Geographical Analysis

ORIGINAL ARTICLE OPEN ACCESS

Focal-Feature Regression Kriging

Peng Luo[1] | Yilong Wu[2] | Yongze Song[2]

[1]Senseable City Lab, Massachusetts Institute of Technology, Cambridge, MA, USA | [2]School of Design and the Built Environment, Curtin University, Perth, Australia

Correspondence: Yongze Song (yongze.song@curtin.edu.au)

Received: 18 August 2025 | **Revised:** 11 January 2026 | **Accepted:** 18 February 2026

Keywords: regression kriging | spatial interpolation | spatial statistics

ABSTRACT

Spatial interpolation is a crucial task in geography. As perhaps the most widely used interpolation methods, geostatistical models-such as Ordinary Kriging (OK)-assume spatial stationarity, which makes it difficult to capture the nonstationary characteristics of geographic variables. A common solution is trend surface modeling (e.g., Regression Kriging, RK), which relies on external explanatory variables to model the trend and then applies geostatistical interpolation to the residuals. However, this approach requires high-quality and readily available explanatory variables, which are often lacking in many spatial interpolation scenarios-such as estimating heavy metal concentrations underground. This study proposes a Focal Feature Regression Kriging (FFRK) method, which automatically extracts geospatial features to construct a regression-based trend surface without requiring external explanatory variables. We conducted experiments on the spatial prediction of three heavy metals in a mining area in Australia. In comparison with 17 classical interpolation methods, the results indicate that FFRK, which relies solely on extracted geospatial features, consistently outperforms both conventional Kriging techniques and machine learning models that depend on explanatory variables. This approach effectively addresses spatial nonstationarity while reducing the cost of acquiring explanatory variables, improving both prediction accuracy and generalization ability.

1 | Introduction

Spatial interpolation is one of the essential tasks in geography (Lam 1983; Goodchild 2004; Luo, Lou, et al. 2025). Based on samples collected from the Earth's surface (e.g., land and ocean), a continuous surface is estimated by appropriately modeling the relationships between the samples (Webster and Oliver 2007). With the new Earth observation technologies, such as remote sensing, have emerged, the availability of many geographical variables has increased dramatically (Campbell and Wynne 2011; Luo et al. 2022). However, there are still scenarios where the distribution of variables can only be obtained through sampling, such as the soil organic matter (Cheng et al. 2024) or the elemental content under the ocean (Luo et al. 2023). In addition, the potential of spatial interpolation method in deep space exploration, such as Mars mineral prediction, has been recognized (Jiao et al. 2025). In such cases, spatial interpolation remains a necessary step.

From the perspective of modern spatial statistics, the earliest interpolation methods are considered to belong to the category of deterministic interpolation, such as Inverse Distance Weighting (IDW) (Panigrahi 2021). They based on an assumption that the value at a certain location can be obtained by inversely weighting the distances from surrounding observed points (Tomczak 1998). However, this approach neglects spatial dependencies and fails to adequately consider the overall spatial distribution characteristics and patterns of observed points, limiting its interpolation accuracy. Geostatistical models were developed with a thorough consideration of spatial dependence (Matheron 1963). For

example, the original geostatistical method was Ordinary Kriging (Cressie 1988). They interpolate based on the spatial variation patterns in the data, resulting in more accurate and unbiased estimation results. This approach has found extensive applications across various fields—from geology and mining, where it is used to predict unknown geological features or the distribution of mineral deposits, to agriculture, environmental science, and ecology. The basic assumption of geostatistical models is second-order stationarity, which assumes that the variance between two samples depends only on their distance and direction, regardless of their absolute locations (Clark and Harper 1979).

However, the distribution of geographic variables can exhibit non-stationary characteristics (Cantrell and Cosner 1991). For example, when a study area contains multiple distinct terrain structures, the distribution patterns of geographic variables often vary (Anselin 2013). This phenomenon contradicts the assumptions of ordinary kriging, highlighting the need for interpolation methods specifically designed for non-stationary features.

Many methods have been developed to enhance the Kriging model to account for second-order spatial non-stationarity and improve interpolation accuracy. The first and most straightforward solution for handling spatial second-order non-stationarity is to divide the non-stationary surface into several homogeneous subregions (Luo et al. 2023). Within each subregion, spatial stationarity holds, allowing for separate application of Ordinary Kriging. A representative method following this approach is Stratified Kriging (StK). However, modeling at partition boundaries can introduce discontinuities in the prediction results. To do the stratified modelling while maintain the smooth of the prediction result over space, some methods have been proposed to comprehensively model spatial relationships, enabling the construction of semivariogram functions for different strata. Notable examples include P-MSN (Gao et al. 2020) and the Generalized Heterogeneous Model (GHM) (Luo et al. 2023). P-MSN and GHM provide effective stratified modeling solutions with improved spatial continuity and accuracy compared to ordinary StK. However, they still suffer from the inherent drawbacks of stratified modeling, namely, the effectiveness of this approach heavily depends on the reliability of spatial partitioning algorithms. Furthermore, partition-based modeling reduces the number of available samples within each subregion, which makes it challenging to fit reliable semivariogram functions (Liu et al. 2021; Luo, Li, et al. 2025).

The second solution is to model spatial non-stationary patterns by decomposing the predicted values of geographic variables into two components: a trend surface and spatial variability (Hengl et al. 2007). The trend surface is assumed to be fitted by a deterministic function, while the spatial variability represents a random process that geostatistical methods can predict. Different methods are developed to model the trend (Hengl et al. 2007). When non-stationarity varies by region, location information can be used to model the spatial trend, leading to the development of Universal Kriging (Stein and Corsten 1991). In this approach, a deterministic polynomial function is used to represent the trend surface, while the residual spatial variability is captured through a stochastic component. Regression Kriging (RK) generalizes this idea by explicitly modeling the trend using a regression model, which can include multiple explanatory variables (Hengl et al. 2007). RK first fits a regression model to predict the trend and then applies Kriging to the residuals (Hengl et al. 2004). Because it integrates diverse information sources and allows for a flexible choice of trend models, regression kriging often achieves the highest predictive accuracy among Kriging-based interpolation methods.

However, incorporating explanatory variables significantly increases the cost of spatial interpolation and introduces more uncertainty (Luo, Li, et al. 2025; Luo, Lou, et al. 2025; Lou et al. 2025a, 2025b). Additionally, for many geostatistical tasks, obtaining high quality explanatory variables remains a challenge (Yao et al. 2023). For example, a classic kriging interpolation task involves predicting mineral distribution. Despite the availability of numerous satellite sensors for obtaining remote sensing data, collecting explanatory variables for subsurface prediction is difficult. This presents a challenge for current spatial interpolation methods, including geostatistical models.

The motivation of this study is to develop a regression kriging method that does not require external variables and can address spatial non-stationarity. When explanatory variables are unavailable, we can leverage the spatial patterns of geographic variables to predict the trend surface—an approach that also serves as a spatial feature engineering technique (Rey et al. 2023). This form of feature engineering follows a long tradition in spatial statistics (Griffith 2003; Tiefelsdorf and Griffith 2007). From a machine learning perspective, it conceptually aligns with Deep Kernel Learning (DKL), a hybrid paradigm that combines the structural flexibility of neural networks with the probabilistic rigor of kernel methods (e.g., Gaussian Processes) (Wilson et al. 2016). The key distinction lies in their inductive biases: spatial feature engineering typically relies on explicitly constructed spatial descriptors or basis functions, often guided by statistical assumptions or domain knowledge, whereas DKL formulates feature extraction as an optimization problem, learning a non-linear mapping from the input space to a high-dimensional latent feature manifold in a data-driven manner. Nevertheless, a research gap remains: DKL is commonly implemented as a standalone predictive model, while spatial engineering is rarely treated as a flexible and learnable trend component within classical spatial prediction frameworks. To date, the synergy between these advanced feature-learning paradigms and the classical Regression Kriging framework has not been fully explored.

In this study, we propose the Focal-Feature Regression Kriging (FFRK) method, motivated by the need to integrate spatial feature engineering techniques to address the lack of high-quality explanatory variables in certain spatial interpolation scenarios. The core premise of FFRK is that the trend component of a geographic variable can be effectively captured by its focal neighborhood. For each observation point, a set of focal features—representing local dependence, spatial heterogeneity, and geosimilarity—is extracted from the values and spatial configurations of surrounding samples. By utilizing these intrinsic focal features to construct a regression model, the method fits a robust trend surface that accounts for non-stationarity without the need for external covariates. Finally, the residuals are employed to construct Ordinary Kriging for spatial interpolation. To validate the performance of FFRK, we conducted a case study in a selected region of Australia, focusing on the prediction of

three heavy metal concentrations. Furthermore, FFRK was compared against 17 classical spatial interpolation models.

2 | Focal Feature Regression Kriging

2.1 | Basic of Regression Kriging

Regression Kriging is a hybrid spatial prediction method that combines a deterministic regression model with a geostatistical interpolation model. The fundamental idea of RK is to separately model the trend component and spatially correlated residuals, allowing for a more flexible and accurate spatial prediction.

