Ecological Processes

RESEARCH

Open Access

Mapping soil nutrients via different covariates combinations: theory and an example from Morocco

Kingsley John[1*], Yassine Bouslihim[2], Isong Abraham Isong[3], Lahcen Hssaini[2], Rachid Razouk[2], Ndiye M. Kebonye[1], Prince C. Agyeman[1], Vit Penížek[1] and Tereza Zádorová[1]

Abstract

Background: Mapping of soil nutrients using different covariates was carried out in northern Morocco. This study was undertaken in response to the region's urgent requirement for an updated soil map. It aimed to test various covariates combinations for predicting the variability in soil properties using ordinary kriging and kriging with external drift.

Methods: A total of 1819 soil samples were collected at a depth of 0–40 cm using the 1-km grid sampling method. Samples were screened for their pH, soil organic matter (SOM), potassium (K_2O), and phosphorus (P_2O_5) using standard laboratory protocols. Terrain attributes (T) computed using a 30-m resolution digital elevation model, bioclimatic data (C), and vegetation indices (V) were used as covariates in the study. Each targeted soil property was modeled using covariates separately and then combined (e.g., pH ~ T, pH ~ C, pH ~ V, and pH ~ T + C + V). *k* = tenfold cross-validation was applied to examine the performance of each employed model. The statistical parameter RMSE was used to determine the accuracy of different models.

Results: The pH of the area is slightly above the neutral level with a corresponding 7.82% of SOM, 290.34 ppm of K_2O, and 100.86 ppm of P_2O_5. This was used for all the selected targeted soil properties. As a result, the studied soil properties showed a linear relationship with the selected covariates. pH, SOM, and K_2O presented a moderate spatial autocorrelation, while P_2O_5 revealed a strong autocorrelation. The cross-validation result revealed that soil pH (RMSE = 0.281) and SOM (RMSE = 9.505%) were best predicted by climatic variables. P_2O_5 (RMSE = 106.511 ppm) produced the best maps with climate, while K_2O (RMSE = 209.764 ppm) yielded the best map with terrain attributes.

Conclusions: The findings suggest that a combination of too many environmental covariates might not provide the actual variability of a targeted soil property. This demonstrates that specific covariates with close relationships with certain soil properties might perform better than the compilation of different environmental covariates, introducing errors due to randomness. In brief, the approach of the present study is new and can be inspiring to decision-makers in the region and other world areas as well.

Keywords: Soil mapping, Environmental variables, Agriculture, Soil properties, Soil management

Introduction

With the growing trend and vast application of the digital soil mapping (DSM) technique, DSM has become an unavoidable tool for creating high-resolution maps of soil properties and classes from spatially understandable soil data and environmental covariates (McBratney et al.

*Correspondence: johnk@af.czu.cz
[1] Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food, and Natural Resources, Czech University of Life Sciences, Kamýcká 129, 16500 Prague, Czech Republic
Full list of author information is available at the end of the article

2003; Agyeman et al. 2021). Thus, it bridges gaps between conventional soil maps and the diversified environment from which soil is developed (Balkovič et al. 2013). DSM tools may incorporate either statistical (machine learning) concepts (Bouslihim et al. 2021a; Hengl et al. 2021), GIS concepts (Carré and Girard 2002; Hengl et al. 2004; Zádorová et al. 2011; John et al. 2020; Bouslihim et al. 2021b; John et al. 2021a, b) or geostatistics (Webster and Oliver 2007) via environmental correlations and SCORPAN-based models (McKenzie and Ryan 1999; McBratney et al. 2003).

In mapping relatively large areas, soil surveyors often use easily measured variables related to the somewhat costly measurements of targeted variables. For example, satellite imagery, climatic data, and terrain attributes are primarily environmental covariates that are readily used to map soil specific properties (Penížek and Boruka 2006; Penížek et al. 2016; Borůvka et al. 2020; John et al. 2021a, b). This method has received much attention as soil scientists keep building a more robust predictive approach in creating quantifiable soil maps with a reduced level of uncertainty (Agyeman et al. 2021; Hengl et al. 2021).

Geostatistic is a DSM approach commonly used in pedology as a predictive method for modeling different properties such as porosity, permeability, soil depth, or thickness (Webster and Oliver 2007; Penížek and Boruka 2006). On the other hand, kriging methods are ranked top as widely used estimators for soil spatial variability (Odeh et al. 2003; Agyeman et al. 2021; Li and Heap 2014). Kriging methods are interpolation models for estimating a regionalized variable at specific grid points that predict values from interpolation without bias and minimum variance. Kriging methods include ordinary kriging (OK), universal kriging (UK) (also known as kriging with external drift), indicator kriging, Co-kriging (Penížek and Boruka 2006; John et al. 2020; Borůvka et al. 2020; John et al. 2021a, b; Hengl et al. 2004). Among the kriging methods, the OK method ranks the top according to usage (Li and Heap 2014; Ageyman et al. 2020). However, the selected kriging method used in a specific study depends on the data structure.

Kriging with external drift (KED) (also known as universal kriging) and regression-kriging have been classed into the group of the so-called 'hybrid' (McBratney et al. 2000), which refers to non-stationary geostatistical approaches (Wackernagel et al. 2002). KED entails kriging with additional information. Wackernagel et al. (2002) and Papritz and Stein (1999) distinguished universal kriging (UK) from KED only when coordinates are employed. While the drift is determined externally rather than utilizing monomials of the coordinates in the UK equations, the term "Kriging with external drift" (KED) or external trend is used (Hengl et al. 2004). These external drifts could be digital terrain (DEM), satellite images, climatic data, or others readily available in DSM. Hengl et al. (2004) have also reported the superiority of KED over regression kriging (RK). Hudson and Wackernagel (1994) outlined the effectiveness of KED in temperature mapping with elevation. Also, Bourennane et al. (2000) presented the accuracy of KED in predicting soil thickness with different sample densities. Using external drift from various sources, Santra et al. (2017) reported that KED performed better than RK, random forest, and OK for sand content modeling.

However, there is still a need to explore external drifts originating from the same source and different sources to estimate a targeted soil property. This may reduce the time spent excavating covariates that may not be necessary to evaluate the spatial variability of a targeted soil property. Therefore, we hypothesize that external drift from the same source may provide a reliable estimate of the spatial variability of a targeted soil property than the combination of all external drifts from different sources that may not be directly influencing the soil property.

Morocco is a northern African country that spreads from the Mediterranean Sea and the Atlantic Ocean towards the north and the west, respectively, into large mountainous areas in the interior, to the Moroccan Sahara desert in the far south. The country experiences hot arid climatic conditions, with rainfall occurring between October and April. According to FAO/WRB soil taxonomical classification, the predominant soils of the region are Leptisols, Regosols, Calcisols, Vertisols, Kastonozems, among others. However, according to Moussadek's (2014) presentation at the Global soil partnership conference, only 30% of the Morocco region is covered by soil maps (scale = 1:500,000) generated via conventional soil mapping. Hence, there is an urgent need for building high-resolution digital soil maps of areas not tapped or soil surveyed.

Furthermore, the soils of the regions vary both in space and time. Therefore in this study, we try to examine the use of different external drift data combinations (i.e., environmental covariates) to improve soil pH and soil nutrient properties mapping (e.g., potassium, phosphorus, and soil organic matter). Although factors such as different farming, fertilizer, and soil input materials influence the spatial variability of soil nutrients, it is worth noting that these factors are built on the modified soil-forming factors (age, climate, terrain, vegetation, and soil) (McBratney et al. 2003). Therefore, knowledge of the relationship between soil nutrients and environmental covariates has guided the choice of different external drift variable combinations. Also, this approach is the first step in exploring the possibilities of using appropriate auxiliary variables for predicting soil property maps via kriging with external drift. Therefore, in clear terms, this study aims to predict soil property with different combinations of auxiliary variables via kriging with external drift and to compare its performance with the ordinary kriging method. This study approach is new and relevant to decision-makers and land managers in Morocco and other world regions. In addition, the research was conducted to develop a digital soil map of the area under investigation.

Materials and methods

Research location and description

The selected area is located in Taounate province in northern Morocco (34°47′N, 4°4.4′W and 34°05′N, 5°10.3′W), and is displayed as a rectangle of 7979 km^2 (101 × 79 km) (Fig. 1). Jbel Oudka is the most important mountain of Taounate, the major in the region, with an elevation that reaches 1587 m and is characterized by significant vegetation cover. The area covered by this study contains a part of the Atlas Mountains in the northwest. In general, the altitude ranges from 78 to 1969 m.

A Mediterranean climate and irregular rainfall characterize this region. The average annual rainfall is about 650 mm, and the mean yearly temperature is 17 °C, with dry summers and rainy winters. As a result, the average maximum temperature of the hottest month is approximately 34.2 °C and the average minimum of the coldest month is 0.5 °C (Allali et al. 2020; Rezouki et al. 2021).

Today, the region shows a great promise for more extensive development opportunities, particularly legalizing cannabis cultivation for industrial and medicinal purposes. This paper's idea will help achieve this objective by providing more insights into soil properties through a satisfactory, rapid, sustainable, and low-cost approach.

Geological formations of the Taounate wrinkle consist of a Jurassic–Cretaceous series of marl overcome molassic formations composed of sandstone and conglomerates (Mesrar et al. 2017). According to World Reference Base for Soil Resources (FAO Classification System), the region's soils are moderately weathered and classified within Xeralfs and Luvisols. They consist of very deep, somewhat poorly drained soils formed in sandy outwash, glaciolacustrine, or eolian deposits on outwash plains, lake plains, and dunes (FAO/ISRIC/ISSS 2006).

