Heliyon 10 (2024) e40128

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Identifying hidden factors influencing soil Olsen-P in an alkaline calcareous soil using machine learning and geostatistical techniques

Moussa Bouray [a,b,*], Mohamed Bayad [b,f], Adnane Beniaich [a,b], Ahmed G. El-Naggar [d], Rebecca Logsdon Muenich [e], Khalil El Mejahed [a,b], Abdallah Oukarroum [b,c], Mohamed El Gharous [a,b]

[a] Agricultural Innovation and Technology Transfer Center (AITTC), Mohammed VI Polytechnic University (UM6P), Ben Guerir, Morocco
[b] College of Agriculture and Environmental Sciences (CAES), Mohammed VI Polytechnic University, Ben Guerir, Morocco
[c] AgroBioSciences (AgBS), Plant Stress Physiology Laboratory, Mohammed VI Polytechnic University (UM6P), Ben Guerir, Morocco
[d] Land and Water Management Department, IHE Delft Institute for Water Education, Delft, the Netherlands
[e] Department of Biological & Agricultural Engineering, University of Arkansas, Fayetteville, AR, USA
[f] Center for Remote Sensing Applications (CRSA), Mohammed VI Polytechnic University, Ben Guerir, Morocco

ARTICLE INFO

Keywords:
Kriging
Soil mapping
Random forest
Cubist regression
Partial least squares regression
Calcareous soil

ABSTRACT

Phosphorus (P) deficiency is one of the major constraints for sustainable crop production in calcareous soils. This study aimed to elucidate the key soil characteristics modulating the variability of soil Olsen P in these typical soils. A comprehensive soil sampling initiative (1.5 samples per hectare) was conducted on a 100-ha farm, considering 31 attributes that included soil physical and chemical properties, and geographic attributes. Three machine learning algorithms—partial least squares regression (PLSR), random forest (RF), and cubist regression (CR)—were employed to understand key variables controlling soil Olsen P. Furthermore, the same data set was used to spatially map the variations in Olsen P levels using ordinary kriging. The results revealed that soil chemical factors, specifically exchangeable manganese and zinc, cation exchange capacity, and carbonate, played a crucial role in controlling P levels. Among the machine learning models, the best performing model was RF ($R^2 = 0.95$, RMSE = 1.30 mg kg^{-1}) followed by CR ($R^2 = 0.92$ and RMSE = 1.43 mg kg^{-1}). Additionally, the analysis using a Gaussian semivariogram model showed a good performance ($R^2 = 0.78$, RMSE = 2.05 m) in visualizing the spatial distribution of Olsen P, revealing its heterogeneity. The resulting pattern of Olsen P distribution may be attributed not only to soil properties but also to external factors, such as sediment transport through watercourses across the study area and atmospheric deposition from a nearby P mining site. Overall, the combination of geostatistical methods and machine learning approach demonstrates a significant potential in understanding the complexity of soil available P (Olsen-P) that could help to develop sustainable and precise P management.

* Corresponding author. Agricultural Innovation and Technology Transfer Center (AITTC), Mohammed VI Polytechnic University (UM6P), Ben Guerir, Morocco.
E-mail address: moussa.bouray@um6p.ma (M. Bouray).

https://doi.org/10.1016/j.heliyon.2024.e40128
Received 29 March 2024; Received in revised form 24 September 2024; Accepted 4 November 2024
Available online 5 November 2024

1. Introduction

Phosphorus (P) is an essential macro-nutrient that often limits plant growth and development in natural and agricultural ecosystems [1], especially in calcareous soils which represent almost one-third of the world's land surface area [2]. These soils are concentrated mainly in arid and semi-arid areas such as the Mediterranean region and are characterized by the presence of calcium carbonate ($CaCO_3$) in the soil parent material and the consequent accumulation of free $CaCO_3$ in the soil profile. The pH of these soils is usually high (between 7 and 8.5) depending on carbonate concentration [3], and sometimes can exceed 9 when these soils contain sodium carbonate. In some soils, $CaCO_3$ can concentrate into very hard layers, termed "caliche" or petrocalic horizons, that are impermeable to water infiltration and plant root penetration [4].

Nutrient availability in calcareous soils can severely be limited leading to reduced yields and nutrient deficiencies such as P deficiency because sorption and precipitation reactions are strongly favored in calcareous soils due to high concentrations of $CaCO_3$. This increases the sorption strength of the soil and subsequently its P buffer capacity, and precipitation reactions forming less soluble compounds like stable calcium (Ca) phosphates [5–7]. In addition to $CaCO_3$, calcareous soils may also contain significant amounts of iron (Fe), aluminium (Al), and manganese (Mn) either as discrete minerals, as coatings on soil particles, or complexed with soil organic matter [8]. These metals may provide strong sorption sites for P and therefore play a more significant role in controlling available P in calcareous soils than $CaCO_3$ itself [8]. Thus, their importance should not be overlooked, especially Mn [9]. Thus, there exists a research gap in identifying key soil properties, other than lime and pH, that contribute to controlling available P in alkaline-calcareous soils.

Phosphorus concentration in the soil is affected by complex nonlinear influences from several pedogenic factors including soil properties, climate, organisms, relief (topography), parent material, age (time) and position (location) [10–12]. Thus, the identification of the most important factors affecting soil available P (Olsen P) and P behavior in calcareous soils necessitates using techniques that prioritize the complex relationships [13], and highlights the poorly visible and nonlinear interactions missed with the simpler statistical approaches. Here, we thus choose to apply machine learning (ML) techniques which are a subset of artificial intelligence and an important component of Agriculture 4.0 or digital agriculture which refers to the integration of advanced technologies and data-driven approaches in the agribusiness production chains [14]. Modelling via ML is widely used in soil science to understand soil properties and processes, formulate prediction models and for digital soil mapping as reviewed by Wang et al. [15] and Padarian et al.

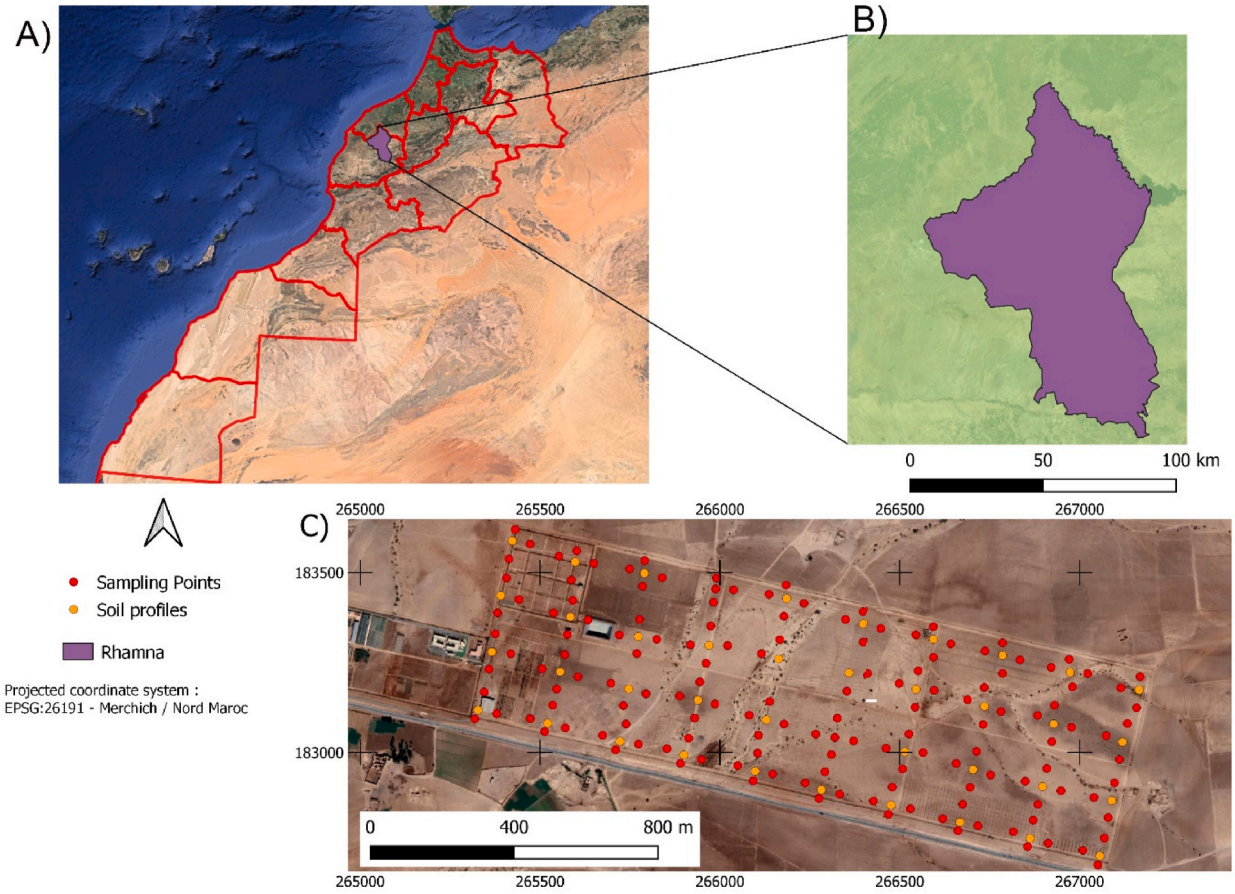

Fig. 1. Location of the study area (A and B) (Rhmana, Center-West of Morocco), and sampling sites (C).

[16]. Among the most frequently utilized ML methods in soil P modelling are partial least squares regression (PLSR) [17], random forest (RF) [18] and cubist regression (CR) [19].

Additionally, identifying spatial distribution patterns and developing accurate digital maps of soil nutrients like P is a crucial step in understanding the underlying factors and mechanisms involved in their spatial variability and examining their origins, providing that P spatial distribution and heterogeneity could also be a result of external exchange with the atmosphere and the biosphere [19,20]. In recent years, geostatistics has been extensively used for spatial prediction of soil properties for unsampled sites from point observations based on the concept that objects that are near each other will be more similar (spatial auto-correlation) [21,22]. This is particularly important because laboratory analyses of some soil properties are expensive and time-consuming. Ordinary kriging (OK) is one of the regularly used geostatistical techniques and was thus applied for the interpolation of various soil properties, including P [19,23,24].

In this study, a detailed soil sampling survey was conducted over a 100-ha area characterized by calcareous soils with high alkalinity. The primary goal was to identify which soil characteristics are most important in explaining the variation in soil Olsen P using a modelling approach. The specific objectives were.