2.1.1 | General Formulation of Regression Kriging

Given an observed spatial variable $Z(x)$ at location x, RK decomposes the spatial variation into two components:

$$Z(x) = m(x) + \varepsilon(x) \tag{1}$$

where $m(x)$ represents the deterministic trend component, which captures large-scale variations and can be modeled using regression techniques, $\varepsilon(x)$ represents the spatially correlated residuals, which account for small-scale variations and are modeled using geostatistical interpolation.

2.1.2 | Trend Modeling via Regression

The trend component $m(x)$ is typically modeled using a regression function:

$$m(x) = g(\mathbf{X}(x)) \tag{2}$$

where $g(\cdot)$ is a regression model, which can be a linear regression, random forest, or any machine learning algorithm. $\mathbf{X}(x)$ is a vector of explanatory variables at location x, which can include environmental, demographic, or geospatial factors.

The regression model is trained using known data points $\{x_i, Z(x_i)\}_{i=1}^{n}$. Once trained, this regression model can predict the trend component at any location x_p:

$$m^*(x_p) = g(\mathbf{X}(x_p)) \tag{3}$$

where $m(x_p)$ is the predicted trend at the location x_p.

2.1.3 | Residual Modeling via Kriging

After estimating the trend component, the residuals at known sample locations x_i are computed as:

$$\varepsilon(x_i) = Z(x_i) - m(x_i) \tag{4}$$

where the $\varepsilon(x_i)$ is the residuals at the location x_i.

These residuals are spatially correlated and modeled using geostatistical techniques. Using the fitted semivariogram, residuals at unknown locations x_p are interpolated using ordinary kriging:

$$\varepsilon^*(x_p) = \sum_{i=1}^{n} w_i \varepsilon(x_i) \tag{5}$$

where w_i are kriging weights derived from the semivariogram model.

2.1.4 | RK Prediction

The predicted value at an unknown location x_p is obtained by combining the regression-predicted trend and the kriging-interpolated residual:

$$Z^*(x_p) = m^*(x_p) + \varepsilon^*(x_p) \tag{6}$$

where $Z^*(x_p)$ is the predicted value at location x_p, and

2.2 | Concept of FFRK

The most important feature of RK is its ability to fully utilize explanatory variables to fit the trend, thereby improving the accuracy of ordinary kriging. However, its limitation is also evident: if explanatory variables are difficult to obtain, the method becomes unusable.

As is shown in Figure 1, the idea behind our proposed FFRK is straightforward: in the absence of explanatory variables, we replace them with the spatial distribution characteristics of the predicted variable itself to perform regression kriging. Therefore, FFRK follows the same algorithmic foundation as RK, with the only difference being in trend fitting-rather than constructing a regression model using explanatory variables, it is built based on geospatial features.

$$Z(x) = m(x) + \varepsilon(x) \tag{7}$$

$$m(x) = g(\mathbf{F}(x)) \tag{8}$$

where: $g(\cdot)$ is a predictive model, $\mathbf{F}(x)$ is the extracted geo-feature vector.

Subsequently, the process of FFRK is entirely consistent with that of RK: the trend surface $m(x)$ is fitted based on geospatial features, followed by residual computation and kriging interpolation of the residuals.

2.3 | Geofeature Extraction

In spatial analysis, the interpolation accuracy highly depends on the feature representations of the data. The framework of FFRK can be combined with any kinds of geofeature, according to the specific application.

In this study, as a showcase of the FFRK framework, we consider three distinct types of geospatial features to capture both local pattern and global pattern in spatial distributions (Figure 2). These features are designed to provide a more comprehensive representation of spatial processes and to enhance interpolation performance. The first is the local trend feature, extracted

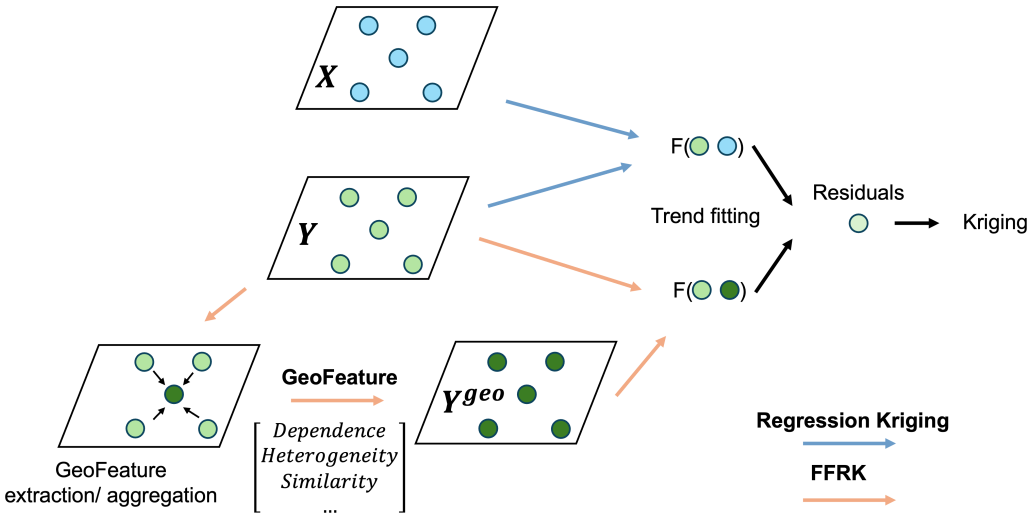

FIGURE 1 | Schematic overview of the developed focal feature regression kriging (FFRK) for spatial interpolation, and its different with the Regression Kriging.

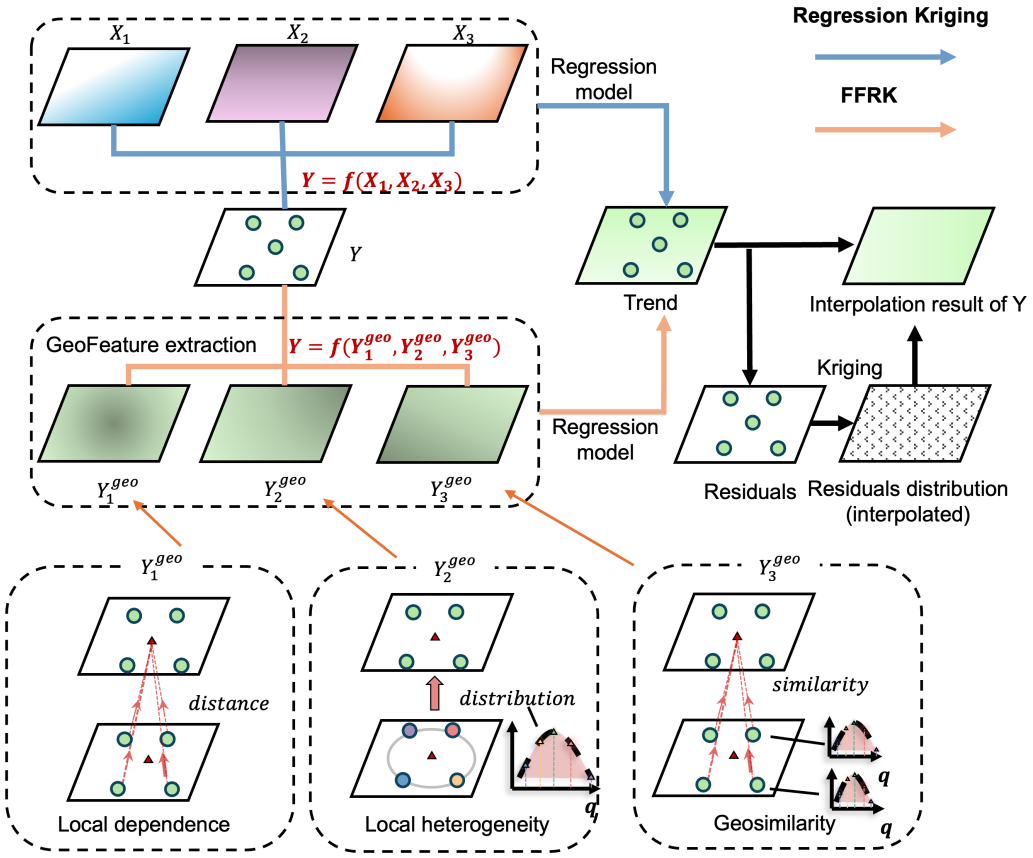

FIGURE 2 | Three types of geofeatures for FFRK-based spatial interpolation used in this study, including local trend, local heterogeneity, and geosimilarity.

using IDW, which describes the general directional pattern of spatial structure in the neighborhood of each sample. This feature reflects the core principle of Tobler's First Law of Geography, and helps characterize the local continuity of spatial variables. The second is the local heterogeneity feature, which captures intra-regional variability and unevenness by computing statistical descriptors (e.g., quantiles) of neighboring observations. This feature highlights the spatial heterogeneity inherent in many geospatial phenomena. The third is the geosimilarity feature, which measures statistical similarity between sample points based on the local distributions used in the second feature. Unlike traditional distance-based methods, this similarity reflects environmental resemblance between distant locations, allowing the model to account for non-local spatial dependencies and enhancing its ability to identify globally consistent patterns.

Each prediction location x_p is associated with a feature vector:

$$\mathbf{F}(x_p) = \left[\mathbf{F}_{\mathrm{IDW}}(x_p), \mathbf{F}_{\mathrm{SVD}}(x_p), \mathbf{F}_{\mathrm{GOS}}(x_p)\right] \tag{9}$$

where the three feature sets are computed as follows.

