



REVIEW ARTICLE

From soil pollution mapping to sustainable crop production: Advances in geostatistical and hybrid modelling techniques

Yogaswathy Damotharan¹, Muthumanickam D^{2*}, Pazhanivelan S³, Kannan P⁴, Prabu P C⁵ & Gnanachitra M⁶

¹Department of Soil Science and Agricultural Chemistry, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

²Department of Remote Sensing and GIS, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

³Directorate of Crop Management, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

⁴Department of Soil Science and Agricultural Chemistry, Agricultural College and Research Institute, Tamil Nadu Agricultural University, Chettinad 630 102, Tamil Nadu, India

⁵Department of Environmental Sciences, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

⁶Department of Agricultural Microbiology, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

*Correspondence email - muthumanickam.d@tnau.ac.in

Received: 04 June 2026; Accepted: 06 July 2026; Available online: Version 1.0: 01 August 2026

Cite this article: Yogaswathy D, Muthumanickam D, Pazhanivelan S, Kannan P, Prabu PC, Gnanachitra M. From soil pollution mapping to sustainable crop production: Advances in geostatistical and hybrid modelling techniques. *Plant Science Today*. 2026; 13(sp5): 1-19. <https://doi.org/10.14719/pst.15865>

Abstract

Soil is a critical component of terrestrial ecosystems and is increasingly exposed to a wide range of pollutants, particularly heavy metals and potentially toxic elements (PTEs), which are intensified by rapid industrialisation, urbanisation and agricultural intensification. Heavy metal pollution is a significant concern in many regions worldwide that requires systematic and regular monitoring of soils to track contamination levels and prevent associated health risks. Traditional soil pollution monitoring methods, while accurate at localised scales, are limited in their ability to capture the spatial complexity and heterogeneity of heavy metal distribution. Thus, recent advances in geostatistical and modelling approaches have enhanced the accuracy, efficiency and scalability of soil pollution assessment. Geostatistical methods such as kriging, cokriging and empirical Bayesian kriging are known for their ability to generate high-resolution spatial maps and quantify uncertainty in unsampled areas. The integration of geostatistics with remote sensing, machine learning and proximal sensing technologies has significantly improved predictive performance and enabled large-scale monitoring. Advanced hybrid models, such as random forest-ordinary kriging (RF-OK) combinations, artificial neural networks (ANN) and empirical Bayesian kriging regression prediction (EBKRP), have demonstrated superior performance in spatial interpolation, with consistent improvements of 20–40 % in accuracy metrics compared to traditional methods. Despite notable advancements, challenges remain regarding computational complexity, data scarcity, implementation in developing regions and the integration of uncertainty quantification. This review synthesises state-of-the-art geostatistical advancements, hybrid modelling approaches and emerging technologies to provide a comprehensive framework for soil pollution monitoring and risk assessment, aligning with global sustainability and environmental protection goals.

Keywords: geostatistics; heavy metals; kriging; machine learning; soil pollution; spatial interpolation

Introduction

Rapid industrialisation and urbanisation have significantly increased the number of contaminated sites worldwide, polluting over 20 million hectares of soil with toxic heavy metals (1). The most prevalent elements include zinc (Zn), copper (Cu), chromium (Cr), lead (Pb), nickel (Ni) and cadmium (Cd). These non-biodegradable contaminants persist in the environment, accumulate in food chains and cause a wide range of diseases and health complications (2). Consequently, this bioaccumulation hinders the progress toward several United Nations sustainable development goals (SDGs), such as clean water and sanitation (SDG 6), good health and well-being (SDG 3) and sustainable land use and ecosystem protection (SDG 15) (3, 4).

Therefore, effective monitoring and management of soil contamination are critical. While conventional analytical techniques such as atomic absorption spectroscopy provide high localised precision and accuracy, they often fail to capture the

critical spatial variability necessary for regional risk management (5, 6). Consequently, reliance on point-based measurements alone often leads to an incomplete representation of contamination patterns and uncertainty in exposure and risk assessment (7).

To overcome these limitations, recent research has increasingly adopted spatially explicit assessment frameworks, enabling contamination mapping at regional and global scales. Current research has expanded towards large-scale regional assessments, with investigations across the Middle East (8), Asia (9) and Europe (10), employing diverse methods including proximal sensing, remote sensing, machine learning algorithms and geostatistical interpolations (11, 12). These approaches facilitate the identification of pollution hotspots, estimation of contaminant migration patterns and evaluation of the effectiveness of remediation techniques (13).

In the past two decades, there has been a shift from exclusively laboratory-based assessments to integrated spatial

assessment frameworks (14). Geostatistical approaches such as kriging and variogram analysis remain foundational for mapping spatial variability of contaminants, but their integration with machine learning (ML) (15), receptor models, fuzzy logic and geographic information system (GIS) (16) has led to more accurate, source-specific and actionable risk assessments (17). Hybrid models routinely combine geostatistics with ML (e.g., random forest, support vector machines), receptor models [e.g., positive matrix factorisation (PMF), absolute principal component scores-multiple linear regression (APCS-MLR)] and probabilistic simulations (e.g., Monte Carlo) to improve spatial prediction (18), source apportionment (19) and risk quantification for heavy metals, PAHs and other pollutants in diverse environments, from agricultural soils to urban and mining-impacted areas (20). These advancements enable not only more precise mapping of contamination but also the identification of pollution sources and quantification of their contributions to ecological and human health risks (21). These hybrid approaches enable the processing of large, multidimensional datasets and facilitate multiscale contamination analysis (22). The integration of geostatistics with GIS, as well as the use of hybrid evaluation models, further enhances the ability to delineate risk zones and inform targeted remediation strategies (23, 24).

Despite the extensive application of geostatistical and hybrid models for heavy metals, current reviews rarely critically compare the gains of hybrid models across contexts, integrate uncertainty into decision frameworks and extend these tools to co-occurring non-metal contaminants such as polycyclic aromatic hydrocarbons (PAHs), pesticides, per- and polyfluoroalkyl substances (PFAs), pharmaceuticals and other emerging contaminants. The existing literature predominantly focuses on either traditional sampling or laboratory techniques (25) or isolated applications of geostatistical methods (26), with fewer studies providing a holistic view of how these approaches can be combined to enhance pollution monitoring accuracy (27). Addressing these gaps, this review provides a critical and integrative synthesis of recent advances in geostatistical and hybrid modelling approaches for soil pollution monitoring and risk assessment. By evaluating methodological evolution, comparative performance, uncertainty handling and practical applicability, this review aims to guide researchers, policymakers and practitioners toward data-driven, scalable and decision-relevant frameworks for sustainable soil management and precision remediation.

Data collection and processing: Methodology

Literature search strategy

A comprehensive, systematic literature review was conducted to examine the evolution and current state of geostatistical methods in soil pollution monitoring. The review focused on peer-reviewed studies published between 2000–25, with particular emphasis on the integration of ML and hybrid modelling approaches. Multiple scientific databases were systematically searched, including Scopus, Web of Science and Google Scholar, using structured keyword combinations. The main search term used was: (“geostatistics” OR “geostatistical methods”) AND (“soil pollution” OR “heavy metal contamination”) AND (“potentially toxic elements”) AND (“machine learning” OR “empirical Bayesian kriging” OR “inverse distance weighting”) AND (“kriging” OR “empirical Bayesian kriging” OR “inverse distance weighting”)

AND (“artificial intelligence”) The following second search string was used: (“plant uptake” OR “phytoaccumulation” OR “crop production” OR “precision agriculture”) AND (“contaminated soil” OR “risk mapping” OR “site-specific management” OR “soil remediation”). Reference lists from relevant review papers and from articles that were cited several times were also scanned to identify further eligible studies.

Study selection and eligibility criteria

Studies were selected if they: (i) included in geostatistical research, spatial interpolation studies or hybrid geostatistical-ML studies for soil pollution assessment; (ii) included studies to assess the spatial distribution of heavy metals, potentially toxic elements (PTEs) or associated contaminants in soils; (iii) reported model validation metrics such as root mean square error (RMSE), mean absolute error (MAE), R^2 , cross validation statistics or equivalent measures of model performance; (iv) included soils contaminated by agricultural, industrial, urban or mining activities; and (v) included studies on plant uptake, phytoaccumulation, precision agriculture, risk mapping, crop management or remediation applications, if present.

The following studies were excluded: (i) those without an analysis of a spatial component; (ii) those which were only laboratory based studies but did not include mapping applications at the field scale; (iii) those that were conference abstracts, editorials, notes or non-peer-reviewed publications; (iv) those focusing solely on water, sediment or atmospheric pollution and did not directly address soil contamination; and (v) those that did not include enough methodology information to be evaluated.

PRISMA screening process

The database search initially identified 642 records. Duplicate records ($n = 148$) and preliminary exclusions ($n = 24$) resulted in 470 records being retained for title and abstract screening. Of these, 318 records were excluded because they did not directly relate to geostatistical soil pollution assessment. The remaining 152 articles were reviewed for full text. After thorough assessment, 66 studies were rejected for being too poorly described, lacking spatial analysis or otherwise being inconsequential to the goals of the review. Finally, 102 studies were chosen for detailed synthesis and critical evaluation. The complete study selection procedure is presented in the preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram (Fig. 1) (28).

Bibliometric and thematic analysis

To understand research trends, knowledge networks and methodological developments a bibliometric analysis was performed using VOSviewer (Fig. 2), which revealed four key research clusters: (i) methodological innovations focusing on geostatistical techniques and spatial variability; environmental applications; (ii) environmental applications emphasising heavy metal contamination and monitoring; (iii) health and risk assessment connecting contamination to human health implications; and (iv) technological integration highlighting convergence of traditional methods with modern analytical approaches. Due to significant variation in study objectives, datasets and modelling techniques, no formal meta-analysis was carried out.

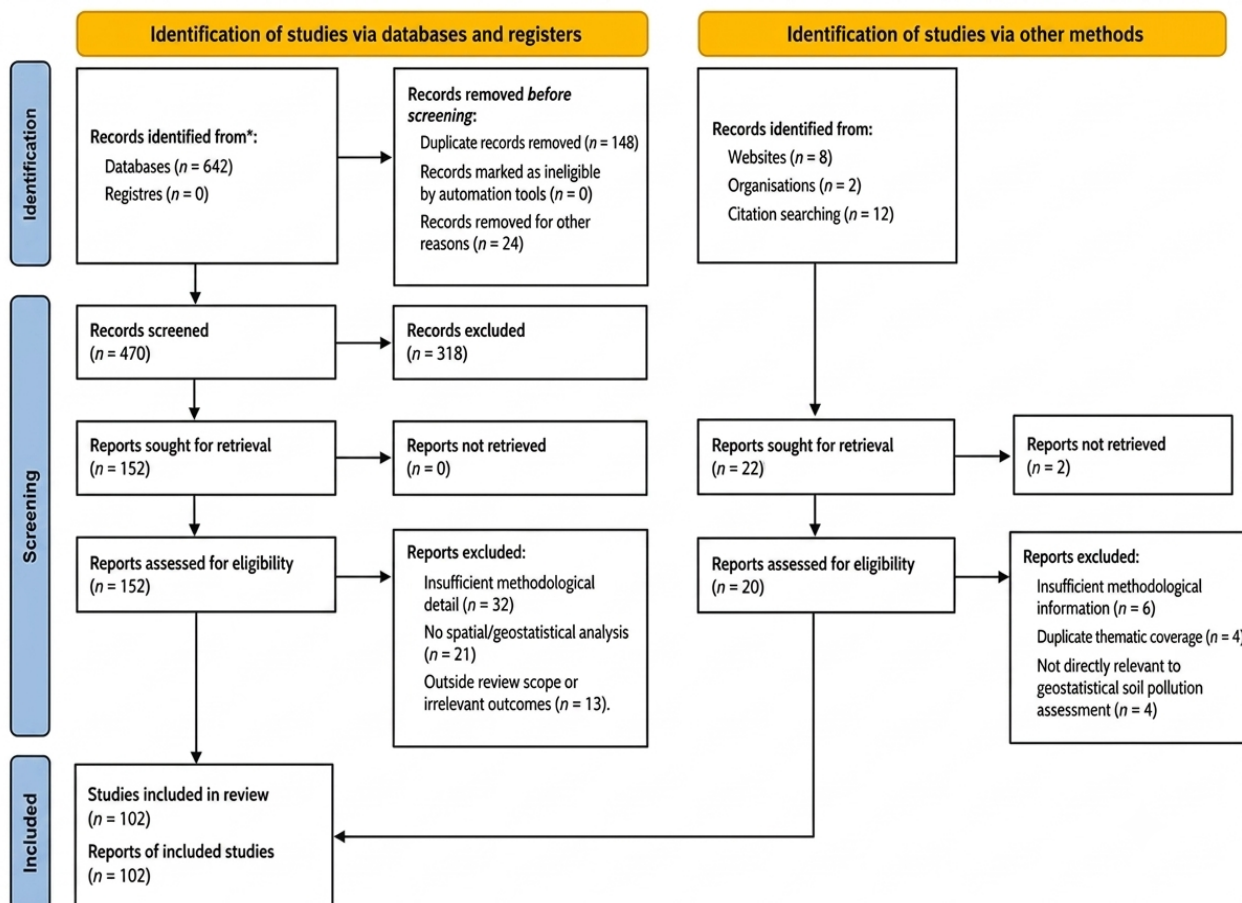


Fig. 1. PRISMA flowchart of the systematic literature search and study selection process for soil pollution assessment.

