Preparing Available Phosphorus Distribution Maps in the Soils Using Various GIS-Based Spatial Interpolation Methods

Abstract

The study,was conducted from January, 2022 to December, 2024 at Jawaharlal Nehru KrishiVishwaVidhyalaya, Jabalpur, Madhya Pradesh (482004), India. We have collected two thousand two hundred sixteen (2216) Global Positioning System (GPS) based soil samples from depth of 0–15 cm from farmer's field of the Kymore Plateau and Satpura Hills zone of Madhya Pradesh, India. Laboratory analysis showed that the available P content ranged from 1.11 to 117.7 kg ha^{-1} (mean: 10.64 kg ha^{-1}). Significant correlations were found between pH, electrical conductivity (EC), organic carbon (OC), and P (r = 0.073).The studyaimed to identify the most suitable interpolation method for mapping available P in soils, employing three geo-statistical (Ordinary Kriging, Simple Kriging, and Empirical Bayesian Kriging) and three deterministic methods (Radial Basis Function, Local Polynomial Interpolation, and Inverse Distance Weighting), as well as two barrier-based methods (Kernel Smoothing and Diffusion Kernel).Geo-statistical results indicated OK(Box-Cox) spatial interpolation method for estimating available P distribution in soils that followed exponential modelwith ranges of 3652.22 meters, nugget values of 16.90 and a N/S ratio of 0.47 which showed moderate spatial dependency.Among the methods tested, Empirical Bayesian Kriging (EBK) provided the most accurate estimates of Olsen P distribution, followed by Ordinary Kriging (OK) and Simple Kriging with Box-Cox transformation.

Keywords: Phosphorus, Spatial Variability Map, Geo-Statistical,GIS, GPS

1. Introduction

Soil consists of physical, chemical, and biological characteristics, all of which play a vital role in maintaining fertility. Soil fertility plays a crucial role in determining agricultural productivity, and the availability of phosphorus is vital for energy transfer, root development, and overall plant metabolism (Saito et al., 2019). Soil, being a complex and heterogeneous medium, exhibits significant variations over short distances.In the domain of soil science (Khallouf et al., 2020; Criado et al., 2021), geo-statistics plays a crucial role in understanding the spatial variability of soil parameters.

The issues of P have become increasingly critical due to their complex reactions and transformations in the soil, and the small amounts available in the soil. The spatial distribution of P is often uneven due to factors like soil type, climate, land management practices, and parent material. Understanding this variability is crucial for developing site-specific nutrient management strategies to optimize fertilization and improve crop production (Kaur et al., 2020). Imbalances in soil nutrients can adversely affect crop productivity, making the systematic assessment crucial for effective soil management which helps farmers determine the type of crop yield that can be expected under specific soil conditions.Understanding nutrient variability in soil provides a scientific basis for effective nutrient management in agriculture.

“The application of spatial interpolation and geostatistical techniques has been recommended in varied scientific fields for parameter distribution in soil sciences”(Brus et al. 1996; Bourennane et al. 2000; Bishop and Mcbratney 2001; Robinson and Metternicht 2006).“The spatial distribution of soil parameters at the unsampled sites could not be determined using the traditional statistical approach. Therefore, geo-statistics is an effective technique for analyzing the spatial distribution of soil attributes and also for significantly reducing the variation of evaluation error and associated costs" (Davis et al., 2009; Nickel et al., 2014; Bhunia et al., 2018; Fischer et al., 2021).Geo-statistical methods such as kriging, are particularly useful in analyzing spatial patterns and modeling the variability of soil properties (Goovaerts, 1997).Geo-statistics also recommended for mappingand identifying areas requiring specific interventions (Kumar et al., 2021, OpenShaw and Clarke, 2019; Wang and Liu, 2023).

The geo-statistical analysis aids in characterizing spatial variability (AbdelRahman et al., 2020 and Jadon et al., 2023), creating spatial models (Zakeri and Mariethoz, 2021), and making reliable predictions (Kingsley et al., 2019) of soil properties at unsampled locations. Modern geo-statistical tools and techniques, such as semi-variograms, spatialauto-correlogram, and various interpolation approaches, are employed to assess the spatialvariability (Gokmen et al., 2023; Khan et al. 2021) of soil properties. In contrast, classical statistical techniques typically rely on descriptive statistical tools like mean, median, mode, coefficient of variation, etc., to measure soil property variability without considering its spatial dependence on the sampling point. However, they fail to adequately explain the continuous spatial variability pattern. Key tools of geostatistics (Gangopadhyay and Reddy, 2022) include variogram, kriging interpolation, spatial uncertainty, and cross validation.

Madhya Pradesh, located in central India, is known for its diverse agro-climatic zones, including the Kymore Plateau and Satpura Hills, which have distinct soil characteristics and agricultural practices. This zone is characterized by varying topography, climate, and soil types, leading to differential nutrient availability across the landscape. However, information on the spatial variability of available P in this region remains limited. Understanding the spatial distribution of P in the area can help farmers and agronomists adopt more efficient nutrient management practices, improving agricultural sustainability and productivity. This study aimed to assess the spatial variability of available P in the soils that can aid in the targeted management of P for improved soil fertility and crop productivity.

2. Methodology:

2.1. Location of study

The investigation was performed from January, 2022-December, 2024 at Jawaharlal Nehru KrishiVishwaVidyalaya, Jabalpur, Madhya Pradesh (482004), India. The Kymore Plateau and Satpura Hill Zone in central Madhya Pradesh, India encompassing the districts of Jabalpur, Katni, Seoni, Panna, Rewa, Sidhi, Singrauli, and Satna, was selected for this study. Located between 21° to 24° north latitude and 79° to 83° east longitude, this region exhibits diverse geographical features (Figure 1). The zone showed significant variation in physiography, soil types, rainfall, irrigation practices, and cropping patterns across different areas. The region receives an annual rainfall ranged from 1000–1400 mm.

