpè, H., Blavet, D., 2011. Géoréférencement dans un système d'information géographique de cartes pédologiques au 1/200 000 élaborées par les pédologues de l'ORSTOM, de l'IRHO et du Bénin-Feuilles 1/200 0000 Porto Novo et Abomey. République du Bénin.
Amonmide, I., Dagbenonbakin, G., Agbangba, C.E., Akponikpe, P., 2019. Contribution à l'évaluation du niveau de fertilité des sols dans les systèmes de culture à base du coton au Bénin. Int. J. Biol. Chem. Sci. 13 (3), 1846–1860.
Amoussou, E., Cledjo, F., Allagbe, Y., 2016. Evolution climatique du Bénin de 1950 à 2010 et son influence sur les eaux de surface, XXIXth Symposium of the International Association of Climatology: Climate and Air Pollution.
Arrouays, D., Leenaars, J.G., Richer-de-Forges, A.C., Adhikari, K., Ballabio, C., Greve, M., Grundy, M., Guerrero, E., Hempel, J., Hengl, T., 2017. Soil legacy data rescue via GlobalSoilMap and other international and national initiatives. GeoResJ 14, 1–19.
Azuka, C.V., Igué, M.A., Diekkrüger, B., 2020. Land use and slope position effect on the hydrological properties of sandy loam soils of Koupendri catchment, North-West of Benin. Trop. Subtrop. Agroecosyst. 23 (1).
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.
Blume, H.-P., Brümmer, G.W., Fleige, H., Horn, R., Kandeler, E., Kögel-Knabner, I., Kretzschmar, R., Stahr, K., Wilke, B.-M., 2016. Threats to the soil functions. In: Scheffer/SchachtschabelSoil Science. Springer, pp. 485–559.
Bonfatti, B.R., Hartemink, A.E., Giasson, E., Tornquist, C.G., Adhikari, K., 2016. Digital mapping of soil carbon in a viticultural region of southern Brazil. Geoderma 261, 204–221.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Breure, A., De Deyn, G., Dominati, E., Eglin, T., Hedlund, K., Van Orshoven, J., Posthuma, L., 2012. Ecosystem services: a useful concept for soil policy making! Curr. Opin. Environ. Sustain. 4 (5), 578–585.
Calzolari, C., Ungaro, F., Filippi, N., Guermandi, M., Malucelli, F., Marchi, N., Staffilani, F., Tarocco, P., 2016. A methodological framework to assess the multiple contributions of soils to ecosystem services delivery at regional scale. Geoderma 261, 190–203.

Čeh, M., Kilibarda, M., Lisec, A., Bajat, B., 2018. Estimating the performance of random forest versus multiple regression for predicting prices of the apartments. ISPRS Int. J. Geo Inf. 7 (5), 168.
Challinor, A., Wheeler, T., Garforth, C., Craufurd, P., Kassam, A., 2007. Assessing the vulnerability of food crop systems in Africa to climate change. Clim. Chang. 83 (3), 381–399.
Charzyński, P., Piotrowska-Długosz, A., Breza-Boruta, B., 2017. Soil sealing influence on some microbiological biochemical and physicochemical properties of Ekranic Technosols of Toruń. SUITMA 9.
Cincotta, M.M., Perdrial, J.N., Shavitz, A., Libenson, A., Landsman-Gerjoi, M., Perdrial, N., Armfield, J., Adler, T., Shanley, J.B., 2019. Soil aggregates as a source of dissolved organic carbon to streams: an experimental study on the effect of solution chemistry on water extractable carbon. Front. Environ. Sci. 7, 172.
Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for automated geoscientific analyses (SAGA) v. 2.1. 4. Geosci. Model Dev. 8 (7), 1991–2007.
Dansi, A., Vodouhè, R., Azokpota, P., Yedomonhan, H., Assogba, P., Adjatin, A., Loko, Y., Dossou-Aminon, I., Akpagana, K.J.T.S.W.J., 2012. Diversity of the neglected and underutilized crop species of importance in Benin. Sci. World J. 2012.
De Groot, R.S., Alkemade, R., Braat, L., Hein, L., Willemen, L., 2010. Challenges in integrating the concept of ecosystem services and values in landscape planning, management and decision making. Ecol. Complex. 7 (3), 260–272.