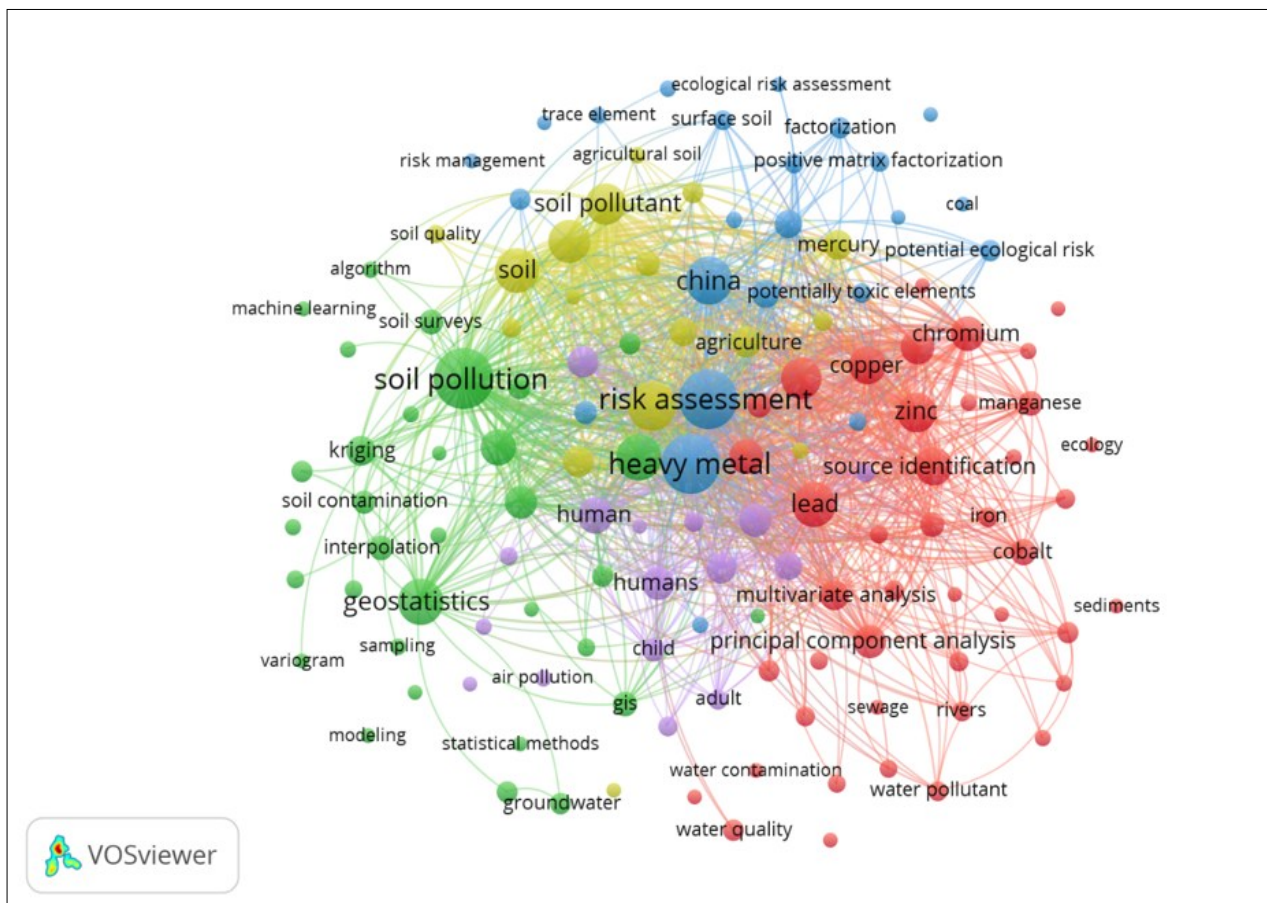


Fig. 2. Bibliometric network visualisation showing major research clusters in geostatistical and hybrid modelling approaches for soil pollution monitoring.

Sources of heavy metal contamination in soil

Heavy metals are hazardous substances characterised by high density ($\geq 5 \text{ g cm}^{-3}$) and atomic weight, with some being carcinogenic. These include both essential (e.g., copper (Cu), Zn, iron (Fe), manganese (Mn), cobalt (Co) and molybdenum (Mo)) and non-essential elements (e.g., Cd, Pb, mercury (Hg) and arsenic (As)). The essential metals have significant physiological roles in plants, animals and humans but are toxic when their concentrations exceed biological requirements. In contrast, non-essential heavy metals have no known biological function and can exert toxic effects even at relatively low concentrations owing to their tendency to accumulate in living organisms and disrupt cellular metabolism. In addition, the term “heavy metals” is increasingly complemented by PTEs in modern environmental science to more accurately reflect the impact and chemical diversity of these soil pollutants, including metalloids such as arsenic (As), which pose significant toxicity despite not being true metals (29). Their occurrence in soil systems and the environment is attributed to both natural processes and anthropogenic activities (Fig. 3) (30).

Natural sources

Geogenic (natural) sources arise from parent rock weathering, volcanic emissions, atmospheric deposition and groundwater-soil interactions. These processes establish the regional base levels of concentration in soil, which vary depending on the geological and climatic conditions (31, 32). The geogenic metal concentrations serve as a critical baseline for evaluating anthropogenic contamination, even if they are present below the threshold of ecological concern (15).

Anthropogenic activities

Beyond natural processes, heavy metal concentrations in agricultural and urban soils are elevated by human activities, especially in industrial and urban regions. The anthropogenic sources are diverse in nature and origin, including urban and industrial activities, mining operations and agricultural practices (33, 34).

Urban and industrial sources: Rapid urbanisation in metropolitan areas has altered the physical and chemical properties of soil, negatively influencing infiltration, organic matter content and contaminant pathways. Owing to population growth, industrial expansion and increasing transportation, urban centres have become hotspots for ecological risks at different scales. Heavy metals are introduced into urban soils via multiple sources, such as vehicle emissions, industrial effluents, municipal waste, improper waste disposal and excessive use of agrochemicals (35). Urban soils adjacent to major traffic routes often have two to five times higher levels of Cu, Zn and Pb than the less-trafficked adjacent areas due to brake and tyre wear, fuel combustion and road dust suspension. Besides heavy metals, urban runoff has been found to have heavy loads of emerging contaminants such as persistent organic pollutants (POPs), PAHs, personal care products, pharmaceuticals, hydrocarbons and human-origin microorganisms (36).

Industrialisation is a primary driver of heavy metal contamination in the environment. Industrial and various manufacturing activities, including electroplating, tanning, petrochemicals, dyes, pigments, textiles and ceramics, release pollutants into the atmosphere, water and soil, which accumulate over time (8). In addition, untreated industrial wastewater introduces heavy metals into the soil when used for irrigation (37). The crops irrigated with this wastewater have recorded the bioaccumulation of potentially toxic concentrations of metals like Cu, Fe, Zn and Pb. Besides metals, emerging contaminants are deposited into the environment by industrial activities, emissions, effluent and wastewater disposal (38).

Mining operations: Mining and ore processing activities significantly contribute to soil pollution by releasing heavy metals such as As, Cd, Cr, Pb, Hg, Zn, Cu and Ni (39). These operations cause the release and deposition of heavy metals in adjacent soils, which remain in the soil even after the closure of the mine area (40, 41). Investigations conducted in mining areas worldwide have quantified high levels of Hg, Mn, Cd, Cr, As and Pb in arable and non-arable lands. Moreover, agricultural soils around mines are contaminated with heavy metal(loid)s, which adversely affect

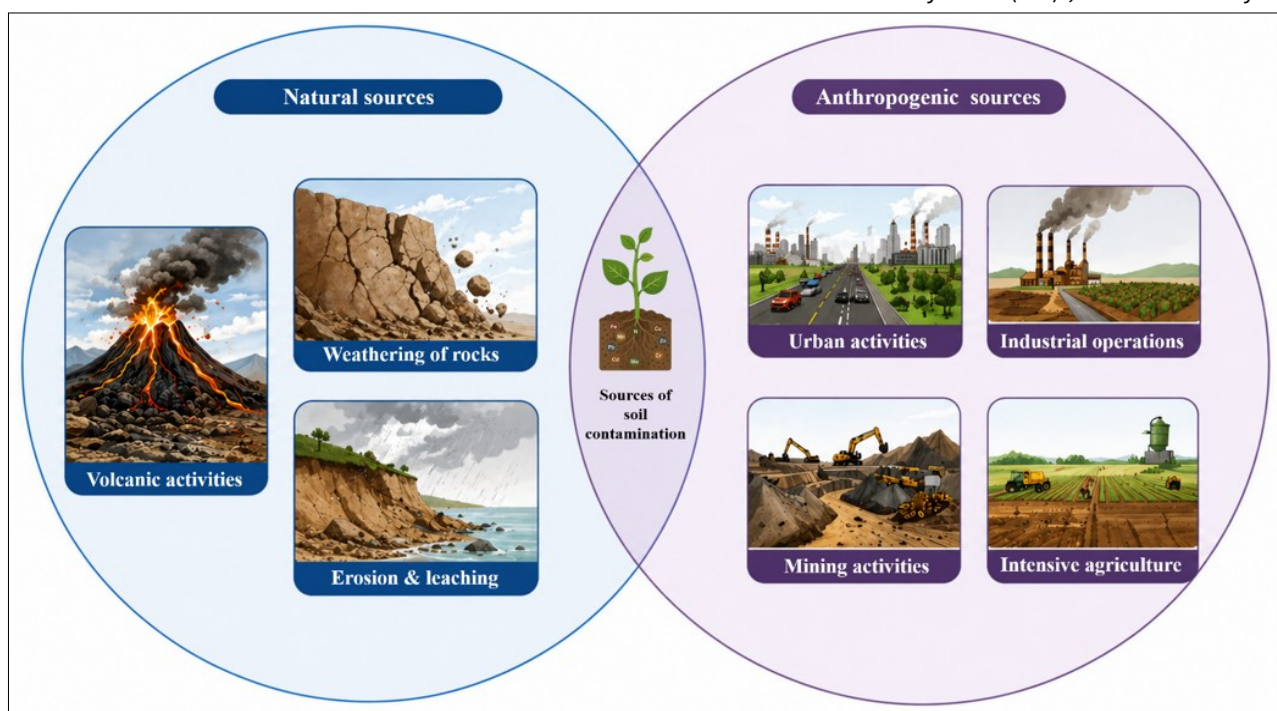


Fig. 3. Natural and anthropogenic sources of heavy metal contamination in soils.

ecosystems and human health (42).

Agricultural sources: Agricultural activities are another major source of heavy metal contamination in soils. The extensive use of fertilisers and chemical pesticides introduce heavy metals into the soil. Although phosphatic fertilisers are vital for crop production, they can introduce Cd and other PTEs into agricultural soils (43). The use of sewage sludge as a soil amendment, although beneficial for organic matter, carries the risk of introducing heavy metals into soils and crops. Many pesticides, such as copper-based fungicides (Bordeaux mixture, copper oxychloride, etc.) and arsenical pesticides (historically used in orchards and cotton cultivation), contain metal-based active ingredients that persist in the soil and generate toxic residual substances (44). These compounds are increasingly being considered emerging contaminants because of their high toxicity, bioaccumulation and biomagnification in soil (45).

Geostatistical principles and interpolation methods

Theoretical basis: Regionalised variables and stationarity

Geostatistics is based on the theory of regionalised variables, which emphasises that the value of a spatial variable is influenced by its location and its neighbouring values. By exploring spatial correlation to produce minimum variance and using regional variables as a foundation, geostatistics has proven to be one of the most effective tools for assessing the spatial pattern and variation of soil characteristics (Table 1) (26). Geostatistics emerged in the mid-20th century and was initially developed to address challenges in predicting ore grades during mining. This field originated from Danie G. Krige's pioneering work on spatial correlation techniques, which laid the foundation for modern geostatistics. Over time, geostatistics evolved beyond mining applications, finding relevance in petroleum geology, environmental science, contamination studies and public health (46).

A regionalised variable, denoted as $z(x)$, is a spatially dependent random variable that assumes different values of z at each spatial location x , varying continuously across a geographic domain. The ability to model and quantify spatial dependence is fundamental to effective pollution monitoring, as it enables more accurate predictions of contamination levels across unmonitored areas (26).

Spatial autocorrelation and the first law of geography

Spatial autocorrelation refers to the degree to which geographically proximate areas exhibit similar characteristics. It represents the inherent spatial dependence of environmental variables and violates the classical assumption of independence among

variables. This concept is formalised by Tobler's First Law of Geography, which states that "everything is related to everything else, but near things are more related than distant things" (47). Spatial autocorrelation is a ubiquitous phenomenon in environmental systems and can be observed at various spatial scales, from microscale soil heterogeneity to regional gradients. In geostatistical analysis, it is vital to factor in spatial autocorrelation to ensure accurate prediction and interpolation. It serves as the theoretical basis for interpolation methods such as kriging and deterministic methods, in which the predictions conform to the spatial structure of the observed data (48, 49).

Semivariogram theory and structure analysis

The semivariogram is a foundational tool for quantifying spatial dependence and the computation of semivariance is the first step in geostatistical analysis. It quantifies spatial variability by computing the average squared difference between sample pairs as the distance between them increases. Mathematically, the semivariance is described as the mean square variability between two neighbouring points of distance h and expressed as (50–52):

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [z(x_i) - z(x_i + h)]^2 \quad (\text{Eqn. 1})$$

Here, h is the separation distance (lag) between two sample points, $n(h)$ is the number of pairs of differences at distance h and $z(x_i)$ and $z(x_i + h)$ are observed values at corresponding locations.