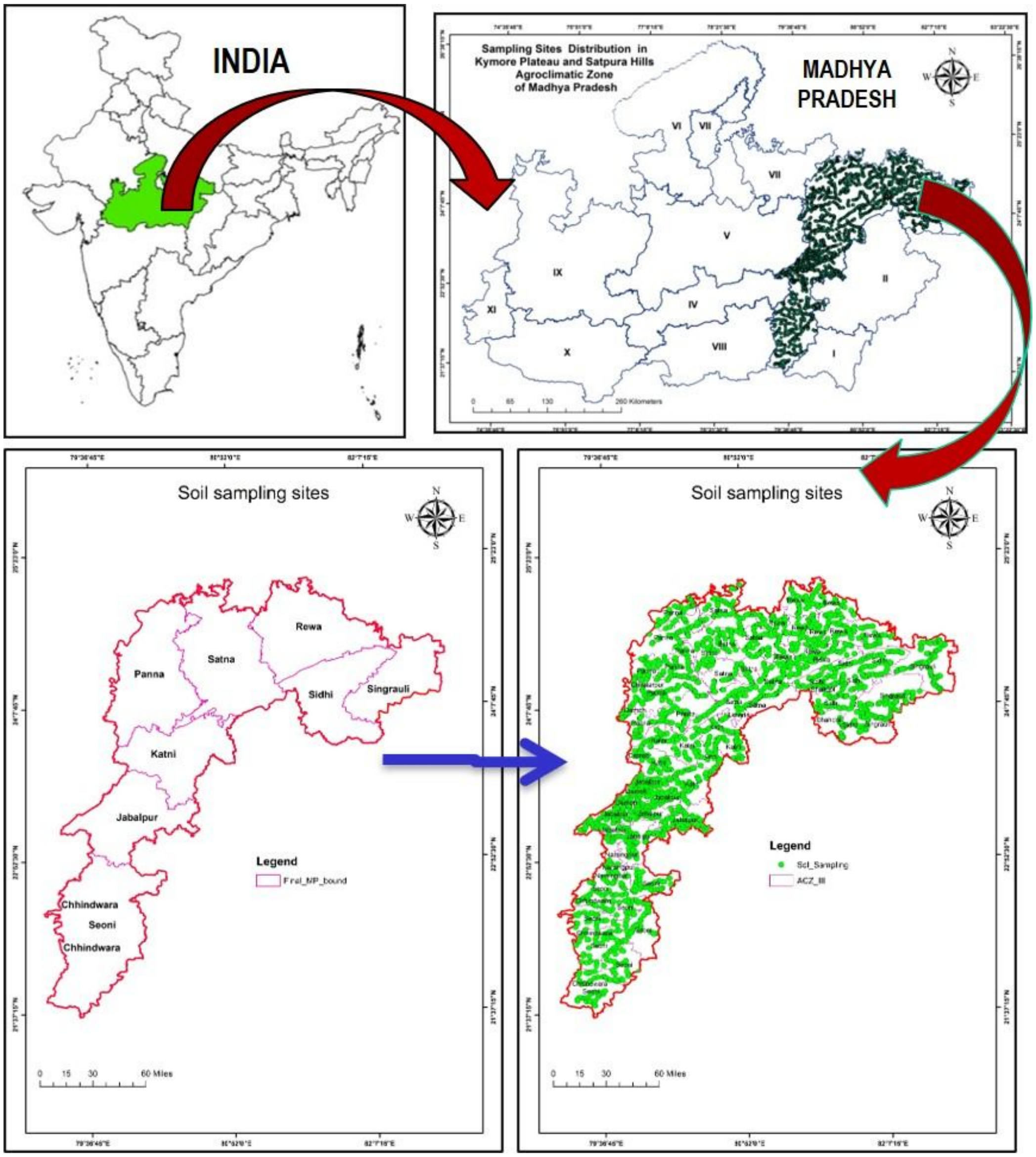


Figure 1: Sampling sites in the study area

2.2. Sample collection and analysis

In the off-season of 2022–2023, a total of 2216 soil samples at 0–15 cm depth was collected from farmers' fields using GPS. The soil samples were air-dried to remove moisture, crushed with a wooden hammer and sieved through a 2 mm mesh. The available phosphorus in soilextracted by 0.5 M sodium bicarbonate ($NaHCO_3$) with a pH of 8.5 and with the help of a spectrophotometer assessed the intensity of the blue color at a wavelength of 660 nm by Olsen et al. (1954).

2.3 Statistical analysis:

Data was calculated such as mean, median, minimum, maximum, standard deviation (SD), skewness, and kurtosis to recognize how data is distributed and each soil characteristics were examined during descriptive statistics.

2.4 Spatial interpolation analysis in GIS-Environment

Geo statistics is a branch of statistics focusing on spatial or spatiotemporal datasets. The main tool in geo statistics is the semi-variogram, which expresses the spatial dependence between neighboring observations. The semi-variogram quantifies the relationship between the semi variance and the distance between sampling pairs by the following equation

$$\lambda(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(X_i) - Z(X_i + h)]^2 \qquad \text{equation} \quad (1)$$

Where N (h) is the number of all pair-wise Euclidean distances, and z (xi) and z (xi+h) are observations of the variable Z at spatial locations xi and xi+h, respectively.

2.4.1 Semi-variogram parameters

The semi-variogram parameters, the nugget, the sill and the range. Nugget (C_0) represents the measurement and data errors or random spatial sources of variation at distances smaller than the sampling interval or both and represents the value of the initial variability. Range (a) is the distance where the semi-variogram reaches the total sill ($C_0 + C_1$) and after that distance, there is no spatial correlation of the data. Sill is the value that the semi-variogram reaches the range and represents the maximum variability, while partial sill is the sill minus the nugget (C_1-C_0). The use of the nugget-sill ratio ($C_0/C_0 + C_1$) was applied for the estimation of the spatial dependence of the variables (Jerosch 2013; Adhikary and Dash 2017; Tziachris et al. 2017). A ratio of less than 25% means strong spatial dependence, while a ratio between 25 and 75% indicates a moderate spatial dependence and a ratio over 75% shows a weak spatial dependence (Cambardella et al. 1994).

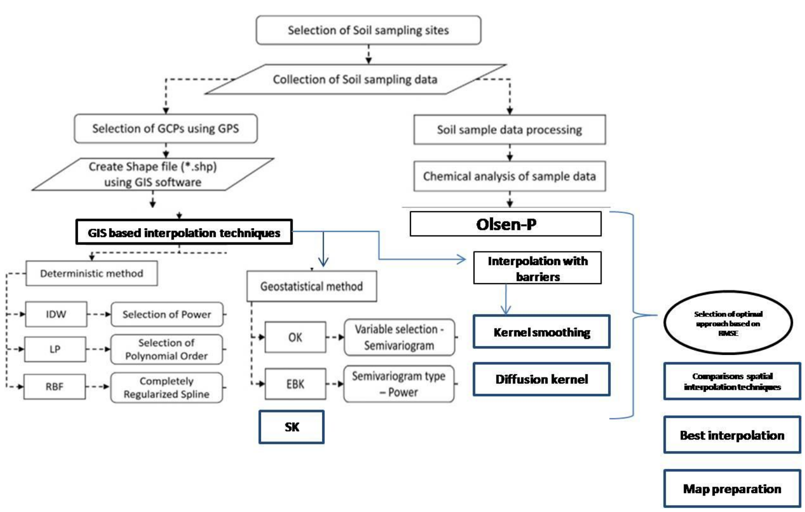


Fig.2 Workflow for sample acquisition and geo-statistical interpolation

2.5 Spatial interpolation analysis in GIS-Interpolation methods

Three geo-statistical techniques (OK, SK and EBK), three deterministic (IDW, RBF and LPI) algorithms and two interpolations with barriers (Kernel smoothing and Diffusion Kernel) were implemented to interpolate the available P in soils. The Box-Cox transformation normalizes the data, stabilizing variance and improving the fitness of the model.

2.5.1Deterministic models

2.5.1.1 Inverse Distance Weighting (IDW):

IDW is one of the most widely used deterministic (mechanical) interpolation methods in soil research which combines multivariate statistical analysis with GIS. This method assumes that the measured values at a closer distance have greater weight than those further away. The influence of a known value is inversely related to the distance from the unknown data point. Consequently, this method gives greater weights to values closest to the prediction position and the weights reduce as a function of distance (Qi et al., 2020). For this work, IDW calculations were performed on adjacent observed points. It is implied that the known observed points regulate themselves independently of one another (Bhuniaet al., 2018; Saha et al., 2022).

$$Z = \sum ni{=}1\ (Zi/dpi) / \sum ni{=}1\ (1/dpi) \qquad \text{equation} \quad (2)$$

Where Z denotes the approximate value at an interpolated point; Zi denotes the computed values at point i; n denotes the total number of values obtained in interpolation; di denotes the distance between interpolated value Z and the computed value Zi, and p denotes the weighting power.

2.5.1.2 Radial Basis Functions (RBF): The RBF (also known as Spline) refers to a set of precise interpolation techniques that are based on artificial neural networks (ANN) i.e. input layer, hidden layers, and output layer (Johnston et al., 2001; Antal et al., 2021, Ali et al. 2021). In addition, RBF can predict values above the maximum and below the minimum. The technique includes five distinct basis functions: thin-plate splines (TPS), spline with tension (ST), inverse multi-quadratic function (IMQ), completely regularized spline (CRS), and multi-quadratic function (MQ). RBF provides predictions about new values based on an operator-specified region, and each predicted value must carry through each measured value (Xie et al., 2011; Antal et al., 2021).

$$(x) = \sum mi{=}1\ aifi(x) + \sum ni{=}1\ bj\psi(\ dj\ )\text{equation} \quad (3)$$

Where dj represents the distance between each observed sample point and the estimated point x, and (dj) represents the radial basis functions. The trend function $f(x)$ is regarded as a component of the basis for polynomials with degree m; n is the total number of known points considered in the interpolation. In this study, we have assessed the completely regularized spline: radial basis functions (RBF-CRS). The following functional equations are used for this radial basis function case (Xie et al., 2011).

$$(d) = \ln\ (\ cd\ 2\ )2 + E1(cd)\ 2 + \gamma\text{equation} \quad (4)$$

Where d represents the difference between the estimated and observed points, c represents the smoothing factor, $E1$ represents the modified Bessel function, and γ denotes the Euler's constant.