2.3.1 | Local Dependence: IDW Feature Extraction

The inverse distance weighting (IDW) method estimates the value of a target point x_p based on the weighted average of its k nearest neighbors x_i, where the weight decreases with distance. The predicted value at x_p is calculated as:

$$f_{\mathrm{IDW}}(x_p) = Z^*(x_p) = \frac{\sum_{i=1}^{k} w_i Z(x_i)}{\sum_{i=1}^{k} w_i} \tag{10}$$

where the weight w_i is defined as:

$$w_i = \frac{1}{d(x_p, x_i)^{\alpha}} \tag{11}$$

Here, $d(x_p, x_i)$ denotes the Euclidean distance between the target point x_p and the neighboring observation x_i, and α is a power parameter that controls the degree of distance decay. The extracted spatial feature $f_{\mathrm{IDW}}(x_p)$ is thus the interpolated value at the target location derived from nearby observations, reflecting the local spatial dependence structure. The extracted IDW-based feature for each point is:

$$\mathbf{F}_{\mathrm{IDW}}(x_p) = [f_{\mathrm{IDW}}(x_p)] \tag{12}$$

2.3.2 | Local Heterogeneity: Spatially Varying Distribution

We refer to the quantile-based characterization of local neighborhoods as the Spatially Varying Distribution (SVD) feature, which captures the local statistical structure of spatial data. The extracted feature vector $\mathbf{F}_{\mathrm{SVD}}(x_p)$ provides a structured representation of local heterogeneity. By incorporating SVD into spatial modeling, we can effectively account for local non-stationarity and spatial variability, making interpolation more adaptive to complex spatial structures.

First, given a target location x_p, we define its neighborhood $\mathcal{N}_k(x_p)$ as the set of its k nearest observed points:

$$\mathcal{N}_k(x_p) = \{x_i | d(x_i, x_p) \text{ is among the } k \text{ smallest distances}\} \tag{13}$$

where the distance $d(x_i, x_p)$ is computed using the Euclidean metric.

Second, for each point x_p, we compute the empirical quantiles of the variable $Z(x)$ within its neighborhood $\mathcal{N}_k(x_p)$:

$$Q_q(x_p) = \mathrm{Quantile}\left(\{Z(x_i) | x_i \in \mathcal{N}_k(x_p)\}, q\right) \tag{14}$$

where $Q_q(x_p)$ represents the qth quantile of the observed values among the k nearest neighbors, with $q \le k$ to ensure valid quantile computation.

Third, the final SVD feature vector for location x_p consists of multiple quantile values across different probability levels:

$$\mathbf{F}_{\mathrm{SVD}}(x_p) = \left[Q_{q_1}(x_p), Q_{q_2}(x_p), \ldots, Q_{q_d}(x_p)\right] \tag{15}$$

where $\{q_1, q_2, \ldots, q_d\}$ represents the predefined quantile levels (e.g., $q \in \{0, 0.05, 0.10, \ldots, 1.0\}$).

2.3.3 | Geosimilarity: GOS Feature Extraction

The Geographically Optimal Similarity (GOS) method is selected to being as the third geofeature to estimate the target variable at unknown locations (Song 2023). GOS identifies spatial configurations with similar structures to make predictions.

First, we define the spatial configuration of each observed location based on extracted features. For each observed location x_i, we define its spatial configuration based on a set of extracted features $\mathbf{F}_{\mathrm{SVD}}(x_p)$.

Second, the similarity between an unknown location x_p and an observed location x_i is defined as:

$$S(x_i, x_p) = P\left\{E_j\left(f_{\mathrm{SVD},j}(x_i), f_{\mathrm{SVD},j}(x_p)\right)\right\} \tag{16}$$

Here, E_j denotes the similarity function for the jth latent feature extracted via SVD, and P is an aggregation function that combines the individual feature-wise similarities into a single similarity score between locations x_i and x_p. Following the design of GOS methods (Zhu et al. 1997; Song 2023), we adopt the minimum operator as the aggregation function P. This conservative aggregation strategy ensures that two locations are considered similar only if they are similar across all dominant features, which has been shown to be effective in spatial similarity modeling (Zhu et al. 2015).

For continuous spatial features, we define E_j as:

$$E_j(x_i, x_p) = \exp\left(-\frac{(F_{\mathrm{SVD},j}(x_i) - F_{\mathrm{SVD},j}(x_p))^2}{2\sigma_j^2}\right) \tag{17}$$

where σ_j represents the standard deviation of feature j.

Third, instead of using all observations for prediction, GOS selects only the most similar locations. The optimal similarity threshold S_λ is determined by minimizing prediction errors:

$$\lambda = \arg\min_{\kappa} \mathrm{RMSE}(\kappa) \tag{18}$$

where κ is the proportion of the most similar samples used for prediction, and RMSE is the root mean square error from cross-validation.

Fourth, the prediction at x_p is computed as:

$$F_{\mathrm{GOS}}(x_p) = \frac{\sum_{i \in \mathcal{N}_\lambda} S_\lambda(x_i, x_p) Z(x_i)}{\sum_{i \in \mathcal{N}_\lambda} S_\lambda(x_i, x_p)} \tag{19}$$

where $\mathcal{N}_\lambda$ is the set of selected observations with similarity above S_λ.

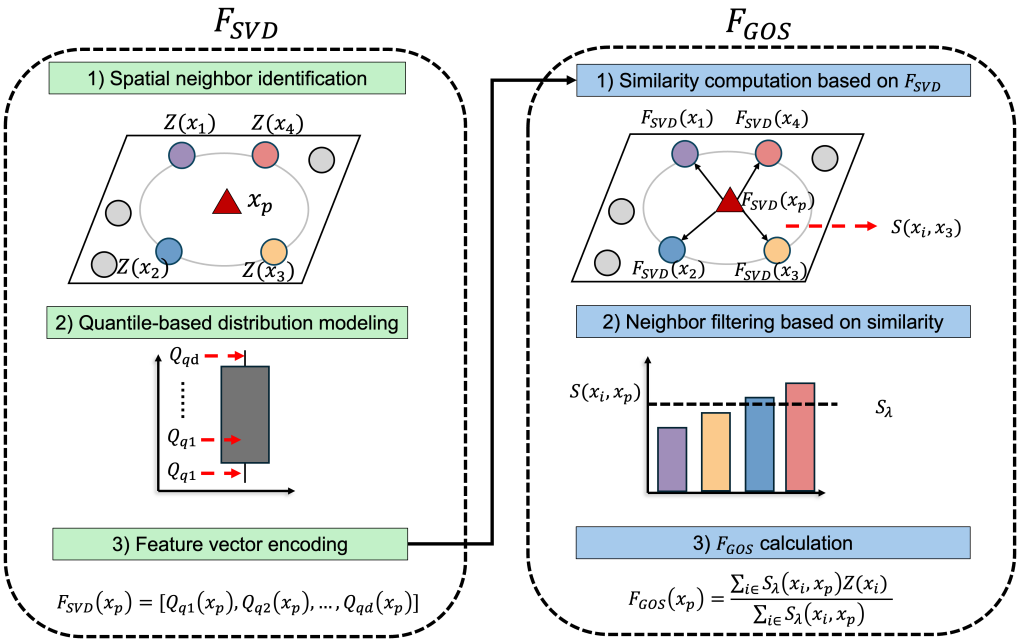

FIGURE 3 | Workflow of the local heterogeneity and geosimilarity features. The left panel captures local heterogeneity via quantile-based distribution modeling, denoted as $F_{\mathrm{SVD}}(x_p)$. The right panel leverages these distributional features to perform geosimilarity-based estimation, denoted as $F_{\mathrm{GOS}}(x_p)$, through similarity matching and weighted aggregation.

Figure 3 illustrates the computational workflows of $F_{\mathrm{SVD}}(x_p)$ and $F_{\mathrm{GOS}}(x_p)$. The F_{SVD} module (Left panel) captures local heterogeneity by extracting a set of quantiles from the attribute values within the neighborhood $\mathcal{N}_k(x_p)$. Each sample point is subsequently represented by a feature vector composed of these quantiles, thereby encoding its local statistical distribution. Right panel demonstrates the estimation of $F_{\mathrm{GOS}}(x_p)$, which builds upon the pre-computed F_{SVD} feature vectors. By evaluating the similarity between the target feature vector and those of its neighbors, the method identifies a subset of the most similar candidates. The final estimate $F_{\mathrm{GOS}}(x_p)$ is obtained through a similarity-weighted aggregation of the selected neighbors.

2.4 | Regression Kriging With Geofeatures

Combining all three components, we define the final spatial feature vector:

$$\mathbf{F}(x_p) = \left[F_{\mathrm{IDW}}(x_p), F_{\mathrm{SVD}}(x_p), F_{\mathrm{GOS}}(x_p)\right] \tag{20}$$

where: $F_{\mathrm{IDW}}(x_p)$ captures local dependence. $\mathbf{F}_{\mathrm{SVD}}(x_p)$ captures local heterogeneity using statistical quantiles. $F_{\mathrm{GOS}}(x_p)$ captures global similarity based on regression over spatially similar regions.

The final feature vector has a total of:

$$\dim(\mathbf{F}) = 1 + d_{\mathrm{SVD}} + 1 \tag{21}$$

where: 1 represents the IDW feature. d_{SVD} represents the number of quantile-based SVD features (which can be adjusted based on resolution). 1 represents the GOS similarity feature.