I To evaluate a set of soil properties from the collected samples, including geographic covariates, to enhance the modelling accuracy.
II To use three different machine learning algorithms to assess the importance of the key covariates and rank them by their impact on explaining Olsen P variability.
III To use ordinary kriging with semi-variogram modelling to map soil Olsen P, analyzing its spatial distribution and identifying external factors influencing its variability.

2. Materials and methods

2.1. Study area

The research was conducted in the Rhmana region which is located in the Center-West of Morocco. (Fig. 1). Rhamna covers 5877 km^2 with a population of 315,077 people. It is at approximately 700 m in altitude and 150 km from the Atlantic Ocean and is characterized by a dry climate with inadequate and poorly distributed rainfall varying between 150 and 350 mm annually [25]. The study area is an experimental farm of Mohammed 6 Polytechnic University (UM6P) in Benguerir city (Fig. 1). The farm covers a total area of 110 ha and falls within the latitude 32° 13′ 6′′ North and longitude 7° 52′ 43′′ West. Before 2016, the study area was primarily utilized for sheep grazing, and there was no application of fertilizers. For the past six years, the study area has been utilized as an experimental farm. Approximately 6–10 % of the surface area is now dedicated to field trials, where various crops such as annuals (quinoa, barely, fababean) or permanent plantations (cactus, carobs, olives) have been cultivated. The soil at the site is classified as a calcimagnesic (Frensh classification after Aubert [26]; the International Union of Soil Sciences (IUSS), classification: Calcisol Petrocalcic [27]). The daily mean temperature and rainfall in the farm during the last five years from May 2019 to June 2023 are given in Fig. S1. During that period, the maximum and minimum average daily temperatures were 28 °C and 13 °C respectively, and the average annual rainfall was 137 mm.

2.2. Soil sampling and analysis

Soil samples were taken from 153 locations (n = 153), and collected from the whole experimental farm to cover an area of 100 ha (Fig. 1C). The sampling was conducted in January 2016 at different depths varying between 15 and 50 cm depending on the thickness of the soil due to the presence of the limestone bedrock known as “caliche”. The sampling was done manually using a hoe and a spade. The samples were collected according to a systematic unaligned sampling protocol [28]: 40 soil profiles were created for soil characterization (not the objective of this study) and each profile represents a grid of 2.5 ha. The soil profiles were systematically distributed, with a horizontal spacing of 200 m and a vertical spacing of 150 m between each profile (Fig. 1C). Two to four soil samples were then randomly collected at 50 m from each profile. Upon collection, the soils were sieved at 5 mm to remove plant material and coarse rock fragments, air-dried at 40 °C and then passed through a 2 mm sieve to separate the gravel (>2 mm) from the rest of the soil. The sieved soils were then stored in plastic bags at room temperature to await the analysis, and the gravel was weighed and discarded.

The collected samples were analyzed for a variety of chemical and physical characteristics. Soil pH was measured with a pH probe (InoLab™ 7310 pH meter, Xylem Analytics, Germany) using a 1:5 soil: deionized water or KCl ratio. Electrical conductivity (EC) was measured in deionized water (1:5 soil: deionized water ratio) using an EC meter (SevenCompact, Mettler Toledo, USA). Soil exchangeable cations (Na, Ca, K, and Mg) were extracted by 1 M ammonium acetate and analyzed using Atomic Absorption Spectroscopy (Agilent Technologies. 200 Series AA, Santa Clara, USA) according to NF X31-108 method. Ammonium, nitrate, sulphate, and chlorine contents were quantified using a continuous flow analyzer (Skalar Analytica, Breda, Netherlands). Available phosphorus (Olsen P) was quantified as described by Olsen [29]. Bioavailable micro-elements (Fe, Cu, Mn, Zn, B) were extracted with ammonium acetate in the presence of DTPA (Diethylene triamine penta-acetic acid), and the concentrations were determined using Atomic Absorption Spectroscopy according to NFX 31–121 method. The hydrometer method proposed by Bouyoucos [30] was used to determine soil particle size distribution. Soil organic matter was estimated from the soil organic carbon content multiplied by the Van Bemmelen factor of 1.724 based on the assumption that organic matter contains 58 % organic C [31]. Determination of soil organic carbon was performed using sulfo-chromic oxidation of carbon in a mixture of potassium dichromate ($K_2Cr_2O_7$) and sulfuric acid (H_2SO_4) at 135 °C according to Walkley and Black [32]. Calcium carbonate was estimated from the carbon dioxide (CO_2) volume obtained from the

reaction between soil carbonate and hydrochloric acid [33]. Soil total nitrogen was determined by the distillation of ammonia gas after mineralization of the sample using the Kjeldahl method [34]. For cation exchange capacity (CEC), the soils were extracted with cobaltihexamine chloride and the cations were analyzed using atomic absorption spectroscopy as described by AFNOR [35].

2.3. Machine learning (ML) algorithms

In the present study, data regarding the potential controlling predictors of soil Olsen P (response variable) were acquired from each sampling site. In total, 31 predictor variables were included in the modelling. The abbreviations and definitions of the 31 variables are summarized in Table 1. These variables were divided into seven classes: geographic, granulometry, macro-nutrients, micro-nutrients, sampling depth, exchangeable cations and other key soil chemical properties like OM, CEC and pH and total carbonate. The three geographic variables (longitude, latitude, and elevation) were determined by using a handheld global positioning system (Table 1). The measurements of soil variables are described in the previous section. Three ML algorithms were used to assess and identify the most important variables in explaining soil Olsen P variation, including partial least squares regression (PLSR) random forest (RF), and cubist regression (CR) models.

Partial least squares regression (PLSR) is a robust multivariate regression approach that allows the users to explore the relationship between two matrices, $Xm \times n$, which consists of m variables (columns) and n objects (rows), and a response vector $Yn \times 1$, thus revealing the relative importance of the different x-variables in explaining y-variable (s) (Olsen P in our case). Moreover, using PLSR based on multivariate statistical projections can overcome the limitations of traditional multivariate regression approaches like collinearity and noise in the data. Detailed information on the theory, principles, and application of the PLSR approach can be found in previous literature [36–38]. In PLSR modelling, the importance of a predictor for both the independent and dependent variables is determined by the variable importance for the projection (VIP) [38]. Terms with high VIP values are the most relevant for explaining the dependent variable. In addition, the PLSR coefficients (RCs) show the direction of the relationship between each of the individual predictors and the dependent variable; if the coefficient is positive, it suggests a positive relationship; if it's negative, it suggests a negative relationship. Moreover, the relevant predictors could be selected according to the magnitude of the absolute values of regression coefficients [37].

Cubist regression (CR) model is a rule-based model, developed by Quinlan [39] as an extension of the M5 tree model. Cubist constructs an unconventional type of regression tree, where the prediction is based on linear regression models instead of discrete values [40]. The final model of Cubist is a set of comprehensible rules, where each rule has an associated multivariate linear model [41]. Whenever a situation satisfies the conditions of the rule, the associated model is used to calculate the predicted value. The main

Table 1
Abbreviations and descriptions of the selected variables.

Variables	Variable	Abbreviation	Unit
Geographic variables	East longitude	E	∘
	Northern latitude	N	∘
	Elevation	Ele	m
Granulometry	Gravel	Gravel	%
	Clay	Clay	%
	Fine silt	Fsilt	%
	Coarse silt	Csilt	%
	Fine sand	Fsand	%
	Coarse sand	Csand	%
Sampling depth	Depth	Depth	cm
Exchangeable cations	Calcium	Ca	mg/kg
	Potassium	K	mg/kg
	Magnesium	Mg	mg/kg
	Sodium	Na	mg/kg
Miro-nutrients	Copper	Cu	mg/kg
	Manganese	Mn	mg/kg
	Iron	Fe	mg/kg
	Zinc	Zn	mg/kg
	Boron	B	mg/kg
Macro-nutrients	Total N	TN	%
	Nitrate	NO3	mg/100g
	Ammonium	NH4	mg/100g
	Available P	Olsen P	mg/kg
	Sulphate	SO3	mg/100g
Other key soil chemical properties	C: N ratio	C:N	–
	Organic matter	OM	%
	Cation exchange capacity	CEC	meq/100g
	Electrical conductivity	EC	ms/cm
	Chloride	Cl	mg/100g
	pH_{water}	pHw	–
	pH_{KCl}	pHKcl	–
	Total carbonate	CO3	%

advantage of the Cubist method is to add multiple training committees and boostings to make the weights more balanced. In CR, neighbors and committees are two hyperparameters that have the largest effect on the performance of the final cubist model.

Random forest (RF) is a ML algorithm that can handle a large subset of continuous and categorical variables, but are more robust with respect to noise and does not require a pre-selection of variables [42]. The RF is constructed of a number of regression trees and each tree performs its own individual prediction, then the prediction of the model is averaged using the prediction of each tree. The number of trees or decision trees needs to be large to reduce overfitting and enhance the performance of the model. In RF, the number of trees (n_{tree}) and the number of variables that are sampled at each split (m_{try}) are the two hyperparameters that users usually modify to regulate the complexity of the models.

These ML algorithms used in the study were performed using R version 4.2.2 [43]. Partial least squares regression was carried out using "pls" package; all variables used in the PLS analysis were centered and scaled to unit variance, and the optimum number of PLS components corresponds to the first minimum for the prediction error from the full cross-validation (leaves out only one sample at the time). The cubist algorithm was performed using the "cubist" package [44], and the following parameters we considered: the number of rules by node was 5, the number of extrapolation was 5 and the number of the committees was 1. Random forest was performed using the "randomForest" package [45], the selected number of trees in the model was 1000 ($n_{tree} = 1000$) and the number of variables tried at each split was 10 ($m_{try} = 10$). These adjustable hyperparameters were selected once the model performed best (minimal RMSE) with as less as possible of computation costs. The optimal hyperparameters values used in our models tuning were similar to those in previous soil modelling studies [46,47].

2.4. Evaluation metrics for ML models

In this study, the dataset (n = 153) was randomly divided into two subsets. The first accounted for 70 % of samples (n = 107) for training, while the remaining subset of 30 % of samples (n = 46) was used for validation. The performances of the ML models herein employed were evaluated using the following statistical metrics: root-mean-square error (RMSE) and coefficient of determination (R^2) according to the following equations (1) and (2), respectively:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2} \tag{1}$$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{2}$$

Where, n = the size of the observations, $\hat{Y}$ = predicted response value, Y = the measured response value for the i-th term observation and $\bar{Y}$ = the average of the response variable. We have used the suggestions of Li et al. [48] to evaluate the accuracy of the predictions, if $R^2 < 0.5$ the prediction will be considered unacceptable, if $0.5 \leq R^2 < 0.75$ the prediction will be considered acceptable and if $R^2 \geq 0.75$ the prediction will be considered as good.