Semivariogram modelling involves fitting a theoretical model to experimental data, commonly spherical, exponential, Gaussian or linear models. The objective was to select the best-fitting model to make reliable predictions (53). The standard variogram model is characterised by three basic parameters: (i) range (a), (ii) nugget (C_0) and (iii) sill ($C_0 + C$) (Fig. 4). The range of impact determines the distance beyond which spatial correlation becomes negligible. The nugget specifies the internal credibility of data, comprising measurement errors and microscale variation. The sill is the value of the variances of all measurements, indicating the total semivariance (54). The degree of spatial independence is determined by the ratio of nugget to sill, which indicates the relative contribution of the random soil variable to the total spatial variability. A higher ratio suggests weaker spatial dependence and vice versa, while the slope reaches zero when the semivariogram is constant, that is, $\gamma(h) = C$, indicating a pure nugget effect and absence of spatial correlation. Knowledge of these parameters is essential for selecting appropriate kriging

Table 1. Summary of the key differences between classical statistics and geostatistics in soil contamination mapping

Aspect	Classical statistics	Geostatistics
Data assumption	Random, independent and identically distributed samples	Spatially correlated samples (autocorrelation)
Objective	Describes global patterns (averages and trends)	Predict values at unsampled locations and produce site-specific predictions
Spatial information	Limited capacity to capture spatial variability in heterogeneous soil	Captures spatial variability and directional patterns of contamination
Typical methods used	Descriptive and inferential statistics (mean, variance, regression and hypothesis testing)	Spatial interpolation, variograms and kriging variants
Data requirements	Any (location-free)	Requires georeferenced (spatial) data
Handling of uncertainty	Limited capacity to capture spatial variability in heterogeneous soils	Maps of prediction error and uncertainty
Risk assessment	Checks if contamination exceeds safe limits, but without indicating the specific location	Creates risk maps that show the probability of contamination at specific sites
Output	Global summary statistics and generalised trends	Continuous spatial distribution maps of heavy metals with uncertainty estimates
Application	Overall contamination evaluation and regulatory reports	Distribution and hotspots of contamination, remediation planning and precision monitoring

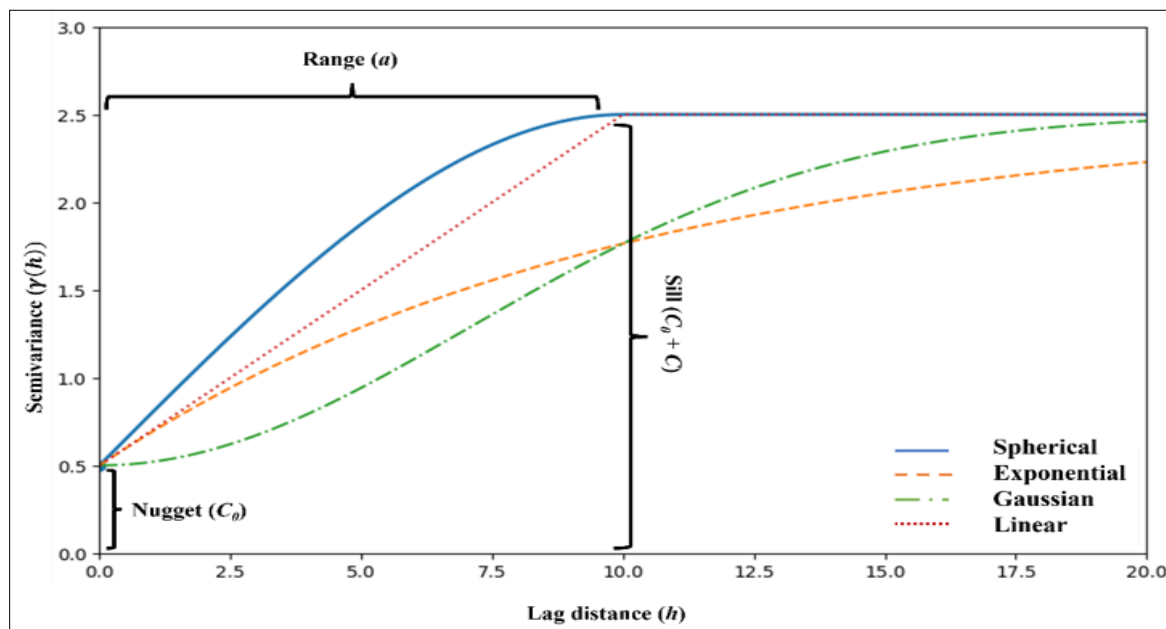


Fig. 4. Conceptual semivariogram showing nugget, sill and range parameters used in spatial analysis.

methods and interpreting spatial patterns of contamination (5, 53, 55).

Kriging: Principles and methodological variants

Basic kriging principles: Kriging is a geostatistical technique developed on the theory of regionalised variables, which yields the best linear unbiased estimators (BLUE). This method is a weighted linear estimator that relies on mathematical models to define spatial covariance. This technique uses a variogram-based model to account for spatial correlation within a generalised least-squares framework and it is a weighted linear estimator that predicts unknown values as a linear combination of observed data (56).

The key characteristics of kriging include the following: the estimates are computed as linear combinations of the available data, the average of the prediction errors is zero and it minimises the variance of errors, ensuring the most accurate linear unbiased estimators (57). To apply kriging, an experimental variogram is calculated from sample data and fitted to a theoretical model, commonly linear, spherical, exponential or Gaussian. This model captures the spatial dependency in the data, enabling kriging to estimate unknown values and assess prediction uncertainty (42).

Variants of kriging: Kriging comprises a collection of interpolation methods based on a stochastic approach.

Ordinary kriging: This is the most common and standard type of kriging, in which the average value of the variable is assumed to be non-uniform across the study area. The ordinary kriging (OK) is expressed as a linear-weighted average of observation data in the neighbourhood of the unsampled location being estimated (58).

Simple kriging: In this type of kriging, the average of the data is assumed to be known and independent of the coordinates (49).

Cokriging: As OK is a univariate interpolation method, it cannot directly utilise auxiliary information from other variables (59). Co-kriging compensates for this deficiency, using covariate information between the primary and secondary variables to obtain more accurate interpolation results. This is particularly useful for mapping properties that are expensive to measure by leveraging their strong spatial correlation with more readily measured attributes. This approach incorporates auxiliary variables to enhance estimation. This method requires the

generation of variograms and cross-variograms to model the spatial relationships between variables (26, 50).

Universal kriging: In the case of the occurrence of a trend in data (e.g., soil contaminant spread influenced by proximity to industrial sources), universal kriging (UK) is often preferred as the best estimation method when spatial trends are present (49).

Regression kriging: This is a hybrid approach that combines kriging with regression analysis. This method uses an auxiliary variable significantly correlated with the variable of interest, enhancing the accuracy of spatial forecasting. The regression kriging (RK) has been reported to outperform several conventional interpolation methods, even when the auxiliary variables exhibit only weak correlations with the variable of interest (52).

Indicator kriging: Indicator kriging (IK) is a non-parametric geostatistical method used to estimate the probability of exceeding threshold value at a given location. Unlike traditional methods, it does not rely on assumptions about the distribution of data. Instead, it estimates the probability of exceeding a specified threshold at unsampled locations, making it particularly useful in soil pollution mapping (60).

Empirical Bayesian kriging: Empirical Bayesian kriging (EBK) differs from classical kriging methods by accounting for the error introduced by estimating the semivariogram model using many semivariogram models rather than one. The EBK provides a more robust framework for uncertainty quantification and is increasingly preferred for contamination assessments where model uncertainty is substantial (61, 62). The mathematical expressions, model hypotheses and common applications of each kriging variant are listed in Table 2.

Inverse distance weighting

Inverse distance weighting (IDW), also known as Shepard's method, is a widely used interpolation technique based on the principle that points closer together are more similar than those farther apart. It assumes that the influence of a data point on the estimated value decreases as its distance from the target location increases (42, 55).

Table 2. Summary of kriging methods in geostatistical interpolation

Method	Model assumption	Model	Estimator	Application	References
Ordinary kriging	Unknown but constant mean	$Z(s) = \mu + \varepsilon(s)$	$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i)$	Most common; no spatial trend	(58)
Simple kriging	Known mean (μ)	$Z(s) = \mu + \varepsilon(s)$	$\hat{Z}(s_0) = \mu + \sum_{i=1}^n \lambda_i (Z(s_i) - \mu)$	Rare; simulation studies; theoretical cases; known stationary mean	(37)
Universal kriging	Mean varies spatially (trend)	$Z(s) = m(s) + \varepsilon(s)$ Where $m(s) = \sum_{k=0}^k \beta_k f_k(s)$	$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i)$	Data shows spatial trend; no external covariates	(50)
Regression kriging	Trend from regression and kriged residuals	$Z(s) = m(s) + \varepsilon(s)$	$\hat{Z}(s_0) = \hat{m}(s_0) + \sum_{i=1}^n \lambda_i \varepsilon(s_i)$	Strong auxiliary predictors available (DEM, NDVI, etc.)	(52)
Co-kriging	Primary and secondary correlated variables	Multivariate spatial model	$\hat{Z}(s_0) = \left(\sum_{i=1}^n \lambda_i Z(s_i) \right) + \left(\sum_{j=1}^m \mu_j Y(s_j) \right)$	Correlated variables exist	(51)
Indicator kriging	Binary transformation at threshold z_k	$I(s; z_k) = \begin{cases} 1, & \text{if } Z(s) \leq z_k, \\ 0, & \text{otherwise} \end{cases} \quad k = 1, 2, \dots, m$	$\hat{I}(s_0) = \sum_{i=1}^n \lambda_i I(s_i)$	Probability/risk mapping; threshold exceedance; contamination studies	(60)

$z(s_0)$: predicted value at location s_0 ; $Z(s_i)$: observed value; λ_i : kriging weights; μ : mean; $m(s)$: trend function; $\varepsilon(s)$: residual; $Y(s_j)$: secondary variable; $I(s, z_k)$: indicator variable; DEM: digital elevation model; NDVI: normalised difference vegetation index.

In IDW, greater weight is assigned to data points nearer to the target location and these weights decrease as an inverse function of distance raised to the power p . The parameter p , a positive real number, controls the rate at which the influence of a point diminishes with distance. Higher values of p increase the influence of nearby points while reducing the impact of distant ones. It is expressed as:

$$\hat{z}(x_0) = \frac{\sum_{i=1}^n z(x_i) d_{ij}^{-p}}{\sum_{i=1}^n d_{ij}^{-p}} \quad (\text{Eqn. 2})$$

Here, $\hat{z}$ is the estimate value of x_0 , which is the estimation point and x_i are the data points within a chosen neighbourhood. The power parameter (p) are related to distance by d_{ij} , which is the distance between the estimation point and the data points (63, 64).

Inverse distance weighting depends on generating surfaces, taking into consideration the distance relationship between the measured values in geographically known sites. The measured values closest to the prediction site have more effect on the predicted values than those further away. Nevertheless, it's important to note that the computational process of IDW relies solely on geometric structural feature data and does not consider the influence of environmental factors on the spatial distribution of pollutants (65). Given the diversity of interpolation methods available for spatial prediction, Table 3 provides a comparative overview of their fundamental principles, advantages, disadvantages and applications in soil contamination mapping to aid in method selection under different environmental scenarios.

Evaluation methods for spatial interpolation accuracy

Cross-validation : A variety of techniques can be implemented to assess the correlations between the measured and predicted data values. For comparing the relation between predicted and measured values, cross-validation requires removing randomly measured data points and interpolating a value from the remaining dataset. The MAE, mean relative error (MRE) and the RMSE were calculated from the measured and predicted values to evaluate the accuracy of different interpolation techniques (14). The lower value of these indices indicates higher accuracy of spatial interpolation (42).

$$MAE = \frac{1}{n} \sum |Z(x_i) - Z^*(x_i)| \quad (\text{Eqn. 3})$$

$$MRE = \frac{1}{n} \sum_{i=1}^n |Z(x_i) - Z^*(x_i)/Z(x_i)| \quad (\text{Eqn. 4})$$

$$RMSE = \sqrt{\frac{1}{n} \sum [Z(x_i) - Z^*(x_i)]^2} \quad (\text{Eqn. 5})$$

where $Z(x_i)$ is the observed value at location x_i , $Z^*(x_i)$ is the interpolated value at location x_i and n is the sample size.

Uncertainty quantification: Modern methods of accuracy assessment go beyond providing point estimates and explicitly quantify and propagate uncertainty into environmental decision-making. Conditional simulation, probabilistic kriging and empirical Bayesian kriging are geostatistical methods that provide several realisations of contaminant distributions and enable the prediction uncertainty to be included in ecological and human health risk assessments instead of only a single deterministic estimate (66). Threshold exceedance probability maps are created to determine areas that are likely to exceed regulatory limits and Monte Carlo simulation can propagate uncertainty in the contaminant concentration and exposure parameters to provide a probabilistic risk estimate (67). Furthermore, adaptive sampling and value-of-information (VoI) strategies utilise spatial uncertainty maps to determine the optimal sampling locations, thereby minimising monitoring expenses while enhancing the reliability of predictions. The combination of uncertainty-aware geostatistical modelling and risk-based decision frameworks allows for a more transparent prioritisation of monitoring, remediation and land-use management, which can assist in more reliable and cost-effective environmental decision-making (68, 69).

Applications of geostatistics in soil pollution

Geostatistical methods have become essential tools for characterising, analysing and predicting soil pollution patterns, offering critical insights for environmental management and remediation strategies (Fig. 5). These techniques address several key aspects of assessing soil contamination through advanced spatial analysis approaches (23, 70).