A machine learning regression model $g(\mathbf{F})$ is trained on known observations, which captures the large-scale trend:

$$Z_{\mathrm{trend}}(x) = g(\mathbf{F}(x)) \tag{22}$$

where $Z_{\mathrm{trend}}(x)$ is the trend value at the location x.

The final step is to interpolate the residuals and combine them with the predicted trend using ordinary kriging.

3 | Case Study: Mapping Trace Elements With FFRK

3.1 | Study Area and Data

In this work, we use trace element data of Cu, Pb, and Zn (measured in parts per million, ppm), from one region of Australia to test the performance of FFRK. We concentrated on the geographic variability of the three elements for the trace element distribution because these elements are well known to be important markers of environmental contamination and are essential for evaluating ecological health. The spatial distributions of trace elements are shown in Figure 4.

We included nine environmental explanatory variables in alphabetical order to examine the factors influencing their spatial patterns: slope, road network density (Road), normalized difference vegetation index (NDVI), distance to major roads (MainRd), distance to lithology (Dlith), distance to fault lines (Dfault), soil organic carbon (SOC), soil pH levels, and water distribution. These variables were selected for their relevance in characterizing and predicting the distribution of trace elements. The calculation of the variables, such as Dlith, Dfault, is referenced in (Song 2023). Table 1 summarizes the descriptive statistics of

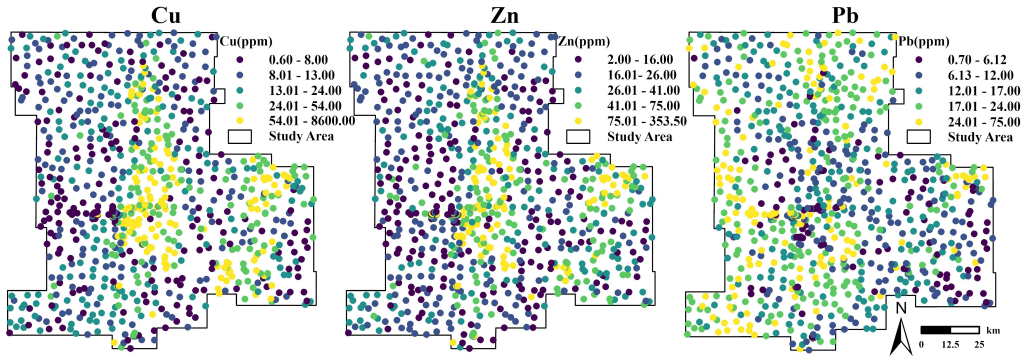

FIGURE 4 | Spatial distributions of trace elements: Cu, Zn, and Pb.

TABLE 1 | Descriptive statistics of environmental variables.

Variable	Code	Mean	Min	Median	Max	SD	CV
Distance to water (km)	Water	1.114	0.000	0.648	7.696	1.188	1.067
Distance to mainroads (km)	MainRoads	19.991	0.010	17.485	58.371	14.219	0.711
Distance to roads (km)	Roads	9.817	0.002	7.953	50.141	8.363	0.852
Distance to mine (km)	MineKm	14.785	0.025	11.425	56.638	12.117	0.820
Slope (degree)	Slope	0.275	0.008	0.246	1.712	0.161	0.585
Normalized difference vegetation index	NDVI	0.178	0.062	0.180	0.251	0.024	0.135
Soil organic carbon	SOC	0.868	0.686	0.870	1.066	0.053	0.061
Soil pH	pH	5.741	5.113	5.753	6.178	0.178	0.031
Distance to lithology (km) for Cu	$Dlith_{Cu}$	9.132	0.000	7.147	39.839	7.769	0.851
Distance to lithology (km) for Zn	$Dlith_{Zn}$	8.150	0.000	6.229	39.839	7.637	0.937
Distance to lithology (km) for Pb	$Dlith_{Pb}$	3.363	0.000	2.627	15.946	3.020	0.898
Distance to fault (km) for Cu	$Dfault_{Cu}$	16.070	0.001	13.285	54.765	12.150	0.756
Distance to fault (km) for Pb	$Dfault_{Pb}$	12.017	0.003	11.111	43.471	8.090	0.673
Distance to fault (km) for Zn	$Dfault_{Zn}$	14.174	0.001	11.844	52.697	10.851	0.766
Elevation (m)	Elevation	482.699	398.050	485.128	588.649	36.700	0.076
Aspect (degree)	Aspect	171.645	0.751	174.126	358.712	90.576	0.528

all explanatory variables, including the mean (Mean), minimum (Min), median, maximum (Max), standard deviation (SD), and coefficient of variation (CV).

3.2 | Experiment Design

Figure 5 illustrates the experimental design of this study, which aims to validate the reliability of the proposed FFRK method. The prediction target is the ln-transformed concentrations of heavy metals (Cu, Pb, and Zn, measured in ppm). For the target variable, we computed its three-dimensional geofeatures, including dependence, heterogeneity, and similarity. Subsequently, we selected four types of machine learning models—linear model (LM), decision tree (DT), random forest (RF), and support vector machine (SVM)—to predict the trend of the target variable. After obtaining the residuals by comparing the predicted trend with observations, we applied ordinary kriging for spatial interpolation of the residuals. Finally, the predicted trend and interpolated residuals were combined to produce the prediction.

We compare our proposed method with several representative baselines, including classical geostatistical approaches, machine learning models, and their stratified and hybrid variants. First, Ordinary Kriging serves as the foundational geostatistical method and is used as a baseline for performance comparison. Second, we consider machine learning models, which directly establish predictive relationships between explanatory variables and the target variable. In this study, we employ four commonly used models: Linear Regression (LM), Decision Tree Regression (DT), Random Forest Regression (RF), and Support Vector Machine Regression (SVM).

Third, stratified models are introduced to better account for spatial heterogeneity. These models partition the study area into multiple subregions based on homogeneity analysis, with separate machine learning models trained within each region. The same four algorithms (LM, DT, RF, and SVM) are applied in each subregion. The spatial partitioning is conducted using a decision tree, which segments the domain based on geographic coordinates (longitude and latitude) as well as trace element values. In

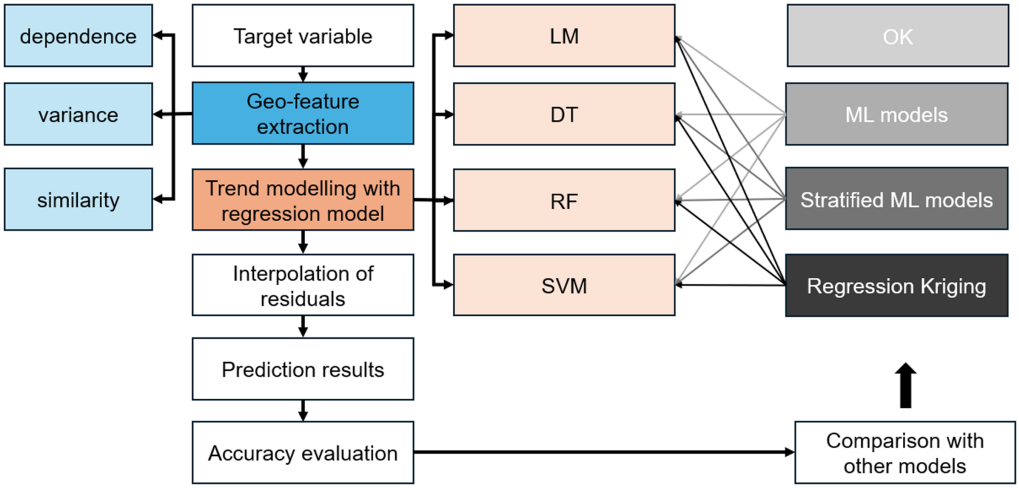

FIGURE 5 | Experimental design for evaluating the performance of FFRK in comparison with other interpolation models.

addition to stratified machine learning models, we also include stratified kriging, resulting in a total of five stratified models.

Fourth, we evaluate RK, a hybrid method that combines trend modeling and spatial interpolation. Specifically, a machine learning model (LM, DT, RF, or SVM) is used to estimate the trend, and the residuals are then interpolated using kriging. Finally, we present our proposed method, FFRK. Similar to RK, FFRK follows a two-step structure but incorporates a feature selection process prior to regression. The four variants of FFRK correspond to the four regression models used: FFRK (LM), FFRK (DT), FFRK (RF), and FFRK (SVM).

In total, we compared 18 models. Among them, only Ordinary Kriging, Stratified Kriging, and four FFRK models do not require any explanatory variables, whereas the other three model categories require nine explanatory variables as inputs.

In the FFRK model, two hyperparameters are involved. The first is K, which defines the number of neighboring sample points used to compute geo-features and fit the semivariogram. The second is the quantile interval, used in the second geo-feature to characterize the statistical information within the neighborhood of each sample point. In this study, we set $K = 15$ and the quantile interval = 5%, following the settings adopted in a previous study (Song 2023). A sensitivity analysis of these two hyperparameters and their impact on the performance of FFRK is presented in Section 3.4.2.