2.5.1.3 Local Polynomial Interpolation (LPI): The LP interpolation method only uses points in the predefined neighborhood to match the specific polynomial order (Saha et al., 2022). This technique adjusts a unique polynomial equation for each region based on the maximum and minimum observed values, regions, observed neighborhood types, and kernel types (Johnston et al., 2001; Antal et al., 2021). The purpose of polynomial interpolation is to identify a polynomial that can access a group of specified observation points. Overall, a global polynomial may cover the entire surface; however, it cannot perfectly match the surface when there is more natural variation (Liao et al., 2018). The LP method provides a number of advantageous characteristics, including efficiency and the ability to successfully detrend data in a variety of geostatistical models (Gribov and Krivoruchko, 2011).

$$Zi = (1 - diR)p \qquad \text{equation (5)}$$

Where Zi is the mean observed values made at the ith measurement point, di represents the difference between observed and predicted points, R denotes the neighboring area carried into consideration, and p is the order of the polynomial function defined by the operator.

Therefore, LPI fits the specified order (zero, first, second and third) using all points only within the defined overlapping neighbourhood, which is used as a value for each prediction in the fitted polynomial at the centre of the neighbourhood (Johnston et al. 2001). GPI is useful for identifying long-range trends in the dataset, whereas LPI can produce surfaces that capture the short-range variation.

2.5.2 Geo-statistical methods:Kriging

The Kriging method is based on a sequential interpolation technique that applies a semi-variogram model to predict unknown values based on distance and variations in measured values (Paramasivam and Venkatramanan, 2019). Kriging method is a precise interpolation estimator used to find the best linear unbiased estimate. The general form of kriging equation:

$$Z(Xp) = \sum_{i=1}^{h} \lambda i \, Z(x_i) \text{equation} \quad (6)$$

This estimator has high application due to minimizing of error variance with unbiased estimation.

2.5.2.1 Ordinary Kriging (OK)

The basis of OK is a statistical model that includes autocorrelation, or the statistical correlations between the observed points. However, geo-statistical algorithms not only have the proficiency to generate a prediction surface but also offer some indication of the reliability or efficiency of the prediction (Oliver and Webster, 1990; Hu et al., (2016) Ghosh et al., (2020).OK emphasizes the function that is spatially associated and represented as the following weighted sum of the data:

$$(x) = \sum ni{=}1 \ \lambda i Z(\ xi\ ) \text{equation} \quad (7)$$

(x) denotes the estimated value at point x, $Z(xi)$ denotes the observed value at position x, λi indicates the weight applied to the residual of $Z(xi)$, and n represents the number of sample data utilized at specific points within the neighborhood.

2.5.2.2 Simple Kriging (SK)

In contrast with OK, the application of Simple Kriging (SK) presupposes the assumption of stationarity. SK considers μ to be known and constant all over the study area, unlike with the OK type, where the μ is unknown and is considered to fluctuate locally, maintaining the stationarity within the local neighbourhood (Moral et al., 2010). The equation used for SK interpolation is:

$$ZSK(xo)=\sum i{=}1N\lambda\iota Z(xi)+(1-\sum i{=}1N\lambda\iota)\mu \text{equation} \quad (8)$$

where μ is a known stationary mean

2.5.2.3 Empirical Bayesian Kriging (EBK)

EBK is a combination of two geo-statistical concepts: intrinsic random function kriging (IRFK) and linear mixed model (LMM) (Schabenberger and Gotway, 2017). In EBK, the stochastic spatial process is represented locally as a stationary or nonstationary random field and the parameters of the locally defined random field are allowed to vary across space (Gribov and Krivoruchko 2020). EBK is a geo-statistical interpolation method that automates the most difficult aspects of building a valid kriging model through a process of sub setting the study area, coupled with multiple simulations to obtain the best fit (Krivoruchko and Gribov 2019). This process finally creates a spectrum of semi-variograms and each of these is an estimate of the true semi-variogram for the subset (Pellicone et al. 2018).

$$Zi = (si\ ) + \varepsilon i,\ i = 1 \text{……equation} \quad (9)$$

2.5.3 Interpolation with Barriers

2.5.3.1 Kernel smoothing interpolation with Barrier is the variance of the first-order local polynomial interpolation method, which uses methods similar to those used in ridge regression. As a moving window predictor, the kernel interpolation model uses the shortest distance between two points, and points located on the arbitrary side of a specified absolute line barrier are connected through a series of straight lines. However, the kernel interpolation method without absolute barriers has higher smoothness at the contour line of the interpolated surface. KIB consists of six different kernel functions, including Exponential, Gaussian, Quartic, Epanechnikov, Polynomial and Constant function. The Polynomial function was used in this study as a kernel function, with the degree of the polynomial being the default value 1, and other parameters remaining default.

2.5.3.2 Diffusion Interpolation with Barrier (DIB) uses a kernel interpolation surface based on the heat equation and allows the distance between input points to be redefined using raster and element barriers. In the absence of barriers, the estimations obtained by diffusion interpolation are approximately identical to those by kernel interpolation with a Gaussian kernel.

2.6 Performance metrics

In this study,the performance metrics like RMSS (Root Mean Square Standardized Residuals), RMS (Root Mean Square Error), and ASE (Average Squared Error). It is ideal to have the RMSSE close to one. If the RMSSE > 1, then it indicates a general under-estimation in the inconsistency of the estimated/ predicted variable. If the RMSSE < 1, then it indicates a general overestimation in the inconsistency of the estimated/predicted variable. ASE measures the arithmetic average of the prediction standarderrors. It represents the error magnitude showing the method's accuracy.MSE provides the average of the standardized errors. The value of MSE is better if it is close to 0

3. Results and Discussion

3.1. Status of available P content in soil

The result presented in Table 1 showed that the Olsen phosphorus content ranged from 1.15 to 95.18 kg ha^{-1}, 1.44 to 27.19 kg ha^{-1}, 1.44 to 53.52 kg ha^{-1}, 1.73 to 50.05 kg ha^{-1}, 1.19 to 50.67 kg ha^{-1}, 1.44 to 117.17 kg ha^{-1}, 1.44 to 67.12 kg ha^{-1}, and 1.11 to 96.3 kg ha^{-1} with a mean value of 13.32, 9.30,9.77,10.88,9.22,11.03,8.92,and 11.20 kg ha^{-1} in soils of Jabalpur, Katni, Panna, Seoni, Rewa, Sidhi, Singrauliand Satna district, respectively. The available Pcontent in districtsexhibited the orderJabalpur>Satna>Sidhi>Seoni>Panna>Katni>Rewa>Singrauli.Overall,availableP content in the Kymore Plateau and Satpura hill zone ranged from 1.11 in Satnato 117.7 kg ha^{-1}in Sidhiwith amean of 10.64 kg ha^{-1}.

The CV of 87.38, 51.83, 64.66, 52.81, 74.29, 92.16,72.90, and 99.61 % indicate substantial heterogeneity in soils ofJabalpur, Katni, Panna,Seoni, Rewa, Sidhi, Singrauli and Satna district, respectively. The high kurtosis value (8.08) suggests that the data distribution is leptokurtic, with a higher peak and more outliers than a normal distribution. The skewness value of 2.24 indicates a right-skewed distribution, meaning that the phosphorus in Jabalpur is generally clustered toward the lower end, but with a few areas exhibiting higher P concentrations. This could be indicative of relatively uniform agronomic practices and soil conditions across the district. In the Sidhidistrict a skewness of 6.20, indicating an extremely right-skewed distribution. The very high kurtosis value (58.13) suggests a distribution with very pronounced peaks and heavy tails. The skewness of 3.83 indicates strong skew. The high kurtosis value (19.56) further suggests a distribution with a sharp peak, implying that while most of the phosphorus concentrations are low, there are several areas with higher values. Panna and Seoni exhibited skewness values of 2.52 and 2.42, respectively, indicating right-skewed distributions;High variability was evident, as reflected by the standard deviation (SD) of 8.61 and a CV of 80.96%, indicating a significant spread of P availability across the region.