2.5. Geostatistical analysis and validation

Soil available P (or Olsen P) data was interpolated at a 10 m spatial resolution using ordinary kriging method to provide the geospatial Olsen P map. Ordinary kriging is an univariate interpolation method based on a weighting scheme widely used in soil science to estimate the unknown primary variable at unsampled location X_0 as a linear combination of neighboring observations (measured values) according to the following formula [49,50]:

$$Z^*(X_0) = \sum_{i=1}^{n}\lambda_i Z(X_i) \tag{3}$$

where $Z^*(X_0)$ is the predicted value at the unsampled position X_0, $Z(X_i)$ is the measured value at position X_i, λ_i is the weighting coefficient for a particular location X_i and n is the number of positions considered in the searching neighborhood.

The structure of the spatial variability was characterized using the semivariogram modelling under QGIS software (version 3.28), using the Smart-Map plugin [51] according to equation (4). The best-suited semivariogram model was then selected based on R^2 and RMSE. The accuracy of the interpretation was validated using the leave-one-out cross-validation method.

$$\gamma(h) = \frac{1}{2N(h)}\sum_{i=1}^{N(h)}[Z(x_i + h) - Z(x_i)]^2 \tag{4}$$

where $\gamma(h)$ is the semi-variance for separate distance class h (the distance between location x_i and x_0), N(h) is the number of data pairs at each distance interval h, $Z(x_i)$ is the value of the variable Z at sampled location x_i, and $Z(x_i + h)$ is the value of the variable Z at a distance h away from x_i. Parameters defining semivariogram models are nugget, sill and range. Nugget is the small-scale variability of the data. The range of the semivariogram is defined as the distance after which the variogram levels off. The sill expresses the distance (range) beyond which samples are not correlated.

3. Results and discussion

3.1. Descriptive statistics

The summary statistics of the geographic and soil properties of the study site are shown in Table 2. The coefficients of variation (CV) indicate that the data of soil variables are characterized by different variability extents. According to Wilding [52], these variables can be divided into three categories: the data with (CV >35 %) is characterized by high variability, and if (15 % <CV < 35 %) the data can be considered as data with moderate variability, else the variability of the data can be considered as low. Thus, most of the variables in the present study are characterized by moderate (e.g., Olsen P and particle size distribution) to high variability (e.g., micro-nutrients such as Mn, B, Z and Fe, and EC and CO3). This indicates that, although from the same farm (100 ha), the selected soil properties are spatially varied. Similar variability pattern were also found in other studies on a larger scale than our farm scale [17,50]. These wide ranges can help to obtain a high predictive accuracy of the modelling and then identify the most important predictors explaining the variability in the soil Olsen P.

Olsen P showed a relatively small skewed distribution, with a mean of 14.8 mg kg^{-1}, a median of 14.5 mg kg^{-1} and a range between 5.3 mg kg^{-1} and 27.3 mg kg^{-1}, which means that the distribution of Olsen P is relatively normal. The resulting variability of Olsen P in the study area could partly be explained by the differences in soil sampling depths, ranging from 15 to 50 cm, because phosphorus has low mobility in the soil [53], especially in calcareous soil due to absorption and precipitation reactions [7]. Moreover, the decrease in microbiological activity with depth due to the decline in soil fertility could also have contributed to reducing soil Olsen P down the soil

Table 2
Descriptive statistics of soil and geographic attributes for all data sets (n = 153 for each variable).

Variables	Mean	Median	SD	Kurtosis	Skewness	Min.	Max.	CV (%)
Olsen P	14.8	14.5	4.8	0.4	0.6	5.3	27.3	32.4
Ele	471.5	470.9	2.8	0.6	1.0	466.2	481.1	0.6
Gravel	29.7	30.3	10.1	0.6	-0.4	5.2	55.0	34.2
Clay	21.0	20.8	4.8	0.2	0.8	13.1	33.1	22.6
Fsilt	11.5	11.1	2.4	0.0	0.6	6.9	17.2	20.7
Csilt	7.8	7.8	2.3	4.1	0.7	3.3	19.9	29.3
Fsand	16.5	16.0	3.5	0.2	0.9	10.8	24.8	21.1
Csand	13.8	13.0	3.5	-0.4	0.6	7.8	21.6	25.3
Depth	30.0	30.0	9.3	-0.5	0.4	15.0	50.0	31.1
Ca	2404.4	2297.9	672.6	-0.1	0.5	1296.3	4084.3	28.0
K	194.4	172.3	82.5	1.0	1.3	99.1	430.1	42.5
Mg	196.7	193.7	67.2	1.2	1.2	108.4	400.3	34.2
Na	60.2	51.8	42.9	18.8	4.0	24.5	287.2	71.2
Cu	0.6	0.5	0.2	0.5	1.0	0.3	1.1	33.9
Mn	3.7	3.2	1.3	-0.5	0.5	1.7	8.1	35.8
Fe	2.1	2.1	0.8	1.0	1.1	1.0	4.4	36.1
Zn	0.5	0.5	0.2	6.0	2.1	0.3	1.6	45.2
B	0.1	0.1	0.1	0.9	1.2	0.1	0.3	43.1
TN	0.1	0.1	0.0	-0.1	0.2	0.1	0.1	11.5
NO3	1.1	1.0	0.5	3.9	1.8	0.5	3.1	46.7
NH4	0.2	0.2	0.1	8.9	2.5	0.1	0.6	31.8
SO4	9.0	8.7	1.6	2.3	0.8	4.7	14.3	18.2
C:N	7.1	7.1	1.0	0.6	0.4	4.6	9.9	14.5
OM	1.5	1.5	0.2	0.5	0.5	1.2	2.0	11.7
CEC	12.0	11.1	2.4	-0.8	0.5	8.9	17.1	19.8
EC	0.1	0.1	0.1	23.6	4.6	0.1	0.5	51.2
Cl	4.8	3.0	7.8	30.6	5.5	1.7	51.2	162.1
pHw	8.6	8.6	0.1	-0.3	0.3	8.5	8.8	0.9
pHKCl	7.7	7.7	0.1	0.2	-0.7	7.3	7.9	1.8
CO3	5.4	5.1	3.1	-0.5	0.3	0.2	12.4	57.5

Colors reperesent the separation of the variables based on their CV(%): Green = low variablity, gold = moderate and orange = high variability. n = number of samples. SD = standard deviation. CV = coefficient of variation. Min. = minimum, Max. = maximum, Ele = elevation, Fsilt = fine silt, Csilt = coarse silt, Fsand = fine sand, Csand = coarse sand, TN = total nitrogen, OM = organic matter, EC = electrical conductivity, pHw = water-extracted pH, pHKCl = KCl-extracted pH.

profile [54,55]. However, this hypothesis cannot be confirmed in the present study, given that Olsen P varied even within the same sampling depth. The presence of caliche below soil surface (40–50 cm) could also have reduced Olsen P due to the adsorption of P on the surface of calcite minerals [56]. Moreover, high alkalinity favors Mg sorption on the surface of calcite resulting from the formation of Mg–P phases [57]. Additionally, the fact that the study area is located close to a rock phosphate mining site, the atmospheric depositions due to dust could also have contributed to the resulting variability in the topsoil [20].

3.2. Pearson correlation matrix

Fig. 3 shows the Pearson correlation between Olsen P and the rest of soil properties and geographic variables. Olsen P was positively correlated with most of the soil attributes, while negatively correlated only with gravel, pH_w and CEC with moderate to weak correlations: $r = -0.39$ ($p < 0.001$), -0.08 ($p > 0.05$) and -0.03 ($p > 0.05$), respectively. The most important positive relationships were found between Olsen P and: NO3 ($r = 0.51$, $p < 0.001$), Mn ($r = 0.50$, $p < 0.001$), K($r = 0.49$, $p < 0.001$), NH4 ($r = 0.46$, $p < 0.001$), Zn ($r = 0.43$, $p < 0.001$), C/N ratio ($r = 0.43$, $p < 0.001$) and Csand ($r = 0.41$, $p < 0.001$). The strength of these correlations is moderate, but the rest of the positive correlations are below 0.40. The geographic variables are poorly correlated with Olsen P: N ($r = 0.03$, $p > 0.05$), E ($r = -0.15$, $p > 0.05$) and Ele ($r = -0.15$, $p > 0.05$). This could be explained by the low-relief feature of the study area and a low-gradient landscape. The correlation matrix is useful for a quick and simple summary of the direction and strength of the co-relationship between the variables. However, it does not provide sufficiently strong evidence to suggest the combined impact of the most important variables influencing and explaining the variability of Olsen P. Hence, the importance of regression methods and ML algorithms in prioritizing complex relationships, especially when dealing with an element known for its chemical complexity such as P, because its concentration in the soil is affected by complex nonlinear influences from several pedogenic factors (soil properties, climate, organisms, relief, parent material, age and position) [10] and agricultural practices [19].

Gravel proportion in the present soil varied largely and went up to a maximum of 55 % (Table 2). This could, to some extent, explain the resulting variability in the soil fertility including Olsen P, because the capacity of soil to supply water and nutrients decreases as gravel content increases [58]. This is evidenced by the negative correlation found between gravel and most of the soil properties (Fig. 2, eg., Clay: $r = -0.63$ ($p < 0.001$), C:N: $r = -0.98$ ($p < 0.001$), SO4: $r = -0.75$ ($p < 0.001$), Olsen P: $r = -0.39$ ($p < 0.001$)). Moreover, the observed low correlation between depth and most of the soil variables could be attributed to the dilution effect due to higher gravel content in the soil profiles. Another important correlation was found between CO3 and pH_{KCl} ($r = 0.75$, $p <$