Dharumarajan, S., Hegde, R., Singh, S.K., 2017. Spatial prediction of major soil properties using random Forest techniques-a case study in semi-arid tropics of South India. Geoderma Reg. 10, 154–162.
Dossou-Yovo, E.R., Brüggemann, N., Ampofo, E., Igue, A.M., Jesse, N., Huat, J., Agbossou, E.K., 2016. Combining no-tillage, rice straw mulch and nitrogen fertilizer application to increase the soil carbon balance of upland rice field in northern Benin. Soil Tillage Res. 163, 152–159.
Flynn, T., De Clercq, W., Rozanov, A., Clarke, C., 2019. High-resolution digital soil mapping of multiple soil properties: an alternative to the traditional field survey? South African J. Plant Soil 36 (4), 237–247.
Forkuor, G., Hounkpatin, O.K., Welp, G., Thiel, M., 2017. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. PLoS One 12 (1), e0170478.
Gomes, L.C., Faria, R.M., de Souza, E., Veloso, G.V., Schaefer, C.E.G., Fernandes Filho, E. I.J.G., 2019. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 340, 337–350.
Grant, C.A., 2016. Soil Fertility and Management. International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology, pp. 1–10.
Greiner, L., Keller, A., Grêt-Regamey, A., Papritz, A., 2017. Soil function assessment: review of methods for quantifying the contributions of soils to ecosystem services. Land Use Policy 69, 224–237.
Günal, H., Korucu, T., Birkas, M., Özgöz, E., Halbac-Cotoara-Zamfir, R., 2015. Threats to sustainability of soil functions in central and Southeast Europe. Sustainability 7 (2), 2161–2188.
Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., Mendes de Jesus, J., Tamene, L., Tondoh, J.E., 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. PLoS One 10 (6), e0125814 (EP -).
Hengl, T., de Jesus, J.M., Heuvelink, G.B., Gonzalez, M.R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., 2017a. SoilGrids250m: global gridded soil information based on machine learning. PLoS One 12 (2), e0169748.
Hengl, T., Leenaars, J.G.B., Shepherd, K.D., Walsh, M.G., Heuvelink, G.B.M., Mamo, T., Tilahun, H., Berkhout, E., Cooper, M., Fegraus, E., 2017b. Soil nutrient maps of sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. Nutr. Cycl. Agroecosyst. 109 (1), 77–102.
Hien, E., Ganry, F., Oliver, R., 2006. Carbon sequestration in a savannah soil in southwestern Burkina as affected by cropping and cultural practices. Arid Land Res. Manag. 20 (2), 133–146.
Holleran, M., Levi, M., Rasmussen, C., 2015. Quantifying soil and critical zone variability in a forested catchment through digital soil mapping. Soil 1 (1), 47.
Holmes, G., Hall, M., Prank, E., 1999. Generating Rule Sets from Model Trees. Advanced Topics in Artificial Intelligence. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 1–12.
Hopkins, B., 1987. Vegetation Map of Africa. A Descriptive Memoir to Accompany the Unesco/AETFAT/UNSO Vegetation map of Africa. JSTOR. The Vegetation of Africa.
Horning, N., 2010. Random Forests: An algorithm for image classification and generation of continuous fields data sets. In: International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences, Osaka, Japan 2010 Dec 9, vol. 911.
Hounkpatin, K.O., Schmidt, K., Stumpf, F., Forkuor, G., Behrens, T., Scholten, T., Amelung, W., Welp, G., 2018a. Predicting reference soil groups using legacy data: a data pruning and random Forest approach for tropical environment (Dano catchment, Burkina Faso). Sci. Rep. 8 (1), 9959.
Hounkpatin, O.K., de Hipt, F.O., Bossa, A.Y., Welp, G., Amelung, W., 2018b. Soil organic carbon stocks and their determining factors in the Dano catchment (Southwest Burkina Faso). Catena 166, 298–309.
Hounnou, E.F., Dedehouanou, H., 2018. Variability of temperature, precipitation and potential evapotranspiration time series analysis in Republic of Benin. IJAER 4, 991–1019.
Hu, Y., Xu, X., Wu, F., Sun, Z., Xia, H., Meng, Q., Huang, W., Zhou, H., Gao, J., Li, W., 2020. Estimating Forest stock volume in Hunan Province, China, by integrating in situ plot data, Sentinel-2 images, and linear and machine learning regression models. Remote Sens. 12 (1), 186.