Table 1: Status of available P (kg ha^{-1}) in soils of Kymore Plateau and Satpura hills zone

Kymore Plateau & Satpura Hills Zone (n)	**Available P (kg ha^{-1})**							
	Min	**Max**	**Mean**	**SE**	**SD**	**CV%**	**Kurtosis**	**Skewness**
Jabalpur (379)	1.15	95.18	13.32	0.6	11.64	87.38	8.08	2.24
Katni (220)	1.44	27.19	9.30	0.32	4.82	51.83	1.54	1.00
Panna (301)	1.44	53.52	9.77	0.36	6.32	64.66	11.09	2.52
Seoni (300)	1.73	50.05	10.88	0.33	5.74	52.81	10.96	2.42
Rewa (284)	1.19	50.67	9.22	0.41	6.85	74.29	6.05	2.07
Sidhi (210)	1.44	117.17	11.03	0.70	10.17	92.16	58.13	6.20
Singrauli (222)	1.44	67.12	8.92	0.44	6.50	72.90	29.51	3.96
Satna (300)	1.11	96.30	11.20	0.64	11.15	99.61	19.56	3.83
Overall (2216)	**1.11**	**117.17**	**10.64**	**0.18**	**8.61**	**80.96**	**26.63**	**3.81**

The high variability might be associated with differences in soil types, land use practices, and fertilizer application patterns. The relatively high phosphorus levels in some areas of Jabalpur could indicate localized soil conditions that promote better phosphorus availability, such as the presence of less P-fixing soil types. Lower phosphorus availability in Katni could be due to specific soil properties, such as high acidity, that limit the phosphorus availability to crops. The relatively uniform phosphorus levels across the district may also reflect uniform agronomic practices. The higher available phosphorus in Sidhi could be due to the application of phosphorus-rich fertilizers or the presence of phosphorus-rich parent material. However, the high variability suggests that not all areas are equally enriched with phosphorus, highlighting the need for localized management. The lower P status may necessitate targeted interventions, such as phosphorus fertilization or soil amendments, to optimize crop yields in SingrauliSatna, Soils in this region likely suffer from high P fixation or low phosphorus content due to poor soil fertility management and the inherent characteristics of the local soils. The results on soil phosphorus availability in the study area highlight the variability in P content due to factors like soil type,and management practices. For instance, a study by Raghuwanshi et al. (2023) reported with mean P values in the range of 9-12 kg ha^{-1} in central Indian soils. Desavathu et al. (2017) Maqbool et al. (2017) Sharma et al. (2021) andVerma et al. (2022) who also revealed that P, contents was low in the cultivated land.

3.2 Nutrient status in soils

Data given in Table 2 showed that the P status in soils were about 52.77, 62.27, 62.13, 71.83, 61.00, 51.00, 60.00 and 70.27; 28.50, 34.09, 32.56, 19.01, 29.00, 44.00, 30.48 and 25.67 and 18.73, 3.64, 5.32, 9.15, 10.00, 5.00, 9.52 and 4.05 soil samples were observed to be low, medium and high, respectively. The nutrient index values of 1.66, 1.41, 1.43, 1.37, 1.49, 1.54, 1.50 and 1.34 for P in Jabalpur, Katni, Panna, Rewa, Satna,Seoni, Sidhi and Singrauli districts, respectively. The P deficiency, is common in tropical and subtropical soils, where phosphorus is often fixed by iron and aluminium oxides, rendering it unavailable to plants, potassium deficiencies, despite its relatively high abundance in many soils, can still limit plant growth due to the high leaching rate in regions with high rainfall,Gupta& Joshi (2020). Similar observations were made by Yadav et al. (2023), who found diverse phosphorus distributions in the region, suggesting the need for targeted nutrient management strategies based on local soil fertility levels.

Table 2. Available P status insoils of Kymore Plateau and Satpura hill zone based on critical limit

Kymore Plateau and Satpurahill(n)	**Percent sample**			
	Low	**Medium**	**High**	**NI**
Jabalpur (n=379)	52.77	28.50	18.73	1.66
Katni (n=220)	62.27	34.09	3.64	1.41
Panna (n=301)	62.13	32.56	5.32	1.43
Rewa (n=284)	71.83	19.01	9.15	1.37
Satna (n=300)	61.00	29.00	10.00	1.49
Seoni (n=300)	51.00	44.00	5.00	1.54
Sidhi (n=210)	60.00	30.48	9.52	1.50
Singrauli (n=222)	70.27	25.67	4.05	1.34
Total(n=2216)	**60.74**	**30.46**	**8.79**	**1.48**

3.3Correlation between soil physico-chemical properties and available P in soil

Data on correlation Table 3revealed significant relationships between soil pH, electrical conductivity (EC), and organic carbon with phosphorus (P) content. Specifically, soil pH, EC, and organic carbon showed positive correlations with P (r = 0.073**, r = 0.068**, r = 0.133*), respectively. These results align with findings from Bania et al. (2024) who reported significant relationships between soil organic carbon and the availability of P in soils, highlighting the importance of organic matter in enhancing nutrient availability. Jadon et al., 2023 and Chen et al., 2020 found a significant relationship among soil physic-chemical properties on available P.

Table 3 Correlation matrixbetween soil physico-chemical properties and available P

Parameter	pH	EC	OC
EC	0.435**		
OC	-0.055**	0.029NS	
P	0.073**	0.068**	0.133**

*= significant at 5%, **= significant at 1%, NS = non-significant

3.4Comparison between different spatial interpolation methods

A comparative study between the deterministic, geo-statistical approaches and interpolation with barrierswas applied and the performance of each interpolation method given in Table 4, 5and 6.Figure 4 provides a comparison of various interpolation methods and highlighted key metrics like Mean, RMSS, RMS, MS and ASE.

Table4 Selection value of input data for each interpolator

Name	Trans.	Kernel Function (order)	Others parameters
Deterministic models			
IDW	None		Min 10 and max 15 neighbors Sector 1
RBF	None	CRS	Min 10 and max 15 neighbors Sector 1
LPI	None	Exponential (1)	
Geo-statistical methods -Kriging			
OK	None		Min 2 and max 5 neighbors and 4 Sectors with 45° offset
OK	Box cox	Polynomial5(2)	Sector 1 Bandwidth47245.12
SK	Box cox	Polynomial5(2)	Bandwidth47245.12
EBK	None		Transformation -Empirical Min 10 and max 15 neighbors' sector 1 Semi variogram types Exponential Detrended Radius (7043.764)
Interpolation with barriers			
Kernel smoothing (KS)	None	Exponential (1)	(ridge 50) smoothing factor 0.2 and radius bandwidth 13881.05
Diffusion Kernel (DK)	None	Gaussian	Bandwidth 19470.42 (Iteration 100)