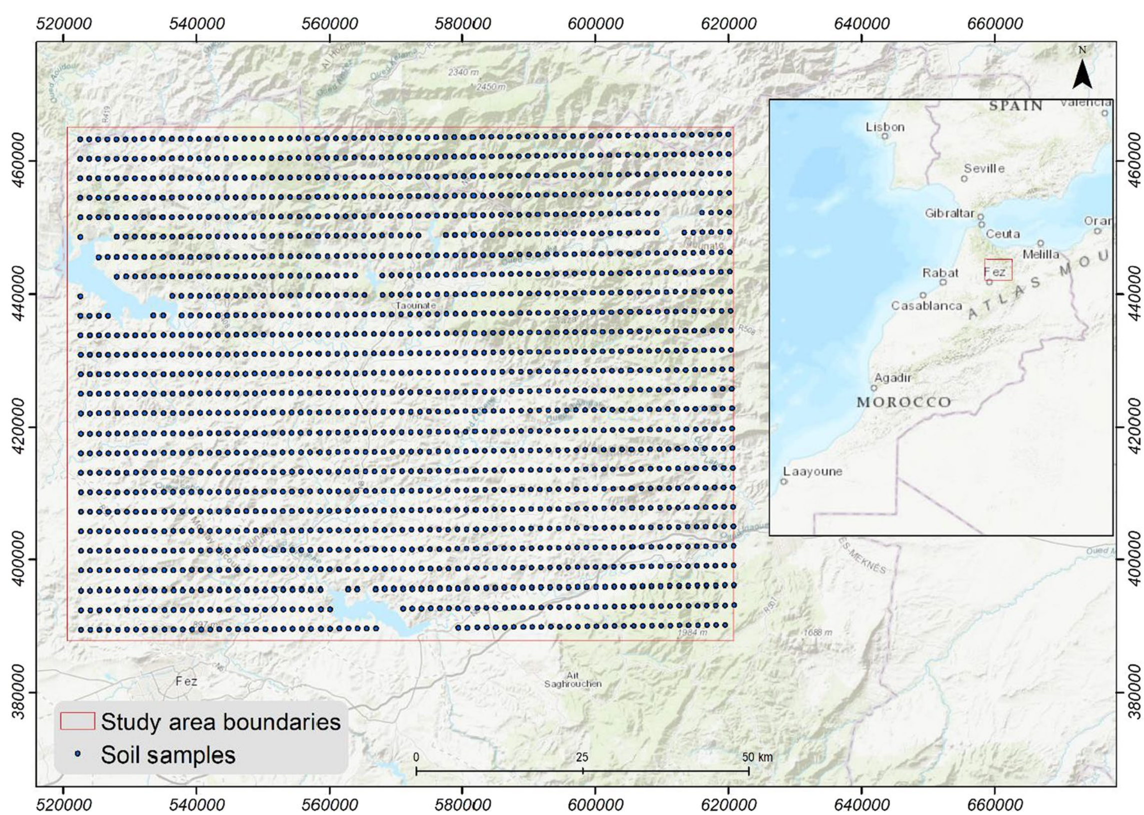

Fig. 1 The study area map shows $n = 1819$ sample points at a depth of 0–40 cm. The study area covers the Atlas Mountains in the northwest and also by the coast of the Mediterranean Sea

Sampling regime

A sample campaign was conducted in November and December 2013 to account for soil diversity and ensure adequate coverage of the study region. A systematic sampling plan based on a 1 km grid was also implemented. As a result, 1819 soil samples were taken at a depth of 0–40 cm, bypassing the urban area of Tounate city and the water reservoirs at the level of El Wahda, Idriss I, and Assfalou dams (Fig. 1).

Sample preparation and laboratory analysis

The samples were analyzed in the laboratory for a set of soil fertility parameters such as organic matter (SOM), available potassium K_2O, available phosphorus P_2O_5, and pH. The SOM content was estimated using the Walkley Black method to determine soil organic carbon (SOC). In addition, to obtain SOM, SOC was multiplied by 1.724 (Udo et al. 2009).

The soil pH was measured in water with a soil/water ratio of 1:2. Available K and P were analyzed by extraction method using ammonium acetate (1:10) and molybdate ammonium (1:20), respectively (Udo et al. 2009).

Environmental covariates sources and preparation

Different auxiliary data were used in this study to consider all the factors that can influence the spatial distribution of the studied parameters. We collected four bioclimatic parameters representing temperature variation (annual mean temperature (bio_1), the max temperature of the warmest month (bio_5), min temperature of the coldest month (bio_6) and annual precipitation (bio_12) obtained from the WorldClim database version 2 (Fick and Hijmans 2017). These data are available in GeoTiff (.tif) format with a resolution of 10 min (~ 340 km^2).

We obtained Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) at 30 m resolution covering the whole area for the terrain attributes. The data were processed using SAGA-GIS (Olaya 2004). Elevation, slope, profile curvature, plan curvature, multi-resolution valley bottom flatness (MrVBF), multi-resolution ridge top flatness (MrRTF), topographic wetness index (TWI), convergence index, and aspect were calculated and included as covariates.

In addition, the Landsat-8 OLI/TIRS image with a spatial resolution of 30 m was used to extract six parameters. The calculated indices are as follows: normalized difference vegetation index (NDVI), transformed normalized difference vegetation index (TNDVI), soil adjusted vegetation index (SAVI), ratio vegetation index (RVI), difference vegetation index (DVI), and chlorophyll vegetation index (CVI). Since the study area is within two scenes' limits (path = 200/201 and row = 36), two Landsat 8 images were used to calculate all required parameters. The pre-processing of the bands used, treatment, and analysis were done using the ArcGIS program. All the covariates with coarser resolution were downscaled to 30 m pixel size using the nearest neighbor function in the ArcGIS program (Table 1).

Geostatistical models and experimental design

Ordinary kriging

This study performed ordinary kriging (OK) on pH, K_2O, P_2O_5, and SOM. The OK technique uses an

Table 1 Environmental covariates obtained from the WorldClim database, ASTER-GDEM, and Landsat-8 OLI/TIRS and applied in the modeling regime

Environmental covariates	Sources	Resolution	Resample resolution
Bioclimatic parameters			
Annual mean temperature (bio_1) Max temperature of the warmest month (bio_5) Min temperature of the coldest month (bio_6) Annual precipitation (bio_12)	WorldClim database version 2	10 min (~ 340 km^2)	30 m
Terrain attributes			
Elevation; slope; profile curvature; plan curvature, Multi-resolution Valley Bottom Flatness (MrVBF); Multi-resolution Ridge Top Flatness (MrRTF); Topographic wetness index (TWI); convergence index and aspect	ASTER-GDEM	30 m	30 m
Remote sensing indices			
Normalized difference vegetation index (NDVI) Transformed normalized difference vegetation index (TNDVI) Soil adjusted vegetation index (SAVI) Ratio vegetation index (RVI) Difference vegetation index (DVI) Chlorophyll vegetation index (CVI)	Landsat OLI 8	30 m	30 m

All the input data were resampled at 30 m pixel size estimated average of a specific soil property (e.g., pH, K_2O, P_2O_5, and SOM) at a known location (x_0) to interpolate the value at an unsampled location (x_i) as outlined by Goovaerts (2001) and Bishop and McBratney (2001) (Eq. 1):

$$Z'(x_0) = \sum_{i=1}^{n} \lambda_i . Z(x_i), \quad (1)$$

where $Z'(x_0)$ is the predicted/interpolated value for point x_O, x_i) is the known value, and λ_i is the kriging weight for the $Z(x_i)$ values. It can be calculated by the semi-variance function of the variables on the condition that the estimated value is unbiased and optimal. The semivariogram model as (Eq. 2):

$$\gamma(h) = \frac{1}{2N(h)} \sum\nolimits_{i=1}^{n} [Z(x_i) - Z(x_i + h)]^2, \quad (2)$$

where γ(h) is the semi-variance, $N(h)$ is the point group number at distance h, $Z(x_i)$ is the numerical value at position x_i, and $Z(x_i + h)$ is the numerical value at a distance $(x_i + h)$.

Theory of kriging with external drift (KED)

Kriging with a trend or kriging in the presence of drift is also known as universal kriging (UK). The KED method also predicts $Z(x)$ at an unsampled area. It splits the random function into a linear combination of deterministic processes, the smoothly varying and non-stationary trend called a drift, and a random component representing the residual random function (Wackernagel et al. 2002). OK assumes a stationary, i.e., constant mean of the underlying real-valued random function $Z(x)$. Nevertheless, in reality, the mean value varies, it is often not consistent across the entire study area, and the variable is said to be non-stationary. A non-stationary regionalized variable can be considered as having two components (Davis 1973): drift (average or expected value of the regionalized variable) and a residual (difference between the actual measurements and the drift).

The KED method allows predicting a variable Z, known only at a small set of sample points of the research area, through another variable s, exhaustively known in the same area.

The external drift method thus consists in incorporating into kriging system addition universality conditions about one or several external drift variables, $s_i(x) = 1, \ldots, M$, measured exhaustively in the spatial domain. The functions $s_i(x)$ need to be known at all locations x_i of the samples of $Z(x_i)$ as well as at nodes of the prediction grid.

In this type of DSM technique, we assume a linear relationship between the response variable and the environmental covariates (climate, terrain derivatives, vegetation indices, soil properties) at the observation points of the response variables (e.g., pH, K_2O, P_2O_5, SOM). This assumption is fundamental in the prediction by the external drift approach. Thus, if a non-linear function describes the relationship between the two variables, this function should first transform the data of the environmental covariates. The transformed data could then be used as an external drift. This method has been applied in different fields of science, as reported by Bourennane et al. (2000). Similarly, the technique is important in soil science due to the increase of data from different sources (digital terrain model (DEM), satellite images, lithological data), which improves conventional pedological data obtained at specific locations.

The approach adopted in this study is explained as: thus, each targeted soil property or soil nutrient was modeled with covariates from a specific source and then evaluated when all the covariates were combined. That is to say, each soil property was modeled with climatic (C) covariates, terrain attributes (T), vegetation indices (V), and then all the covariates (A) (i.e., pH ~ C + T + V).

All this was performed in an R environment with the following packages: *gstat, sp, rgdal, rgeos, MASS.*

Model prediction accuracy

The model performance was accessed on the entire data via k = tenfold cross-validation. The data were randomly split into n = 10 partitions or folds; nine of these partitions were engaged to fit and evaluate the model at each step. This procedure was repeated for each partition sequentially. Averaged over all k = tenfold held back, the performance delivers the cross-validated performance assessment—all the processes were rendered automatically in the R environment.