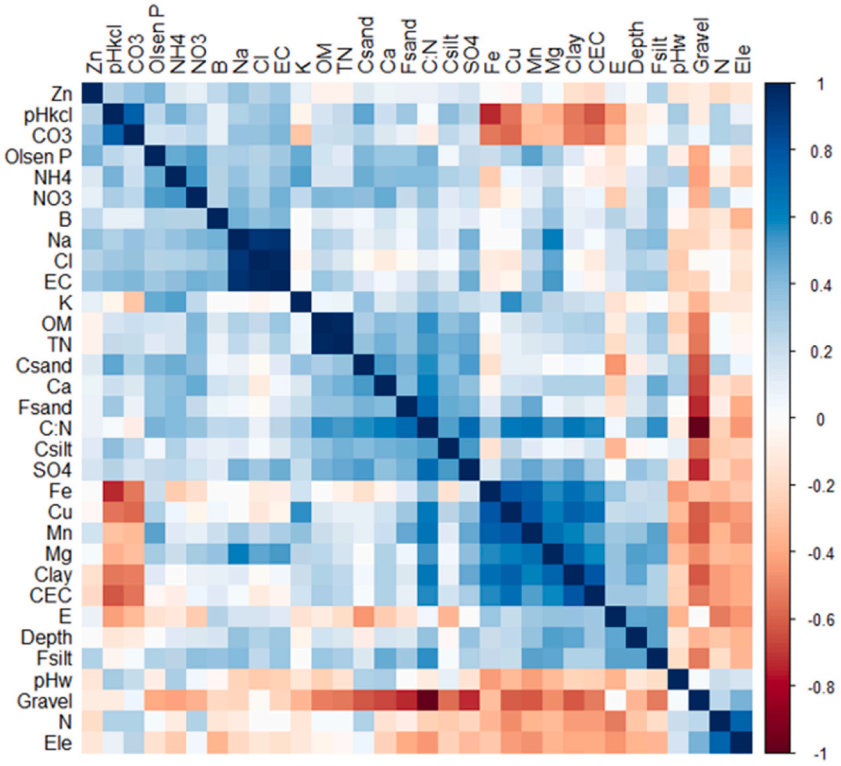

Fig. 2. Correlation matrix between soil Olsen P, soil properties and geographic variables. Ele = elevation, N = northern altitude, E = east longitude, CO3 = total carbonate, Zn = zinc, B = boron, Cl = chloride, EC = electrical conductivity, C:N = carbon to nitrogen ratio, Csilt = coarse silt, Fsilt = fine silt, Csand = coarse sand, Fsand = fine sand, SO4 = sulphate, Mg = magnesium, OM = organic matter, Cu = copper, TN = total nitrogen, Fe = iron, Ca = calcium, Na = sodium, K = potassium, NH4 = ammonium, Mn = manganese, CEC = cation exchange capacity, pHKCl = KCl-extracted pH, pHw = water-extracted pH.

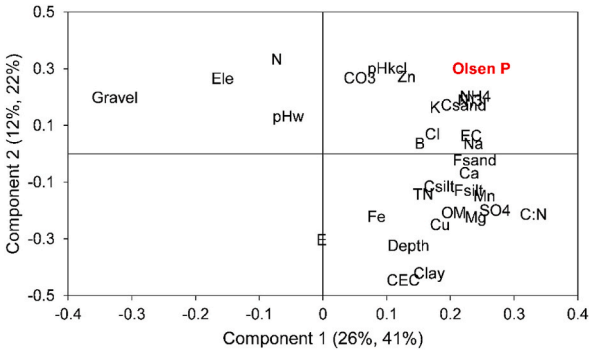

Fig. 3. Loading plot of PLS regression showing the loadings of x-variables including soil properties and geographic attributes on Olsen P (y-variable). Ele = elevation, N = northern altitude, E = east longitude, CO3 = total carbonate, Zn = zinc, B = boron, Cl = chloride, EC = electrical conductivity, C:N = carbon to nitrogen ratio, Csilt = coarse silt, Fsilt = fine silt, Csand = coarse sand, Fsand = fine sand, SO4 = sulphate, Mg = magnesium, OM = organic matter, Cu = copper, TN = total nitrogen, Fe = iron, Ca = calcium, Na = sodium, K = potassium, NH4 = ammonium, Mn = manganese, CEC = cation exchange capacity, pHKCl = KCl-extracted pH, pHw = water-extracted pH.

0.001). This confirms the role of carbonate in controlling soil pH [59], especially in alkaline-calcareous soils such as ours where pH_w varied between 8.5 and 8.8.

3.3. Modelling results

The PLS regression loading plot (Fig. 3) shows that most of the x-variables had positive loadings within the first PLS component which explained 44 % of Olsen P variability. The first component also brings in the information on gravel being the only soil variable that showed a negative effect on Olsen P, if we exclude the negative effect of soil pH_w which was negligible (very close to the origin). Among soil variables, C:N ratio showed the highest degree of positive co-variance with Olsen P with a loading of +0.33. This suggests the importance of C and N stoichiometric ratios in modulating soil available P. The soil C:N is considered a predictor of OM decomposition [60]. Thus, a higher C:N ratio indicates an accumulation of OM in the soil as evidenced by the resulting positive correlation between OM and C:N (r = 0.55) in the present study. The accumulation of OM could be attributed to plant growth (root biomass) enhancement following a higher available P concentration. On the other hand, higher carbon content (source of energy for microbes) in the soil could have potentially mineralized soil OM, boosted the microbial cycling of organic P, and subsequently increased plant available P [61]. A strong co-variance was also obtained between soil clay content and CEC (r = 0.80), indicating that soil CEC originates mainly from clay minerals in the study area. Interestingly, clay content and CEC showed the highest negative load (−0.42 and −0.44, respectively) on PLS component 2. This suggests that the presence of clay could have contributed to P-binding in this soil [62], owing to a good amount of clay varying from 13 to 33 %. However, the P fixed by clays is more easily desorbed than that fixed by Al and Fe oxides and hydroxides (e.g., goethite and gibbsite), thereby indicating that the stability of P adsorption to clay is weaker than that of Al/Fe oxides and hydroxides; thus, P is more readily available to plants. This is evidenced, to some extent, by the positive loading of clay on PLS component 1 which mainly describes the soil fertility. Alternatively, the high net negative charge that clays minerals have (e.g.: montmorillonite: a dominant clay mineral in the study area [63,64]) may have prevented their interaction

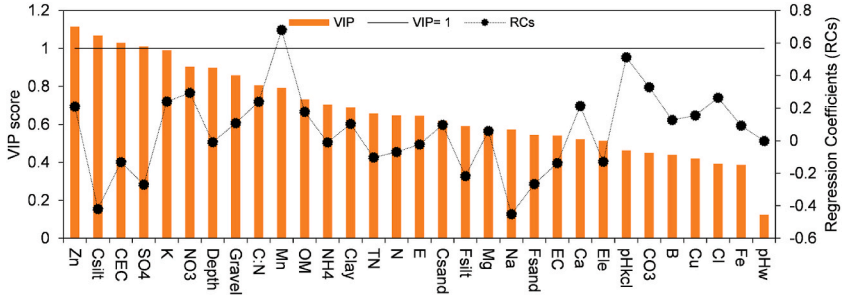

Fig. 4. Variable importance of the projection (bars) and regression coefficients (black line) of each predictor of soil Olsen P. Regression coefficients show the direction in which the predicted response (Olsen P) depends on the predictors. The straight solid grey line indicates the VIP threshold (VIP >1) above which the predictors are considered as important for predictions. Ele = elevation, N = northern altitude, E = east longitude, CO3 = total carbonate, Zn = zinc, B = boron, Cl = chloride, EC = electrical conductivity, C:N = carbon to nitrogen ratio, Csilt = coarse silt, Fsilt = fine silt, Csand = coarse sand, Fsand = fine sand, SO4 = sulphate, Mg = magnesium, OM = organic matter, Cu = copper, TN = total nitrogen, Fe = iron, Ca = calcium, Na = sodium, K = potassium, NH4 = ammonium, Mn = manganese, CEC = cation exchange capacity, pHKCl = KCl-extracted pH, pHw = water-extracted pH.

with phosphate leading to higher Olsen P in clay-rich samples. Nonetheless, the capacity of clay minerals (e.g., kaolinite, montmorillonite, illite) to bind P depends on different other factors including the crystallinity of these minerals and soil pH [65].

Although the PLS regression loading values showed the important factors influencing soil Olsen P, a more comprehensive and convenient expression of the relative importance of the variables could be acquired by exploring VIP and Regression Coefficients (RCs) [37,38]. The VIP and RCs are presented in Fig. 4. Interestingly, based on VIP values, Zn (1.11), Csilt (1.06), and CEC (1.03) were the key factors responsible for the variation in Olsen P. This result implies that soil texture exerted a substantial influence on soil available P, because clay and silt are negatively charged which increases soil CEC. Higher CEC could increase the soil available P [66] through the fixation of the soil solution cations such as Ca^{2+}, Zn^{2+} and Mn^{2+}, which are known to interact negatively with P in calcareous soils [67]. This is evidenced, somewhat, by the resulting positive Pearson correlation between soil DTPA-extractable Zn and Olsen P (Fig. 2, $r = 0.43$). Similarly, Olsen P appeared to increase with higher DTPA-extractable Mn ($r = 0.50$ and $RC_{PLSR} = 0.68$). These results suggest that the presence of Zn and Mn together on soil exchange complex could have likely reduced the solubility of Ca compounds like calcite in the soil solution and subsequently reduced the amounts of Ca-P precipitates increasing the Olsen P. However, this view is questionable considering that the concentration of DTPA-extractable Mn and Zn, being trace elements, in the soils is far below that of Ca. Alternatively, the formation of P-Mn and P-Zn compounds in the current alkaline soil is possible and the sorbed P was readily removed by ''$NaHCO_3$'' in the Olsen method, indicating that it was extremely labile. However, the reaction mechanisms by which exchangeable Zn and Mn could affect Olsen P measurement in calcareous soils warrants further investigations, especially at the current pH range 8.5–8.8 due to the prevalence of highly reactive orthophosphate species (HPO_4^{2-}) [68,69]. Thus, the characterization of phosphate species and minerals in the present soils is necessary to better elucidate the mechanisms. For instance, $MnHPO_4$ has been found to be one of the Mn complexes that precipitates in calcareous soils [70]. Furthermore, it has been reported in some other studies conducted on rice managed paddy soils that the reduction of Mn-associated P could be a significant source of P for plants [71,72]. Moreover, exchangeable Mg has been found to reduce the formation of Ca-P precipitates and positively affect the available P in calcareous soils with low Ca/Mg saturation [6]. For instance, Manimel Wadu et al. [6] explained this to be due to two possible reactions: (1) formation of ion pairs between Mg and P, and (2) cation exchange reactions on the soil exchange sites. The formation of ion pairs between Mg and phosphate ions could possibly decrease the effective concentration of free phosphate ions in solution, thereby reducing its precipitation with Ca. We postulate that similar mechanisms could also be relevant to Mn and Zn in our soils.