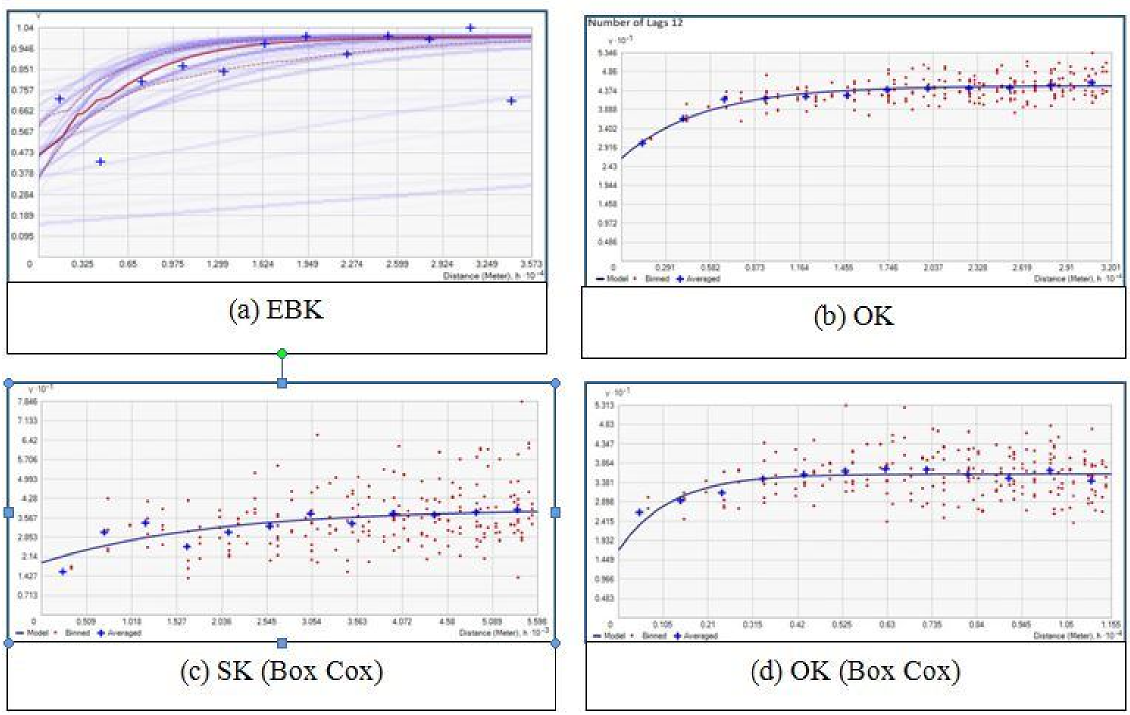
In the Figure 3 indicatesthat standard OK, the exponential semi-variogram structure a significant range (16,020.91 m). The nugget-to-sill(N/S) ratio of 0.58 is relatively high. The OK with Box-Cox transformation showed a reduced range of 3,652.22 m. The nugget

(16.907) is also smaller, while the partial sill is slightly higher (19.055). The N/S ratio of 0.47 is lower than that of the untransformed OK, suggesting that the Box-Cox transformation has improved the model's ability to account for spatial variability and reduced the influence of unmodeled random noise.Molla et al., (2023) include the exponential model which aid in capturing the spatial correlation structure and are fundamental for accurate predictions and spatial analysis (Mondal et al., 2021).

Table 5 Semi-variogram parameters for each interpolation method

Kriging Type	Model	Range(a) (m)	Nugget (C_o)	Partial sill (C_1)	Sill	Nugget-Sill ratio
OK (none)	Exponential	16020.91	26.428	18.550	44.978	0.588
OK(Box-cox)	**Exponential**	**3652.22**	**16.907**	**19.055**	**35.962**	**0.470**
SK(Box-cox)	Exponential	5598.37	19.213	19.711	38.924	0.494

Fig.3 (a-d): Variogram parameters fitted to the data of available P in soils



The SK method with Box-Cox transformation exhibits a range of 5,598.37 m, which is intermediate between the untransformed OK and OK-Box-Cox models. The nugget (19.213) is slightly higher than that in OK-Box-Cox, and the partial sill (19.711) is slightly greater as well. The N/S ratio of 0.494 reflects a moderate proportion of spatial variability being explained by the model.The nugget-to-sill ratios, which quantify the proportion of unexplained variation relative to the total spatial variability, are lower in the transformed models (0.47 and 0.49 for OK-Box-Cox and SK-Box-Cox, respectively) compared to the untransformed OK (0.588). This suggests that the Box-Cox transformation helps in reducing the influence of noise and unexplained variability, leading to more reliable and interpretable predictions.

Table 6 Parameters of accuracy of prediction

Method	Mean	RMSS	RMS	MS	ASE	Regression function
Deterministic models						
IDW	-0.0210		6.182			0.2446 x + 7.1080
RBF	**-0.0557**		**5.957**			**0.1828 x + 7.8105**
Local polynomial	0.0119		5.972			0.1603 * x + 8.1764
Geo-statistical methods - Kriging						
OK	-0.0150	0.969	5.951	-0.0021	6.149	0.1862* x + 7.8329
OK (Box cox)	0.0105	1.007	5.96	0.0016	5.917	0.2109 * x + 7.6234
SK(Box-cox)	**0.0112**	**0.986**	**5.900**	**0.0018**	**5.986**	0.1891* x + 7.8425
EBK	**-0.0327**	**0.989**	**5.774**	**-0.0050**	**5.784**	**0.2110* x + 7.6629**
Interpolation with barriers						
KS	**0.0460**	**1.021**	**5.986**	**0.0080**	**5.888**	0.1873* x + 7.9432
DK	0.0230		6.163			0.0617* x + 9.1502

Deterministic models

IDW showed a mean of -0.021, with an RMS of 6.18, suggesting it has a modest level of accuracy. The regression equation derived is 0.2446x + 7.1080. This means it tends to over-predict at higher values.RBF demonstrated a mean of -0.0557 and an RMS of 5.957, which indicates improved prediction accuracy compared to IDW. Its regression function, 0.1828x + 7.8105, is indicative of a better fit with a steeper slope than IDW, making it more suitable for certain spatial patterns.With a mean of 0.0119 and an RMS of 5.972, the Local Polynomial method offers the smallest mean, and it shows improved performance with a regression function of 0.1603 * x + 8.1764.