Spatial mapping and hotspot identification

Geostatistical interpolation provides robust frameworks for spatially mapping contamination patterns and identifying potential pollution sources. The OK remains widely used for interpolating

Table 3. Comparative summary of geostatistical and deterministic interpolation methods for soil contamination mapping

Method	Principle	Strengths	Limitations	Recommended applications
Ordinary kriging	Uses spatial autocorrelation and an estimated variogram	Unbiased, quantifies prediction variance and strong for spatially structured data	Assumes stationarity; may smooth hotspots	Regional soil pollution mapping with moderate-to-high sampling density
Universal kriging	Accounts for spatial trend plus local variation	Handles broad directional gradients and source-driven trends	More assumptions and model complexity	Areas exhibiting large-scale contamination gradients
Co-kriging	Uses the primary variable plus correlated secondary variable	Improves prediction when auxiliary variables are informative	Requires cross-variogram and well-correlated covariates	Soil pollution mapping with auxiliary environmental variables
Regression kriging	Combines regression with kriging of residuals	Captures environmental drivers and spatial dependence simultaneously	Requires reliable predictor variables and model tuning	Digital soil mapping and contamination prediction using environmental covariates
Indicator kriging	Models the probability of exceeding a threshold	Useful for compliance and risk zoning	Loses information on exact concentration values	Regulatory compliance, probability mapping and hotspot delineation
Empirical Bayesian kriging	Repeats the semivariogram estimation to reflect uncertainty	More robust under uncertainty and sparse sampling	Computationally heavier than standard kriging	Sparse datasets and uncertainty-aware environmental assessments
Inverse distance weighting	Weights nearby points more strongly than distant points	Simple, transparent and easy to implement	No explicit spatial structure or uncertainty estimation	Preliminary mapping and exploratory spatial analysis

heavy metal concentrations across unsampled locations, though its smoothing effect may underestimate peak contamination levels (64). More sophisticated approaches like IK overcome this limitation by generating probability maps of threshold exceedance, proving particularly valuable for regulatory compliance assessments (60). Empirical studies consistently demonstrate the effectiveness of these methods, with industrial zones typically showing 2–5 times higher concentrations of Cu, Zn and Pb compared to adjacent areas (71). Table 4 provides a global overview of reported heavy metal concentrations, geostatistical methodologies and major pollution sources across various regions.

High-resolution mapping with sensing technologies

High-resolution mapping has been revolutionised through the integration of geostatistics with modern sensing technologies. Portable X-ray fluorescence (pXRF) sensing, when coupled with regression kriging, can achieve prediction accuracy improvements of 30–40 % compared to traditional laboratory-based methods (72). Hyperspectral remote sensing extends these capabilities to landscape scales, with ML-enhanced geostatistical models achieving sub-meter resolution in mining-impacted areas (15). These approaches are particularly valuable for initial site assessments, allowing rapid identification of contamination gradients and prioritising areas for detailed investigation.

Source apportionment

Source apportionment represents one of the most sophisticated applications of geostatistics in soil pollution studies. Multivariate techniques like factor kriging analysis (FKA) can discriminate between geogenic and anthropogenic sources with >85 % accuracy in well-characterised sites (73). Isotopic fingerprinting combined with spatial interpolation has proven particularly effective for lead source identification, distinguishing between historical gasoline emissions, industrial releases and natural background (74). Bayesian geostatistical models provide quantitative estimates of source contributions while accounting for measurement uncertainty, revealing that agricultural inputs may account for 15–32 % of heavy metal loads in intensively farmed regions (75).

Geostatistical simulation and uncertainty analysis

Geostatistical modelling can generate realistic and equiprobable 3D images of contaminant grades, enabling the identification of potential contamination sources and areas of uncertainty (53). The treatment of spatial heterogeneity and uncertainty has seen significant methodological advances. Conditional simulation approaches like sequential Gaussian simulation generate multiple equally probable realisations of contamination patterns, enabling probabilistic risk assessment (69). Modern implementations leverage high-performance computing to process millions of spatial data points, making these methods practical for large-scale assessments. Ensemble modelling techniques combine predictions from multiple geostatistical algorithms, reducing overall uncertainty by 20–30 % compared to single-model approaches (27).

Geostatistical risk mapping for site-specific crop management

Contamination of agricultural lands is a major risk factor affecting sustainable crop production and risk mapping is an important tool for integrating geostatistical techniques. This process is not limited to diagnosis of contamination but encompasses a complete workflow beginning with georeferenced sampling in the field, through hybrid spatial modelling, to site-specific agronomic management practices (Fig. 6). In this respect, the uptake and

phytoaccumulation of contaminants mapped using geostatistical methods are important factors. The major pathways of accumulation of the heavy metals and other PTEs in plant tissues involve root uptake, foliar deposition of atmospheric fallout and irrigation with contaminated water sources. The resulting spatial variability of plant-available fractions of these contaminants in agricultural soils is characteristic of the type of contamination that geostatistical methods are well suited to describe. Concentration maps generated from the kriging interpolation results provide continuous estimates of bioavailable metal pools in unsampled areas, which can be useful for input into plant uptake models. The validity of using interpolated soil data to assess crop quality has been demonstrated through studies of a significant positive correlation between predicted soil Cd concentrations from OK and the actual grain Cd concentrations in crops such as rice and wheat. In addition, IK maps have been successfully used to delineate areas where phytoaccumulation risk exceeds regulatory limits to allow for targeted interventions before cultivation (76, 77).

The most direct contribution of geostatistics to plant science lies in aiding the site-specific application of soil amendments and nutrients on polluted farmlands. A combination of geostatistical risk maps and variable rate technology (VRT) systems in moderate-risk areas will enable the optimised application of lime, subsequently lowering metal uptake by up to 40 % (78). Similarly, biochar prescriptions based on RF-kriging residual maps have shown promise for immobilising contaminants such as Pb within fields with varying contamination levels. The incorporation of the metal-nutrient interaction maps further supports the ability to tailor nutrient management strategies to reduce the risk of metal phyto-availability increasing because of interaction with different nutrients, such as phosphate (79). Geostatistical outputs inform a strategic placement of phytoremediation crops in high-risk areas, with a special focus on hyperaccumulator species to remove contaminants from the soil. These spatial models help identify optimal locations for deploying hyperaccumulators like *Thlaspi caerulescens*, *Pteris vittata* or *Sedum alfredii*, based on the severity and extent of contamination (80). Geostatistical monitoring is conducted continuously over several seasons to monitor the decrease in topsoil metal loads, thereby documenting the remediation efforts and when a restored area is safe for food crops. From contamination mapping to the production of safe crops, this cycle demonstrates the utility of geostatistical methods in sustainable agriculture by integrating advanced spatial modelling with crop quality protection, enhanced yield productivity and resource conservation.

Soil-to-plant transfer factors, bioavailability modelling and validation with crop contamination data

The conversion of geostatistically predicted soil contamination into crop-level risk depends on the soil-to-plant transfer factor (TF), defined as the ratio of metal concentration in plant tissue to that in the corresponding soil. Since TF differs among crop species and cultivars and is influenced by soil characteristics, it is not possible to reliably extrapolate dietary exposure from total soil metal concentrations. Therefore, the use of geostatistical approaches is increasingly combined with the use of crop-specific transfer factors and bioavailability models to estimate the spatial distribution of plant-available metals, instead of total metal concentrations (34).

Bioavailability modelling further enhances prediction by

Table 4. Global overview of heavy metal contamination in soils: pollution characteristics, geostatistical methods and health implications

Location	Contaminant elements	Concentration (mg kg ⁻¹)	Background/threshold value (mg kg ⁻¹)	Geostatistical method	Primary pollution source	References
Carsington Pasture, Derbyshire, UK	Pb, Zn, Ba, Cu	Pb: up to 8000 (mean 3512); Zn: up to 3100 (mean 975); Ba: 3702–3734; Cu: 84–100	Pb: 2000; Zn: 300; Cu: 130; Ba: 400	OK	Anthropogenic (past heavy metal mining); geogenic (sulphide mineralisation)	(68)
Chilmari Upazila, Northern Bangladesh	Pb, Cr, Mn, Cu, Zn, Co, Ni, As	Pb: 105–174; Cr: 68–280; Mn: 287–321; As: 25–33 (means)	Pb: 50; Cr: 100; Mn: 2,000; Cu: 100; Zn: 300; Co: 50; Ni: 100; As: 20	EBK	Industrial effluents, brickfields and agrochemicals	(22)
Douala, Cameroon	Cr, Cd	Cr: 2.18–168.70; Cd: 0.02–34.13	Control sites: Cr: 20; Cd: 10	OK; IDW	E-waste recycling	(64)
Dahetan subbasin, Xiangjiang River, China	Cd, Hg, As, Pb, Cr	Cd: 1.04; Hg: 0.16; As: 41.57; Pb: 126.25; Cr: 75.27 (means)	Cd: 0.07; Hg: 0.07; As: 13.41; Pb: 30; Cr: 68	IDW	Anthropogenic (industrial/mining); Geogenic (Hg, Cr)	(55)
Sheikhupura, Pakistan	Pb, Cr, Cd	Pb: 6.16–14.1; Cr: 17.55–26.0; Cd: 5.22–6.0 (means)	FAO/WHO: Cr: 30; Pb: 100; Cd: 300	Kriging	Untreated industrial effluents used for irrigation	(37)
Isfahan Industrial Zone, Iran	Cu, Pb, Zn, Cd, Ni, Co, Cr	Agricultural - Cu: 35.7; Pb: 34.6; Zn: 111.5; Ni: 66.2; Co: 14.7; Cd: 0.43; Cr: 85.9 (means)	Thresholds: Cu: 30.3; Pb: 37.5; Zn: 120; Ni: 83; Co: 17.9; Cd: 0.42; Cr: 105.8	NR	Anthropogenic (mining/industrial/pesticides); geogenic (Ni, Co, Cr)	(40)
Zhoucun County, Shandong Province, China	Cd, Cr, Cu, Hg, Ni, Pb, Zn, As, Co, Mn	As: 8.51; Cd: 0.16; Cr: 71.91; Hg: 0.07; Pb: 28.60; Zn: 73.27 (means)	As: 8.90; Cd: 0.075; Cr: 64.30; Hg: 0.016; Pb: 24.50; Zn: 60.90	EBK	Industrial emissions, traffic and fertilisers	(70)
Hamedan Province, Iran	As, Cd, Cr, Co, Ni, Pb, V, Zn, Cu	As: 10.61; Cd: 0.15; Cr: 76.95; Ni: 53.50; Pb: 10.06; Zn: 63.35 (means)	Critical: As: 10; Cr: 51; Co: 10; Ni: 50; Cu: 23; Zn: 60; Pb: 25; Cd: 3	OK; RBF (for Cu and Cd alone)	Geogenic (bedrocks); anthropogenic (fertilisers)	(44)
Cerrito Blanco, Matehuala, Mexico	Ag, Cd, Co, Cr, Li, Ni	Ag: 0.60–11.31; Cd: 0–2.18; Cr: 0.28–11.10; Ni: 0.24–8.37	Cd: 0.8; Cr: 100; Li: 25; Ni: 35; Ag: 4.4	OK; IDW; RBF; LP	Anthropogenic (metallurgical waste); geogenic (Co, Cr, Li, Ni)	(14)
Wuhan, Yangtze River Delta, China	Cd, Hg, As, Pb, Cr, Cu, Zn, Ni	Cd: 0.120; Hg: 0.051; As: 9.93; Pb: 27.8; Cr: 71.71; Zn: 70.77; Cu: 27.01 (means)	Background: Cd: 0.08; Hg: 0.02; As: 11.05; Pb: 24.32; Cr: 73.59; Zn: 67.99	OK	Zinc-plating, coal combustion and agriculture	(58)
Aksaray University Campus, Turkey	Pb, Cu, Ni, Cr, Zn	Pb: 10–39; Cu: 8–36; Ni: 24–214; Cr: 12–56; Zn: 98–220	NR	IDW	Geogenic (57.49 %); anthropogenic (industrial, traffic and agricultural)	(33)
North Nile Delta, Egypt	Ni, Co, Cr, B	Ni: 0–2720; Co: 0–2694; Cr: 0–2327; B: 0–3551	NR	IDW	Industrial drainage; agricultural wastewater	(63)
Northern Industrial Zone, Kazakhstan	Fe, Mn, Cr, Zn, Pb, Cu, Ni, Hg	Fe: 20406.9; Mn: 553.8; Pb: 46.9; Hg: detected	NR	OK; B-spline; ML models	Petrochemical and metallurgical industries; mercury ponds	(31)
Ledo Coal Mining Area, India	Cr, Cd, Ni, Pb	Cr: >100; Cd: >3; Pb: >50 (exceed FAO limits)	FAO permissible limits	OK; IK	Acid mine drainage	(32)
Abteymour Oilfield, Iran	Cr, Ni, V, Cu, Pb	Heavy contamination (values not specified)	NR	Co-kriging	Petroleum drilling and oil extraction	(34)
Policoro Coast, Italy	Sb, As, Cd, Hg, Pb, Zn	Not fully quantified	NR	IDW; spline interpolation	Agricultural runoff; pesticide use	(41)

Pb: lead; Zn: zinc; Ba: barium; Cu: copper; Cr: chromium; Mn: manganese; Co: cobalt; Ni: nickel; As: arsenic; Cd: cadmium; Hg: mercury; V: vanadium; Ag: silver; Li: lithium; B: boron; Fe: iron; Sb: antimony; NR: not reported; OK: ordinary kriging; IK: indicator kriging; EBK: empirical Bayesian kriging; IDW: inverse distance weighting; ML: machine learning; RBF: radial basis functions; LP: local polynomial; FAO: Food and Agriculture Organization of the United Nations; WHO: World Health Organization.