Statistical metrics

The performance of the different models was compared via these criteria, RMSE (root mean square error), which gives an estimate of the standard deviation of the errors:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} p_i - o_i)^2}, \quad (3)$$

where p_i = predicted values, o_i = observed values.

Results

Summary statistics of soil variables

The characteristics of response variables are shown in Table 2. The pH of the studied region ranged from 6.00 to 8.40, with the mean value slightly above neutral. The SOM ranged from 0.51 to 92.11%, with a mean of 7.82%. SOM content obtained here was higher than the 0.22% reported

Table 2 Summary statistics of the response variables ($n=1819$) for spatial modeling

Parameters	pH	SOM (%)	K_2O (ppm)	P_2O_5 (ppm)
Mean	7.32	7.82	290.34	100.86
Standard error	0.01	0.27	6.73	3.60
Median	7.2	2.44	240	28.66
Standard deviation	0.33	11.65	286.99	153.54
Kurtosis	0.85	15.25	− 0.06	6.46
Skewness	0.14	3.35	0.78	2.46
Range	2.40	91.61	1258.13	1012.35
Minimum	6.00	0.51	0.67	1.65
Maximum	8.40	92.11	1258.8	1014
CV (%)	4.52	148.95	98.85	152.23
Confidence level (95.0%)	0.02	0.54	13.20	7.06

CV coefficient of variation

by Laghrour et al. (2016) for surface soils under conservation. K_2O ranged from 0.67–1258.8 ppm with a mean of 290.34 ppm, while P_2O_5 ranged 1.65–1014 ppm. The values of P_2O_5 and K_2O were higher than that of Nabyl et al. (2020).

Correlations between soil properties and studied covariates

Figures 2, 3, 4, 5 show the correlation between the studied soil properties and respective environmental covariates. The strength of the correlations between pH, SOM, K_2O, and P_2O_5 with vegetation covariates, soil properties, terrain covariates, and climatic data is shown in Figs. 2 and 3. The result indicated that soil pH is significantly correlated with vegetation index (CVI) ($r=0.05$). This implies that pH could be increased with a rise in CVI in the area. Conversely, soil pH negatively and significantly correlated with silt ($r=-0.07$) and bulk density (− 0.05). However, terrain covariates did not show any correlation with soil pH. Similarly, soil pH displayed a significant positive relationship with climate data, bio_12 ($r=0.10$).

SOM showed a significant positive relationship with vegetation indices including TNDVI ($r=0.08$), NDVI ($r=0.07$) and CVI ($r=0.12$). This result implies that the increase in these vegetation variables could lead to an increase in soil SOM. Conversely, SOM negatively and significantly correlated with RVI ($r=0.08$). SOM showed positive correlation with terrain covariates such slope ($r=0.09$) and elevation ($r=0.1$) and negative correlation with profile curvature ($r=-0.07$) and MRVBF ($r=-0.07$). John et al. (2021a, b) also observed SOC in their study to be positively correlated with slope ($r=0.44$) and elevation ($r=0.47$). The result obtained therein corroborated this report and confirmed Askoy et al. (2012). They conducted a similar study and concluded that elevation is the primary terrain attribute of soil properties. The results also showed that SOM was positively correlated with climate data, bio_12 ($r=0.22$) and negatively correlated with bio_6 ($r=-0.09$), bio_5 ($r=-0.10$) and bio_1 ($r=-0.11$). This result is in line with the study of Okon et al. (2019), where climate variables were negatively correlated with soil nutrient indicators. Amalu and Isong (2015) pointed out that excessive rainfall amounts could leach out virtually all nutrient elements from the rhizosphere zones in line with this report. There is always a problem of heavy leaching, erosion, and generally poor performance in arable crops' growth during the rainy season in areas with poor soil conservation measures.

K_2O showed a significant negative relationship with vegetation indices including TNDVI ($r=-0.13$), NDVI ($r=-0.13$), CVI ($r=-0.21$) and SAVI ($r=-0.05$). These results imply that an increase in these vegetation variables could possibly lead to a decrease soil K_2O. However, K_2O is positively and significantly correlated to RVI ($r=0.14$). K_2O showed positive correlation with terrain covariates such MRVBF ($r=0.07$) and MRRTF ($r=0.05$) and negative correlation with slope ($r=-0.12$) and elevation ($r=-0.20$). The results also showed that K_2O was positively correlated with climate data, boi_6 ($r=0.18$), bio_5 ($r=0.21$) and bio_1 ($r=0.21$), and negatively correlated with bio_12 ($r=-0.34$).

P_2O_5 only showed a significant positive relationship with vegetation index CVI ($r=0.07$). P_2O_5 also positively and significantly correlated with elevation ($r=0.07$) and bio_12 ($r=0.18$), and negatively correlated with bio_1 ($r=-0.06$).

Spatial dependency of soil properties

Table 3 shows the parameters of the studied soil properties via semivariogram. Spherical models were the most efficient for modeling soil pH, SOM, K_2O, and P_2O_5. However, John et al. (2020) also found pH and SOC best modeled using spherical models, while available phosphorus was fitted best with a stable model. The higher Nugget/Sill connotes that spatial variability is mainly caused by stochastic factors, including fertilization, farming management practices, and other human activities.

In contrast, a lower ratio suggests that structural factors, such as climate, parent material, topography, soil properties, and other natural factors play a significant role in spatial variability. The studied soil properties had diverse spatial dependence due to their nugget to sill ratios. Soil pH, SOM, and K_2O estimated via OK and KED had moderate spatial dependence, whereas P_2O_5 estimated via OK and KED had solid spatial dependence, as shown in Table 3.

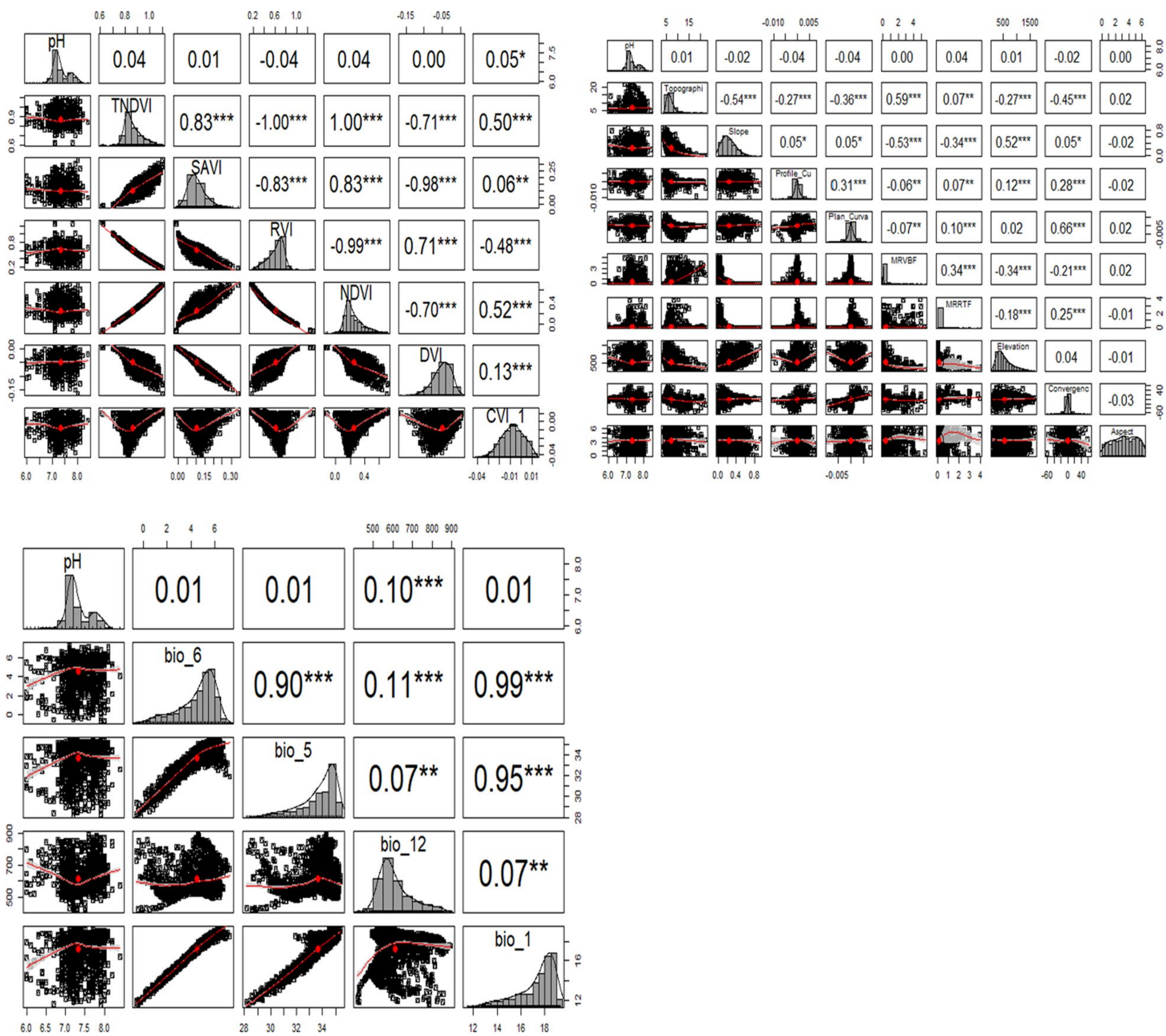

Fig. 2 Correlation between pH and bioclimatic data, terrain attributes, and vegetation indices, respectively. The total number of observation points used = 1819

Cross-validation

From the analysis of the developed semivariogram, the best fit model was selected through cross-validation technique based on the RMSE to assess the spatial distribution of soil pH, SOM, K_2O, and P_2O_5. The detailed outcomes about the fitness and selection of different models for the interpolation of the studied soil properties are given in Table 2. Soil pH was best predicted with climate data an external drift (KED + C). This is consistent with the finding of Chytry et al. (2007), who reported that climate conditions at a regional level significantly influenced pH. The result shows this relationship at the lowest RMSE (0.281). Similarly, SOM was also the best model using KED + C with an RMSE value of 9.505%. Likewise, K_2O was the best model when utilizing terrain covariates (KED + T) with an RMSE value of 209.764 ppm.