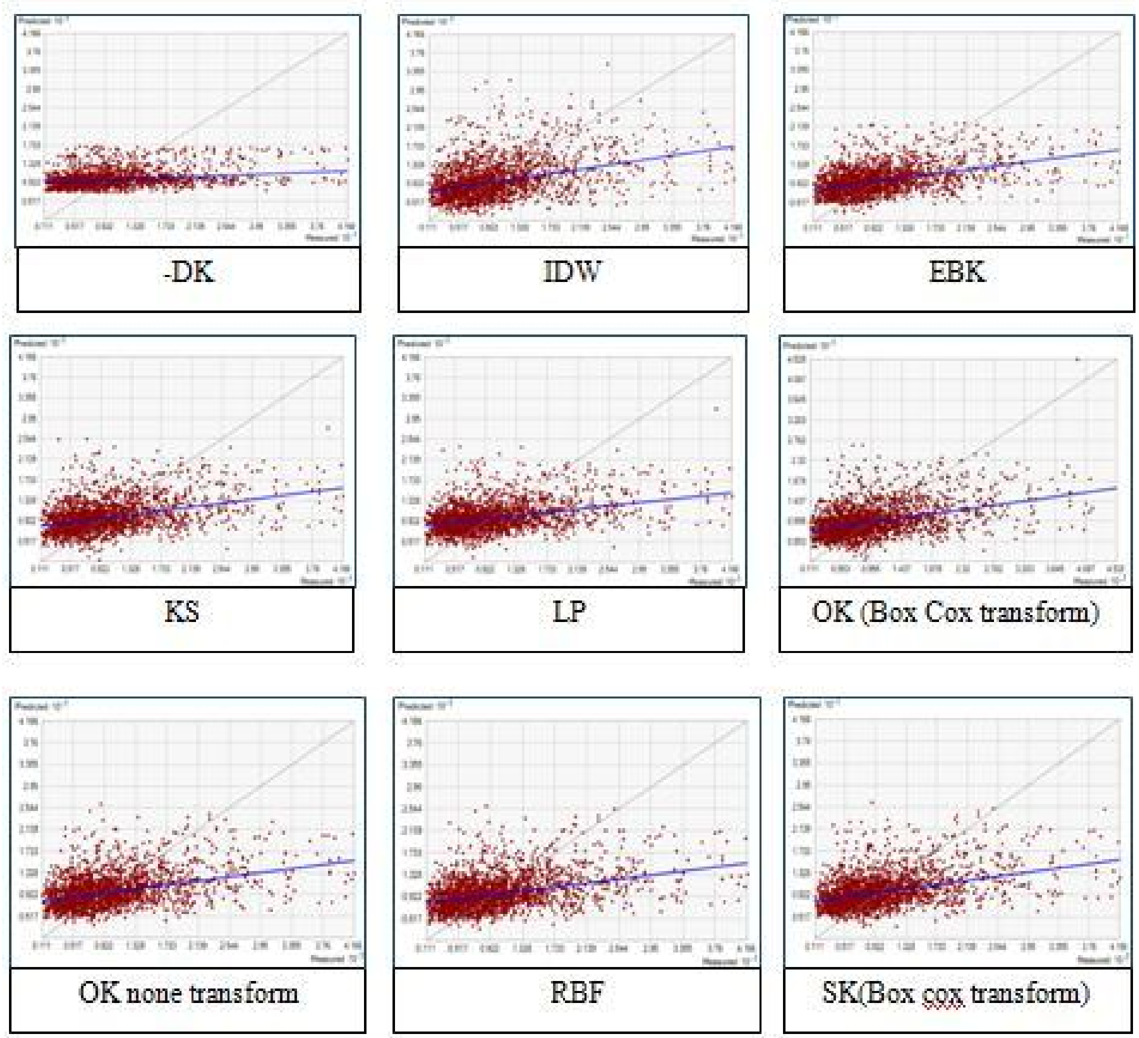
Geo-statistical methods

OK showed a mean of -0.0150 and RMS of 5.951 shows that it performs comparably to the deterministic models. The regression function of 0.1862x + 7.8329 indicates that Kriging accounts for spatial correlation better than deterministic models.With a mean of 0.0105 and an RMS of 5.960, OK with Box-Cox model slightly outperforms the standard OK method. The regression function 0.2109x + 7.6234 shows a higher sensitivity to changes in x. The simple Kriging(SK) method with Box-Cox transformation performed well, with a mean of 0.0112 and an RMS of 5.900. Its regression function, 0.1891 * x + 7.8425, shows a slight improvement over ordinary kriging due to the transformation's ability to better handle skewed data.EBK (Empirical Bayes Kriging) provides a robust alternative with a mean of -0.0327 and an RMS of 5.774, both showing improved accuracy over the traditional Kriging methods. Its regression function of 0.2110 * x + 7.6629 suggests a highly adaptive model that might be preferable when dealing with uncertain or heteroscedastic data.

Interpolation with barriers

With a mean of 0.0460 and an RMS of 5.986, the KS method provides a slightly higher mean error. The regression equation 0.1873 * x + 7.9432 reflects the importance of barriers in improving model accuracy in heterogeneous environments. The DK method presents a mean of 0.0230 and an RMS of 6.163, which indicates it performs reasonably well. The regression function of 0.0617 * x + 9.15029 demonstrates a lower sensitivity to changes in x compared to other methods, making it potentially useful in less variable data sets.

Figure 4. Scatter plots of measured (y-axis) *vs.* predicted (x-axis) available P values from all interpolation methods

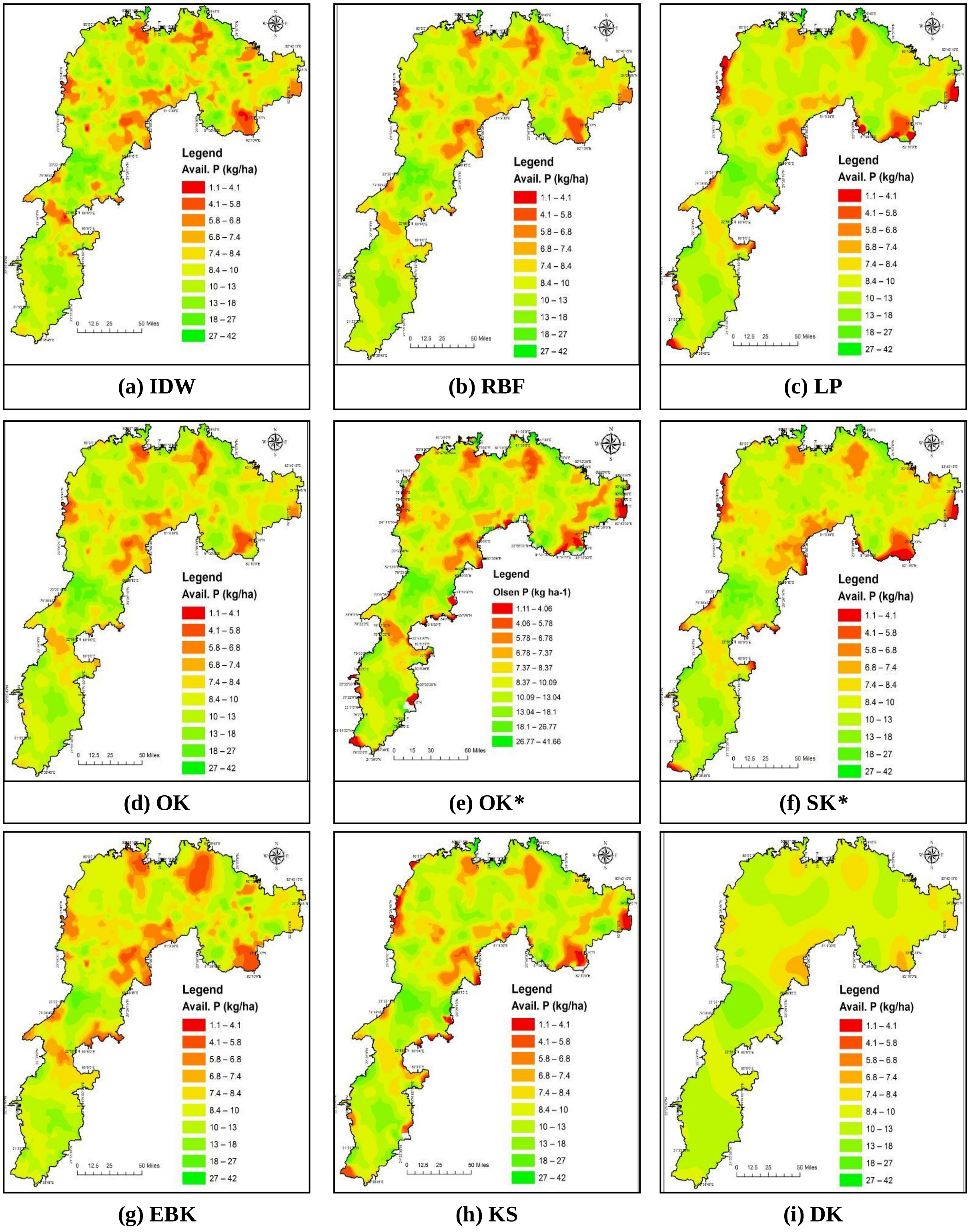


From the results, it is evident that EBK stands out due to its ability to provide robust estimates under uncertain conditions. Geo-statistical methods usually have a better performance than the deterministic ones (Li and Heap 2011; Yao et al. 2014), According to Adhikary and Dash (2017) and Wen et al. (2022) RBF outperforms the OK method and IDW. Zarco-Perello and Simões, (2017) estimated that the IDW outperforms the OK method. Qiao et al., (2018) recommended that the prediction accuracy of IDW was higher than OK in the spatial prediction of As concentration in the soils of Beijing.Guan et al., (2017), Ramzan et al., (2017) and Valera et al., (2023) also reported a moderate spatial dependence of P contents indicating that P were controlled by both intrinsic and extrinsic factors. Tuncay et al., 2021 found the OK methods yielded the lowest RMSE values.Bhunia et al., (2018) also proved that LPI had higher prediction accuracy than IDW and RBF in the spatial distribution of soil organic carbon (SOC). Chen et al (2016) and Daya and Bejari, (2015)calculated that OK is more efficient as an interpolator compared with the SK

3.5 Spatial variability map of available P

Geo-statistical analysis was performed using the kriging interpolation technique within the spatial analyst extension module in Arc-GIS software package to determine the spatial dependency and spatial variability of soil properties. In this analysis, geo-statistical tools were applied to model the available P across different methods Figure 5.***Box-Cox transformed for normality***

Fig.5.(a-i) Spatial distribution of available Pfrom different spatial interpolation methods



4. Conclusion

This study demonstrates phosphorus availability in the Kymore Plateau and Satpura Hills Zone, is medium statusand highly variable with some districts exhibitedhigh Pstatus. The Box-Cox transformation improved spatial data fitting by reducing unexplained variability. Empirical Bayes Kriging (EBK) yielded the most accurate results, followed by Ordinary Kriging (OK) and Simple Kriging (SK) with Box-Cox Overall, geo-statistical tools provide valuable insights for managing soil fertility and informing agricultural decisions.