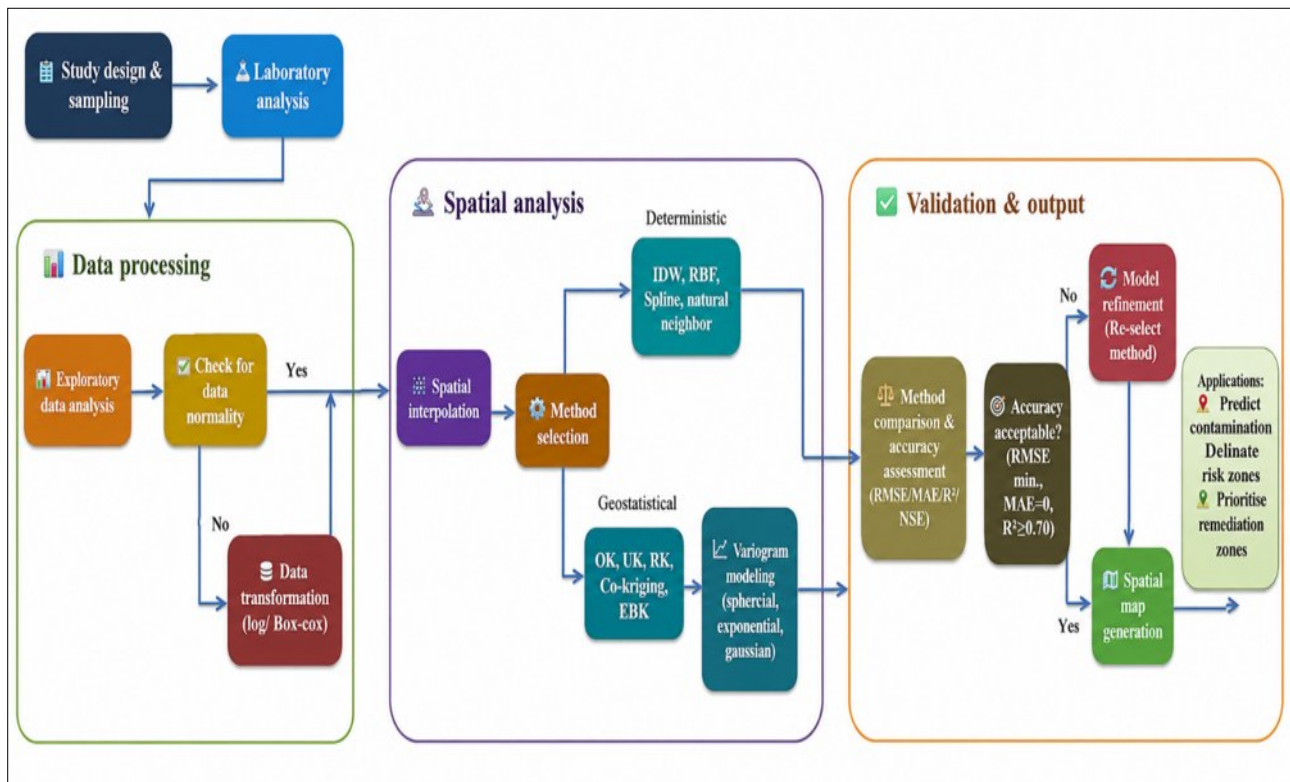


Fig. 5. Methodology for geostatistical spatial interpolation of soil heavy metal contamination. GIS: geographic information system; IDW: inverse distance weighting; RBF: radial basis function; OK: ordinary kriging; UK: universal kriging; RK: regression kriging; EBK: empirical Bayesian kriging; RMSE: root mean square error; MAE: mean absolute error; R²: coefficient of determination; NSE: Nash-Sutcliffe efficiency.

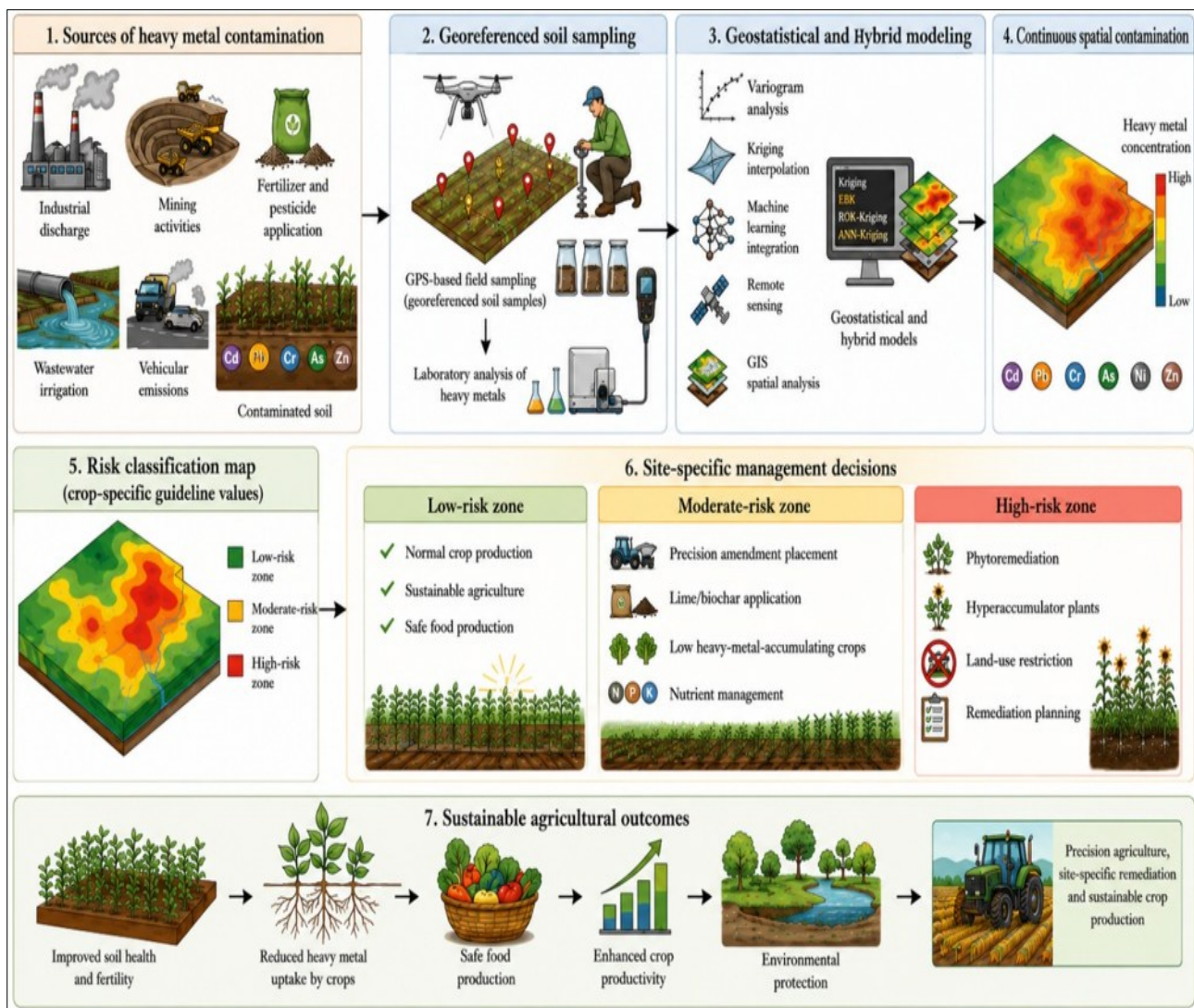


Fig. 6. Geostatistical risk mapping for contaminated agricultural lands and its application in sustainable crop production.

including the exchangeable fraction and other readily available soil fractions of metal, as well as key soil properties including pH, cation exchange capacity, organic matter content and clay content. These variables may be used as auxiliary covariates for regression kriging and ML-geostatistical hybrid models and produce more realistic predictions of contaminant uptake than traditional concentration-based interpolations (20).

Geostatistical predictions must be validated against independently collected data on crop contamination before use in food safety or precision agriculture. Bioavailable soil metal concentrations should therefore be correlated and regressed against measured metal concentrations of harvested crops. Studies on rice and wheat have reported moderate to strong correlations between predicted soil Cd and measured grain Cd, demonstrating the potential of geostatistical models for identifying crop contamination hotspots while highlighting the influence of crop genotype and soil chemistry on transfer efficiency (76, 77).

Advanced and emerging geostatistical methods

Soil contamination assessment has seen substantial methodological advances through the integration of ML with traditional geostatistical approaches, yielding hybrid modelling frameworks that markedly enhance predictive accuracy and spatial resolution. These approaches exploit the strength of geostatistics in modelling spatial dependence while leveraging ML algorithms to capture complex, nonlinear interactions among soil contaminants and environmental covariates (81). Algorithms such as random forest (RF), support vector machine (SVM), extreme gradient boosting (XGBoost) and artificial neural networks (ANN) have consistently demonstrated superior performance over conventional interpolation techniques across a wide range of soil prediction applications (26).

Each geostatistical method relies on specific assumptions and often excludes key environmental explanatory variables that influence the spatial prediction and distribution of PTEs in soils. Hence, no single method has shown better performance for soil PTE prediction and distribution studies (9); thus, combining different techniques can neutralise their weaknesses. A combination of different techniques, hereafter called hybrid methods, refers to combining two conceptually different approaches to model the spatial variation of soil properties (20, 24, 82).

Hybrid machine learning-geostatistical models

Hybrid ML-geostatistical models have shown remarkable improvements in contamination mapping. The RF-kriging approach exemplifies this synergy, wherein RF captures complex feature interactions and kriging accounts for spatial autocorrelation in model residuals (83). Studies have reported that RF-kriging outperformed standalone methods in abandoned mine sites, achieving 15–25 % improvements in prediction accuracy (23). Similarly, support vector machine kriging (SVMK) has proven effective in complex terrains by handling multivariate nonlinear relationships. More recently, deep learning integrations, particularly convolutional neural networks (CNNs) for spatial pattern recognition and recurrent neural networks (RNNs) for temporal dynamics, have expanded the capacity of geostatistical frameworks to handle high-dimensional covariate datasets (62, 84).

The hybrid models have also improved the predictive power of phytoaccumulation models by adding other variables, including soil pH, soil organic matter content, CEC and heavy metal concentrations. The models, such as random forest, have

resulted in good prediction accuracy for estimating concentrations of Pb and Cd in crops such as spinach and lettuce in contaminated fields (85). They can be used to map phytoaccumulation risk continuously on different scales, which can lead to valuable insights that cannot be obtained by using discrete sampling only. Furthermore, the results of a sequential Gaussian simulation can be employed to produce probabilistic estimates of contamination of crops, which can be incorporated into food safety risk assessment and to support regulators in establishing safe zones around industrial and mining sites (81).

Influence of sampling density and sampling strategy on geostatistical performance: The predictive performance of both conventional geostatistical and hybrid ML-geostatistical models is strongly influenced by sampling density and sampling strategy. The density of sampling is crucial to the successful estimation of the semivariogram; too few observations lead to more uncertainty in prediction and less accuracy in interpolation. Although increasing sample density generally improves model performance, accuracy gains diminish beyond a site-specific threshold, indicating that an optimal rather than maximum sampling density should be achieved within practical and economic constraints (85, 86).

Sampling strategy is also significant. Regular grid sampling provides uniform spatial coverage, whereas stratified and risk-based sampling designs are better suited to detect contamination gradients and localised hotspots that are likely to be found around industrial facilities, mining sites or waste sites. Adaptive sampling approaches, in which an initial survey guides subsequent sampling in areas of high uncertainty, further improve prediction accuracy while reducing unnecessary sampling effort. Thus, variogram analysis, nested sampling and cross-validation can be useful tools for optimising sampling intensity and sampling spacing.

Conventional kriging models typically need less dense and representative data sets than hybrid ML-geostatistical models, as the ML component requires adequate amounts of training data and environmental variables. It is therefore important to interpret the prediction accuracy of the hybrid models in the light of the sampling design used in the original studies.

Innovative interpolation techniques

The genetic algorithm-neural network (GANN) model incorporates multiple environmental covariates, including soil properties, atmospheric deposition patterns and agricultural practices, resulting in 20–30 % error reduction compared to OK (86). Similarly, the two-point ML method simultaneously utilises spatial neighbour information and high-dimensional covariates through a difference-based modelling strategy (87). These innovative interpolation techniques have emerged to address longstanding limitations of traditional geostatistical methods (Table 5).

Advanced geostatistical frameworks and uncertainty handling

Advanced geostatistical frameworks are increasingly overcoming traditional limitations in spatial analysis. The variance-octree kriging (VOK) method addresses variance non-stationarity through three-dimensional spatial deformation, achieving a 9 % improvement in interpolation accuracy via anisotropic-to-isotropic transformation (88). Bayesian maximum entropy (BME) approaches have also emerged as powerful tools for spatiotemporal modelling, enabling reliable forecasting of contaminant distributions across both space and time (42). These methods are particularly valuable for long-

term monitoring and remediation planning. Uncertainty quantification and model optimisation have become integral components of modern geostatistical practice. Cross-validation and hyperparameter tuning are now standard procedures, with evidence suggesting that proper optimisation can enhance model performance by 15–20 % (13, 89). Recent studies have further emphasised the need to address interpolation uncertainty, particularly smoothing effects that may obscure contamination hotspots critical for remediation decision-making (26, 66, 69).

Data integration and spatiotemporal modelling

The integration of multi-source data is expanding the boundaries of soil pollution assessment. The combination of proximal sensing data with geostatistical modelling has proven particularly effective, with RF-empirical Bayesian kriging regression prediction (EBKRP) models achieving RMSEs 30–40 % lower by incorporating soil organic carbon, nutrient status and topographic variables (24). Remote sensing integration has further enabled landscape-scale contamination mapping, with hyperspectral data providing unprecedented spatial resolution (15). Temporal dynamics modelling represents the next frontier in geostatistical analysis. Dynamic Bayesian networks and space-time kriging approaches now allow simulation of contamination evolution over time, offering critical insights for long-term environmental management, remediation effectiveness assessment and future contamination forecasting (46, 90).