Maps production

Figures 6, 7, 8, 9 are the interpolated maps of OK and KED, respectively. The soil pH map produced from the OK map revealed that the pH value in the range of 7.0–7.6 dominated the surveyed area. However, 7.6–7.8 pH values are seen traveling from east to west and patches northward. The KED pH map with terrain attributes showed more calibrations in the scale, ranging from 6.6 – 8.0. Patches of pH of 7.4 were seen clearly in the

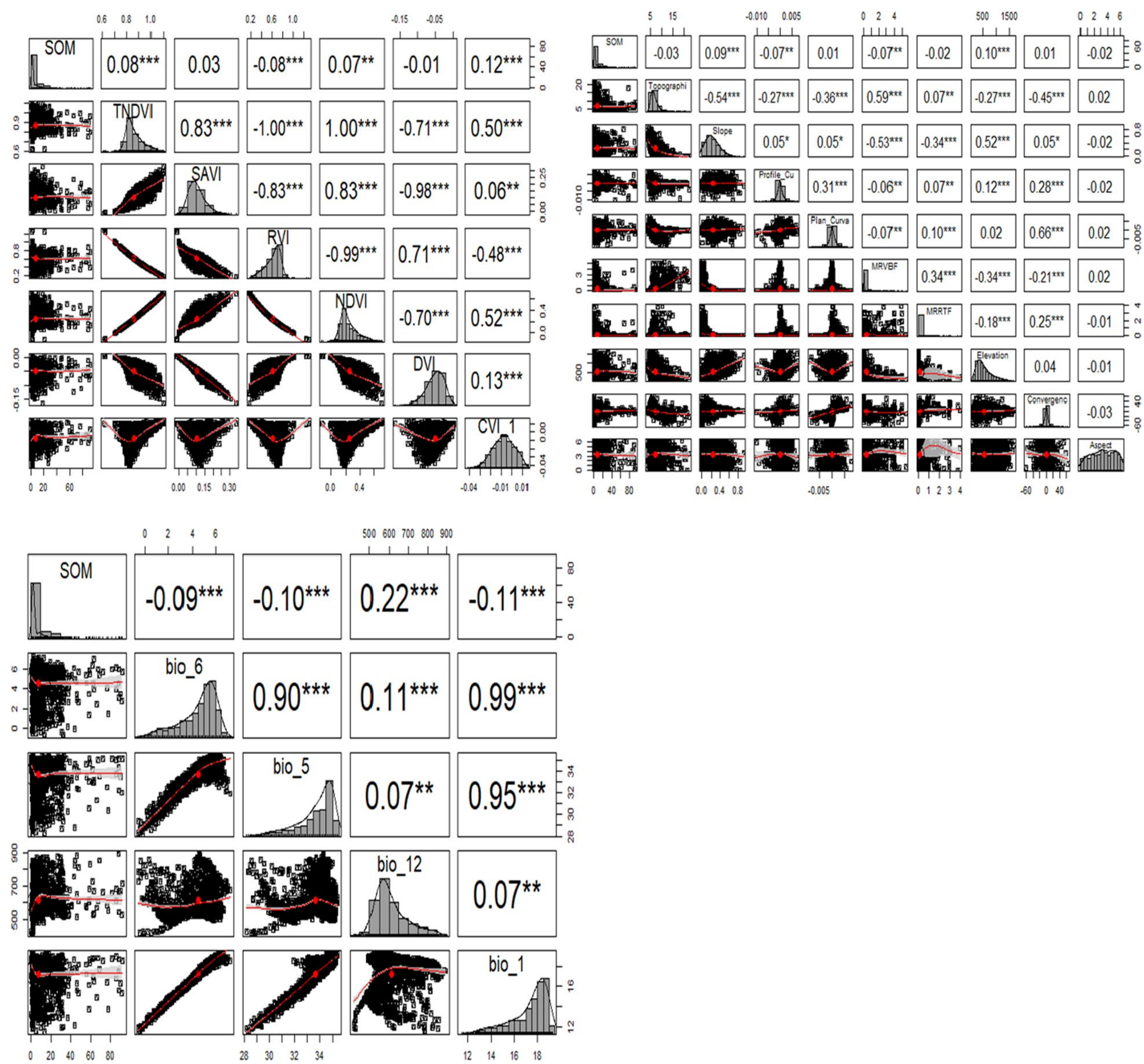

Fig. 3 Correlation between SOM and bioclimatic data, terrain attributes, and vegetation indices, respectively. The total number of observation points used = 1819

southern part of the area. Also, a pH of 7.8 was observed in the northern part of the area. The result obtained here is consistent with Smith et al. (2002) and Moore et al. (1993). They outlined that elevation gradient results in an increase or decrease in pH.

Similarly, the pH map modeled with terrain as an external drift showed noise in the spatial distribution. However, we observed a conspicuous patch of pH of 7.6 in the southern part. pH map via climate followed a similar distribution pattern with terrain and soil data. However, the soil pH map produced with vegetation data was utterly different from pH mapped with OK, KED (terrain), and KED (climate), respectively. The pH legend ranged from 0–80, with the pH of the area ranging from 0–20. Thus, the map does not visually look good as the others.

SOM map by OK model revealed lower SOM contents are dominated in the southern part (0–10%), while patches of SOM content between 10–20% are dominated in the central region of the area. Only a few sites are revealed to have SOM (30–40%). Nevertheless, the SOM map created with terrain attributes highlighted places in the south having SOM (10–20%) even without

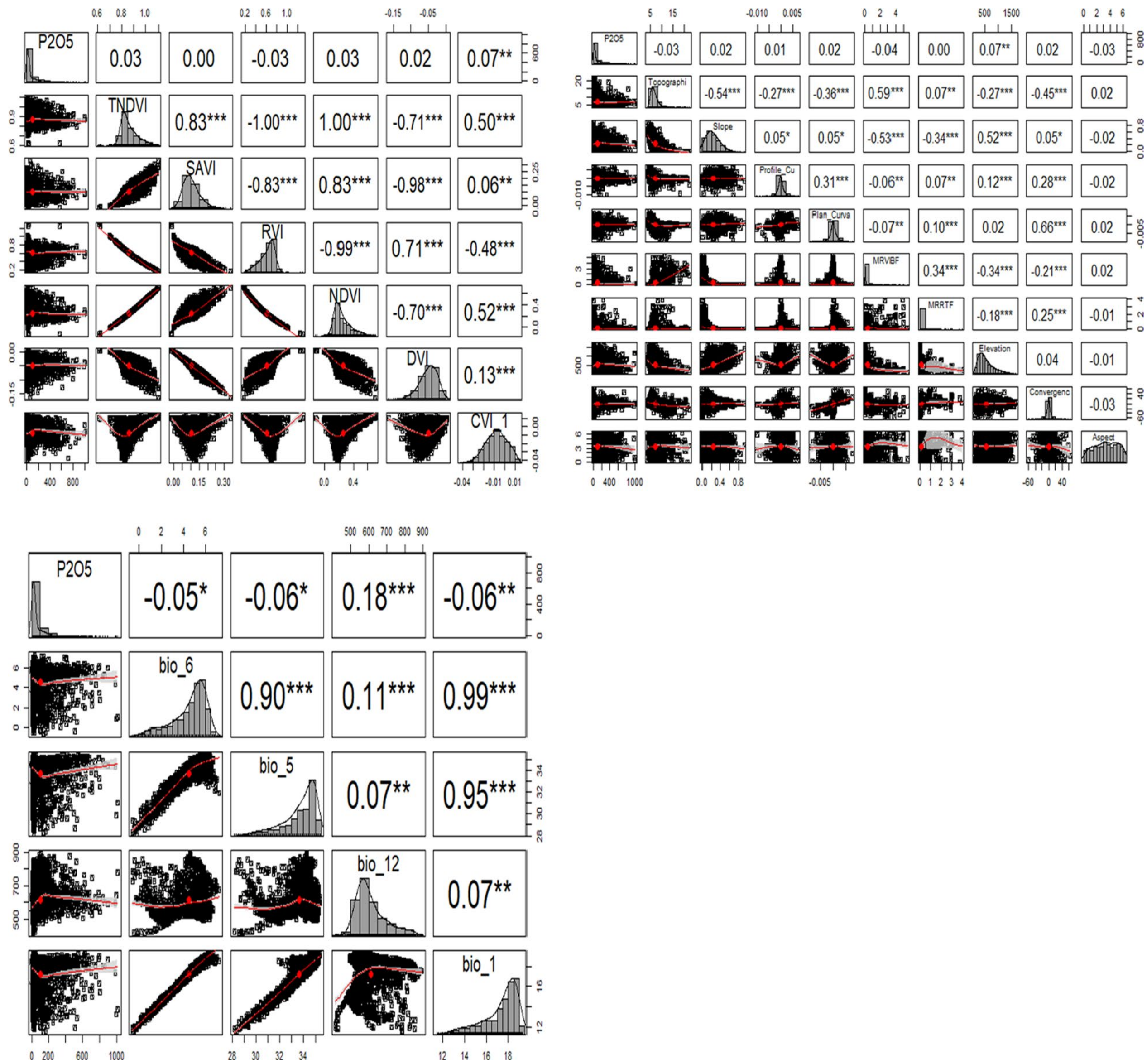

Fig. 4 Correlation between P_2O_5 and bioclimatic data, terrain attributes, and vegetation indices, respectively. The total number of observation points used = 1819

the smoothing effect. Areas with SOM (30–40%) are also clearly indicated compared to the OK map.

OK's interpolated potassium (K_2O) map revealed a smooth effect of soil potassium distribution in the region. K_2O ranged from 0–600 ppm. The highest K_2O concentration was found in the southern part, where lower pH in the area is concentrated. Besides that, the pH level in the south part is the optimum pH (6.6–7.2) for most nutrient elements. The K_2O map obtained via terrain attributes revealed that low K_2O (0–200 ppm) spreads from the central part of the studied region to the northern angles. K_2O map generated with climatic data showed that 500 ppm is the maximum K_2O in the study area. Similarly, the K_2O map by vegetation data produced a plain map with K_2O distribution showing a maximum prediction of 60,000 ppm. And also, by using all the covariates, a similar map was obtained using vegetation data was also obtained here.