Igué, A.M., Adjanohoun, A., 2014. Projet « Evaluation du statut nutritionnel des sols des différentes zones agroécologiques du Bénin ». Rapport technique final. In: Institut National des Recherches Agricoles du Bénin (INRAB), Centre National de Spécialisation sur le maïs (CNS-Maïs).
Igue, A.M., Saidou, A., Adjanohoun, A., Ezui, G., Attiogbe, P., Kpagbin, G., Gotoechan-Hodonou, H., Youl, S., Pare, T., Balogoun, I., 2013. Evaluation de la fertilité des sols au sud et centre du Bénin. Bull. Rech. Agron. Bénin, Spécial numéro, Fertilisation du maïs, pp. 12–23.
Igue, M.A., Oga, A.C., Balogoun, I., Saidou, A., Ezui, G., Youl, S., Kpagbin, G., Mando, A., Sogbedji, J.M., 2016. Détermination Des Formules D'engrais Minéraux Et Organiques Sur Deux Types De Sols Pour Une Meilleure Productivité De Maïs (Zea mays l.) Dans La Commune De Banikoara (Nord-Est Du Bénin). Eur. Sci. J. 12, 16.
Igué, A.M., Balogoun, I., Saidou, A., Oga, A.C., Ezui, G., Youl, S., Kpagbin, G., Mando, A., Sogbedji, J.M., 2018. Recommendations of fertilizer formulas for the production of the EVDT 97 maize variety in northern Benin, improving the profitability, sustainability and efficiency of nutrients through site specific fertilizer recommendations in West Africa agro-ecosystems. Springer 105–123.
INSAE, 2017. Synthèse des analyses sur l'état et la structure de la population (RGPH-4 2013), Cotonou (20 p).
Kasraei, B., Heung, B., Saurette, D.D., Schmidt, M.G., Bulmer, C.E., Bethel, W., 2021. Quantile regression as a generic approach for estimating uncertainty of digital soil maps produced from machine-learning. Environ. Model Softw. 144, 105139.
Kempen, B., Dalsgaard, S., Kaaya, A.K., Chamuya, N., Ruipérez-González, M., Pekkarinen, A., Walsh, M.G., 2019. Mapping topsoil organic carbon concentrations and stocks for Tanzania. Geoderma 337, 164–180.
Kogo, B.K., Kumar, L., Koech, R., Kariyawasam, C.S., 2019. Modelling climate suitability for rainfed Maize cultivation in Kenya using a Maximum Entropy (MaxENT) approach. Agronomy 9 (11), 727.
Koné, B., Saïdou, A., Camara, M., Diatta, S., 2010. Effet de différentes sources de phosphate sur le rendement du riz sur sols acides. Agron. Afr. 22 (1), 55–63.
Kouelo, F.A., Houngnandan, P., Dedehouanou, H., Tossou, R., Orou Bello, D.O., Joël Bekou, K.A., Tchetangni, A.Y., 2016. Soil conservation practices in three watersheds of Benin: farmers cropping systems characterization. Afr. J. Agric. Res. 11 (7), 507–515.
Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling, 26. Springer.
Kuhn, M., 2017. caret: Classification and Regression Training. https://CRAN.R-project.org/package=caret. R package version 6.0-76. [p424].
Kumhálová, J., Kumhála, F., Kroulík, M., Matějková, Š., 2011. The impact of topography on soil properties and yield and the effects of weather conditions. Precis. Agric. 12 (6), 813–830.
Lawrence, I., Lin, K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 255–268.
Leenaars, J., 2013. Africa Soil Profiles Database, Version 1.1. A compilation of georeferenced and standardised legacy soil profile data for Sub-Saharan Africa (with dataset). Africa Soil Information Service (AfSIS) Project, ISRIC-World Soil Information.
Leenaars, J., Van Oostrum, A., Ruiperez Gonzalez, M., 2012. Africa Soil Profiles Database, Version 1.2. ISRIC Report, 3.
Leenaars, J., Kempen, B., van Oostrum, A., Batjes, N., 2014. In: Arrouays, et al. (Eds.), Africa Soil Profiles Database: A compilation of georeferenced and standardised legacy soil profile data for Sub-Saharan Africa, pp. 51–57, 2014b.