Emerging applications and decision-support implications

Emerging applications include the integration of geostatistics with ML for dynamic and real-time pollution monitoring. Real-time sensor networks combined with streaming geostatistical analysis enable near-instantaneous detection of contamination events in industrial areas, with adaptive sampling strategies driven by uncertainty metrics to optimise monitoring costs while maintaining data quality (91). These innovations are shifting soil pollution management from reactive approaches to proactive, early-warning systems. Practical implementations of advanced geostatistical methods have demonstrated tangible benefits. In brownfield redevelopment projects, geostatistically optimised sampling designs have reduced investigation costs by 25–40 % while improving data quality (92). Remediation programs guided by geostatistical analysis have achieved 50–70 % greater efficiency in contaminant removal compared to uniform treatment strategies (3, 93, 94). Consequently, regulatory frameworks in several countries now explicitly recommend geostatistical approaches for site characterisation, reflecting their growing importance in evidence-based environmental decision-making. Overall, hybrid geostatistical-ML models frequently outperform traditional methods and often provide substantial improvements in predictive accuracy (24, 95). These methodological advances not only enhance scientific understanding of soil pollution dynamics but also directly inform more effective, cost-efficient and sustainable remediation strategies.

Challenges and future directions

Current challenges in implementation

Despite rapid methodological progress, several challenges continue to limit the widespread implementation of advanced geostatistical and hybrid modelling approaches. A primary constraint is computational complexity and data demand, as hybrid ML and geostatistical models often require substantial computational resources and dense, high-quality datasets. Many

advanced models still rely on dense, high-quality samples; performance under sparse or biased sampling is less studied. Many developing regions lack sufficient historical soil contamination data to train robust ML models, limiting model transferability. Although transfer learning, adapting models trained in data-rich regions to data-poor environments, shows promise, its reliability across diverse climatic, geological and land-use conditions remains insufficiently validated (37, 96).

Despite the potential for significant data reduction, the utility of transfer learning across regions is difficult to assess, as the spatial relationships between environmental predictor variables and contaminant concentrations may vary across climates, soils and pollution sources. Differences in rainfall, temperature, parent material, soil texture, organic matter and contaminant pathways can substantially reduce model performance when transferred to new environments (53, 82). Thus, transferable geostatistical frameworks need to account for domain adaptation, local recalibration and region-specific validation before operation. However, transfer learning should supplement, not supersede, site-specific sampling and field validation for accurate environmental assessment (96).

A key gap in digital soil mapping is that ML and hybrid models often optimise accuracy but provide limited and poorly calibrated estimates of local uncertainty. Few frameworks jointly map concentration, risk and uncertainty in an interpretable way for regulators (97). Explainable AI approaches, such as SHapley Additive exPlanations (SHAP) are increasingly recognised as essential tools for addressing this limitation by quantifying the contribution of individual environmental variables to spatial predictions. However, systematic integration of XAI into geostatistical workflows remains limited. Many studies are limited to concentration mapping; fewer integrate full ecological/human health risk assessments, linking them quantitatively to sources (source-oriented risk) (98, 99). Besides, current research is dominated by heavy metals, while organic and emerging pollutants such as POPs, PAHs, microplastics and other emerging contaminants are underrepresented.

Emerging solutions and future research directions

Emerging technological and methodological innovations are steadily addressing the abovementioned challenges. Further exploration of geostatistics–ML hybrids (e.g., RF-kriging, RF-EBKRP, RF-OK) should be carried out for different contaminants and sampling densities, including formal guidelines on when hybrids outperform single models (23, 24, 83). ML-assisted source attribution, combined with isotopic analysis and elemental fingerprinting, is enabling more precise differentiation between anthropogenic and geogenic metal sources, supporting targeted remediation that preserves natural background levels. Ensemble and stacking models with embedded feature selection and interpretability should be developed to tailor high-dimensional soil data. Model transferability should be investigated across regions, years and pollution types, including transfer learning strategies and domain adaptation. Integrative frameworks that couple spatial prediction, receptor/source models and ecological and health risks in a single workflow should be extended to include explicit source-oriented risk mapping. Such frameworks should be applied beyond heavy metals to organic pollutants and mixed contaminant systems.

Table 5. Comparison of geostatistical and hybrid models for predicting soil pollution

Model category	Methods evaluated	R ²	RMSE/MAE	Key findings	References
Hybrid + probabilistic ML	Multivariate statistics (PCA, hierarchical clustering); Bayesian networks; IK; SGS	NR	NR	SGS and IK improved probabilistic contamination estimation, while Bayesian networks enhanced modelling of inter-variable relationships. The framework is well-suited for multi-contaminant pollution scenarios	(84)
	OK-PMF	0.76–0.99	EBK-PMF recorded prediction efficiency up to 92 % with consistently lower RMSE and MAE than OK-PMF	EBK-PMF outperformed OK-PMF by enhancing source apportionment accuracy; the hybrid approach demonstrated superior spatial prediction and contaminant source attribution	(19)
	EBK-PMF	0.97–0.99			
Hybrid kriging + regression	EBK	NR	EBK-SVMR: RMSE = 95.48 mg kg ⁻¹ , MAE = 77.37 mg kg ⁻¹	EBK-SVMR achieved the best overall performance; hybrid models significantly reduced prediction errors compared to independent models; EBK-MLR performed poorly, indicating incompatibility of linear assumptions with the dataset	(61)
	SVMR	0.66			
	EBK-SVMR	0.63			
	EBK-MLR	<0.10			
Deterministic vs. geostatistical	IDW (p = 1–4); RBF; OK	NR	Based on RMSE/MRE, IDW (especially IDW-4 and IDW-2) emerged as the best performing interpolation method	IDW-4 and IDW-2 yielded the best performance for Ag and Co, respectively; OK did not consistently outperform deterministic methods across all elements	(14)
	MLR-OK	Zn (37.3 %); Ni (37.9 %); Pb (31.1 %); As (35.8 %); Cr (53.5 %)			
Hybrid ML-kriging	SVM-OK	Zn (46.0 %); Ni (42.1 %); Pb (54.2 %); As (49.0 %); Cr (52.3 %)	RF-OK recorded the lowest RMSE among all variants	RF-OK explained 60–78 % of spatial variance across five heavy metals, substantially outperforming other models (mean R ² < 8 %); RF-based hybrid models are strongly recommended for multi-metal contamination mapping	(83)
	RF-OK	Zn (71.4 %); Ni (61.5 %); Pb (74.7 %); As (65.8 %); Cr (77.9 %)			
	IDW; OK; spline with barriers; GANN	0.79–0.98			
Hybrid AI-kriging			GANN RMSE: 0.05–77.4	GANN outperformed all conventional interpolation methods; prediction accuracy was greatly improved by accounting for soil and environmental variables	(86)

Hybrid RF-Bayesian kriging	EBKRP	0.40–0.66	Compared with EBKRP, RF-EBKRP improves accuracy by 122.5 % (RMSE/MAPE) and raises R ² up to 48.5 %	RF-EBKRP achieved R ² 8.9–15.6 % higher than RF alone; the hybrid model demonstrated the highest accuracy, confirming the synergistic advantage of combining ML with Bayesian geostatistics	(24)
	RF	0.77–0.90			
	RF-EBKRP	0.89–0.98			
Bayesian kriging + risk integration	EBK; RBF; HHRA	NR	ME: 0–1.29; RMSE: 0.86–1.29	EBK reduced interpolation errors and improved spatial prediction accuracy; the integrated EBK-HHRA framework effectively coupled geostatistical mapping with quantitative health risk assessment	(62)
Multi-method source apportionment	PCA; PMF; OK; AERMOD model	NR	NR	Integration of PCA, PMF and geostatistical interpolation improved source identification and quantification; coupling with AERMOD reduced source attribution uncertainty and enhanced interpretive robustness	(21)
Hybrid kriging-regression	Co-kriging + GTWR model	NR	NR	The hybrid model improved prediction accuracy by incorporating socio-economic variables; GTWR effectively captured non-stationarity in contaminant distribution across space and time	(18)
Hybrid ML-kriging	OK	0.87	RF-OK RMSE: 52.884 mg kg ⁻¹	RF-OK achieved the highest R ² (0.91) and the lowest RMSE, outperforming both RF and OK; the hybrid framework demonstrated consistent superiority over individual models in urban soil pollution mapping	(23)
	RF	0.87			
	RF-OK	0.91			
Deep learning-kriging hybrid	OK + 1D CNN and BiLSTM	NR	Outperformed conventional geostatistical benchmarks	The deep learning-kriging framework captured complex spatial heterogeneity and improved high-resolution prediction accuracy; BiLSTM integration enabled the model to exploit sequential spatial patterns beyond the capability of conventional kriging	(82)

NR: not reported; OK: ordinary kriging; IK: indicator kriging; EBK: empirical Bayesian kriging; PMF: positive matrix factorisation; IDW: inverse distance weighting; RBF: radial basis functions; ML: machine learning; RF: random forest; SVM: support vector machine; SVMR: support vector machine regression; SGS: sequential Gaussian simulation; MLR: multiple linear regression; RMSE: root mean square error; MAE: mean absolute error; ME: mean error; MAPE: mean absolute percentage error; PCA: principal component analysis; GANN: genetic algorithm-neural network; HHRA: human health risk assessment; AERMOD: atmospheric transport modelling; GTWR: geographically and temporally weighted regression; 1D CNN: one-dimensional convolutional neural network; BiLSTM: bidirectional long short-term memory.

Sampling designs should be optimised for hybrid models (how many, where and when to sample) under budget constraints. Adaptive sampling algorithms can identify optimal sampling locations that maximise information gain while balancing cost and logistical constraints (100). Early studies indicate that such approaches can reduce sampling costs by 30–40 % while improving predictive accuracy by 15–25 % (101, 102). The focus should be on developing strategies that safely combine coarse open data with targeted field sampling, given high misclassification when open data are used alone. These directions collectively aim at robust, transferable, interpretable hybrid systems that fuse diverse data sources and deliver actionable, uncertainty-aware risk maps for management and policy.

Conclusion

Soil contamination, particularly from heavy metals and PTEs, remains a critical and multifaceted environmental issue with severe implications for human health, ecological integrity and sustainable development. This review demonstrates that the integration of geostatistical methods with ML and sensing technologies has fundamentally transformed soil pollution monitoring. Hybrid frameworks consistently achieved significant improvements in prediction accuracy, reduced uncertainty and enabled high-resolution hotspot identification. Geostatistically optimised sampling and remediation strategies can lower investigation and cleanup costs by 25–40 % while enhancing environmental protection outcomes. Integrating geostatistics with advanced computational and sensing technologies represents a paradigm shift in soil pollution monitoring and risk assessment.

Future endeavour should lie in embracing interdisciplinary, data-rich and computationally empowered frameworks that bridge the gap between scientific knowledge and practical environmental stewardship. The following key topics are identified for future research: (i) development of standardised benchmark datasets and shared validation protocols for fair cross-study comparison, (ii) extension of the hybrid spatial models to other contaminants, such as PAHs, microplastics and pharmaceuticals, (iii) integration of explainable AI tools (such as SHAP) to enhance interpretability for regulators and practitioners, (iv) design of adaptive, uncertainty-guided sampling campaigns instead of uniform grid protocols, (v) development of transferable, domain-adapted models validated across contrasting climatic and pedological conditions, especially in data-poor tropical and global South environments and (vi) Coupling multi-contaminant spatial models with real-time sensor networks for proactive early-warning monitoring. Achievement of these priorities will be crucial for converting methodological developments into tools for environmental management that are actionable, equitable and policy relevant.

Acknowledgements

The authors are grateful to Tamil Nadu Agricultural University for providing library resources for the collection of literature.

Authors' contributions

YD conceptualised the study and prepared the original manuscript. MD reviewed and edited the manuscript and supervised the study.

PS, KP, PPC and GM reviewed the manuscript and contributed to the visualisation of the study. All authors have read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interest to declare.