In P_2O_5, OK prediction revealed that low P_2O_5 was located in the southern part, northeast and northwest direction. P_2O_5 between 400–600 ppm was observed traveling from west to east. Some patches of 200 ppm of P_2O_5 were found in the west direction. Applying KED,

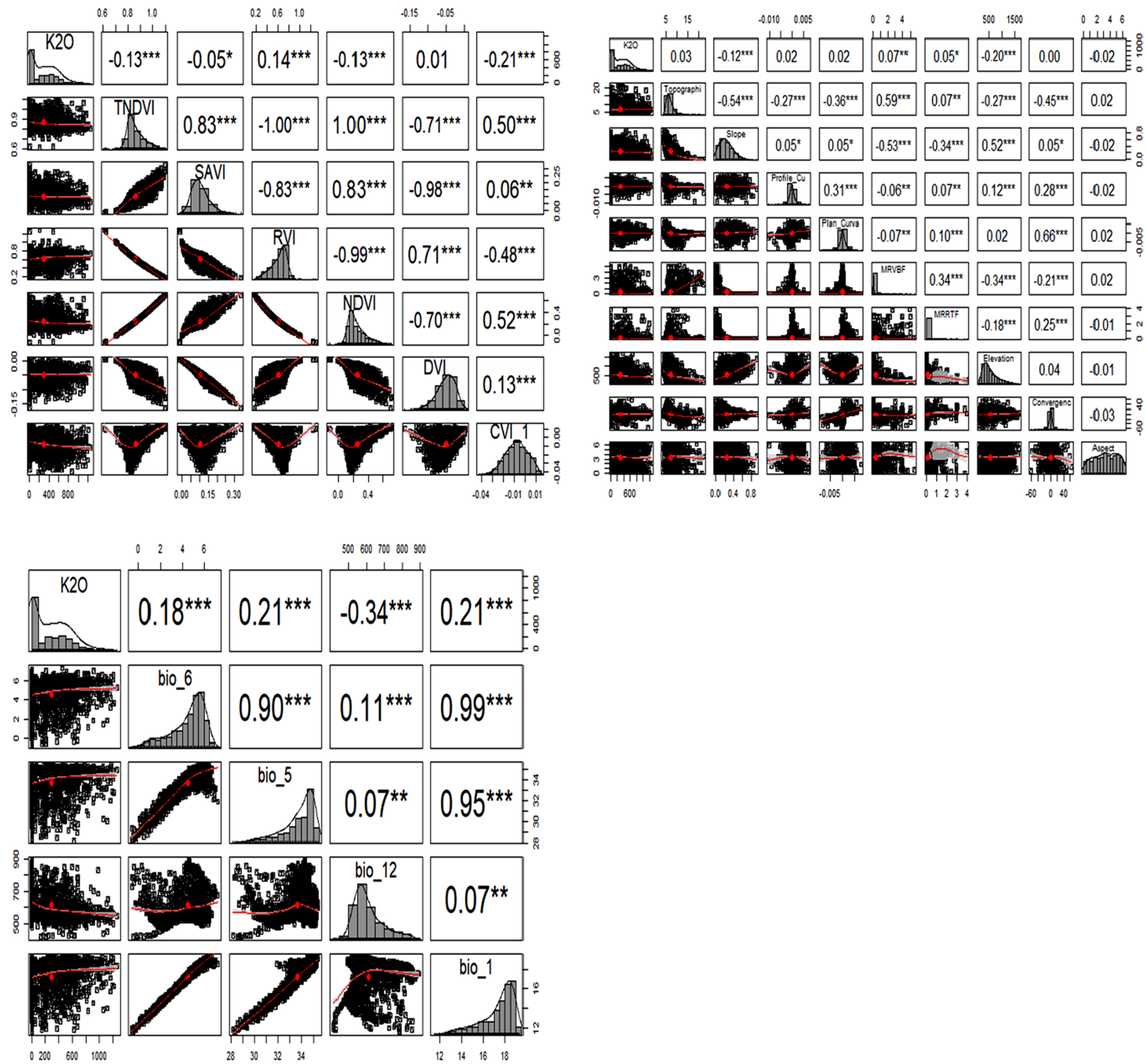

Fig. 5 Correlation between K_2O and bioclimatic data, terrain attributes, and vegetation indices, respectively. The total number of observation points used = 1819

P_2O_5 generated via terrain and climatic covariates followed OK's similar spatial distribution pattern. While using vegetation and all covariates (i.e., terrain, climate, and vegetation), P_2O_5 generated was plain with unclear patches.

Discussion

The results revealed that soils of the arid climates are predominantly alkaline with a high soil pH. In addition, Chytrý et al. (2007) reported that climatic factors regulate pH variation at a regional level. Therefore, the high SOM content in this study may considerably influence soil nutrients and physical properties (Bruun et al. 2013; Deng and Shangguan 2017).

Consequently, the high SOM explains an increase in phosphate mineralization and the absorption of K on the exchangeable site. On the other hand, the high kurtosis and skewness obtained in this study may be due to heavy tails or outliers in the dataset. Nevertheless, the result is similar to that estimated in Heihe River Basin, China (Li and Heap 2014). Also, this study's low but significant correlation is not surprising, as John et al. (2020) similarly reported low to no significant correlation between SOC,

Table 3 The semivariogram parameters of studied soil properties via OK (spatial autocorrelation) and KED (with external drifts from bioclimatic data, terrain attributes, and vegetative indices)

Variables	Modeling techniques	Model	Nugget (C_0)	Partial sill (C)	Sill (C_0+C)	Range (m)	Nugget/Sill	Spatial class	RMSE
pH	OK	Spherical	0.0324	0.0754	0.1078	5517.81	0.301	Moderate	0.283
	KED + C	Spherical	0.0320	0.0755	0.1075	5445.87	0.298	Moderate	**0.281**
	KED + V	Spherical	0.0324	0.0757	0.1081	5495.52	0.300	Moderate	0.281
	KED + T	Spherical	0.0319	0.0753	0.1072	5463.41	0.298	Moderate	0.284
	KED + A	Spherical	0.0320	0.0731	0.1051	5320.63	0.304	Moderate	0.329
SOM	OK	Spherical	42.400	77.387	119.787	5868.03	0.354	Moderate	9.54
	KED + C	Spherical	42.162	78.233	120.395	5690.85	0.350	Moderate	**9.505**
	KED + V	Spherical	44.535	76.677	121.212	5707.12	0.367	Moderate	9.653
	KED + T	Spherical	42.488	76.803	119.291	5891.19	0.356	Moderate	9.695
	KED + A	Spherical	43.614	74.576	118.19	5580.74	0.369	Moderate	9.722
K_2O	OK	Spherical	34,448.86	34,961.36	69,410.22	24,678.26	0.496	Moderate	211.79
	KED + C	Spherical	25,794.52	34,949.68	60,744.2	8667.65	0.425	Moderate	211.25
	KED + V	Spherical	31,936.70	32,486.23	64,422.93	12,001.12	0.496	Moderate	211.13
	KED + T	Spherical	32,597.23	33,524.03	66,121.26	17,689.55	0.493	Moderate	**209.764**
	KED + A	Spherical	27,326.64	33,242.60	60,569.24	7967.91	0.451	Moderate	212.952
P_2O_5	OK	Spherical	981.493	21,708.082	22,689.575	6492.08	0.043	Strong	106.737
	KED + C	Spherical	1024.801	21,660.461	22,685.262	6334.52	0.045	Strong	**106.511**
	KED + V	Spherical	998.211	21,804.537	22,802.748	6391.46	0.044	Strong	107.70
	KED + T	Spherical	1015.802	21,592.194	22,607.996	6479.01	0.045	Strong	106.894
	KED + A	Spherical	1305.767	21,051.336	22,357.103	6252.80	0.058	Strong	123.91

Bold figures are the best model in each category

Keys: KED + C = kriging with climate covariates as external drift; KED + V = kriging with vegetation covariates as external drift; KED + T = kriging with terrain derivatives as external drift; KED + A = kriging with all covariates as external.

P, and other soil nutrients with soil properties. John et al. (2020) and other studies have shown soil properties to correlate with environmental covariates, especially in low-relief areas.

The stronger spatial correlation of soil P_2O_5 may be attributed to structural factors, whereas a moderate spatial correlation of pH, SOM, and K_2O results from random factors. In their research, John et al. (2020) also obtained moderate spatial autocorrelation for pH and strong spatial autocorrelation for available phosphorus.

As presented in Table 3, the maximum distance in which spatial dependence or autocorrelation exists was defined as the range value of the semivariogram. The range values of soil properties in this study ranged between 5,320.63 m for pH predicted through universal kriging with all the covariates (KED + A) and 24,678.26 m for K_2O predicted via OK. Therefore, larger than the obtained range values in this study, spatial dependence does not exist for these soil properties in the study area.

According to Lopez-Granados et al. (2002), large range value indicated that estimated soil properties were significantly influenced by anthropogenic and natural factors over larger distances than the other soil properties which have smaller ranges. Therefore, the obtained result for ranges can be used for planning future soil sampling in the study area for geostatistical research by taking samples at interval distances less than half the obtained range values of the studied soil properties. Kerry and Oliver (2004) had reported that the distance between soil samples should be below half the semivariogram range value.

The pH map generated by combining all covariates (KED + A) was ultimately off, just like the map obtained from vegetation data. The pH legend scale > 14. However, the pH of the area with OK, KED + C, KED + V, and KED + T, respectively, fell between 0 and 10 (see Fig. 7A–D). On the global scale, rainfall and potential evapotranspiration have affected soil pH fluctuations (Slessarev et al. 2016). As a result, regional-scale effects of climate conditions on soil pH fluctuations have been documented (Brady and Weil 2002; Ji et al. 2014; Chytr et al. 2007). According to Ji et al. (2014), soil pH negatively correlates with mean temperature and mean precipitation. Chytrý et al. (2007) also discovered that as precipitation increases, soil pH decreases.