Funding: The research was conducted with the supports from NAHEP-CAAST on “Skill Development to Use Data for Natural Resources Management in Agriculture”, College of Agricultural Engineering Jawaharlal Nehru KrishiVishwaVidyalaya, Jabalpur.

Acknowledgement: This research was supported by the National Agriculture Higher Education Project, Centre for Advanced Agriculture Science and Technology (NAHEP-CAAST) on “Skill Development to Use Data for Natural Resources Management in Agriculture”, College of Agricultural Engineering Jawaharlal Nehru Krishi Vishwa Vidyalaya, Jabalpur for providing financial, and the Department of Soil science and Agricultural Chemistry, laboratory facility and other support.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

Option 2:

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc. have been used during the writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1.
2.
3.

5. References

1. AbdelRahman, M. A. Zakarya, Y. M.Metwaly, M. M. and Koubouris, G.(2020).Deciphering soil spatial variability through geostatistics and interpolation techniques. Sustainability. 13(1):194.
2. Adhikary, P. P., & Dash, C. J. (2017). Comparison of deterministic and stochastic methods to predict spatial variation of groundwater depth. Applied Water Science, 7, 339-348.
3. Ali, G., Sajjad, M., Kanwal, S., Xiao, T., Khalid, S., Shoaib, F., &Gul, H. N. (2021). Spatial–temporal characterization of rainfall in Pakistan during the past half-century (1961–2020). Scientific reports, 11(1), 1-15.

4. Antal, A., Guerreiro, P. M., & Cheval, S. (2021). Comparison of spatial interpolation methods for estimating the precipitation distribution in Portugal. Theoretical and Applied Climatology, 145(3), 1193-1206.
5. Bania, J. K., Sileshi, G. W., Nath, A. J., Paramesh, V., & Das, A. K. (2024). Spatial distribution of soil organic carbon and macronutrients in the deep soil across a Chrono sequence of tea agroforestry. Catena, 236, 107760.
6. Bhunia, G. S., Shit, P. K., &Maiti, R. (2018). Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC). Journal of the Saudi Society of Agricultural Sciences, 17(2), 114-126.
7. Bishop, T. F. A., &McBratney, A. B. (2001). A comparison of prediction methods for the creation of field-extent soil property maps.Geoderma, 103(1-2), 149-160.
8. 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.
9. Brus, D. J., De Gruijter, J. J., Marsman, B. A., Visschers, R., Bregt, A. K., Breeuwsma, A., &Bouma, J. (1996). The performance of spatial interpolation methods and choropleth maps to estimate properties at points: a soil survey case study. Environmetrics, 7(1), 1-16.
10. Cambardella, C. A., Moorman, T. B., Novak, J. M., Parkin, T. B., Karlen, D. L., Turco, R. F., &Konopka, A. E. (1994). Field□scale variability of soil properties in central Iowa soils. Soil science society of America journal, 58(5), 1501-1511.
11. Chen, S., Lin, B., Li, Y., & Zhou, S. (2020). Spatial and temporal changes of soil properties and soil fertility evaluation in a large grain-production area of subtropical plain, China. Geoderma, 357, 113937.
12. Criado, M. Martínez-Graña, A. Santos-Francés, F. and Merchán, L. (2021). Improving the Management of a Semi-Arid Agricultural Ecosystem through Digital Mapping of Soil Properties: The Case of Salamanca (Spain). Agronomy. 11(6).1189.
13. Davis, H. T., Aelion, C. M., McDermott, S., & Lawson, A. B. (2009). Identifying natural and anthropogenic sources of metals in urban and rural soils using GIS-based data, PCA, and spatial interpolation. Environmental Pollution, 157(8-9), 2378-2385.
14. Daya, A. A., &Bejari, H. (2015). A comparative study between simple kriging and ordinary kriging for estimating and modeling the Cu concentration in Chehlkureh deposit, SE Iran.Arabian journal of Geosciences, 8, 6003-6020.
15. Desavathu, R.N., Nadipena, A.R., Peddada, J.R., 2017. Assessment of soil fertility status in PaderuMandal, Visakhapatnam district of Andhra Pradesh through Geospatial techniques. The Egyptian Journal of Remote Sensing and Space Sciences, 21(1), 73–81.
16. Fischer, A., Lee, M. K., Ojeda, A. S., & Rogers, S. R. (2021). GIS interpolation is key in assessing spatial and temporal bioremediation of groundwater arsenic contamination. Journal of Environmental Management, 280, 111683.
17. Gangopadhyay, S.K. and Reddy, G.P. (2022). Spatial variability of soil nutrients under the rice-fallow system of eastern India using geostatistics and Geographic Information System. Journal of Soil and Water Conservation. 21(1):55-66.
18. Ghosh, M., Pal, D. K., &Santra, S. C. (2020). Spatial mapping and modeling of arsenic contamination of groundwater and risk assessment through geospatial interpolation technique. Environment, Development and Sustainability, 22, 2861-2880.
19. Gokmen, V. Sürücü, A. Budak, M. and Bilgili, A. V. (2023). Spatial modeling and mappingof the spatial variability of soil micronutrients in the Tigris basin. Journal of King Saud University- Science. 102724.
20. Goovaerts, P. (1997). Geostatistics for natural resources evaluation (Vol. 483). Oxford University Press.
21. Gribov, A., &Krivoruchko, K. (2011). Local polynomials for data detrending and interpolation in the presence of barriers. Stochastic environmental research and risk assessment, 25, 1057-1063.
22. Gribov, A., &Krivoruchko, K. (2020). Empirical Bayesian kriging implementation and usage. Science of the Total Environment, 722, 137290.