3.3 | Cross-Validation and Model Evaluation

In this study, we adopted 10-fold cross-validation to evaluate the prediction performance and generalization capability of the proposed FFRK method and baseline models. Specifically, the dataset was randomly divided into ten subsets of equal size. In each iteration, nine subsets served as the training set, while the remaining subset was reserved as the validation set. This training-validation procedure was repeated ten times, with the averaged results across the ten iterations representing the final model performance. This approach effectively mitigates the instability arising from a single random partition and enhances the robustness of model evaluation.

In particular, for Stratified Models, we employed a regression tree to partition the study area spatially, and subsequently performed stratified sampling based on the distribution of the target variable. This ensured consistent internal data distribution across each of the ten folds constructed for cross-validation.

Moreover, the cross-validation process was utilized for model parameter optimization. Optimal parameters for each machine learning model were determined by minimizing the mean RMSE obtained from cross-validation. Finally, each model was retrained once using the complete dataset to derive globally optimal parameters, and the resulting optimized models were preserved for subsequent global spatial predictions.

To further evaluate the generalization performance of the proposed FFRK, we conducted an additional 80/20 train-test split experiment. The results are presented in Appendix A.

3.4 | Results

3.4.1 | Accuracy Evaluation

Figure 6 presents the spatial interpolation results of FFRK together with two stratified models, Stratified RK and STK. We chose LM as the representative machine learning regression method. For other categories of ML prediction results, readers may refer to (Song 2023). Our results show that the STK method produces predictions with pronounced artificial boundaries and the least local spatial detail. Stratified RK provides richer spatial details, though boundary effects remain, albeit less prominently than in STK. In contrast, FFRK eliminates artificial boundary issues while still offering sufficient spatial detail. The subsequent accuracy assessment further demonstrates that FFRK also achieves the best predictive performance.

Figure 7 presents the accuracy results of 13 global modeling approaches for predicting the concentrations of three heavy metals, evaluated using Rš, RMSE, and MAE. Among these

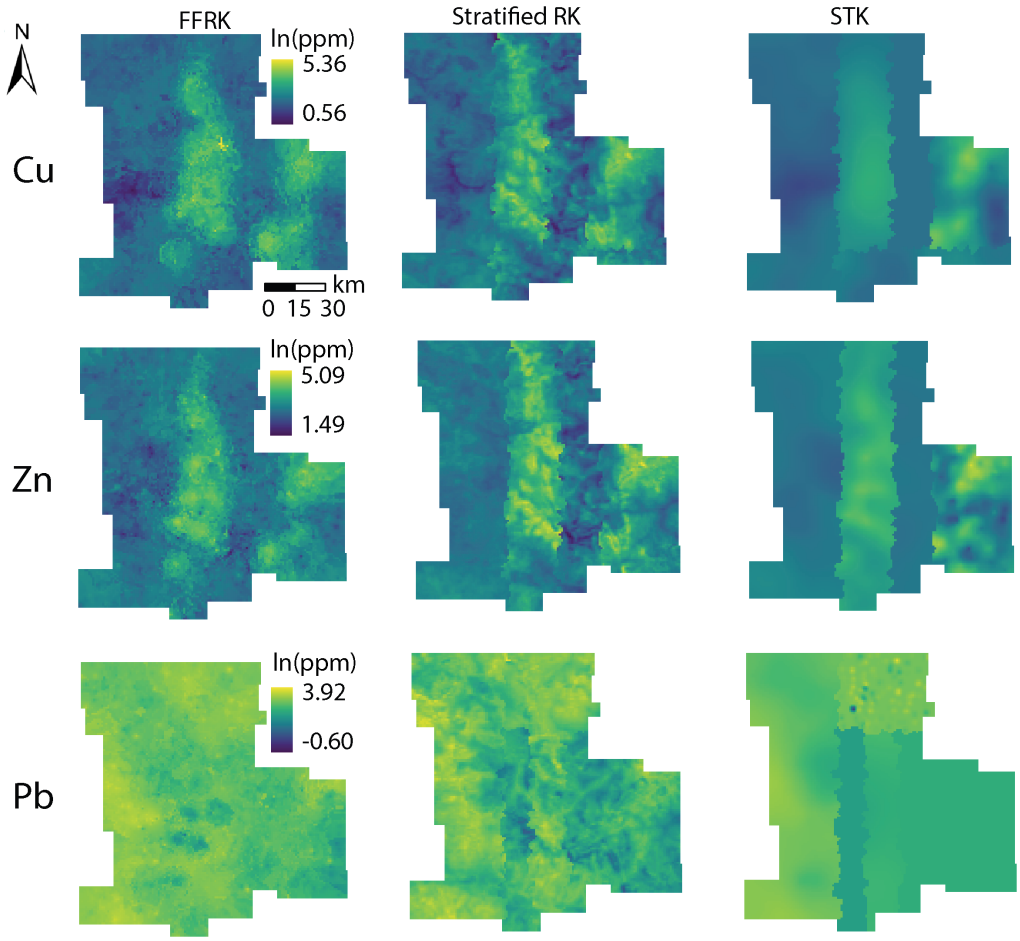

FIGURE 6 | Spatial distributions of the predicted concentrations of three trace elements (Cu, Zn, and Pb) obtained from FFRK, Stratified RK, and STK.

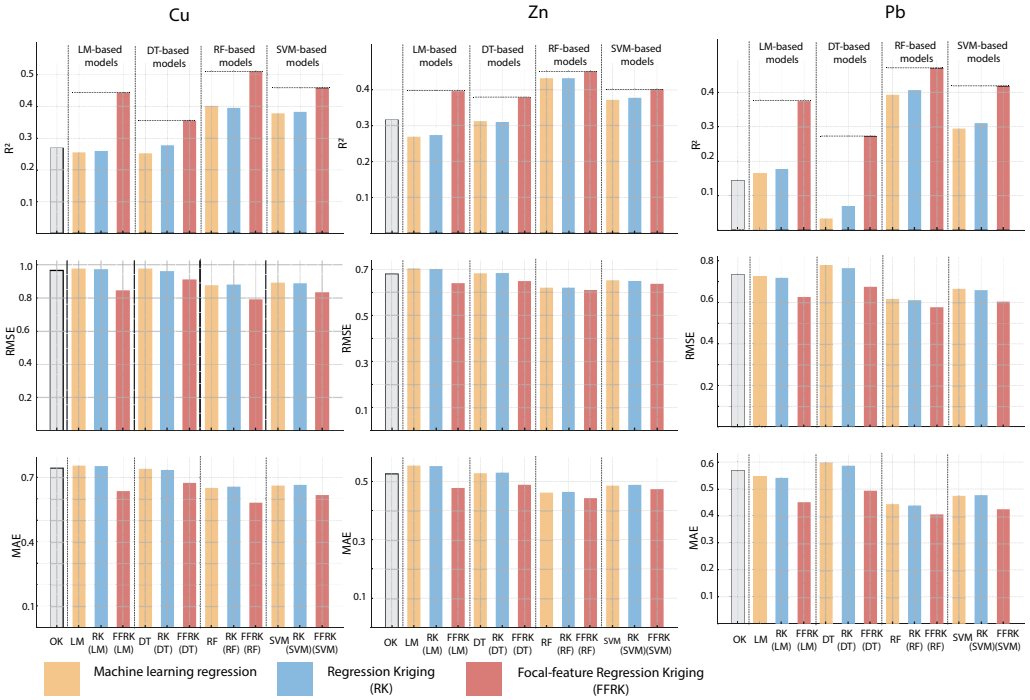

FIGURE 7 | Evaluation results (R^2, $RMSE$, and MAE) for the 13 global models.

TABLE 2 | Mean R^2, MAE, and RMSE for different models and elements ($K = 2$ to 10).

Element	Model	Mean R^2	Mean MAE	Mean RMSE
Cu	Stratified_Kriging	0.3584	0.6720	0.8964
	Stratified_LM	0.3738	0.6699	0.8978
	Stratified_DT	0.3009	0.7114	0.9486
	Stratified_RF	0.4115	0.6406	0.8705
	Stratified_SVM	0.4205	0.6318	0.8637
Zn	Stratified_Kriging	0.3791	0.4830	0.6287
	Stratified_LM	−6.4581	0.5559	1.5290
	Stratified_DT	0.3546	0.5054	0.6613
	Stratified_RF	0.4367	0.4575	0.6179
	Stratified_SVM	0.4147	0.4725	0.6298
Pb	Stratified_Kriging	0.2747	0.5277	0.6958
	Stratified_LM	0.1043	0.5040	0.7462
	Stratified_DT	0.2178	0.5296	0.7041
	Stratified_RF	0.4063	0.4399	0.6137
	Stratified_SVM	0.3776	0.4474	0.6283

methods, 12 involve the use of ML models, with the exception of ordinary kriging. To ensure a fair comparison, we categorized the methods into four groups based on the ML model used for trend modeling: LM-based, DT-based, RF-based, and SVM-based.

The results indicate that the FFRK method, despite not incorporating any explanatory variables, achieves significantly higher accuracy than both standalone ML models and regression models based on ML modeling. The latter two approaches utilized nine explanatory variables and a large dataset but failed to yield higher prediction accuracy. For example, in Cu prediction, the Rš values of FFRK with LM-, DT-, RF-, and SVM-based trend modeling exceeded those of regression kriging using the same ML models by 73.72%, 40.31%, 26.87%, and 21.27%, respectively. The most significant improvement was observed in Pb prediction, where FFRK with DT-based trend modeling achieved an Rš of 0.27, surpassing regression kriging and the DT model by 0.07 and 0.03, respectively.