SOM map modeled with soil properties data followed a similar spatial distribution pattern with SOM (terrain data). SOM maps modeled with climate datasets followed

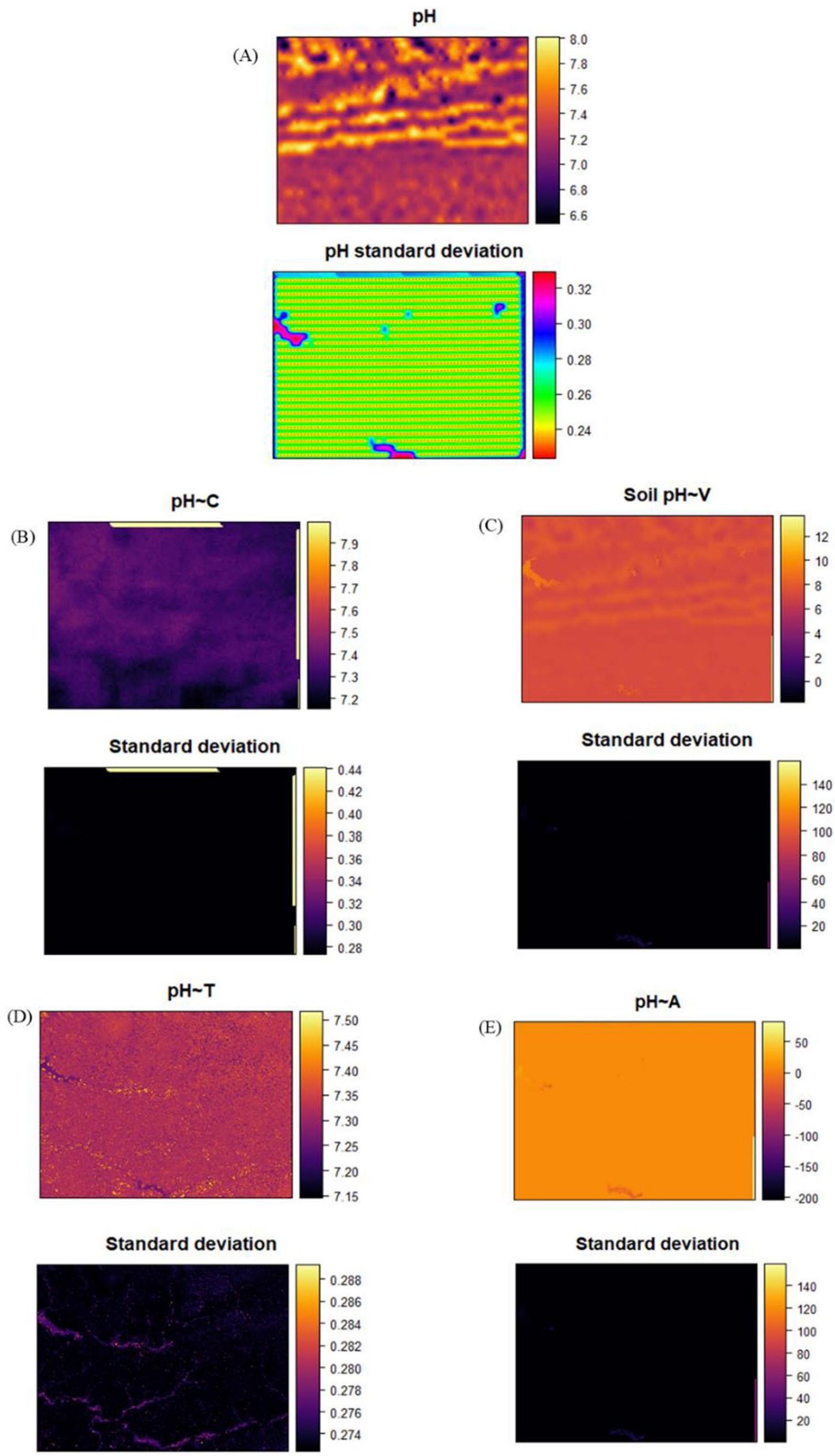

Fig. 6 pH maps via OK (spatial autocorrelation) and KED (with external drifts from bioclimatic data, terrain attributes and vegetative indices). Keys: KED + C = kriging with climate covariates as external drift; KED + V = kriging with vegetation covariates as external drift; KED + T = kriging with terrain derivatives as external drift; KED + A = kriging with all covariates as external.

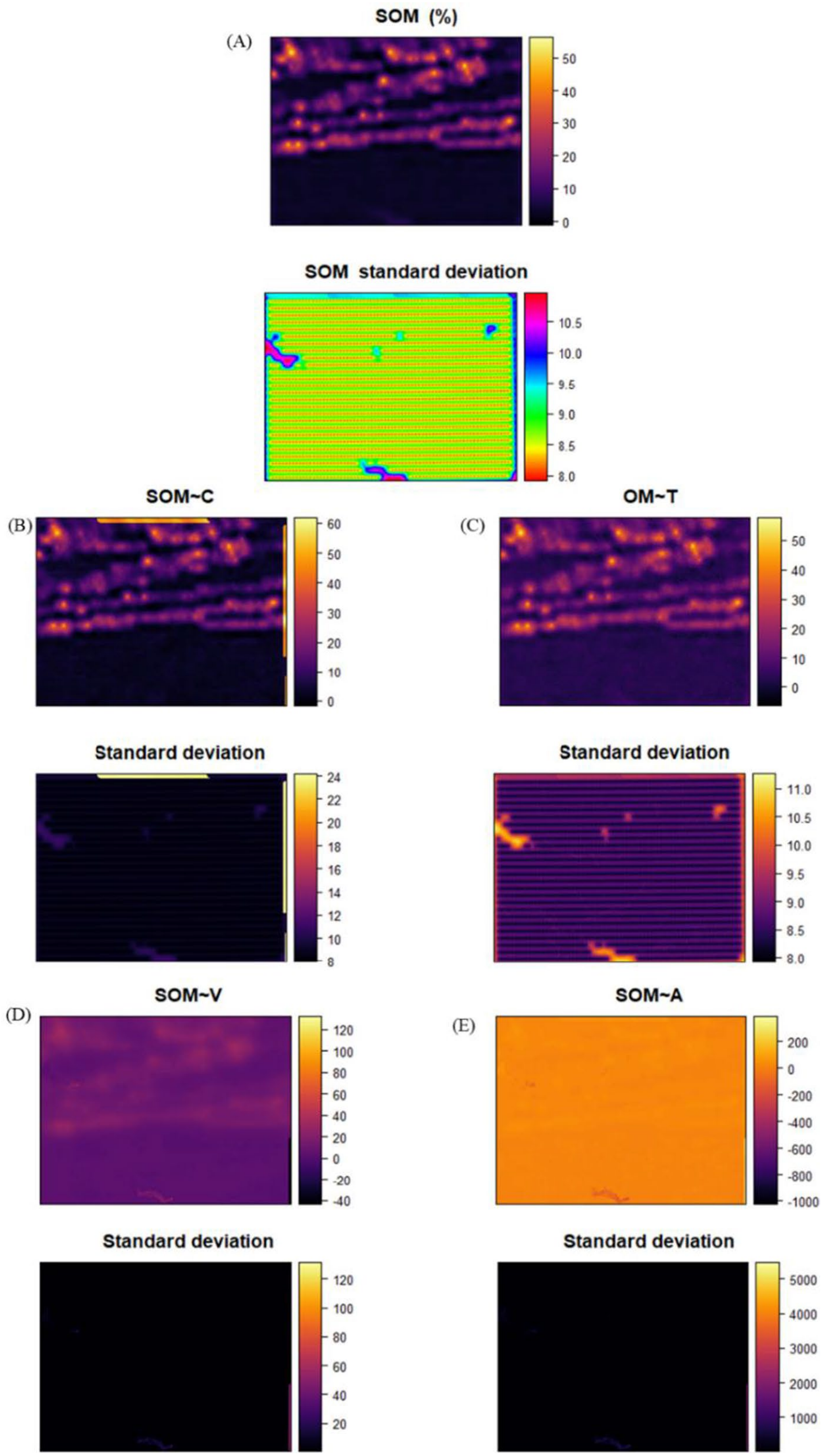

Fig. 7 SOM maps via OK (spatial autocorrelation) and KED (with external drifts from bioclimatic data, terrain attributes, and vegetative indices). Keys: KED + C = kriging with climate covariates as external drift; KED + V = kriging with vegetation covariates as external drift; KED + T = kriging with terrain derivatives as external drift; KED + A = kriging with all covariates as external