Ethical issues: None

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Grammarly and Quillbot software to improve language quality, grammar, readability and overall flow of the text. After using this tool/service, the authors reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

- Swain CK. Environmental pollution indices: a review on concentration of heavy metals in air, water and soil near industrialization and urbanisation. *Discov Environ.* 2024;2(1):5. <https://doi.org/10.1007/s44274-024-00030-8>
- Siddig MMS, Brevik EC, Sauer D. Human health risk assessment from potentially toxic elements in the soils of Sudan: a meta-analysis. *Sci Total Environ.* 2025;958:178196. <https://doi.org/10.1016/j.scitotenv.2024.178196>
- Ali S. Environmental contaminants and remediation within the framework of the UN sustainable development goals. In: Ali S, Negm A, editors. *Environmental contaminants and health*. Cham: Springer Nature Switzerland; 2025. p. 205–27. <https://doi.org/10.1007/978-2025-1226>
- Meftah S, Meftah K, Drissi M, Radah I, Malous K, Amahrous A, et al. Heavy metal polluted water: effects and sustainable treatment solutions using bio-adsorbents aligned with the SDGs. *Discov Sustain.* 2025;6(1):137. <https://doi.org/10.1007/s43621-025-00895-6>
- Kumar V, Sharma A, Kaur P, Singh Sidhu GP, Bali AS, Bhardwaj R, et al. Pollution assessment of heavy metals in soils of India and ecological risk assessment: a state-of-the-art. *Chemosphere.* 2019;216:449–62. <https://doi.org/10.1016/j.chemosphere.2018.10.066>
- Liu S, Wang L, Liu D, Diao J, Jiang Y. A novel spatial prediction method for soil heavy metal based on unbiased conditional kernel density estimation. *Sci Total Environ.* 2024;952:175843. <https://doi.org/10.1016/j.scitotenv.2024.175843>
- Li Y, Chen B, Yang S, Jiao Z, Zhang M, Yang Y, et al. Advances in environmental pollutant detection techniques: enhancing public health monitoring and risk assessment. *Environ Int.* 2025;197:109365. <https://doi.org/10.1016/j.envint.2025.109365>
- Bux RK, Batool M, Shah SM, Solangi AR, Shaikh AA, Haider SI, et al. Mapping the spatial distribution of soil heavy metals pollution by principal component analysis and cluster analyses. *Water Air Soil Pollut.* 2023;234(6):330. <https://doi.org/10.1007/s11270-023-06361-1>
- Xiang J, Xu P, Chen W, Wang X, Chen Z, Xu D, et al. Pollution characteristics and health risk assessment of heavy metals in agricultural soils over the past five years in Zhejiang, Southeast China. *IJERPH.* 2022;19(22):14642. <https://doi.org/10.3390/ijerph192214642>
- Štrbac S, Randelović D, Gajica G, Hukić E, Stojadinović S, Veselinović G, et al. Spatial distribution and source identification of heavy metals in European mountain beech forests soils. *Chemosphere.*

- 2022;309:136662. <https://doi.org/10.1016/j.chemosphere.2022.136662>
11. Al-Qawasmeh O, Ghrefat H. Integrating pollution indices, spatial interpolation and machine learning for soil contamination analysis along the Zarqa River, Jordan. *Environ Monit Assess.* 2025;197(10):1137. <https://doi.org/10.1007/s10661-025-14586-2>
 12. Hussain A, Bangroo SA, Muslim M. Assessment of soil contamination using remote sensing and spatial techniques. In: Mushtaq F, Farooq M, Mukherjee AB, Ghosh Nee Lala M, editors. *Geospatial analytics for environmental pollution modeling*. Cham: Springer Nature Switzerland; 2023. p. 249–66. https://doi.org/10.1007/978-3-031-45300-7_10
 13. Nkinahamira F, Feng A, Zhang L, Rong H, Ndagijimana P, Guo D, et al. Machine learning approaches for monitoring environmental metal pollutants: recent advances in source apportionment, detection, quantification and risk assessment. *TrAC Trends Anal Chem.* 2024;180:117980. <https://doi.org/10.1016/j.trac.2024.117980>
 14. Saha A, Gupta BS, Patidar S, Martínez-Villegas N. Spatial distribution based on optimal interpolation techniques and assessment of contamination risk for toxic metals in the surface soil. *J South Am Earth Sci.* 2022;115:103763. <https://doi.org/10.1016/j.jsames.2022.103763>
 15. Zhang B, Jia T, Peng S, Yu X, She D. Spatial distribution, source identification and risk assessment of heavy metals in the cultivated soil of the Qinghai–Tibet Plateau region: case study on Huzhu County. *Global Ecol Conserv.* 2022;35:e02073. <https://doi.org/10.1016/j.gecco.2022.e02073>
 16. Dash PK. Remote sensing as a potential tool for advancing digital soil mapping. In: *Remote sensing of soils*. Elsevier; 2024. p. 357–70. <https://doi.org/10.1016/B978-0-443-18773-5.00021-1>
 17. Biswas A, Biswas A. Geostatistics in soil science: a comprehensive review on past, present and future perspectives. *J Indian Soc Soil Sci.* 2024;72(1):1–22. <https://doi.org/10.5958/0974-0228.2024.00017.1>
 18. Wang H, Zhao M, Huang X, Song X, Cai B, Tang R, et al. Improving prediction of soil heavy metal(loid) concentration by developing a combined Co-kriging and geographically and temporally weighted regression (GTWR) model. *J Hazard Mater.* 2024;468:133745. <https://doi.org/10.1016/j.jhazmat.2024.133745>
 19. Agyeman PC, John K, Kebonye NM, Borůvka L, Vašát R, Drábek O. A geostatistical approach to estimating source apportionment in urban and peri-urban soils using the Czech Republic as an example. *Sci Rep.* 2021;11(1):23615. <https://doi.org/10.1038/s41598-021-02968-8>
 20. Huang G, Wang X, Chen D, Wang Y, Zhu S, Zhang T, et al. A hybrid data-driven framework for diagnosing contributing factors for soil heavy metal contaminations using machine learning and spatial clustering analysis. *J Hazard Mater.* 2022;437:129324. <https://doi.org/10.1016/j.jhazmat.2022.129324>
 21. Qi C, Xu M, Liu J, Li C, Yang B, Jin Z, et al. Source analysis and contribution estimation of heavy metal contamination in agricultural soils in an industrial town in the Yangtze River Delta, China. *Minerals.* 2024;14(3):279. <https://doi.org/10.3390/min14030279>
 22. Al Yeamin A, Mia MdY, Khan SR, Rahman MS, Senapathi V, Islam ARMdT, et al. Innovative strategies for pollution assessment in Northern Bangladesh: mapping pollution areas and tracing metal (loid)s sources in various soil types. *PLoS One.* 2025;20(2):e0311270. <https://doi.org/10.1371/journal.pone.0311270>
 23. Han H, Suh J. Spatial prediction of soil contaminants using a hybrid random forest–ordinary kriging model. *Appl Sci.* 2024;14(4):1666. <https://doi.org/10.3390/app14041666>
 24. Molla A, Zhang W, Zuo S, Ren Y, Han J. A machine learning and geostatistical hybrid method to improve spatial prediction accuracy of soil potentially toxic elements. *Stoch Environ Res Risk Assess.* 2023;37(2):681–96. <https://doi.org/10.1007/s00477-022-02284-1>
 25. Fan Y, Wang X, Funk T, Rashid I, Herman B, Bompoti N, et al. A critical review for real-time continuous soil monitoring: advantages, challenges and perspectives. *Environ Sci Technol.* 2022;56(19):13546–64. <https://doi.org/10.1021/acs.est.2c03562>
 26. De Caires SA, Martin CS, Atwell MA, Kaya F, Wuddivira GA, Wuddivira MN. Advancing soil mapping and management using geostatistics and integrated machine learning and remote sensing techniques: a synoptic review. *Discov Soil.* 2025;2(1):53. <https://doi.org/10.1007/s44378-025-00082-z>
 27. Khan ZR, Jehangir A. Geostatistical methods and framework for pollution modelling. In: Mushtaq F, Farooq M, Mukherjee AB, Ghosh Nee Lala M, editors. *Geospatial analytics for environmental pollution modeling*. Cham: Springer Nature Switzerland; 2023. p. 33–56. https://doi.org/10.1007/978-3-031-45300-7_2
 28. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ.* 2021;372:n160. <https://doi.org/10.1136/bmj.n160>
 29. Degvekar JV, Gadekar EU, Darshana O, Chand J, Amrish VN, Jose J, et al. Tracking soil health and potentially toxic elements (PTEs) across land-use types using physico-chemical, magnetic and geochemical proxies: a case study from Manipal, Southwestern India. *Environ Geochem Health.* 2025;47(9):397. <https://doi.org/10.1007/s10653-025-02665-9>
 30. Nyiramigisha P, Komariah, Sajidan. Harmful impacts of heavy metal contamination in the soil and crops grown around dumpsites. *Rev Agri Sci.* 2021;9(0):271–82. https://doi.org/10.7831/ras.9.0_271
 31. Safarov R, Shomanova Z, Nossenko Y, Mussayev Z, Shomanova A. Digital visualization of environmental risk indicators in the territory of the urban industrial zone. *Sustainability.* 2024;16(12):5190. <https://doi.org/10.3390/su16125190>
 32. Reza SK, Baruah U, Singh SK, Das TH. Geostatistical and multivariate analysis of soil heavy metal contamination near coal mining area, Northeastern India. *Environ Earth Sci.* 2015;73(9):5425–33. <https://doi.org/10.1007/s12665-014-3797-1>
 33. Gök G, Tulun Ş, Çelebi H. Mapping of heavy metal pollution density and source distribution of campus soil using geographical information system. *Sci Rep.* 2024;14(1):29918. <https://doi.org/10.1038/s41598-024-78961-8>
 34. Parizanganeh AH, Bijnavand V, Zamani AA, Hajabolfath A. Concentration, distribution and comparison of total and bioavailable heavy metals in top soils of Bonab District in Zanjan Province. *OJSS.* 2012;2(2):123–32. <https://doi.org/10.4236/ojss.2012.22018>
 35. Müller A, Österlund H, Marsalek J, Viklander M. The pollution conveyed by urban runoff: a review of sources. *Sci Total Environ.* 2020;709:136125. <https://doi.org/10.1016/j.scitotenv.2019.136125>
 36. Hanif MN, Benitez IB. Soil pollution in urban environments: sources, consequences, potential mitigation strategies and the importance of sustainable urban development. *Water Air Soil Pollut.* 2025;236(14):945. <https://doi.org/10.1007/s11270-025-08421-0>
 37. Shaheen A, Iqbal J. Spatial distribution and mobility assessment of carcinogenic heavy metals in soil profiles using geostatistics and random forest, Boruta algorithm. *Sustainability.* 2018;10(3):799. <https://doi.org/10.3390/su10030799>
 38. Chakraborty TK, Islam MS, Ghosh GC, Ghosh P, Zaman S, Hossain MR, et al. Receptor model-based sources and risks appraisal of potentially toxic elements in the urban soils of Bangladesh. *Toxicol Rep.* 2023;10:308–19. <https://doi.org/10.1016/j.toxrep.2023.02.011>
 39. Chandra GV, Golla SY, Ghosh PK. Review of soil environment quality in India near coal mining regions: current and future predictions. *Environ Geochem Health.* 2024;46(6):194. <https://doi.org/10.1007/s10653-024-01968-7>
 40. Esmaeili A, Moore F, Keshavarzi B, Jaafarzadeh N, Kermani M. A geochemical survey of heavy metals in agricultural and background soils of the Isfahan industrial zone, Iran. *CATENA.* 2014;121:88–98. <https://doi.org/10.1016/j.catena.2014.05.003>
 41. Cagnazzo C, Potente E, Rosato S, Mastronuzzi G. Geostatistics and structure from motion techniques for coastal pollution assessment along the Policoro Coast (Southern Italy). *Geosciences.* 2020;10