Leenaars, J.G., Claessens, L., Heuvelink, G.B., Hengl, T., González, M.R., van Bussel, L.G., Guilpart, N., Yang, H., Cassman, K.G., 2018. Mapping rootable depth and root zone plant-available water holding capacity of the soil of sub-Saharan Africa. Geoderma 324, 18–36.
Li, N., Zare, E., Huang, J., Triantafilis, J., 2018. Mapping soil cation-exchange capacity using Bayesian Modeling and proximal sensors at the field scale. Soil Sci. Soc. Am. J. 82 (5).
Ma, Z., Zhou, Y., Hu, B., Liang, Z., Shi, Z., 2017. Downscaling annual precipitation with TMPA and land surface characteristics in China. Int. J. Climatol. 37 (15), 5107–5119.
Madena, K., Bormann, H., Giani, L., 2012. Soil functions-Today's situation and further development under climate change. Erdkunde 221–237.
Malone, B., McBratney, A., Minasny, B., 2011. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. Geoderma 160 (3–4), 614–626.
McBratney, A.B., Santos, M.M., Minasny, B., 2003. On digital soil mapping. Geoderma 117 (1), 3–52.
Miller, B.A., Koszinski, S., Wehrhan, M., Sommer, M., 2015a. Comparison of spatial association approaches for landscape mapping of soil organic carbon stocks. Soil 1 (1), 217.
Miller, B.A., Koszinski, S., Wehrhan, M., Sommer, M., 2015b. Impact of multi-scale predictor selection for modeling soil properties. Geoderma 239, 97–106.
Minai, J.O., 2019. Utilization of Legacy Soil Data for Digital Soil Mapping and Data Delivery for the Busia Area. Purdue University Graduate School, Kenya.
Minasny, B., McBratney, A.B., 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. Chemom. Intell. Lab. Syst. 94 (1), 72–79.
Minasny, B., McBratney, A.B., 2016. Digital soil mapping: a brief history and some lessons. Geoderma 264, 301–311.
Mosleh, Z., Salehi, M.H., Jafari, A., Borujeni, I.E., Mehnatkesh, A., 2016. The effectiveness of digital soil mapping to predict soil properties over low-relief areas. Environ. Monit. Assess. 188 (3), 195.

Nawar, S., Buddenbaum, H., Hill, J., 2015. Digital mapping of soil properties using multivariate statistical analysis and ASTER data in an arid region. Remote Sens. 7 (2), 1181–1205.
Nelson, M., Bishop, T., Triantafilis, J., Odeh, I., 2011. An error budget for different sources of error in digital soil mapping. Eur. J. Soil Sci. 62 (3), 417–430.
Neuenschwander, P., Sinsin, B., Goergen, G., 2011. Protection de la Nature en Afrique de l'Ouest: Une Liste Rouge pour le Bénin Nature Conservation in West Africa: Red List for Benin. IITA, Ibadan.
Padarian, J., Minasny, B., McBratney, A.B., 2019. Using deep learning to predict soil properties from regional spectral data. Geoderma Reg. 16, e00198.
Phachomphon, K., Dlamini, P., Chaplot, V., 2010. Estimating carbon stocks at a regional level using soil information and easily accessible auxiliary variables. Geoderma 155 (3–4), 372–380.
Piikki, K., Wetterlind, J., Söderström, M., Stenberg, B., 2015. Three-dimensional digital soil mapping of agricultural fields by integration of multiple proximal sensor data obtained from different sensing methods. Precis. Agric. 16 (1), 29–45.
Poggio, L., de Sousa, L.D.M., Heuvelink, G., Kempen, B., Batjes, N., Leenaars, J., Mantel, S., Bai, Z., Turdukulov, U., Gonzalez, M.R., 2019. SoilGrids: consistent soil information to assess and map soil functions at global scale, Wageningen soil conference 2019. ISRIC 50–51.
Poppiel, R.R., Lacerda, M.P.C., Safanelli, J.L., Rizzo, R., Oliveira, M.P., Novais, J.J., Demattê, J.A.M., 2019. Mapping at 30 m resolution of soil attributes at multiple depths in Midwest Brazil. Remote Sens. 11 (24), 2905.
Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., Thompson, J., 2018. Soil property and class maps of the conterminous United States at 100-meter spatial resolution. Soil Sci. Soc. Am. J. 82 (1), 186–201.