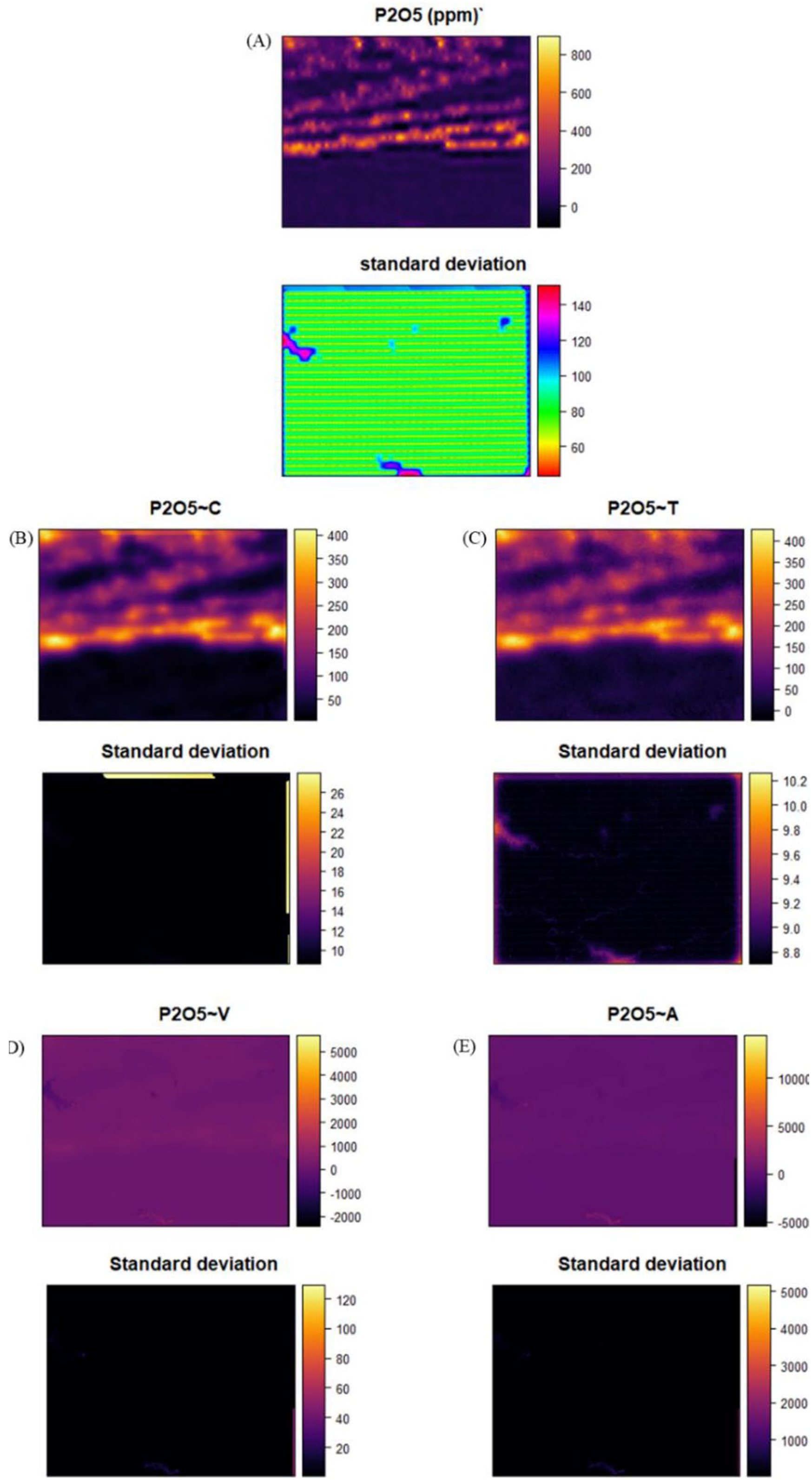

Fig. 8 P_2O_5 maps via OK (spatial autocorrelation) and KED (with external drifts from bioclimatic data, terrain attributes, and vegetative indices). Keys: KED + C = kriging with climate covariates as external drift; KED + V = kriging with vegetation covariates as external drift; KED + T = kriging with terrain derivatives as external drift; KED + A = kriging with all covariates as external

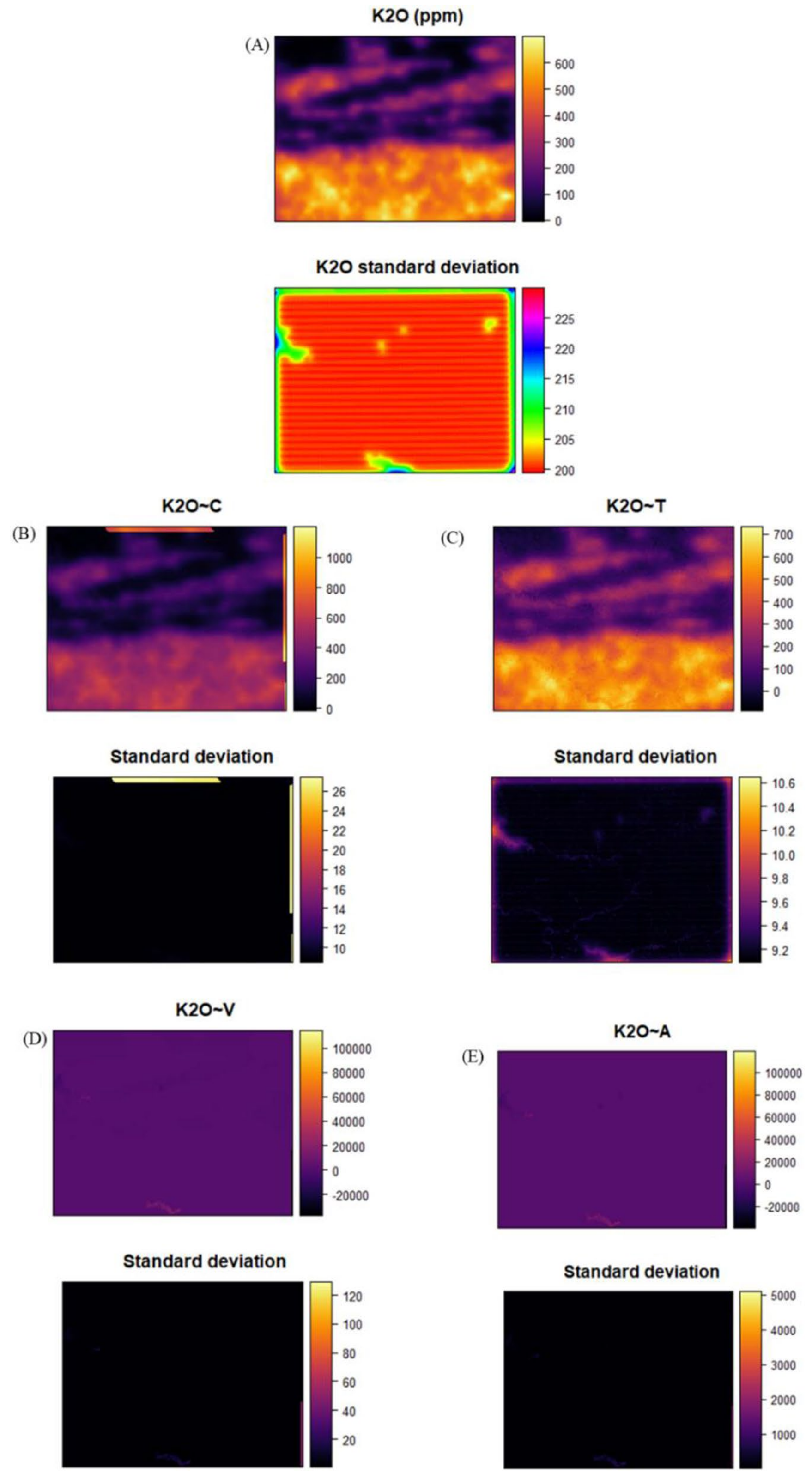

Fig. 9 K_2O maps via OK (spatial autocorrelation) and KED (with external drifts from bioclimatic data, terrain attributes, and vegetative indices). Keys: KED + C = kriging with climate covariates as external drift; KED + V = kriging with vegetation covariates as external drift; KED + T = kriging with terrain derivatives as external drift; KED + A = kriging with all covariates as external

a similar distribution trend. However, the conspicuous SOM of 10% revealed by terrain was not seen. The SOM resembled the OK map. In comparison, the prediction of SOM with all the covariates produced the worse map.

The maps obtained for the targeted soil properties, terrain, and climatic covariates showed consistency in their produced maps. In contrast, vegetation and the combination of all the covariates from different sources had similar maps. Therefore, we infer that terrain attributes may perform equally as soil properties and climate covariates in spatial prediction and mapping. While instead of combining covariates from all sources, importing radiometric errors, vegetation covariates may provide similar output.

Soil management recommendations

The generated predictive maps show that the region displays distinct variations of pH, SOM, K_2O, and P_2O_5 over space. Moreover, to appropriately manage the soil for continuous cropping, we recommend dividing areas into two independent divisions (northern and southern) based on the most accurate map. The region characterized by moderate acid conditions should be engaged for forage crops. And the general application of reduced or increased quantities of lime, based on average soil pH values to stabilize uniform soil acidity would not be an appropriate amendment approach to this field. On the other hand, SOM would increase pH value when properly managed. Furthermore, elevated SOM would increase soil nutrient delivery and soil buffering capability. Likewise, one method for improving SOM is to use organic fertilizers, charcoal, compost, and crop rotation.

Conclusion

The study observed that pH, SOM, and P_2O_5 spatial variation with the climatic dataset as external drift yielded the best model. K_2O was accurately modeled with terrain attributes. In addition, the study demonstrated that the relationship between targeted soil property and the covariates was linear, and the prediction was feasible because the covariates were known in all the locations where targeted soil property was unsampled.

In conclusion, specific covariates from the same source other than many covariates can be successfully applied to estimate the Mediterranean region's spatial variation in soil properties. Also, this study would guide the establishment of national and regionalized specific raster maps for soil nutrient management programs.

Acknowledgements

The present research data belong to the National Institute of Agricultural Research (INRA). The authors are grateful to the Department of Soil Science and Soil Protection Staff, Faculty of Agrobiology, Food, and Natural Resources, Czech University of Life Sciences, Prague, and Head of Department, Professor Luboš Borůvka. In addition, the support from the Ministry of Education, Youth and Sports of the Czech Republic (project No. CZ.02.1.01/0.0/0.0/16_019/0000845) is also acknowledged.

Authors' contributions

JK wrote the manuscript. YB, HL and RR conceived the experiments. JK, YB, HL, RR, KN, IAI and AP conducted the experiments. JK and YB interpreted data. PV and ZT guided the research. PV and ZT revised the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by an internal Ph.D. grant no. SV20-5-21130 of the Faculty of Agrobiology, Food and Natural Resources of the Czech University of Life Sciences Prague (CZU). And also, NutRisk grant: European Regional Development Fund, project Center for the investigation of synthesis and transformation of nutritional substances in the food chain in interaction with potentially harmful substances of anthropogenic origin: comprehensive assessment of soil contamination risks for the quality of agricultural products, number CZ.02.1.01/0.0/0.0/16_019/0000845.

Availability of data and materials

The datasets used for the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food, and Natural Resources, Czech University of Life Sciences, Kamýcká 129, 16500 Prague, Czech Republic. [2]National Institute of Agricultural Research, Rabat, Morocco. [3]Department of Soil Science, University of Calabar, Calabar, Nigeria.