In Table 2, we present the results of five stratified modeling approaches. First, we partitioned the spatial domain into multiple homogeneous subregions based on the values of explanatory variables. Within each subregion, we decomposed and fitted separate models to perform spatial predictions. Finally, we evaluated the overall prediction accuracy across the entire study area. To assess the impact of stratification, we experimented with different numbers of partitions, ranging from 2 to 10, and calculated the average accuracy. The results indicate that compared to global modeling approaches, including OK and various machine learning models, stratified modeling can effectively improve prediction accuracy, as discussed in the Introduction. However, its performance still falls short of that achieved by the FFRK method. In some cases, stratification resulted in over-segmentation, where certain subregions contained too few observations, leading to model underfitting and significantly reduced predictive performance. Additionally, stratified modeling introduced abrupt and unrealistic discontinuities in the final spatial interpolation results, which arose due to artificial boundaries between partitions.

3.4.2 | Sensitivity Analysis for the Hyperparameter in FFRK

In Kriging-based methods, an hyperparameter is the number of neighboring points used when fitting the semivariogram and performing interpolation. In the previous experiments, to ensure a fair comparison among different methods, we set the number of nearest neighbors (k) to 15 for all models. In this section, we assess the sensitivity of different methods to k. Figure 8 presents the accuracy of FFRK and RK models under varying numbers of neighboring points.

The results indicate that for the same type of ML model used in trend modeling, FFRK (dashed line) consistently outperforms RK (solid line) across most k values, demonstrating higher Rš and lower RMSE. However, there are a few exceptions. For instance, when k is less than 11 in Zn prediction, the RF-based RK model achieves higher accuracy than the RF-based FFRK model. Despite this, FFRK based on LM and DT still significantly outperforms RK when k is below 11. This discrepancy may be attributed to the distribution characteristics of Zn: when k is small, the geofeatures used by FFRK may not provide sufficient information to properly fit the RF model. In contrast, the RK method, which utilizes nine explanatory variables, has enough information for RF to achieve better fitting. However, for the relatively simpler LM and DT models, the geofeatures in FFRK provide sufficient information, enabling FFRK to surpass RK in predictive performance.

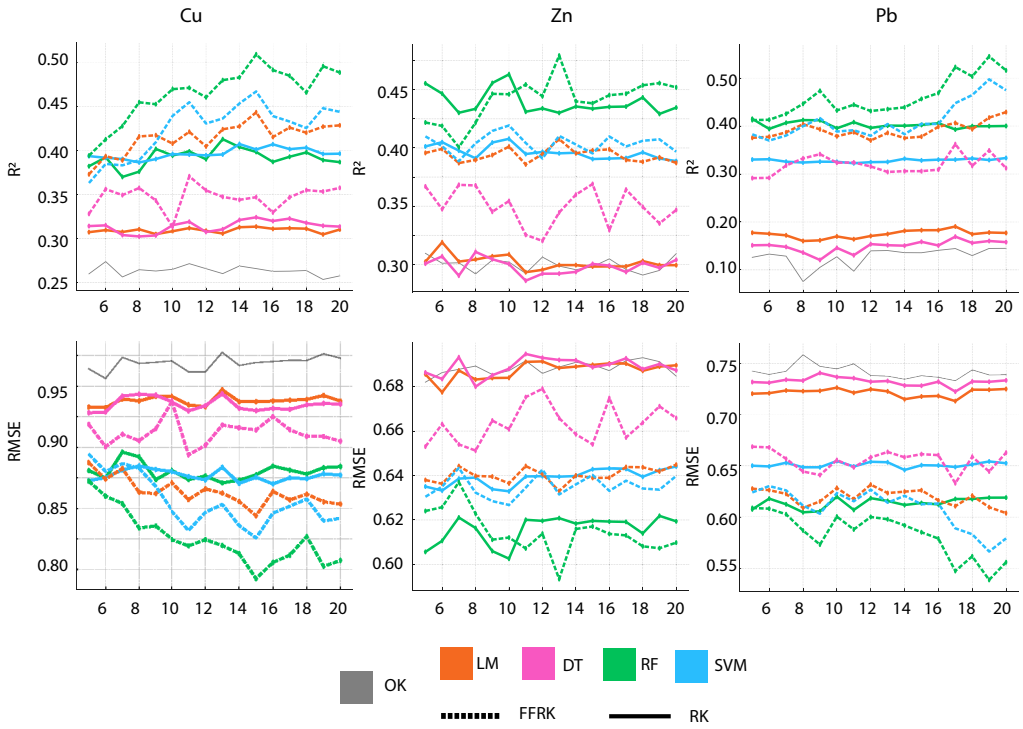

FIGURE 8 | The sensitivity analysis for the *k* value.

In the second category of geofeatures, we sample the value distribution of surrounding sample points for each predicted location. This sampling is based on fixed quantile intervals, such as 1% or 5%. In the previous experiments, we selected a 5% quantile interval, resulting in 20 features in this geofeature category. To assess the sensitivity of FFRK to this parameter, we varied the quantile interval from 5% to 50% and evaluated its impact on the results. Figure 9 shows that the FFRK method is relatively robust to changes in the quantile sampling interval. The maximum fluctuation in Rš is approximately 0.05, while the highest variation in RMSE is only around 0.02.

3.4.3 | Generalized Regression Kriging

In previous experiments, we have demonstrated that FFRK, despite not requiring any explanatory variables, can outperform RF models that utilize multiple explanatory variables solely through the extraction of geofeatures. Here, we propose a new question: if geofeatures and explanatory variables are integrated together for trend modeling, can the predictive performance be further improved? We define this approach, which incorporates both geofeatures and explanatory variables, as generalized regression kriging (GRK).

The results, shown in Figure 10, indicate that across three different datasets and 12 prediction tasks (based on four ML models for trend modeling), GRK outperforms FFRK in 10 cases and significantly surpasses RK in all cases. The two exceptions occur in Cu and Zn prediction tasks when DT is used for trend modeling, where FFRK, relying solely on geofeatures, achieves a slightly higher Rš than GRK, which includes explanatory variables.

These findings suggest that incorporating more information into trend modeling enhances the performance of regression kriging models. The proposed GRK model is particularly suitable for spatial interpolation tasks where explanatory variables are available. Compared to conventional regression kriging, which relies solely on explanatory variables, integrating geofeatures into trend modeling leads to substantial improvements in predictive accuracy.

4 | Discussion

Regression Kriging is a highly effective spatial interpolation model for addressing second-order spatial non-stationarity, which fits a trend surface based on explanatory variables and then performs kriging on the residuals. However, in many spatial interpolation tasks, it is often difficult to obtain explanatory variables, or to find variables that are sufficiently suitable, which greatly limits the applicability of regression kriging.

This work proposes a novel interpolation method for regression kriging under spatial non-stationarity without relying on explanatory variables, called FFRK. FFRK substitutes the traditional need for explanatory variables in trend surface fitting by creating features—termed geofeatures—based on the spatial distribution patterns of geographic variables. In this study, we introduce three categories of geofeatures, derived respectively from spatial dependence, spatial heterogeneity, and geographic similarity.

We apply FFRK to a case study involving the spatial distribution of three heavy metal concentrations in a region of Australia. We also compare the performance of FFRK against 17 other interpolation methods of various types, demonstrating that FFRK achieves the highest predictive accuracy. In Appendix B, we present a collinearity analysis between the

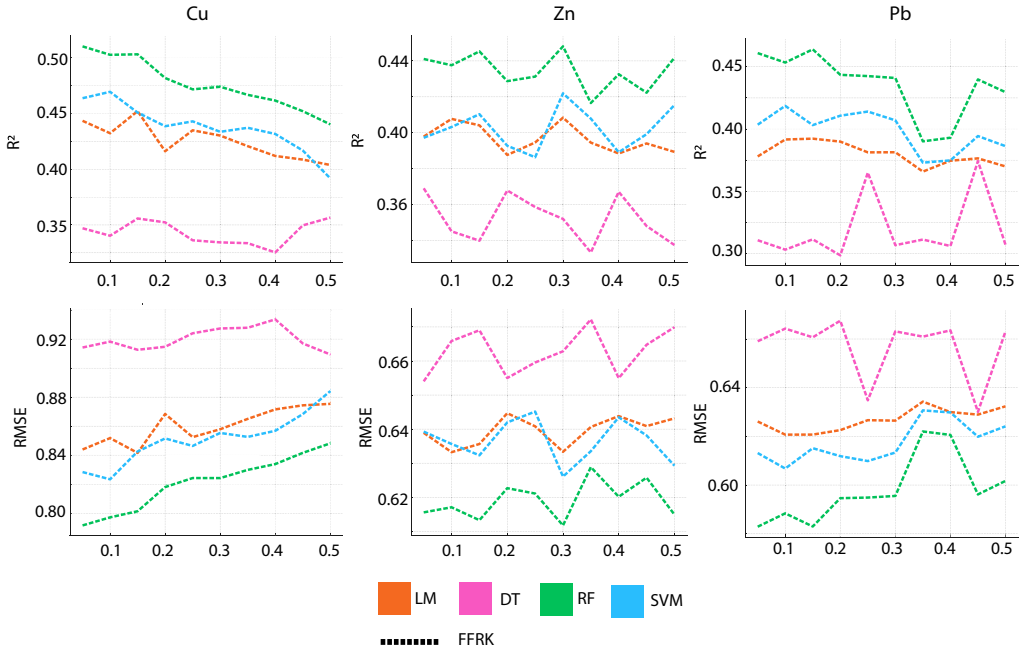

FIGURE 9 | The sensitivity analysis for the quantile step.