For the cubist model, 10 soil variables were selected to be the most important contributors in the prediction of soil Olsen P variability, while none of the geographic variables were important (Fig. 5A). Manganese and CO_3 were the best predictors with a relative importance of 92 and 80 % respectively, followed by NH_4, CEC, Csand and Ca with a relative importance of 35.5, 33.5, 33.5

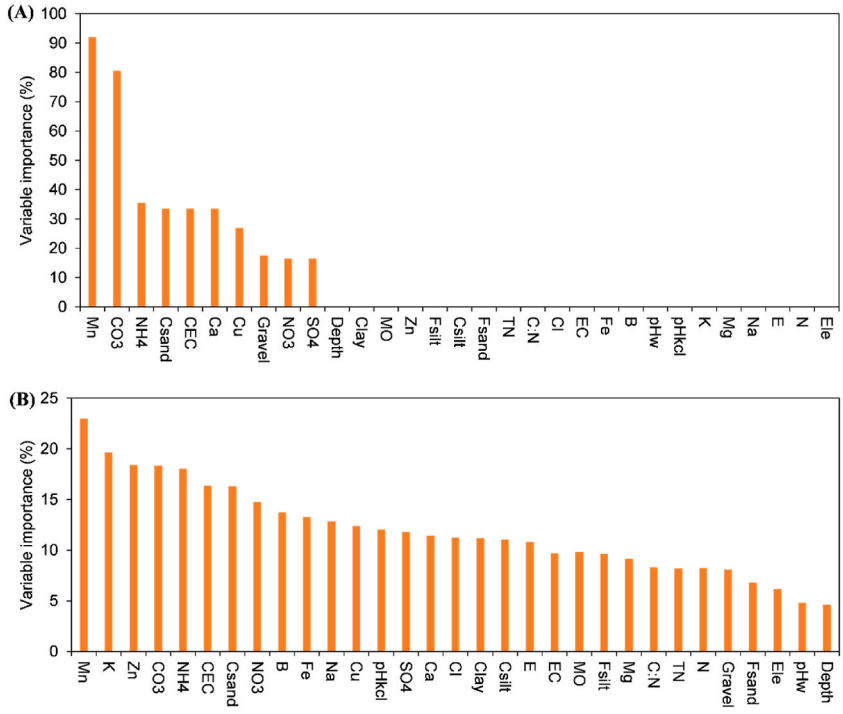

Fig. 5. Variable Importance for Olsen P using Cubist regression model (A) and Random Forest model (B). Ele = elevation, N = northern altitude, E = east longitude, CO3 = total carbonate, Zn = zinc, B = boron, Cl = chloride, EC = electrical conductivity, C:N = carbon to nitrogen ratio, Csilt = coarse silt, Fsilt = fine silt, Csand = coarse sand, Fsand = fine sand, SO4 = sulphate, Mg = magnesium, OM = organic matter, Cu = copper, TN = total nitrogen, Fe = iron, Ca = calcium, Na = sodium, K = potassium, NH4 = ammonium, Mn = manganese, CEC = cation exchange capacity, pHKCl = KCl-extracted pH, pHw = water-extracted pH.

and 33.5 % respectively, then comes Cu, Gravel, NO_3 and SO_4 with a relative importance of 27, 17.5, 16.5 and 16.5 % respectively. These results are in line with the output of the PLS regression and confirm the role of soil cations (Mn^{2+}, NH_4^+, Cu^{2+} and Ca^{2+}) and soil texture (given that CEC is negatively influenced by gravel and sand contents) in modulating soil Olsen P. Soil anions namely CO_3^{2-}, SO_4^{2-} and NO_3^- seemed also to play a key role in influencing Olsen P in our soils. This is because these anions compete for sorption sites on soil particles and influence the surface charge of the metal hydroxides, which could have reduced the phosphate adsorption [73,74]. This is supported by the resulting positive correlation between the aforementioned anions and soil Olsen P (Fig. 2). Furthermore, the role of SO_4^{2-} in explaining the Olsen P variability has also been confirmed in the PLS regression results showing a VIP>1 (Fig. 4).

The outputs of the RF model (Fig. 5B) also corroborate with the previous models as it reveals the dominance of soil cations (Mn^{2+}, K^+, Zn^{2+}, NH_4^+), soil anions (NO_3^-, CO_3^{2-} …) and CEC in predicting soil Olsen P. However, in the RF model, all variables either soil or geographic ones have contributed to the prediction, but at different levels of importance. For instance, Mn, K and Zn were the most important predictors, while elevation, pH_w and depth were the less important ones. Soil pH is considered to have a profound effect on soil P chemistry and solubility, especially in calcareous soil where the alkaline pH promotes the insoluble forms of P like Ca-P compounds. Nonetheless, in our soils, the fact that pH did not contribute that much to explaining Olsen P variability, could be because soil pH_w did not vary largely within the study area ranging only between 8.5 and 8.8. Likewise, the elevation ranged between 466 and 481 m above sea level, which is still a very narrow range and did not allow for a sufficient elevational gradient [75]. Soil sampling depth ranged between 15 and 50 cm; nevertheless, it is difficult to isolate the effect of depth on soil Olsen P variability in this study because, even within each soil depth, the Olsen P varied. Nonetheless, P is known to be a less mobile element in soils and therefore remains generally near the surface [53], especially in the no-till management such as our farm where soils were not frequently tilled prior to the soil sampling. Therefore, we hypothesize that the adopted sampling depths might have diluted to some extent the accumulated P in the near-surface (0–5 cm). This suggests that future studies may consider a higher resolution of sampling depth to take into consideration the P stratification effects. Also, the P fertilizer recommendations could be affected if this stratified nature of P is not considered in the soil testing [76].

3.4. Model evaluation

The results of the performances of the three adopted models showed that RF and CR were the best in predicting Olsen P with an R^2 of 0.95 and 0.92 respectively, and an RMSE of 1.30 and 1.43 mg kg^{-1} respectively, using the validation dataset (Fig. 6). This is an interesting result because Olsen P is usually reported to be a difficult parameter to predict. For instance, Al Masmoudi et al. [18] collected 400 samples from different locations in the Doukkala region in Morocco to predict soil fertility parameters including Olsen P, organic matter and K, using three ML models: RF, linear regression, and multiple linear regression. However, the results were not satisfactory regarding Olsen P prediction regardless of the model type ($R < 0.5$ for validation). On the other hand, Reda et al. [17] collected 660 samples from four different regions of Morocco (Marrakech, Agadir, Oriental and Tadla) and managed to improve the prediction accuracy of Olsen P by combining ML with near-infrared spectroscopy reaching an R^2 of 0.77. The resulting higher performance of our models against the previous studies could be due to the high number of explanatory variables we have used (31 variables). Though, it would be interesting to test if these variables could yield the same modelling performance on a larger scale. Moreover, the heterogeneity of the collected soils from different locations/regions in the previous studies could also have affected the prediction. This is evidenced by the resulting increase in the prediction quality of Olsen P after soil classification and clustering based on texture (clayey and loamy soils got better accuracy than sandy soils) by Reda et al. [17]. Similarly, Al Masmoudi et al. [18] showed that texture was one of the main variables besides $CaCO_3$ and Mn that contributed to the prediction of Olsen P. This aligns interestingly well with our modelling outputs, even though the scale and spatial scope of our study are comparatively smaller. It is also worth noting that even though RF performed slightly better than cubist regression in our study, the cubist model was more selective in identifying the most important variables as it detected 10 important variables only versus 31 variables in RF. The performance of PLS regression is also good ($0.5 \leq R^2 \leq 0.75$) based on the classification criteria proposed by Li et al. [48].

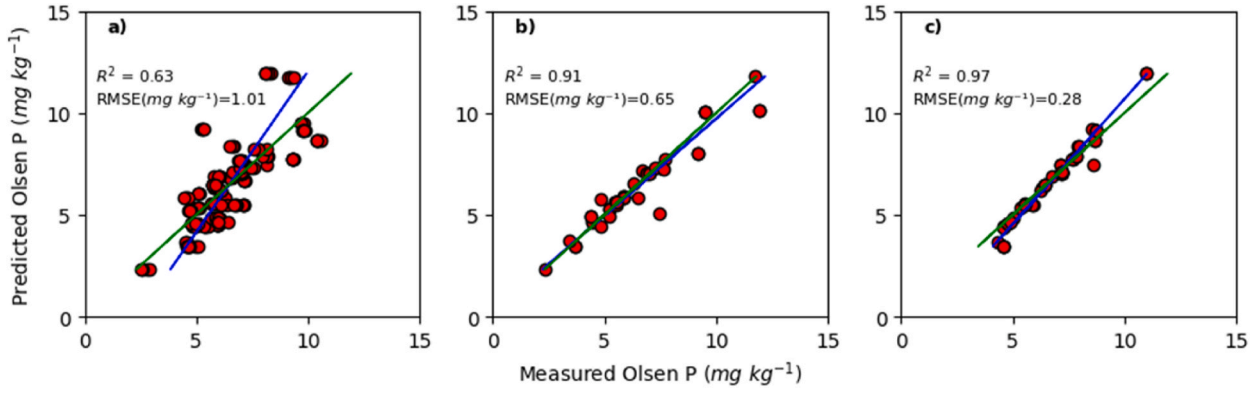

Fig. 6. Plots of measured and predicted soil Olsen P using validation dataset: (a) partial least squares regression, (b) random forest and (c) cubist regression model. R^2: regression coefficient, RMSE: root-mean-square error.

3.5. Soil Olsen-P map

Geostatistical models are highly effective for modelling geospatial data by considering spatial information from geographic locations and other non-spatial covariates, allowing them to address the spatial dependence and subsequently minimizing the spatial autocorrelation. In contrast, ML models are not as effective for modelling and predicting geospatial data because of spatial autocorrelation issues since they do not inherently consider spatial relationships, and incorporating spatial information may require additional preprocessing, which can introduce additional uncertainty, especially from small data sets [77]. Hence the use of geostatistical interpolation techniques, specifically ordinary kriging in the present study.

To identify the spatial variability of Olsen P across the study site, the measured Olsen P data were kriged into an Olsen P map. The parameters of the geostatistical models fitted to the semi-variogram are shown in Fig. 7 and Table 3. Gaussian and spherical models were the best-fitted models using the cross-validation method with an R^2 values of 0.99 and 0.98 respectively, while the RMSE of Gaussian was lower 4.4 m versus 5.7 m for spherical. Thus, the Gaussian model was selected for the predictive mapping of spatial variation of Olsen P ($R^2_{validation} = 0.78$, $RMSE_{validation} = 2.05$ m). These two models were also found to better assess the spatial variability of soil Olsen P in previous studies [78,79]. The value of Nugget/Sill ratio for Gaussian model was 10 %. This means that Olsen P could be classified as a strongly spatially dependent variable [80], with a high degree of autocorrelation between the sampling points, leading to a strongly patchy distribution despite the fact that our farm's spatial scale is small compared to regional or larger scales. According to Cambardella et al. [80], the spatial dependency (Nugget/Sill ratio) is generally classified into three categories: strong spatially dependent if the ratio is <25 %, moderately spatially dependent if the ratio is between 25 and 75 %, while it is classified as weak spatial dependent if it >75 %. The range value (the distance beyond which the variable was spatially independent) for Gaussian model was 250 m, which is greater than our sampling interval (50 m), indicating that our sampling system was sufficient to detect the spatial heterogeneity of soil Olsen P within our study area.