- (1):28. <https://doi.org/10.3390/geosciences10010028>
42. Cui M, Hou E, Lu T, Hou P, Feng D. Study on spatial interpolation methods for high precision 3D geological modeling of coal mining faces. *Appl Sci*. 2025;15(6):2959. <https://doi.org/10.3390/app15062959>
 43. Hou D, Jia X, Wang L, McGrath SP, Zhu YG, Hu Q, et al. Global soil pollution by toxic metals threatens agriculture and human health. *Science*. 2025;388(6744):316–21. <https://doi.org/10.1126/science.adr5214>
 44. Soffianian AR, Bakir HB, Khodakarami L. Evaluation of heavy metals concentration in soil using GIS, RS and geostatistics. *J Environ Sci Toxicol Food Technol*. 2015;9:61–72.
 45. Nawaz R, Aslam M, Nasim I, Irshad MA, Ahmad S, Latif M, et al. Air pollution tolerance index and heavy metals accumulation of tree species for sustainable environmental management in megacity of Lahore. *Air*. 2022;1(1):55–68. <https://doi.org/10.3390/air1010004>
 46. Singh A, Golui K, Rajrana Y, Pannu RS. Modern soil mapping and spatial analysis. *SSRN Electronic J*. 2026. <https://doi.org/10.2139/ssrn.6322479>
 47. Waters N. Tobler's first law of geography. In: Richardson D, Castree N, Goodchild MF, Kobayashi A, Liu W, Marston RA, editors. *International encyclopedia of geography*. Wiley; 2017. p. 1–13.
 48. Oliver MA, Webster R. A tutorial guide to geostatistics: computing and modelling variograms and kriging. *CATENA*. 2014;113:56–69. <https://doi.org/10.1016/j.catena.2013.09.006>
 49. Webster R, Oliver M. *Geostatistics for environmental scientists*. 2nd ed. Chichester: Wiley; 2007. <https://doi.org/10.1002/9780470517277>
 50. Bostan P. Jeo-İstatistik'te Kullanılan Temel Kriging Yöntemleri. *Yüzüncü Yıl Üniversitesi Tarım Bilimleri Dergisi*. 2017;27(1):10–20. <https://doi.org/10.29133/yyutbd.305093>
 51. Goovaerts P. *Geostatistics for natural resources evaluation*. New York: Oxford University Press; 1997. <https://doi.org/10.1093/oso/9780195115383.001.0001>
 52. Hengl T, Heuvelink GBM, Rossiter DG. About regression-kriging: from equations to case studies. *Comput Geosci*. 2007;33(10):1301–15. <https://doi.org/10.1016/j.cageo.2007.05.001>
 53. Hilal A, Bangroo SA, Kirmani NA, Wani JA, Biswas A, Bhat MI, et al. Geostatistical modeling—a tool for predictive soil mapping. In: *Remote sensing in precision agriculture*. Elsevier; 2024. p. 389–418. <https://doi.org/10.1016/B978-0-323-91068-2.00011-4>
 54. Abuzaid AS, Jahin HS, Shokr MS, El Baroudy AA, Mohamed ES, Rebouh NY, et al. A novel regional-scale assessment of soil metal pollution in arid agroecosystems. *Agronomy*. 2023;13(1):161. <https://doi.org/10.3390/agronomy13010161>
 55. Liu Z, Fei Y, Shi H, Mo L, Qi J. Prediction of high-risk areas of soil heavy metal pollution with multiple factors on a large scale in industrial agglomeration areas. *Sci Total Environ*. 2022;808:151874. <https://doi.org/10.1016/j.scitotenv.2021.151874>
 56. Freulon X, Desassis N. Kriging. In: Sagar BSD, Cheng Q, McKinley J, Agterberg F, editors. *Encyclopedia of mathematical geosciences*. Cham: Springer International Publishing; 2023. p. 705–13. https://doi.org/10.1007/978-3-030-85040-1_18
 57. Thompson JA, Roecker S, Grunwald S, Owens PR. Digital soil mapping. In: *Hydropedology*. Elsevier; 2012. p. 665–709. <https://doi.org/10.1016/B978-0-12-386941-8.00021-6>
 58. Qu MK, Li WD, Zhang CR, Wang SQ, Yang Y, He LY. Source apportionment of heavy metals in soils using multivariate statistics and geostatistics. *Pedosphere*. 2013;23(4):437–44. [https://doi.org/10.1016/S1002-0160\(13\)60036-3](https://doi.org/10.1016/S1002-0160(13)60036-3)
 59. Khodoli Zangeneh D, Amanipoor H, Battaleb-Looie S. Evaluation of heavy metal contamination using cokriging geostatistical method (case study of Abteymour oilfield in southern Iran). *Appl Water Sci*. 2023;13(10):200. <https://doi.org/10.1007/s13201-023-01980-9>
 60. Adhikary PP, Dash ChJ, Behera B, Mohanty S, Shit PK. Indicator kriging and its usefulness in assessing spatial suitability of groundwater for drinking. In: *Case studies in geospatial applications to groundwater resources*. Elsevier; 2023. p. 11–26. <https://doi.org/10.1016/B978-0-323-99963-2.00011-0>
 61. Agyeman PC, Kebonye NM, John K, Borůvka L, Vašát R, Fajemisim O. Prediction of nickel concentration in peri-urban and urban soils using hybridized empirical Bayesian kriging and support vector machine regression. *Sci Rep*. 2022;12(1):3004. <https://doi.org/10.1038/s41598-022-06843-y>
 62. Wang L, Liu R, Liu J, Qi Y, Zeng W, Cui B. A novel regional-scale human health risk assessment model for soil heavy metal(loid) pollution based on empirical Bayesian kriging. *Ecotoxicol Environ Saf*. 2023;258:114953. <https://doi.org/10.1016/j.ecoenv.2023.114953>
 63. Abowaly ME, Belal A-AA, Abd Elkhalek EEA, Elsayed S, Abou Samra RMA, Alshammari AS, et al. Assessment of soil pollution levels in North Nile Delta, by integrating contamination indices, GIS and multivariate modeling. *Sustainability*. 2021;13(14):8027. <https://doi.org/10.3390/su13148027>
 64. Ouabo RE, Sangodoyin AY, Ogundiran MB. Assessment of ordinary kriging and inverse distance weighting methods for modeling chromium and cadmium soil pollution in E-waste sites in Douala, Cameroon. *J Health Pollut*. 2020;10(26):200605. <https://doi.org/10.5696/2156-9614-10.26.200605>
 65. Qiao P, Lei M, Yang S, Yang J, Guo G, Zhou X. Comparing ordinary kriging and inverse distance weighting for soil as pollution in Beijing. *Environ Sci Pollut Res*. 2018;25(16):15597–608. <https://doi.org/10.1007/s11356-018-1552-y>
 66. Wang J, Zuo R. Uncertainty quantification in geochemical mapping: a review and recommendations. *Geochim Geophys Geosyst*. 2024;25(3):e2023GC011301. <https://doi.org/10.1029/2023GC011301>
 67. Ekoa Bessa AZ. Spatial variation, ecological risk and point sources of environmental trace metals in lacustrine ecosystems: an assessment of natural and urban inputs. *Soil Sediment Contam*. 2024;33(7):1062–86. <https://doi.org/10.1080/15320383.2023.2293866>
 68. Ersoy A, Yunsel TY, Cetin M. Characterization of land contaminated by past heavy metal mining using geostatistical methods. *Arch Environ Contam Toxicol*. 2004;46(2):162–75. <https://doi.org/10.1007/s00244-003-2269-5>
 69. Juang K-W, Chen Y-S, Lee D-Y. Using sequential indicator simulation to assess the uncertainty of delineating heavy-metal contaminated soils. *Environ Pollut*. 2004;127(2):229–38. <https://doi.org/10.1016/j.envpol.2003.07.001>
 70. Jin Z, Zhang L, Lv J, Sun X. The application of geostatistical analysis and receptor model for the spatial distribution and sources of potentially toxic elements in soils. *Environ Geochem Health*. 2021;43(1):407–21. <https://doi.org/10.1007/s10653-020-00729-6>
 71. Moghimi Dehkordi M, Pournuroz Nodeh Z, Soleimani Dehkordi K, Salmanvandi H, Rasouli Khorjistan R, Ghaffarzadeh M. Soil, air and water pollution from mining and industrial activities: sources of pollution, environmental impacts and prevention and control methods. *Results Eng*. 2024;23:102729. <https://doi.org/10.1016/j.rineng.2024.102729>
 72. Yang M, Wang C, Yang Z-P, Yan N, Li F-Y, Diao Y-W, et al. Use of portable X-ray fluorescence spectroscopy and geostatistics for health risk assessment. *Ecotoxicol Environ Saf*. 2018;153:68–77. <https://doi.org/10.1016/j.ecoenv.2018.01.050>
 73. Castrignanò A, Buttafuoco G, Conforti M, Maesano M, Moresi FV, Mugnoz GS. Improving the spatial prediction of sand content in forest soils using a multivariate geostatistical analysis of LiDAR and hyperspectral data. *Remote Sens*. 2023;15(18):4416. <https://doi.org/10.3390/rs15184416>
 74. Liang X, Wang C, Song Z, Yang S, Bi X, Li Z, et al. Soil metal(loid)s pollution around a lead/zinc smelter and source apportionment using isotope fingerprints and receptor models. *Appl Geochem*.

- 2021;135:105118. <https://doi.org/10.1016/j.apgeochem.2021.105118>
75. Hossain M, Patra PK. Contamination zoning and health risk assessment of trace elements in groundwater through geostatistical modelling. *Ecotoxicol Environ Saf*. 2020;189:110038. <https://doi.org/10.1016/j.ecoenv.2019.110038>
 76. Huang B-Y, Lü Q-X, Tang Z-X, Tang Z, Chen H-P, Yang X-P, et al. Machine learning methods to predict cadmium (Cd) concentration in rice grain and support soil management at a regional scale. *Fundam Res*. 2024;4(5):1196–205. <https://doi.org/10.1016/j.fmre.2023.02.016>
 77. Zhao B, Zhu W, Hao S, Hua M, Liao Q, Jing Y, et al. Prediction heavy metals accumulation risk in rice using machine learning and mapping pollution risk. *J Hazard Mater*. 2023;448:130879. <https://doi.org/10.1016/j.jhazmat.2023.130879>
 78. Cui L, Li L, Zhang A, Pan G, Bao D, Chang A. Biochar amendment greatly reduces rice Cd uptake in a contaminated paddy soil: a two-year field experiment. *BioRes*. 2011;6(3):2605–18. <https://doi.org/10.15376/biores.6.3.2605-2618>
 79. Cui L, Pan G, Li L, Yan J, Zhang A, Bian R, et al. The reduction of wheat Cd uptake in contaminated soil via biochar amendment: a two-year field experiment. *BioResources*. 2012;7(4):5666–76. <https://doi.org/10.15376/biores.7.4.5666-5676>
 80. Ma LQ, Komar KM, Tu C, Zhang W, Cai Y, Kennelley ED. A fern that hyperaccumulates arsenic. *Nature*. 2001;409(6820):579. <https://doi.org/10.1038/35054664>
 81. Sylva L, Izah SC, Hait M. Geostatistical insights into trace metal dynamics: techniques for spatial mapping, variability analysis and source characterization in the environment. *ES General*. 2024. <https://doi.org/10.30919/esg1107>
 82. Anvari K, Benndorf J, Gerber G, Alisch U. Hybrid geostatistical and deep learning framework for geochemical characterization in historical mine tailings. *Sci Rep*. 2025;15(1):35004. <https://doi.org/10.1038/s41598-025-19441-5>
 83. Song Y, Kang L, Lin F, Sun N, Aizezi A, Yang Z, et al. Estimating the spatial distribution of soil heavy metals in oil mining area using air quality data. *Atmos Environ*. 2022;287:119274. <https://doi.org/10.1016/j.atmosenv.2022.119274>
 84. Boente C, Albuquerque MTD, Gerassis S, Rodríguez-Valdés E, Gallego JR. A coupled multivariate statistics, geostatistical and machine-learning approach to address soil pollution in a prototypical Hg-mining site in a natural reserve. *Chemosphere*. 2019;218:767–77. <https://doi.org/10.1016/j.chemosphere.2018.11.172>
 85. Han Y, Zhang S, Kang D, Hao N, Peng J, Zhou Y, et al. A decade review of human health risks from heavy metal contamination in industrial sites. *Water Air Soil Pollut*. 2025;236(2):144. <https://doi.org/10.1007/s11270-025-07759-9>
 86. Yin G, Chen X, Zhu H, Chen Z, Su C, He Z, et al. A novel interpolation method to predict soil heavy metals based on a genetic algorithm and neural network model. *Sci Total Environ*. 2022;825:153948. <https://doi.org/10.1016/j.scitotenv.2022.153948>
 87. Gao B, Stein A, Wang J. A two-point machine learning method for the spatial prediction of soil pollution. *Int J Appl Earth Obs Geoinf*. 2022;108:102742. <https://doi.org/10.1016/j.jag.2022.102742>
 88. Li Y, Hou Y, Tao H, Cao H, Liu X, Wang Z, et al. An improved non-stationary geostatistical method for three-dimensional interpolation of Benzo(a)pyrene at a contaminated site. *Sci Total Environ*. 2022;838:156169. <https://doi.org/10.1016/j.scitotenv.2022.156169>
 89. Mahdavinjad MS, Rezvan M, Barekatain M, Adibi P, Barnaghi P, Sheth AP. Machine learning for internet of things data analysis: a survey. *Digit Commun Netw*. 2018;4(3):161–75. <https://doi.org/10.1016/j.dcan.2017.10.002>
 90. Tang S, Wang C, Song J, Ihenetu SC, Li G. Advances in studies on heavy metals in urban soil: a bibliometric analysis. *Sustainability*. 2024;16(2):860. <https://doi.org/10.3390/su16020860>
 91. Chapman J, Truong VK, Elbourne A, Gangadoo S, Cheeseman S, Rajapaksha P, et al. Combining chemometrics and sensors: toward new applications in monitoring and environmental analysis. *Chem Rev*. 2020;120(13):6048–69. <https://doi.org/10.1021/acs.chemrev.9b00616>
 92. Tao H, Liao X, Cao H, Zhao D, Hou Y. Three-dimensional delineation of soil pollutants at contaminated sites: progress and prospects. *J Geogr Sci*. 2022;32(8):1615–34. <https://doi.org/10.1007/s11442-022-2013-6>
 93. Khan S, Naushad M, Lima EC, Zhang S, Shaheen SM, Rinklebe J. Global soil pollution by toxic elements: current status and future perspectives on the risk assessment and remediation strategies – a review. *J Hazard Mater*. 2021;417:126039. <https://doi.org/10.1016/j.jhazmat.2021.126039>
 94. Rajeswari S, Saravanan P, Linkesver M, Rajeshkannan R, Rajasimman M. Identifying global status and research hotspots of heavy metal remediation: a phase upgrade study. *J Environ Manage*. 2022;324:116265. <https://doi.org/10.1016/j.jenvman.2022.116265>
 95. Agyeman PC, Ahado SK, Borůvka L, Biney JKM, Sarkodie VYO, Kebonye NM, et al. Trend analysis of global usage of digital soil mapping models in the prediction of potentially toxic elements in soil/sediments: a bibliometric review. *Environ Geochem Health*. 2021;43(5):1715–39. <https://doi.org/10.1007/s10653-020-00742-9>
 96. Sun X, Wang H, Forristal D, Fu W, Tunney H, Zhang C. Limited spatial transferability of the relationships between kriging variance and soil sampling spacing in some grasslands of Ireland: implications for sampling design. *Pedosphere*. 2019;29(5):577–89. [https://doi.org/10.1016/S1002-0160\(19\)60801-5](https://doi.org/10.1016/S1002-0160(19)60801-5)
 97. Sani SM, Kharazmi R. Evaluating conventional and digital soil mapping for updating soil maps in Iran: opportunities, challenges and implications for sustainable land management. *Eurasian Soil Sci*. 2026;59(4):69. <https://doi.org/10.1134/S1064229325604433>
 98. Chang H-C, Zhang W. How AI methods enhance the design and performance of nanophotonic environmental sensors: a systematical review. *Discov Sens*. 2026;2(1):7. <https://doi.org/10.1007/s44397-026-00039-z>
 99. Zareba M, Danek T. A novel methodology for explainable artificial intelligence integrated with geostatistics for air pollution control and environmental management. *Ecol Inform*. 2025;92:103450. <https://doi.org/10.1016/j.ecoinf.2025.103450>
 100. Cressie N. Statistics for spatial data. *Terra Nova*. 1992;4(5):613–7. <https://doi.org/10.1111/j.1365-3121.1992.tb00605.x>
 101. Schmidinger J, Schröter I, Bönecke E, Gebbers R, Ruehlmann J, Kramer E, et al. Effect of training sample size, sampling design and prediction model on soil mapping with proximal sensing data for precision liming. *Precis Agric*. 2024;25(3):1529–55. <https://doi.org/10.1007/s11119-024-10122-3>
 102. Zhu AX, Liu J, Du F, Zhang SJ, Qin CZ, Burt J, et al. Predictive soil mapping with limited sample data. *Eur J Soil Sci*. 2015;66(3):535–47. <https://doi.org/10.1111/ejss.12244>

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonpublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonpublishing.com/journals/index.php/PST/indexing_abstracting

Copyright: © The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited (<https://creativecommons.org/licenses/by/4.0/>)

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.