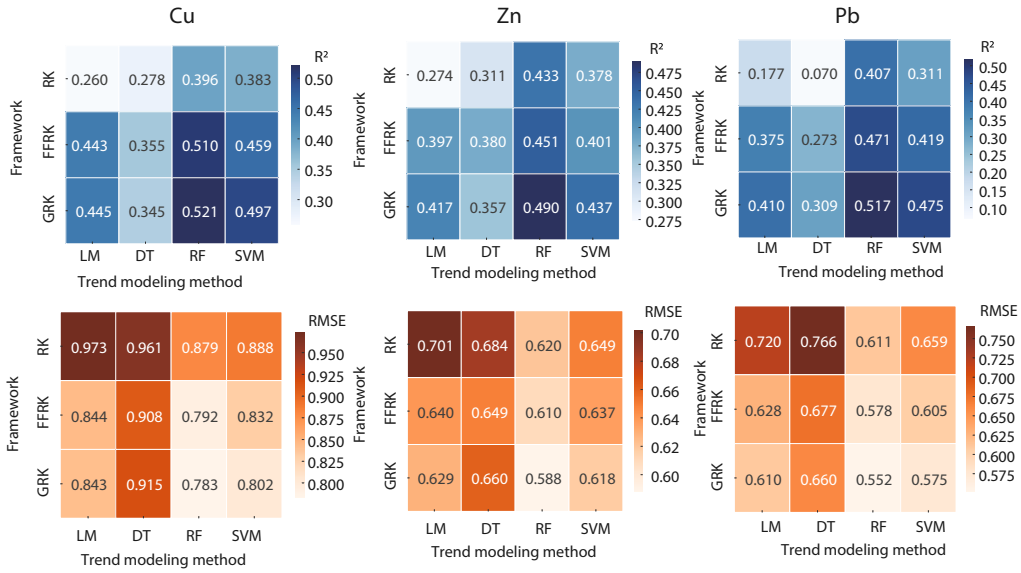

FIGURE 10 | Accuracy of three frameworks: regression kriging, FFRK, and generalized regression kriging (GRK).

constructed geofeatures and the target variable (i.e., heavy metal concentrations), including Pearson correlation and variance inflation factor (VIF). The results demonstrate that the derived features provide informative yet non-redundant signals, as evidenced by their relatively low VIF values. This highlights the value of spatial feature engineering in enhancing the predictive performance of kriging-based models.

FFRK achieves superior performance without using any explanatory variables, outperforming other methods that do use explanatory variables, including machine learning models and regression kriging. This showcases the advantage of using features derived from the spatial distribution itself in regression kriging. However, the advantage does not imply that such spatial distribution features can entirely replace information from traditional explanatory variables. For the specific case study in this paper-predicting heavy metal concentrations-it is particularly difficult to identify suitable, large-scale, readily available explanatory variables. In most spatial prediction tasks, large-scale remote sensing observations are typically relied upon. However, subsurface distributions are often poorly represented by surface-level features detectable through remote sensing. This is precisely the application scenario that FFRK aims to address. As shown in Figure 10, when both explanatory variables and geofeatures are used together (a model we refer to as GRK), the prediction accuracy generally improves further compared to FFRK alone. In addition, to assess the individual contribution of different geofeatures, we conducted an ablation study by removing or combining

feature types. As shown in Appendix D, the full-feature FFRK consistently achieved the best predictive performance across metals. We also examined the necessity of the regression kriging component in FFRK through an additional ablation analysis. Specifically, we compared FFRK with its kriging-free counterpart, Focal Feature Regression (FFR), which directly uses the extracted geofeatures for prediction without residual interpolation. The detailed results are reported in Appendix E.

FFRK is a general spatial feature engineering framework embedded within RK, allowing users to design task-specific geospatial features based on their domain knowledge and research goals. In this study, we demonstrate one implementation using three types of features-local dependence, local heterogeneity, and geosimilarity. While these features have shown consistent effectiveness in the context of heavy metal concentration prediction. However, the construction of local heterogeneity and geosimilarity features requires a sufficient number of neighboring observations (e.g., more than 10) around the target location. When this condition is not met, the resulting quantile-based descriptors may lack the statistical robustness needed for accurate prediction. In future work, we plan to develop adaptive spatial feature extraction strategies that dynamically determine the neighborhood size and composition based on local data density and spatial context.

5 | Conclusion

We propose FFRK, a generalizable regression kriging framework that mitigates the need for high-resolution covariates by engineering spatial features directly from the target variable's spatial distribution. Addressing key limitations in existing methods—such as reliance on external variables and assumptions of spatial stationarity—FFRK captures spatial dependence, local heterogeneity, and geographic similarity to construct a robust trend surface. We demonstrate the effectiveness of FFRK through a spatial prediction task involving the distribution of three heavy metal concentrations in a region of Australia. FFRK outperforms traditional kriging, machine learning, and stratified methods, offering a practical solution for large-scale, non-stationary spatial interpolation tasks where explanatory data are scarce or unavailable. This work contributes to GIScience by demonstrating that spatial features alone can effectively support predictive modeling. However, FFRK's performance depends on the spatial density of observations. Future work will explore its extension to spatiotemporal settings and automated feature learning for broader applicability.

Acknowledgements

Open access publishing facilitated by Curtin University, as part of the Wiley - Curtin University agreement via the Council of Australasian University Librarians.

Disclosure

No conflict of interest exists in this manuscript, and the manuscript was approved by all authors for publication. The authors employed the OpenAI ChatGPT-4 model exclusively as a language-editing aide to enhance readability, grammar, and stylistic clarity. ChatGPT-4 was not used to create, expand, or alter the scientific ideas, data analyses, or conclusions presented herein.

Data Availability Statement

The data and codes that support the findings of the present study are available on https://github.com/pengtum/FFRK.

References

Anselin, L. 2013. *Spatial Econometrics: Methods and Models*. Vol. 4. Springer Science & Business Media.

Campbell, J. B., and R. H. Wynne. 2011. *Introduction to Remote Sensing*. Guilford press.

Cantrell, R. S., and C. Cosner. 1991. "The Effects of Spatial Heterogeneity in Population Dynamics." *Journal of Mathematical Biology* 29: 315–338.

Cheng, S., W. Zhang, P. Luo, L. Wang, and F. Lu. 2024. "An Explainable Spatial Interpolation Method Considering Spatial Stratified Heterogeneity." *International Journal of Geographical Information Science* 39: 1–27.

Clark, I., and W. Harper. 1979. *Practical Geostatistics, Volume 3*. Applied Science Publishers London.

Cressie, N. 1988. "Spatial Prediction and Ordinary Kriging." *Mathematical Geology* 20: 405–421.

Gao, B., M. Hu, J. Wang, et al. 2020. "Spatial Interpolation of Marine Environment Data Using p-Msn." *International Journal of Geographical Information Science* 34: 577–603.

Goodchild, M. F. 2004. "Giscience, Geography, Form, and Process." *Annals of the Association of American Geographers* 94: 709–714.

Griffith, D. A. 2003. "Spatial Filtering." In *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization*, 91–130. Springer.

Hengl, T., G. B. Heuvelink, and D. G. Rossiter. 2007. "About Regression-Kriging: From Equations to Case Studies." *Computers & Geosciences* 33: 1301–1315.

Hengl, T., G. B. Heuvelink, and A. Stein. 2004. "A Generic Framework for Spatial Prediction of Soil Variables Based on Regression-Kriging." *Geoderma* 120: 75–93.

Jiao, L., P. Luo, R. Huang, et al. 2025. "Modeling Hydrous Mineral Distribution on Mars With Extremely Sparse Data: A Multi-Scale Spatial Association Modeling Framework." *ISPRS Journal of Photogrammetry and Remote Sensing* 222: 16–32.

Lam, N. S. N. 1983. "Spatial Interpolation Methods: A Review." *American Cartographer* 10: 129–150.

Liu, Y., Y. Chen, Z. Wu, B. Wang, and S. Wang. 2021. "Geographical Detector-Based Stratified Regression Kriging Strategy for Mapping Soil Organic Carbon With High Spatial Heterogeneity." *Catena* 196: 104953.

Lou, X., P. Luo, Z. Li, S. Gao, and L. Meng. 2025a. "Geoxcp: Uncertainty Quantification of Spatial Explanations in Explainable Ai." *International Journal of Geographical Information Science*: 1–31.

Lou, X., P. Luo, and L. Meng. 2025b. "Geoconformal Prediction: A Model-Agnostic Framework for Measuring the Uncertainty of Spatial Prediction." *Annals of the American Association of Geographers* 115, no. 8: 1–1998.

Luo, P., Y. Li, Y. Song, Z. Li, and L. Meng. 2025. "Measuring Univariate Effects in the Interaction of Geographical Patterns." *International Journal of Geographical Information Science* 40, no. 2: 1–413.