Soil properties characterization is a crucial factor in agricultural field experimental research and agronomic interventions. Moreover, soil P research is highly important for Morocco and globally seeing the vital role of P in the agricultural intensification. Thus, for efficient use of this finite resource and for better informing the agronomist and researchers about the soil Olsen P status and its spatial variation to adapt and design their experiments (the main purpose of our farm) and reduce the soil sampling efforts and costs, it is necessary to develop a detailed predicted map to spatially identify the nutrient-deficient areas and demarcate the areas like low, medium, and high fertility zones. Fig. 8 illustrates the spatial distribution of Olsen P in the study area (experimental farm) using ordinary kriging approach based on the Gaussian semi-variogram model. It shows a patchy distribution of Olsen P (confirming the Nugget/Sill ratio results) and visualizes its variability. In addition to the effects of soil properties, this distribution could also be attributed to the atmospheric deposition of ground dust coming from a neighboring mining site (1 km away from the farm). Moreover, the P transport from the mining site and surroundings through runoff most likely contributed to the localized high concentration of Olsen P, because they coincided with some of the watercourse's beds and sediment depositions in water crossings at the southern edge of the farm next to the road, where some native plants and weeds grew naturally possibly due to (1) higher soil moisture because of water stagnation and spreading around the crossings and (2) better soil fertility owing to sedimentation and nutrients accumulation. This naturally occurring vegetation explains partly some of the PLSR outputs such as the resulting high loading of C:N ratio on Olsen P by reason of OM accumulation.

To accurately guide the P fertilizer application, establish field experiments and collect soil samples for research purposes, the classification of the spatial variability within the farm could be a potentially effective strategy. Thus, one useful approach to address this is to categorize the homogeneous sub-areas or management zones, as depicted in Fig. 9. Therefore, the predicted Olsen P values using ordinary kriging were divided into four main classes: very low $\leq$4 mg P kg^{-1}, low (4–7 mg P kg^{-1}), moderate (7–9 mg P kg^{-1})

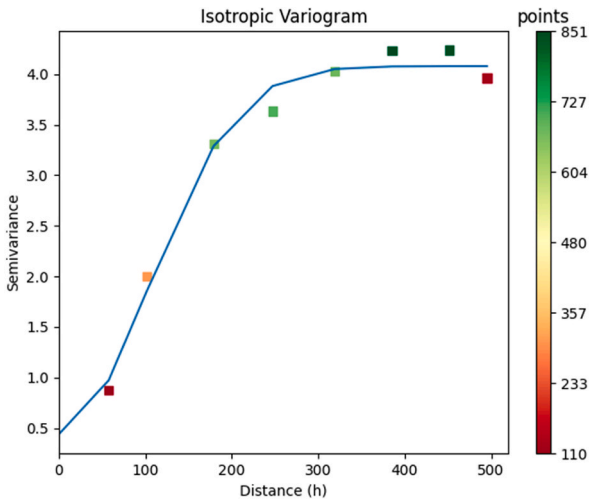

Fig. 7. Semi-variogram analysis of Olsen P.

Table 3
Geostatistical parameters of the best-fitted semi-variogram models.

Model	Nugget	Sill	Nugget/Sill	Range	RMSE	R^2
Gaussian	2.3	21.5	0.1	250.7	4.4	0.99
Spherical	0.0	21.7	0.0	315.1	5.7	0.98
Exponential	0.0	22.3	0.0	417.1	16.3	0.94

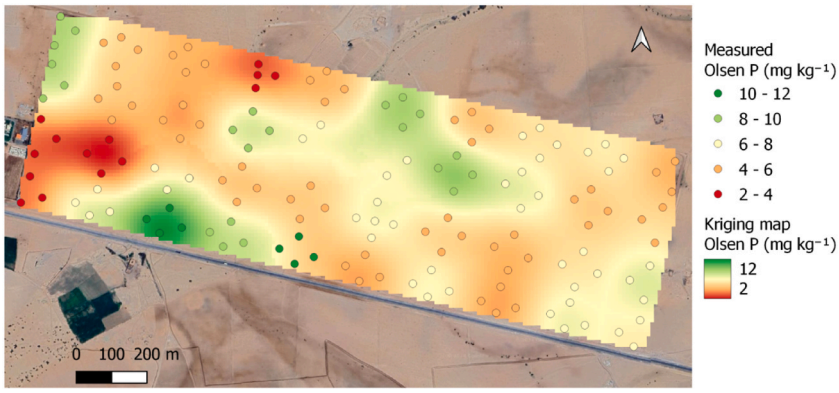

Fig. 8. Spatial distribution map with 10 m resolution of soil Olsen P using the ordinary kriging method.

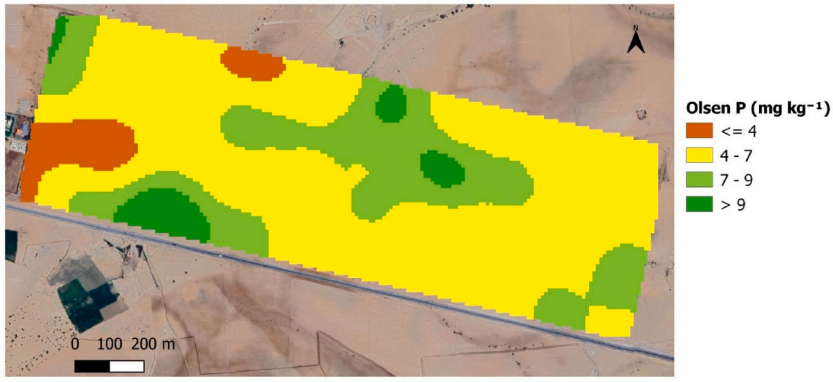

Fig. 9. Soil P management zones delineated based on kriging estimated Olsen P values. Olsen P values (mg P kg^{-1}) were classified into four main homogenous classes (≤4, 4–7, 7–9, and >9 mg P kg^{-1}).

and adequate 9 <P < 12 mg kg^{-1}. The utilization of this categorization was chosen because it enabled a more effective differentiation among the homogeneous zones. The appropriate Olsen P level typically varies based on factors such as crop requirements, soil types, and the level of agricultural intensity. For example, numerous soil test calibration studies conducted in Morocco propose a critical range of 7 mg P kg^{-1} for rainfed cereal and pulse crops. Below this threshold, phosphorus fertilizer is necessary, while above it, phosphorus levels are considered adequate. In irrigated areas, however, the critical level needs to be adjusted to 15 mg P kg^{-1} to accommodate the higher crop yields resulting from irrigation [81,82].

4. Conclusion

The findings of this study revealed that soil chemistry played a significant role in controlling soil Olsen P, particularly in relation to soil CEC and soil cations (Mn^{2+}, Zn^{2+}, NH_4^+, Ca^{2+} and K^+). Additionally, certain soil anions SO_4^{2-}, NO_3^- and CO_3^{2-} also exerted an influence. These findings highlight the importance of considering the interactions between P and other essential nutrient elements in the adoption of fertilization practices. However, based on the results of ML models (PLSR, RF and CR), exchangeable Mn and Zn, total carbonate and CEC were the most important soil attributes in explaining soil P variability in the study area. This can be attributed to the competition between Mn and Zn with Ca on soil exchange site, which subsequently reduced the solubility of Ca and its ability to interact with P in the soil solution. Alternatively, it is possible that P forms compounds with manganese (P-Mn) and zinc (P-Zn) in the alkaline soil and the sorbed P was readily removed by Olsen P extraction, indicating its lability. This is new and valuable information.

However, the two suggested hypotheses cannot be verified in this study and further investigations are warranted. Moreover, our study is only a single field-scale one and the investigation of these interactions on larger scale is needed.

The spatial modelling using ordinary kriging and Gaussian semi-variogram model demonstrated strong performance ($R^2 = 0.78$, RMSE = 2.05 m) in depicting the spatial distribution of Olsen P in the study area (100 ha), revealing its patchy nature. The heterogeneous distribution of P on the farm is likely a result of not only soil properties but also the depositions of atmospheric materials and sediment transport in the watercourses. The use of geostatistical algorithms for interpolating unsampled locations can yield an accurate representation of the spatial distribution of soil fertility. This information can be utilized to delineate homogeneous zones in the field, a particularly crucial aspect in farms like the one examined in this case study, where zonal differences are apparent. Therefore, the application of inputs can be optimized, and a more cost-effective field management can be achieved.

CRediT authorship contribution statement

Moussa Bouray: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Data curation. **Mohamed Bayad:** Writing – review & editing, Visualization, Validation, Software, Data curation. **Adnane Beniaich:** Writing – review & editing, Software, Methodology. **Ahmed G. El-Naggar:** Validation, Conceptualization. **Rebecca Logsdon Muenich:** Writing – review & editing. **Khalil El Mejahed:** Writing – review & editing, Supervision. **Abdallah Oukarroum:** Writing – review & editing. **Mohamed El Gharous:** Writing – review & editing, Project administration.

Data availability

Data will be made available upon request to the corresponding author.

Funding

No funding was received to assist with the preparation of this manuscript.

Declaration of competing interest

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

Acknowledgment

We thank the AITTC team, especially the soil-plant-water analysis laboratory for their technical support regarding soil sampling and analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e40128.

References

[1] E. Hou, Y. Luo, Y. Kuang, C. Chen, X. Lu, L. Jiang, X. Luo, D. Wen, Global meta-analysis shows pervasive phosphorus limitation of aboveground plant production in natural terrestrial ecosystems, Nat. Commun. 11 (2020) 637, https://doi.org/10.1038/s41467-020-14492-w.

[2] N. Bolan, P. Srivastava, C.S. Rao, P.V. Satyanaraya, G.C. Anderson, S. Bolan, G.P. Nortjé, R. Kronenberg, S. Bardhan, L.K. Abbott, H. Zhao, P. Mehra, S. V. Satyanarayana, N. Khan, H. Wang, J. Rinklebe, K.H.M. Siddique, M.B. Kirkham, Distribution, characteristics and management of calcareous soils, in: D. L. Sparks (Ed.), Advances in Agronomy, Academic Press, 2023, pp. 81–130.