Saïdou, A., Balogoun, I., Ahoton, E.L., Igué, A.M., Youl, S., Ezui, G., Mando, A., 2018. Fertilizer recommendations for maize production in the South Sudan and Sudano-Guinean zones of Benin, improving the profitability, sustainability and efficiency of nutrients through site specific fertilizer recommendations in West Africa agro-ecosystems. Springer 215–234.
Saiz, G., Bird, M.I., Domingues, T., Schrodt, F., Schwarz, M., Feldpausch, T.R., Veenendaal, E., Djagbletey, G., Hien, F., Compaore, H., Diallo, A., Lloyd, J., 2012. Variation in soil carbon stocks and their determinants across a precipitation gradient in West Africa. Glob. Chang. Biol. 18 (5), 1670–1683.
Shahbazi, F., Hughes, P., McBratney, A.B., Minasny, B., Malone, B.P., 2019. Evaluating the spatial and vertical distribution of agriculturally important nutrients—nitrogen, phosphorous and boron—in North West Iran. Catena 173, 71–82.
Silatsa, F.B., Yemefack, M., Tabi, F.O., Heuvelink, G.B., Leenaars, J.G., 2020. Assessing countrywide soil organic carbon stock using hybrid machine learning modelling and legacy soil data in Cameroon. Geoderma 367, 114260.
Solomatine, D.P., Shrestha, D.L., 2009. A novel method to estimate model uncertainty using machine learning techniques. Water Resour. Res. 45 (12).
Somarathna, P., Malone, B., Minasny, B., 2016. Mapping soil organic carbon content over New South Wales, Australia using local regression kriging. Geoderma Reg. 7 (1), 38–48.
Sonneveld, B., Keyzer, M.A., Adegbola, P., Pande, S., 2012. The impact of climate change on crop production in west Africa: an assessment for the oueme river basin in Benin. Water Resour. Manag. 26 (2), 553–579.
Sulaeman, Y., Sarwani, M., Minasny, B., McBratney, A., Sutandi, A., Barus, B., 2012. Soil-landscape models to predict soil pH variation in the Subang region of West Java, Indonesia. In: Digital Soil Assessment and Beyond, Taylor and Francis Group, London, England, pp. 317–325.
Sys, D., 1976. Principes de classification et d'evaluation des terres pour la Republique Populaire du Benin. Rapport de la mission de (consultation).
Szatmári, G., Pásztor, L., 2019. Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms. Geoderma 337, 1329–1340.
Taghizadeh-Mehrjardi, R., Minasny, B., Sarmadian, F., Malone, B., 2014. Digital mapping of soil salinity in Ardakan region, central Iran. Geoderma 213, 15–28.
UN, 2016. Transforming our world: The 2030 agenda for sustainable. https://stg-wedocs.unep.org/bitstream/handle/20.500.11822/11125/unepswiosm1inf7sdg.pdf?sequence=1 accessed Sept 14, 2019.
Vasenev, V., Dovletyarova, E., Cheng, Z., Prokof`eva, T.V., Morel, J.L., Ananyeva, N.D., 2018. Urbanization: Challenge and Opportunity for Soil Functions and Ecosystem Services: Proceedings of the 9th SUITMA Congress. Springer.
Vaysse, K., Lagacherie, P., 2017. Using quantile regression forest to estimate uncertainty of digital soil mapping products. Geoderma 291, 55–64.
Volkoff, B., Willaime, P., 1976. Carte pédologique de reconnaissance de la République Populaire du Bénin à 1/200 000: feuille de Porto-Novo.
Walkley, A., Black, I.A., 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Sci. 37 (1), 29–38.
Wilson, C.H., Caughlin, T.T., Rifai, S.W., Boughton, E.H., Mack, M.C., Flory, S.L., 2017. Multi-decadal time series of remotely sensed vegetation improves prediction of soil carbon in a subtropical grassland. Ecol. Appl. 27 (5), 1646–1656.
Wubie, M.A., Assen, M., 2020. Effects of land cover changes and slope gradient on soil quality in the Gumara watershed, Lake Tana basin of North–West Ethiopia. Modeling Earth Syst. Environ. 6 (1), 85–97.
Yigini, Y., Olmedo, G., Reiter, S., Baritz, R., Viatkin, K., Vargas, R., 2018. Soil Organic Carbon Mapping: Cookbook.
Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., Finke, P., 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. Geoderma 338, 445–452.