23. Guan, F., Xia, M., Tang, X., and Fan, S. 2017. Spatial variability of soil nitrogen, phosphorus and potassium contents in Moso bamboo forests in Yong'an City, China. Catena, 150, 161-172.
24. Gupta,S., & Joshi, D. (2020). "Soil Potassium Management for Sustainable Crop Production." Journal of Soil Science and Environmental Management, 11(4), 123-135.
25. Hu, Y., Jia, Z., Cheng, J., Zhao, Z., & Chen, F. (2016). Spatial variability of soil arsenic and its association with soil nitrogen in intensive farming systems.Journal of soils and sediments, 16, 169-176.
26. Jadon, P., Singh, V., Kamle, S., Kamle, R., &Mandloi, S. (2023). Spatial variability of soil macronutrients and chemical properties in Ujjain Tehsil of Ujjain District of Madhya Pradesh, India. Int. J. Plant Soil Sci, 35(21), 325-335.
27. Jadon P, Mewada R, Gehlot Y, and Jamra S. (2023). GIS based spatial variability of available micro nutrients in soil of district Indore, Madhya Pradesh, India. Int. J. Plant Soil Sci. 35(21):196-203.
28. Jerosch, K. (2013). Geostatistical mapping and spatial variability of surficial sediment types on the Beaufort Shelf based on grain size data. Journal of Marine Systems, 127, 5-13.
29. Johnston, K., VerHoef, J. M., Krivoruchko, K., & Lucas, N. (2001). Using ArcGIS geostatistical analyst (Vol. 380). Redlands: Esri.
30. Kaur, H., Kaur, A., Singh, B., & Bhatt, R. (2020). Application of geospatial technology in assessment of spatial variability in soil properties: a review. Curr J ApplSciTechnol, 39(39), 57-71.
31. Khallouf, A. AlHinawi, S. AlMesber, W.Shamsham, S. and Idries, Y.(2020) DigitalMapping of Soil Properties in the Western-Facing Slope of Jabal Al-Arab at SuwaydaaGovernorate, Syria. Jordan Journal of Earthand Environmental Sciences. 11(3).193-201.
32. Khan, M. Z. Islam, M. R. Salam, A. B. A. and Ray, T. (2021). Spatial variability and geo-statistical analysis of soil properties in the diversified cropping regions of Bangladesh using geographic information system techniques. Applied and Environmental Soil Science. 1-19.
33. Kingsley, J. Lawani, S. O. Esther, A. O. Ndiye, K. M. Sunday, O. J. and Penížek, V. (2019). Predictive mapping of soil properties for precision agriculture using geographic information system (GIS) basedgeostatisticsmodels. Modern Applied Science. 13(10):60-77.
34. Krivoruchko, K., &Gribov, A. (2019). Evaluation of empirical Bayesian kriging. Spatial Statistics, 32, 100368.
35. Kumar, M., Kar, A., Raina, P., Singh, S. K., Moharana, P. C., &Chauhan, J. S. (2021). Assessment and mapping of available soil nutrients using GIS for nutrient management in hot arid regions of North-Western India. Journal of the Indian Society of Soil Science, 69(2), 119-132.
36. Li, J., & Heap, A. D. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. Ecological Informatics, 6(3-4), 228-241.
37. Mandal, S., &Ghosh, G. K. (2021).Depth wise distribution of macronutrients (N, P, K, S) and their correlation with soil properties in selected soil profiles of Birbhum district of West Bengal, India. International Journal of Chemical Studies, 9(1), 2801-2809.
38. Maqbool, M., Rehman, N. Z., Rasool, R., Akhtar, F., (2017). Available macronutrient status of soil under different land use systems of district Ganderbal, Jammu and Kashmir, India. Journal of the Indian Society of Soil Science, 65(3), 256–263.
39. Moral FJ (2010) Comparison of di□erent geo-statistical approaches to map climate variables: Application to precipitation. Int J Climatol 30:620–631
40. Nickel, S., Hertel, A., Pesch, R., Schröder, W., Steinnes, E., &Uggerud, H. T. (2014). Modelling and mapping spatio-temporal trends of heavy metal accumulation in moss and natural surface soil monitored 1990–2010 throughout Norway by multivariate generalized linear models and geo-statistics. Atmospheric Environment, 99, 85-93.
41. Oliver, M. A., & Webster, R. (1990). Kriging: a method of interpolation for geographical information systems. International Journal of Geographical Information System, 4(3), 313-332.
42. Olsen S.R., 1954. Estimation of available phosphorus in soils by extraction with sodium bicarbonate No. 939.US Department of Agriculture.

43. Openshaw, S. and Clarke, G. (2019). Developing spatial analysis functions relevant to GIS environments. In Spatial analytical perspectives on GIS, Routledge, London.
44. Pellicone, G., Caloiero, T., Modica, G., &Guagliardi, I. (2018). Application of several spatial interpolation techniques to monthly rainfall data in the Calabria region (southern Italy). International Journal of Climatology, 38(9), 3651-3666.
45. Qi, Z., Gao, X., Qi, Y., & Li, J. (2020). Spatial distribution of heavy metal contamination in mollisol dairy farm. Environmental Pollution, 263, 114621.
46. Qiao, P., Lei, M., Yang, S., Yang, J., Guo, G., & Zhou, X. (2018). Comparing ordinary kriging and inverse distance weighting for soil as pollution in Beijing. Environmental Science and Pollution Research, 25, 15597-15608.
47. Ramzan, S., Wani, M. A., and Bhat, M. A. (2017).Assessment of spatial variability of soil fertility parameters using geospatial techniques in temperate Himalayas. International journal of Geosciences, 8(10), 1251-1263.
48. Robinson, T. P., &Metternicht, G. (2006). Testing the performance of spatial interpolation techniques for mapping soil properties.Computers and electronics in agriculture, 50(2), 97-108.
49. Saha, A., Gupta, B. S., Patidar, S., &Martínez-Villegas, N. (2022). Spatial distribution based on optimal interpolation techniques and assessment of contamination risk for toxic metals in the surface soil. Journal of South American Earth Sciences, 115, 103763.
50. Saito, K., Vandamme, E., Johnson, J. M., Tanaka, A., Senthilkumar, K., Dieng, I., Wopereis, M.C., (2019). Yield-limiting macronutrients for rice in sub-Saharan Africa.Geoderma, 338, 546–554.
51. Schabenberger, O., &Gotway, C. A. (2017). Statistical methods for spatial data analysis.Chapman and Hall/CRC.
52. Sharma, S.K., Chouhan, N., Ae, N., Sinha, D.K., Dwivedi, B.S., (2021). Rhizosphere effect on phosphorus availability to soybean in Vertisols. Soil Science and Plant Nutrition 67(6), 653–661.
53. Tziachris, P., Metaxa, E., Papadopoulos, F., &Papadopoulou, M. (2017). Spatial modelling and prediction assessment of soil iron using kriging interpolation with pH as auxiliary information. ISPRS International Journal of Geo-Information, 6(9), 283.
54. Valera, Á. R. V. and Rodríguez, E. R. A. (2023).Spatial analysis of soil fertility using geostatistical techniques and artificial neural networks.Qeios, 1-18.
55. Verma, P., Deb, P., Nema, A.K., Sheetpal, M., Patle, T., Tiwari, R., (2022). Distribution and characterization of Nitrogen, Phosphorus and Potassium in different depth of soil profile in ITM research farm at Gwalior Madhya Pradesh. Biological Forum– An International Journal, 14(3), 1259–1264.
56. Wang, F. and Liu, L. (2023). Computational Methods and GIS Applications in Social Science. CRC Press. Boca Raton.
57. Wen, L., Zhang, L., Bai, J., Wang, Y., Wei, Z., & Liu, H. (2022). Optimizing spatial interpolation method and sampling number for predicting cadmium distribution in the largest shallow lake of North China. Chemosphere, 309, 136789.
58. Wu, C. Y., Mossa, J., Mao, L., &Almulla, M. (2019). Comparison of different spatial interpolation methods for historical hydrographic data of the lowermost Mississippi River. Annals of GIS, 25(2), 133-151.
59. Xie, Y., Chen, T. B., Lei, M., Yang, J., Guo, Q. J., Song, B., & Zhou, X. Y. (2011). Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: Accuracy and uncertainty analysis. Chemosphere, 82(3), 468-476.
60. Yadav, T.C., Singh, Y.P., Yadav, S.S., Singh, A., Patle, T., (2023). Spatial variability in soil properties, delineation site-specific management division based on soil fertility using fuzzy clustering in Gwalior, Madhya Pradesh, India. International Journal of Plant & Soil Science, 35(6), 49–78.
61. Yao, L., Huo, Z., Feng, S., Mao, X., Kang, S., Chen, J. &Steenhuis, T. S. (2014). Evaluation of spatial interpolation methods for groundwater level in an arid inland oasis, northwest China. Environmental earth sciences, 71, 1911-1924.

62. Zakeri, F. and Mariethoz, G. (2021). A review ofgeostatistical simulation models applied to satellite remote sensing: Methods and applications. Remote Sensing of Environment. 259. 112381.
63. Zarco-Perello, S., &Simões, N. (2017). Ordinary krigingvs inverse distance weighting: spatial interpolation of the sessile community of Madagascar reef, Gulf of Mexico. PeerJ, 5, e4078.