[3] N.P. Rogovska, A.M. Blackmer, A.P. Mallarino, Relationships between soybean yield, soil pH, and soil carbonate concentration, Soil Sci. Soc. Am. J. 71 (2007) 1251–1256, https://doi.org/10.2136/sssaj2006.0235.

[4] M. Meixia, C. Yupeng, W. Xiaogang, W. Yushen, Do caliche nodules in loessial profiles affect root growth? Plant Soil 473 (2022) 369–387, https://doi.org/10.1007/s11104-021-05290-4.

[5] I. Bertrand, R.E. Holloway, R.D. Armstrong, M.J. McLaughlin, Chemical characteristics of phosphorus in alkaline soils from southern Australia, Soil Res. 41 (2003) 61–76, https://doi.org/10.1071/SR02021.

[6] M.C. Manimel Wadu, V.K. Michaelis, S. Kroeker, O.O. Akinremi, Exchangeable calcium/magnesium ratio affects phosphorus behavior in calcareous soils, Soil Sci. Soc. Am. J. 77 (2013) 2004–2013, https://doi.org/10.2136/sssaj2012.0102.

[7] R. von Wandruszka, Phosphorus retention in calcareous soils and the effect of organic matter on its mobility, Geochem. Trans. 7 (2006) 6, https://doi.org/10.1186/1467-4866-7-6.

[8] A. Leytem, R. Mikkelsen, The nature of phosphorus in calcareous soils, Better Crops 89 (2005) 11–13.

[9] C.K. Anthonio, H. Jing, C. Jin, M.N. Khan, D. Jiangxue, H.N. Garba, L. Dongchu, L. Guangrong, L. Shujun, L. Lisheng, Z. Huimin, Impact of long-term fertilization on phosphorus fractions and manganese oxide with their interactions in paddy soil aggregates, J. Environ. Manag. 333 (2023) 117440, https://doi.org/10.1016/j.jenvman.2023.117440.

[10] X. Yang, W.M. Post, Phosphorus transformations as a function of pedogenesis: a synthesis of soil phosphorus data using Hedley fractionation method, Biogeosciences 8 (2011) 2907–2916, https://doi.org/10.5194/bg-8-2907-2011.

[11] E. Hou, D. Wen, Y. Kuang, J. Cong, C. Chen, X. He, M. Heenan, H. Lu, Y. Zhang, Soil pH predominantly controls the forms of organic phosphorus in topsoils under natural broadleaved forests along a 2500km latitudinal gradient, Geoderma 315 (2018) 65–74, https://doi.org/10.1016/j.geoderma.2017.11.041.
[12] X. He, L. Augusto, D.S. Goll, B. Ringeval, Y. Wang, J. Helfenstein, Y. Huang, K. Yu, Z. Wang, Y. Yang, E. Hou, Global patterns and drivers of soil total phosphorus concentration, Earth Syst. Sci. Data 13 (2021) 5831–5846, https://doi.org/10.5194/essd-13-5831-2021.
[13] A. Keshavarzi, E.-S.E. Omran, S.M. Bateni, B. Pradhan, D. Vasu, A. Bagherzadeh, Modeling of available soil phosphorus (ASP) using multi-objective group method of data handling, Model Earth Syst Environ 2 (2016) 157, https://doi.org/10.1007/s40808-016-0216-5.
[14] D.D. Mühl, L. de Oliveira, A bibliometric and thematic approach to agriculture 4.0, Heliyon 8 (2022) e09369, https://doi.org/10.1016/j.heliyon.2022.e09369.
[15] X. Wang, Y. Yang, J. Lv, H. He, Past, present and future of the applications of machine learning in soil science and hydrology, Soil Water Res. 18 (2023) 67–80, https://doi.org/10.17221/94/2022-SWR.
[16] J. Padarian, B. Minasny, A. McBratney, Machine learning and soil sciences: a review aided by machine learning tools, SOIL 6 (2020) 35–52, https://doi.org/10.5194/soil-6-35-2020.
[17] R. Reda, T. Saffaj, S.E. Itqiq, I. Bouzida, O. Saidi, K. Yaakoubi, B. Lakssir, N. El Mernissi, E.M. El Hadrami, Predicting soil phosphorus and studying the effect of texture on the prediction accuracy using machine learning combined with near-infrared spectroscopy, Spectrochim. Acta A Mol. Biomol. 242 (2020) 118736, https://doi.org/10.1016/j.saa.2020.118736.
[18] Y. Al Masmoudi, Y. Bouslihim, K. Doumali, L. Hssaini, K. Ibno Namr, Use of machine learning in Moroccan soil fertility prediction as an alternative to laborious analyses, Model. Earth Syst. Environ. (2021) 1–11, https://doi.org/10.1007/s40808-021-01329-8.
[19] M. Matos-Moreira, B. Lemercier, R. Dupas, D. Michot, V. Viaud, N. Akkal-Corfini, B. Louis, C. Gascuel-Odoux, High-resolution mapping of soil phosphorus concentration in agricultural landscapes with readily available or detailed survey data, Eur. J. Soil Sci. 68 (2017) 281–294, https://doi.org/10.1111/ejss.12420.
[20] F.T. Da Conceição, T. Litholdo, D. de Souza Sardinha, R.B. Moruzzi, G.R.B. Navarro, L.H. Godoy, The influence of phosphate mining on the chemical composition of annual atmospheric deposition in Catalão (GO) and Tapira (MG), Brazil, Water Air Soil Pollut. 227 (2016) 1–13, https://doi.org/10.1007/s11270-015-2731-9.
[21] C. Ahmed, A. Mohammed, A. Tahir, Geostatistics of strength, modeling and GIS mapping of soil properties for residential purpose for Sulaimani City soils, Kurdistan Region, Iraq, Model Earth Syst Environ 6 (2020) 879–893, https://doi.org/10.1007/s40808-020-00715-y.
[22] S.K. Ahado, P.C. Agyeman, L. Borůvka, R. Kanianska, C. Nwaogu, Using geostatistics and machine learning models to analyze the influence of soil nutrients and terrain attributes on lead prediction in forest soils, Model Earth Syst Environ (2023), https://doi.org/10.1007/s40808-023-01890-4.
[23] M. Glendell, S.J. Granger, R. Bol, R.E. Brazier, Quantifying the spatial variability of soil physical and chemical properties in relation to mitigation of diffuse water pollution, Geoderma 214–215 (2014) 25–41, https://doi.org/10.1016/j.geoderma.2013.10.008.
[24] J. Hong, S. Grunwald, G.M. Vasques, Soil phosphorus landscape models for precision soil conservation, J. Environ. Qual. 44 (2015) 739–753, https://doi.org/10.2134/jeq2014.09.0379.
[25] H. Briak, F. Kebede, Wheat (*Triticum aestivum*) adaptability evaluation in a semi-arid region of Central Morocco using APSIM model, Sci. Rep. 11 (2021) 23173, https://doi.org/10.1038/s41598-021-02668-3.
[26] G. Aubert, La classification des sols: la classification pédologique française. International Symposium on Soil Classification STIBOKA, 1962, p. 10. Wageningen.
[27] IUSS, World reference base for soil resources. International Soil Classification System for Naming Soils and Creating Legends for Soil Maps, 4, International Union of Soil Sciences, Vienna, Austria, 2022.
[28] N. Wollenhaupt, R. Wolkowski, Grid Soil Sampling, Better Crops, vol. 78, 1994, pp. 6–9.
[29] S.R. Olsen, Estimation of Available Phosphorus in Soils by Extraction with Sodium Bicarbonate, US Department of Agriculture, Washington, USA, 1954.
[30] G.J. Bouyoucos, Hydrometer method improved for making particle size analyses of soils, Agron. J. 54 (1962) 464–465, https://doi.org/10.2134/agronj1962.00021962005400050028x.
[31] D.a. Nelson, L.E. Sommers, Total carbon, organic carbon, and organic matter. Methods of Soil Analysis: Part 2 Chemical and Microbiological Properties, ASA-SSSA, Madison, 1983.
[32] A. Walkley, I.A. Black, An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method, Soil Sci. 37 (1934) 29–38.
[33] B. Horváth, O. Opara-Nadi, F. Beese, A simple method for measuring the carbonate content of soils, Soil Sci. Soc. Am. J. 69 (2005) 1066–1068, https://doi.org/10.2136/sssaj2004.0010.
[34] J.M. Bremner, Determination of nitrogen in soil by the Kjeldahl method, J. Agric. Sci. 55 (2009) 11–33, https://doi.org/10.1017/s0021859600021572.
[35] AFNOR, Détermination de la capacité d'échange cationique et des cations extractibles, Qualité des sols, Association Française de Normalisation, Paris, France, 1993, pp. 103–106.
[36] H. Abdi, Partial least squares regression and projection on latent structure regression (PLS Regression), Wiley Interdiscip. Rev.: Comput. Stat. 2 (2010) 97–106, https://doi.org/10.1002/wics.51.
[37] L. Duan, H. Xie, Z. Li, H. Yuan, Y. Guo, X. Xiao, Q. Zhou, Use of partial least squares regression to identify factors controlling rice yield in Southern China, Agron. J. 112 (2020) 1502–1516, https://doi.org/10.1002/agj2.20161.
[38] J. Usman, S. Begum, SOC stocks prediction on the basis of spatial and temporal variation in soil properties by using partial least square regression, Sci. Rep. 13 (2023) 7949, https://doi.org/10.1038/s41598-023-34607-9.
[39] J.R. Quinlan, Learning with continuous classes, in: A. Adams, L. Sterling (Eds.), 5th Australian Joint Conference on Artificial Intelligence, World Scientific, Tasmania, Australia, 1992, pp. 343–348.
[40] B. Minasny, A.B. McBratney, Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy, Chemom. Intell. Lab. 94 (2008) 72–79, https://doi.org/10.1016/j.chemolab.2008.06.003.
[41] M. Kuhn, K. Johnson, Applied Predictive Modeling, Springer, New York, NY, Berlin, Germany, 2013.
[42] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[43] R.C. Team, A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2022.
[44] M. Kuhn, S. Weston, C. Keefer, M.M. Kuhn, Package 'Cubist: Rule-And Instance-Based Regression Modeling, 2023.
[45] A. Liaw, M. Wiener, Package 'randomforest': Breiman and Cutler's Random Forests for Classification and Regression, University of California, Berkeley, Berkeley, CA, USA, 2018.
[46] A. Beniaich, W. Otten, H.-C. Shin, H.V. Cooper, J. Rickson, A. Soulaimani, M. El Gharous, Evaluation of pedotransfer functions to estimate some of soil hydraulic characteristics in North Africa: a case study from Morocco, Front. Environ. Sci. 11 (2023) 120.
[47] B.P. Malone, B. Minasny, A.B. McBratney, Continuous Soil Attribute Modeling and Mapping, Using R for Digital Soil Mapping, Springer, Switzerland, 2017, pp. 133–143.
[48] L. Li, J. Lu, S. Wang, Y. Ma, Q. Wei, X. Li, R. Cong, T. Ren, Methods for estimating leaf nitrogen concentration of winter oilseed rape (*Brassica napus L.*) using in situ leaf spectroscopy, Ind. Crops Prod. 91 (2016) 194–204, https://doi.org/10.1016/j.indcrop.2016.07.008.
[49] A. Walche, W. Haile, A. Kiflu, D. Tsegaye, Spatial analysis and mapping of intensity and types of agricultural salt-affected soils around Abaya and Chamo Lakes, South Ethiopia Rift Valley, Heliyon 10 (2024) e33410, https://doi.org/10.1016/j.heliyon.2024.e33410.
[50] D. Vasu, N. Sahu, P. Tiwary, P. Chandran, Modelling the spatial variability of soil micronutrients for site specific nutrient management in a semi-arid tropical environment, Model. Earth Syst. Environ. 7 (2021) 1797–1812, https://doi.org/10.1007/s40808-020-00909-4.
[51] G.W. Pereira, D.S.M. Valente, D.M.d. Queiroz, A.L.d.F. Coelho, M.M. Costa, T. Grift, Smart-map: an open-source QGIS plugin for digital mapping using machine learning techniques and ordinary kriging, Agronomy 12 (2022) 1350, https://doi.org/10.3390/agronomy12061350.
[52] L. Wilding, Spatial variability: its documentation, accomodation and implication to soil surveys, in: J. Bouma, D.R. Nielsen (Eds.), Soil Spatial Variability Pudoc, Wageningen, Las Vegas, USA, 1985, pp. 166–194.
[53] M. Bouray, J.L. Moir, L.M. Condron, D. Paramashivam, Early effects of surface liming on soil P biochemistry and dynamics in extensive grassland, Nutrient Cycl. Agroecosyst. (2021) 1–15, https://doi.org/10.1007/s10705-021-10163-4.