Received: 26 December 2021 Accepted: 6 February 2022
Published online: 19 February 2022

References

Agyeman PC, Ahado SK, Borůvka L, Biney JKM, Sarkodie VYO, Kebonye NM, Kingsley J (2021) Trend analysis of global usage of digital soil mapping models in the prediction of potentially toxic elements in soil/sediments: a bibliometric review. Environ Geochem Health 43:1715–1739. https://doi.org/10.1007/s10653-020-00742-9

Aksoy E, Panagos P, Montanarella L (2012) Spatial prediction of soil organic carbon of Crete by using geostatistics. In: Minasny B, Malone BP, McBratney AB, eds. Digital soil assessments and beyond—proceedings of the Fifth Global Workshop on digital soil mapping, pp 149–153. https://doi.org/10.1201/b12728-31

Allali A, Rezouki S, Lougraimzi H, Touati N, Eloutassi N, Fadli M (2020) Agricultural traditional practices and risks of using insecticides during seed storage in Morocco. Plant Cell Biotechnol Mol Biol 21(40):29–37

Amalu UC, Isong IA (2015) Land capability and soil suitability of some acid sand soil supporting oil palm (*Elaeis guinensis* Jacq) trees in Calabar, Nigeria. Niger J Soil Sci 25:92–109

Balkovič J, Rampašeková Z, Hutár V, Sobocká J, Skalský R (2013) Digital soil mapping from conventional field soil observations. Soil Water Res 8:13–25

Borůvka L, Vašát R, Němeček K, Novotný R, Šrámek V, Vacek O, Pavlů L, Fadrhonsová V, Drábek O (2020) Application of regression-kriging and sequential Gaussian simulation for the delineation of forest areas potentially suitable

for liming in the Jizera Mountains region, Czech Republic. Geoderma Reg 21:e00286
Bourennane H, King D, Couturier A (2000) Comparison of kriging with external drift and simple linear regression for predicting soil horizon thickness with different sample densities. Geoderma 97(3–4):255–271
Bouslihim Y, Rochdi A, El Amrani Paaza N (2021a) Machine learning approaches for the prediction of soil aggregate stability. Heliyon 7(3):e06480. https://doi.org/10.1016/j.heliyon.2021.e06480
Bouslihim Y, Rochdi A, Aboutayeb R, El Amrani-Paaza N, Miftah A, Hssaini L (2021b) Soil aggregate stability mapping using remote sensing and GIS-based machine learning technique. Front Earth Sci 9:748859
Brady NC, Weil RR (2002) The nature and properties of soils, 15th edn. Pearson Education. pp 375–419
Bruun T, Elberling B, Neergaard A, Magid J (2013) Organic carbon dynamics in different soil types after conversion of forest to agriculture. Land Degrad Dev 26(3):272–283
Carré F, Girard MC (2002) Quantitative mapping of soil types based on regression kriging of taxonomic distances with landform and land cover attributes. Geoderma 110(3–4):241–263
Chytrý M, Danihelka J, Ermakov N, Hájek M, Hájková P, Kočí M, Kubešová S, Lustyk P, Otýpková Z, Popov D, Roleček J (2007) Plant species richness in continental southern Siberia: effects of pH and climate in the context of the species pool hypothesis. Glob Ecol Biogeogr 16(5):668–678
Davis JC (1973) Statistics and data analysis in geology. Wiley Interscience, New York, p 550
Deng L, Shangguan ZP (2017) Afforestation drives soil carbon and nitrogen changes in China. Land Degrad Dev 28(1):151–165
FAO/ISRIC/ISSS (2006) World Reference Base for Soil Resources. World Soil Resources Report. No. 84. FAO, Rome
Fick SE, Hijmans RJ (2017) WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. Int J Climatol 37(12):4302–4315
Goovaerts P (2001) Geostatistical modelling of uncertainty in soil science. Geoderma 103(1–2):3–26
Hengl T, Miller MA, Križan J, Shepherd KD, Sila A, Kilibarda M, Antonijević O, Glušica L, Dobermann A, Haefele SM, McGrath SP (2021) African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. Sci Rep 11:6130
Hengl T, Geuvelink GBM, Stein A (2004) Comparison of kriging with external drift and regression-kriging. Technical note, ITC, http://www.itc.nl/library/Academic.output
Hudson G, Wackernagel H (1994) Mapping temperature using kriging with external drift: theory and an example from Scotland. Int J Climatol 14(1):77–91
Ji CJ, Yang YH, Han WX, He YF, Smith J, Smith P (2014) Climatic and edaphic controls on soil pH in alpine grasslands on the Tibetan Plateau, China: a quantitative analysis. Pedosphere 24(1):39–44
John K, Isong IA, Kebonye MN, Ayito EO, Agyeman PC, Afu SM (2020) Using machine learning algorithms to estimate soil organic carbon variability with environmental variables and soil nutrient indicators in an alluvial soil. Land 9:487. https://doi.org/10.3390/land9120487
John K, Afu SM, Isong IA, Aki EE, Kebonye NM, Ayito EO, Chapman PA, Eyong MO, Penížek V (2021a) Mapping soil properties with soil-environmental covariates using geostatistics and multivariate statistic. Int J Environ Sci Technol 18(11):3327–3342. https://doi.org/10.1007/s13762-020-03089
John K, Isong IA, Kebonye MN, Agyeman CP, Ayito EO, Kudjo AS (2021b) Soil organic carbon prediction with terrain derivatives using geostatistics and sequential Gaussian simulation. J Saudi Soc Agric Sci 20(6):379–389. https://doi.org/10.1016/J.JSSAS.2021.04.005
Kerry R, Oliver M (2004) Average variograms to guide soil sampling. Int J Appl Earth Obs Geoinf 5:307–325
Laghrour M, Moussadek R, Mrabet R, Dahan R, El-Mourid M, Zouahri A, Mekkaoui M (2016) Long and midterm effect of conservation agriculture on soil properties in dry areas of Morocco. Appl Environ Soil Sci 2016:6345765. https://doi.org/10.1155/2016/6345765
Li J, Heap AD (2014) Spatial interpolation methods applied in the environmental sciences: a review. J Envsoft 53:173–189
López-Granados F, Jurado-Expósito M, Atenciano S, García-Ferrer A, de la Orden MS, García-Torres L (2002) Spatial variability of agricultural soil parameters in southern Spain. Plant Soil 246:97–105
McBratney AB, Odeh IO, Bishop TF, Dunbar MS, Shatar TM (2000) An overview of pedometric techniques for use in soil survey. Geoderma 97(3–4):293–327
McBratney AB, Santos MM, Minasny B (2003) On digital soil mapping. Geoderma 117(1–2):3–52
McKenzie NJ, Ryan PJ (1999) Spatial prediction of soil properties using environmental correlation. Geoderma 89:67–94
Mesrar H, Sadiki A, Faleh A, Quijano L, Gaspar L, Navas A (2017) Vertical and lateral distribution of fallout ^{137}Cs and soil properties along representative toposequences of central Rif, Morocco. J Environ Radioact 169:27–39
Moore ID, Gessler PE, Nielsen GAE, Peterson GA (1993) Soil attribute prediction using terrain analysis. Soil Sci Soc Am J 57(2):443–452
Moussadek R (2014) Status of soil survey and soil information system in Morocco. Global soil partnership. http://www.fao.org/fileadmin/user_upload/GSP/docs/NENA2014/Morocco.pdf. Accessed 18 Jul 2021
Nabyl B, Hanane L, Houda E, Zakaria A, Otman H, Rahali K, Aouane E (2020) Evaluation of the physico-chemical properties of soil and apple leaves (*Malus domestica*) in Beni MellalKhenifra Region, Morocco. Adv Sci Technol Eng Syst J 5(6):1103–1108
Odeh IO, Todd AJ, Triantafilis J (2003) Spatial prediction of soil particle-size fractions as compositional data. Soil Sci 168(7):501–515
Okon PB, Nwosu NJ, Isong IA (2019) Response of soil sustainability indicators to the changing weather patterns in Calabar, Southern Nigeria. Niger J Soil Sci 29(1):52–61
Olaya V (2004) A Gentle Introduction to SAGA GIS. The SAGA User Group Press, Gottingen, pp 1–216
Papritz A, Stein A (1999) Spatial prediction by linear kriging. In: Stein A, Van der Meer F, Gorte B (eds) Spatial statistics for remote sensing. Springer, Dordrecht, pp 83–113
Penížek V, Boruka L (2006) Soil depth prediction supported by primary terrain attributes: a comparison of methods. Plant Soil Environ 52:424–430
Penížek V, Zádorová T, Kodešová R, Vaněk A (2016) Influence of elevation data resolution on spatial prediction of colluvial soils in a Luvisol region. PLoS ONE 11(11):e0165699
Rezouki S, Allali A, Louasté B, Eloutassi N, Fadli M (2021) Physico-chemical evaluation of soil resources in different regions of Taza-Taounate, Morocco. Mediterr J Chem 11(1):1–9
Santra P, Kumar M, Panwar N (2017) Digital soil mapping of sand content in arid western India through geostatistical approaches. Geoderma Reg 9:56–72
Slessarev EW, Lin Y, Bingham NL, Johnson JE, Dai Y, Schimel JP, Chadwick OA (2016) Water balance creates a threshold in soil pH at the global scale. Nature 540(7634):567–569
Smith JL, Halvorson JJ, Bolton H Jr (2002) Soil properties and microbial activity across a 500 m elevation gradient in a semi-arid environment. Soil Biol Biochem 34(11):1749–1757
Udo EJ, Ibia TO, Ogunwale JA, Ano AO, Esu IE (2009) Manual of soil, plant and water analysis. Sibon Books Publishers Ltd, Nigeria, p 183
Wackernagel H, Bertino L, Sierra JP, del Río JG (2002) Multivariate kriging for interpolating with data from different sources. In: Anderson CW, Barnett V, Chatwin PC, El-Shaarawi AH (eds) Quantitative methods for current environmental issues. Springer, London, pp 57–75
Webster R, Oliver MA (2007) Geostatistics for environmental scientists. John Wiley & Sons, UK
Zadorova T, Jakšík O, Kodešová R, Penížek V (2011) Influence of terrain attributes and soil properties on soil aggregate stability. Soil Water Res 6(3):111–119

Publisher's Note