Luo, P., X. Lou, Y. Zheng, Z. Zheng, and S. Ermon. 2025. "Geoevolve: Automating Geospatial Model Discovery via Multi-Agent Large Language Models." *arXiv Preprint arXiv:2509.21593*.

Luo, P., Y. Song, X. Huang, et al. 2022. "Identifying Determinants of Spatio-Temporal Disparities in Soil Moisture of the Northern Hemisphere Using a Geographically Optimal Zones-Based Heterogeneity Model." *ISPRS Journal of Photogrammetry and Remote Sensing* 185: 111–128.

Luo, P., Y. Song, D. Zhu, J. Cheng, and L. Meng. 2023. "A Generalized Heterogeneity Model for Spatial Interpolation." *International Journal of Geographical Information Science* 37: 634–659.

Matheron, G. 1963. "Principles of Geostatistics." *Economic Geology* 58: 1246–1266.

Panigrahi, N. 2021. "Inverse Distance Weight." In *Encyclopedia of Mathematical Geosciences*, 1–7. Springer.

Rey, S., D. Arribas-Bel, and L. J. Wolf. 2023. *Geographic Data Science With Python*. Chapman and Hall/CRC.

Song, Y. 2023. "Geographically Optimal Similarity." *Mathematical Geosciences* 55: 295–320.

Stein, A., and L. Corsten. 1991. "Universal Kriging and Cokriging as a Regression Procedure." *Biometrics* 47: 575–587.

Tiefelsdorf, M., and D. A. Griffith. 2007. "Semiparametric Filtering of Spatial Autocorrelation: The Eigenvector Approach." *Environment and Planning A* 39: 1193–1221.

Tomczak, M. 1998. "Spatial Interpolation and Its Uncertainty Using Automated Anisotropic Inverse Distance Weighting (Idw)-cross-Validation/Jackknife Approach." *Journal of Geographic Information and Decision Analysis* 2: 18–30.

Webster, R., and M. A. Oliver. 2007. *Geostatistics for Environmental Scientists*. John Wiley & Sons.

Wilson, A. G., Z. Hu, R. Salakhutdinov, and E. P. Xing. 2016. "Deep Kernel Learning." In *Artificial Intelligence and Statistics*, 370–378. PMLR.

Yao, Y., A. Dong, Z. Liu, et al. 2023. "Extracting the Pickpocketing Information Implied in the Built Environment by Treating It as the Anomalies." *Cities* 143: 104575.

Zhu, A., J. Liu, F. Du, et al. 2015. "Predictive Soil Mapping With Limited Sample Data." *European Journal of Soil Science* 66: 535–547.

Zhu, A. X., L. Band, R. Vertessy, and B. Dutton. 1997. "Derivation of Soil Properties Using a Soil Land Inference Model (Solim)." *Soil Science Society of America Journal* 61: 523–533.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** Supporting Information.

Appendix A

Generalization Evaluation with 80/20 Split

In the Section 3, we presented model evaluation results based on 10-fold cross-validation. To further assess the generalization capability of the proposed FFRK method in real-world applications, we conducted an additional 80/20 holdout experiment, in which the entire dataset was randomly split into 80% for training and 20% as an independent test set. Within the training set, we performed cross-validation to tune hyperparameters and identify the optimal configurations for each machine learning model. The final models were then fitted on the full training set and evaluated on the independent test set. As shown in Supporting Information: Figure S1, FFRK consistently outperformed other approaches on the test set, achieving the lowest errors and highest Rš scores for most metals. In many cases, the predictive accuracy was substantially improved. The only exceptions were in Cu prediction using RF and DT models, where the original machine learning models and RK-based approaches slightly outperformed FFRK.

Appendix B

Multicollinearity Analysis of Spatial Features

In this study, we constructed several neighborhood-based spatial statistical features derived from the distribution of the target variable within local neighborhoods. This design raises potential concerns about collinearity or redundancy between the derived features and the response variable. To assess the validity and stability of these spatial features, we conducted a collinearity analysis. Specifically, we computed the Pearson correlation coefficients between each feature and the target variable (i.e., heavy metal concentrations), as well as the variance inflation factors (VIF) to evaluate multicollinearity among predictors. As shown in Supporting Information: Figure S2, most spatial features exhibit moderate correlations with the target variable, typically in the range of 0.3 to 0.6, suggesting that they contribute relevant but non-redundant information. All VIF values are below 2, indicating a low risk of multicollinearity. Moreover, the neighborhood information used to compute these features is entirely drawn from other training samples, and the target value of the prediction location is never used in feature construction. As such, this modeling approach does not involve data leakage and remains applicable to real-world prediction scenarios.

Appendix C

Multicollinearity Analysis of Covariates

To ensure the fairness and rigor of the comparative experiments, a systematic validity analysis was conducted on the nine selected external covariates. This analysis aims to verify whether these environmental factors provide substantial explanatory power for the benchmark models-such as Regression Kriging and Machine Learning-rather than introducing redundant information.

A multi-dimensional statistical testing framework was employed for this evaluation. First, the VIF was utilized to assess multicollinearity, ensuring the independence of the input features. Second, the contribution of each feature was quantified using both Linear Model and Random Forest: the LM estimates weight coefficients via ordinary least squares and tests statistical significance through P-values; meanwhile, the RF employs an ensemble learning mechanism to measure feature importance by calculating the Increase in Mean Squared Error (IncMSE) after permuting specific features, thereby capturing their roles in complex non-linear relationships.

The results, as summarized in Supporting Information: Table S1, demonstrate that all selected external covariates performed robustly across the prediction tasks for the three elements. First, the VIF values for all variables ranged from 1.07 to 3.78, remaining well below the conservative threshold of 4. This indicates the absence of severe multicollinearity and confirms the independence and non-redundancy of the selected environmental features. Regarding statistical significance, the majority of variables exhibited high explanatory power within the linear framework ($P < 0.05$). For instance, in Cu prediction, Dlith ($P < 0.001$) and NDVI ($P < 0.001$) showed highly significant linear correlations. Similarly, for Pb and Zn, factors such as SOC and pH maintained consistently high significance across all tests. Notably, for the few variables that did not reach statistical significance in the linear models-such as Road density in Cu prediction ($P = 0.1679$) or Dfault in Zn prediction ($P = 0.2348$)-they nonetheless displayed substantial contributions in the Random Forest model, with IncMSE values of 15.27 and 30.86, respectively. This disparity suggests that these factors may influence heavy metal distribution through complex non-linear mechanisms that LM fails to capture, yet RF effectively utilizes. In conclusion, this multi-dimensional analysis confirms that the benchmark models (RK and ML) are operating under high-quality, high-relevance feature inputs. This evidence provides a rigorous baseline for comparison, further highlighting the architectural advantage of the proposed FFRK, which achieves superior accuracy without relying on these external environmental covariates.

Appendix D

Ablation Study on the Contribution of Spatial Feature Types

In our study, we developed the FFRK framework for spatial prediction of heavy metal concentrations by incorporating three types of spatial features: local dependence, local heterogeneity, and geo-similarity. To assess the individual contribution of each feature type, we conducted an ablation study involving four combinations of spatial features.

The results are presented in Figure S3. Across all three metals (Cu, Zn, and Pb), the full-feature FFRK consistently achieved the best performance in both RMSE and Rš metrics (see comparison with Figure 10), suggesting that the three spatial components offer complementary information and jointly enhance model generalization. Notably, we also observed that local dependence alone already provides strong predictive power and often outperforms combinations of local heterogeneity and geo-similarity. However, in more complex machine learning models such as RF, relying solely on local dependence leads to a marked drop in performance, indicating the necessity of incorporating higher-dimensional spatial features in such cases.

Appendix E

Ablation Study on the Role of Residual Kriging

FFRK utilizes a two-stage architecture that decouples large-scale deterministic trends from local spatial residuals. Four ML models were employed to estimate the trend based on extracted geofeatures; the resulting residuals were then interpolated via OK and reintegrated with the trend predictions. To isolate the specific contribution of the kriging component, we introduced Focal Feature Regression (FFR) as an ablation baseline. In FFR, the extracted geofeatures are directly used as inputs to machine learning models for prediction, without performing kriging on residuals. This additional experiment allows us to isolate the contribution of the kriging component and to evaluate the trade-off between model complexity and predictive performance.

The performance of FFR and FFRK was compared across four base learners-LM, RF, DT, and SVM-using identical three-dimensional geofeatures and cross-validation protocols. Evaluation metrics (R^2, MAE, and RMSE) were averaged across three heavy metals (Cu, Zn, and Pb). As shown in Supporting Information: Table S2, the utility of residual kriging is highly dependent on the base learner. For LM and DT, FFRK consistently yielded higher accuracy, suggesting that kriging effectively captures spatial dependencies that these simpler learners overlook. Conversely, for Random Forest, FFR performed slightly better, indicating that powerful nonlinear models may inherently internalize the spatial structures embedded within geofeatures, thereby marginalizing the benefits of additional kriging. FFRK also provided consistent, albeit moderate, improvements for SVM. Collectively, these findings confirm that residual kriging is a complementary rather than redundant step, enhancing robustness across diverse learners-particularly those with limited capacity to model spatial autocorrelation-without introducing systematic overfitting.