[54] D.L. Achat, L. Augusto, M.R. Bakker, A. Gallet-Budynek, C. Morel, Microbial processes controlling P availability in forest spodosols as affected by soil depth and soil properties, Soil Biol. Biochem. 44 (2012) 39–48, https://doi.org/10.1016/j.soilbio.2011.09.007.
[55] N. Fierer, J.P. Schimel, P.A. Holden, Variations in microbial community composition through two soil depth profiles, Soil Biol. Biochem. 35 (2003) 167–176, https://doi.org/10.1016/S0038-0717(02)00251-1.
[56] L. Wang, E. Ruiz-Agudo, C.V. Putnis, M. Menneken, A. Putnis, Kinetics of calcium phosphate nucleation and growth on calcite: implications for predicting the fate of dissolved phosphate species in alkaline soils, Environ. Sci. Technol. 46 (2012) 834–842, https://doi.org/10.1021/es202924f.
[57] N. Xu, H. Yin, Z. Chen, S. Liu, M. Chen, J. Zhang, Mechanisms of phosphate retention by calcite: effects of magnesium and pH, J. Soils Sediments 14 (2014) 495–503, https://doi.org/10.1007/s11368-013-0807-y.
[58] C.A. Scanlan, K.W. Holmes, R.W. Bell, Sand and gravel subsoils, in: T.S.d. Oliveira, R.W. Bell (Eds.), Subsoil Constraints for Crop Production, Springer International Publishing, Cham, Switzerland, 2022, pp. 179–198.
[59] C. Wang, W. Li, Z. Yang, Y. Chen, W. Shao, J. Ji, An invisible soil acidification: critical role of soil carbonate and its impact on heavy metal bioavailability, Sci. Rep. 5 (2015) 12735, https://doi.org/10.1038/srep12735.
[60] J. Leifeld, K. Klein, C. Wüst-Galley, Soil organic matter stoichiometry as indicator for peatland degradation, Sci. Rep. 10 (2020) 7634, https://doi.org/10.1038/s41598-020-64275-y.
[61] B. Arruda, D.J. Dall'orsoletta, J.C. Heidemann, L.C. Gatiboni, Phosphorus dynamics in the rhizosphere of two wheat cultivars in a soil with high organic matter content, Arch. Agron Soil Sci. 64 (2018) 1011–1020, https://doi.org/10.1080/03650340.2017.1407028.
[62] J. Xiong, Z. Liu, Y. Yan, J. Xu, D. Liu, W. Tan, X. Feng, Role of clay minerals in controlling phosphorus availability in a subtropical Alfisol, Geoderma 409 (2022) 115592, https://doi.org/10.1016/j.geoderma.2021.115592.
[63] A. Bahhou, Y. Taha, Y. El Khessaimi, H. Idrissi, R. Hakkou, J. Amalik, M. Benzaazoua, Use of phosphate mine by-products as supplementary cementitious materials, Mater. Today: Proc. 37 (2021) 3781–3788, https://doi.org/10.1016/j.matpr.2020.07.619.
[64] M. Loutou, Y. Taha, M. Benzaazoua, Y. Daafi, R. Hakkou, Valorization of clay by-product from moroccan phosphate mines for the production of fired bricks, J. Clean. Prod. 229 (2019) 169–179, https://doi.org/10.1016/j.jclepro.2019.05.003.
[65] F. Gérard, Clay minerals, iron/aluminum oxides, and their contribution to phosphate sorption in soils — a myth revisited, Geoderma 262 (2016) 213–226, https://doi.org/10.1016/j.geoderma.2015.08.036.
[66] F. Amery, E. Smolders, Unlocking fixed soil phosphorus upon waterlogging can be promoted by increasing soil cation exchange capacity, Eur. J. Soil Sci. 63 (2012) 831–838, https://doi.org/10.1111/j.1365-2389.2012.01478.x.
[67] H. He, M. Wu, R. Su, Z. Zhang, C. Chang, Q. Peng, Z. Dong, J. Pang, H. Lambers, Strong phosphorus (P)-zinc (Zn) interactions in a calcareous soil-alfalfa system suggest that rational P fertilization should be considered for Zn biofortification on Zn-deficient soils and phytoremediation of Zn-contaminated soils, Plant Soil 461 (2021) 119–134, https://doi.org/10.1007/s11104-020-04793-w.
[68] N. Barrow, Modelling the effects of pH on phosphate sorption by soils, Eur. J. Soil Sci. 35 (1984) 283–297, https://doi.org/10.1111/j.1365-2389.1984.tb00283.x.
[69] D.P. Schachtman, R.J. Reid, S.M. Ayling, Phosphorus uptake by plants: from soil to cell, Plant Physiol 116 (1998) 447–453, https://doi.org/10.1104/pp.116.2.447.
[70] S. Moharami, M. Jalali, Effects of cations and anions on iron and manganese sorption and desorption capacity in calcareous soils from Iran, Environ. Earth Sci. 68 (2013) 847–858, https://doi.org/10.1007/s12665-012-1787-8.
[71] C.K. Anthonio, H. Jing, C. Jin, M.N. Khan, D. Jiangxue, H.N. Garba, L. Dongchu, L. Guangrong, L. Shujun, L. Lisheng, Z. Huimin, Impact of long-term fertilization on phosphorus fractions and manganese oxide with their interactions in paddy soil aggregates, J. Environ. Manag. 333 (2023) 117440, https://doi.org/10.1016/j.jenvman.2023.117440.
[72] H. Shahandeh, L. Hossner, F. Turner, Phosphorus relationships to manganese and iron in rice soils, Soil Sci. 168 (2003) 489–500.
[73] J.S. Arias, P.G. Fernández, Changes in phosphorus adsorption in a Palexerult amended with limestone and/or gypsum, Commun. Soil Sci. Plant Anal. 32 (2001) 751–758, https://doi.org/10.1081/CSS-100103906.
[74] A. Samuel Kwesi, Processes and factors affecting phosphorus sorption in soils, in: K. George, L. Nikolaos (Eds.), Sorption in 2020s, IntechOpen, Rijeka, 2020. Ch. 3.
[75] A.G. Vincent, M.K. Sundqvist, D.A. Wardle, R. Giesler, Bioavailable soil phosphorus decreases with increasing elevation in a subarctic tundra landscape, PLoS One 9 (2014) e92942, https://doi.org/10.1371/journal.pone.0092942.
[76] V. Reed, B. Finch, J. Souza, P. Watkins, B. Arnall, Soil sampling depth impact on phosphorus yield response prediction in winter wheat, Agric. Environ. Lett. 7 (2022) e20067, https://doi.org/10.1002/ael2.20067.
[77] G.B. Heuvelink, R. Webster, Spatial statistics and soil mapping: a blossoming partnership under pressure, Spatial statistics 50 (2022) 100639, https://doi.org/10.1016/j.spasta.2022.100639.
[78] I. Bogunovic, M. Mesic, Z. Zgorelec, A. Jurisic, D. Bilandzija, Spatial variation of soil nutrients on sandy-loam soil, Soil Tillage Res. 144 (2014) 174–183, https://doi.org/10.1016/j.still.2014.07.020.
[79] D. Panday, B. Maharjan, D. Chalise, R.K. Shrestha, B. Twanabasu, Digital soil mapping in the Bara district of Nepal using kriging tool in ArcGIS, PLoS One 13 (2018) e0206350, https://doi.org/10.1371/journal.pone.0206350.
[80] C.A. Cambardella, T. Moorman, J. Novak, T. Parkin, D. Karlen, R. Turco, A. Konopka, Field-scale variability of soil properties in central Iowa soils, Soil Sci. Soc. Am. J. 58 (1994) 1501–1511, https://doi.org/10.2136/sssaj1994.03615995005800050033x.
[81] J. Ryan, H. Ibrikci, A. Delgado, J. Torrent, R. Sommer, A. Rashid, Significance of phosphorus for agriculture and the environment in the west Asia and North africa region, in: D.L. Sparks (Ed.), Advances in Agronomy, Elsevier, Amsterdam, 2012, pp. 91–153.
[82] J. Ryan, Soil fertility enhancement in mediterranean-type dryland agriculture: a prerequisite for development, in: S.C. Rao, J. Ryan (Eds.), Challenges and Strategies of Dryland Agriculture, Wiley, USA, 2004, pp. 275–290.