THREE-DIMENSIONAL SOIL PROPERTY PREDICTION IN PEACH ORCHARDS BASED ON REGRESSION KRIGING AND BP NEURAL NETWORKS: A CASE STUDY OF NITROGEN

/

基于回归克里格与 BP 神经网络的桃园三维土壤养分推测：以氮元素为例

Xibing LI[1)], Dexiang CHEN[1)], Xizhao LI[2)], Tao XU*[3)]

[1)] College of Mechanical and Electrical Engineering, Fujian Agriculture and Forestry University, Fuzhou 350100, China;
[2)] Department of Intelligent Manufacturing, Shangdong Labor Vocational and Technical College, Jinan 250022, China,
[3)] Institute of Agricultural Facilities and Equipment, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China
Tel: +8617702500115; E-mail: xutao@jaas.ac.cn
Corresponding author: Tao Xu

DOI: https://doi.org/10.35633/inmateh-78-110

Keywords: *RK-BP hybrid model, soil nitrogen, three-dimensional spatial distribution, Kriging, machine learning*

ABSTRACT

Soil nitrogen content plays a critical role in fruit tree growth and fruit quality. To achieve high-precision estimation of the three-dimensional spatial distribution of soil nitrogen in peach orchards, this study proposes a hybrid model integrating a Backpropagation (BP) neural network with Regression Kriging (RK). The model is designed to overcome the limitations of traditional single-method approaches when addressing complex spatial nonlinear problems. Using soil data from peach orchards in Fujian Province as the research subject, the predictive performance of five models—BP neural network, Ordinary Kriging (OK), Regression Kriging (RK), Co-Kriging (CK), and the proposed RK-BP hybrid model—was systematically compared. The results indicate that the RK-BP hybrid model outperforms all baseline models, achieving the highest coefficient of determination (R^2 = 0.97) and the lowest root mean square error (RMSE = 7.0). Compared with the BPNN and RK models, the proposed model improved prediction accuracy by 74.17% and 34.95%, and enhanced precision by 77.29% and 77.46%, respectively. The RK-BP model successfully integrates nonlinear relationship modeling with spatial structural analysis, achieving complementary advantages. This study confirms the effectiveness and great potential of hybrid modeling frameworks combining machine learning and geostatistics for 3D digital soil mapping, offering significant theoretical value and promising prospects for precision agriculture.

摘要

土壤中的氮含量对果树的生长和果实品质起着至关重要的作用。为了实现对桃园中土壤氮元素的三维空间分布高精度估计，本研究提出了一种将反向传播（BP）神经网络与回归克里金（RK）方法相结合的混合模型。该模型旨在克服传统单一方法在处理复杂空间非线性问题时的局限性。以福建省桃园的土壤数据作为研究对象，对五个模型——BP 神经网络、普通克里金（OK）、回归克里金（RK）、协同克里金（CK）以及所提出的 RK-BP 混合模型——的预测性能进行了系统比较。结果表明，RK-BP 混合模型优于所有基准模型，实现了最高的决定系数（R^2 = 0.97）和最低的均方根误差（RMSE=7.0）。与 BPNN 和 RK 模型相比，所提出的模型将预测准确度提高了 74.17%和 34.95%，精度分别提高了 77.29% 和 77.46%。RK-BP 模型成功地将非线性关系建模与空间结构分析相结合，实现了优势互补。本研究证实了将机器学习和地质统计学相结合的混合建模框架在三维数字土壤制图中的有效性和巨大潜力，为精准农业提供了重要的理论价值和广阔的发展前景。

INTRODUCTION

China is the origin center of peach and remains the world's leading producer, ranking first globally in both cultivation area and total yield. Fujian Province is one of the major peach-producing regions in southern China. Multivariate analyses of soil nutrients and fruit quality have shown that among the many factors influencing peach fruit quality, soil nutrient content in orchards exerts a significant impact on fruit attributes *(Sun et al., 2022)*; among these, nitrogen content is essential for peach trees and fruits, as it is a key element influencing tree health, yield, and quality. However, determining soil element content remains a major challenge *(Yu et al., 2016)*.

Traditional analytical methods, though highly accurate, suffer from limitations such as high computational demands, insufficient capability for nonlinear problem handling, and long processing times, making them unsuitable for rapid and timely field investigations. To meet the needs of soil fertility diagnosis and precision agriculture, it is essential to develop fast and cost-effective methods for estimating soil element content while reducing time and economic costs.

Three-dimensional (3D) soil spatial prediction is a key technology in precision agriculture, ecological restoration, and environmental monitoring. By integrating multi-source data—such as soil properties, terrain features, and environmental variables—it enables the construction of high-accuracy 3D models to reveal soil spatial heterogeneity. Soil is a continuous three-dimensional entity in space, and not only the surface layer but also the subsurface layers significantly influence overall soil properties and functions. As two-dimensional digital soil mapping increasingly fails to meet the demands of production and scientific research, 3D digital soil mapping has gained growing attention. In recent years, with the widespread adoption of Geographic Information Systems (GIS) and remote sensing, geostatistical and machine learning methods have become essential tools for 3D soil modeling. Among these, Regression Kriging (RK) and Back Propagation (BP) neural networks are particularly notable for their broad application and strong performance.

Visible near-infrared reflectance spectroscopy (VNIRS) has been used for the rapid characterization of organic materials. Studies have shown that VNIRS has good predictive ability for total carbon content, excellent predictive performance for total nitrogen content, but poor predictive performance for pH *(Mervin et al., 2016)*. Spectroscopic techniques together with a Si-SVR (support vector regression) model were employed to establish predictive models for key soil nutrients (N, P, and K) *(Yu et al., 2022)*. An enhanced variational inference technique was applied for spectroscopic prediction of soil properties using a Bayesian convolutional neural network (Bayesian CNN) to capture uncertainty, and superior performance was demonstrated compared to GAM and PLS-BS *(Omondiagbe et al., 2024)*. However, these methods are limited to two-dimensional soil information prediction and are unable to characterize soil properties at greater depths.

An automated mapping procedure using three-dimensional Regression Kriging (3D-RK) was developed to generate global maps of various soil properties *(Hengl et al., 2014).* SoilML, a method that integrates machine learning models with soil science knowledge to improve the relevance of ML-based predictions in soil science, was introduced *(Minasny et al., 2024).* A combined prediction model integrating five single learners (DTR, GKR, RF, LASSO, and MLP) with weights optimized via generalized reduced gradient (GRG) algorithm achieved a predictive effectiveness of 0.880 and a 12% improvement in stability for estimating soil total nitrogen content from Vis-NIR hyperspectral data *(Zhang et al., 2024)*. Near-infrared (NIR) spectroscopy coupled with partial least squares regression (PLSR) and spectral preprocessing has been demonstrated as a promising tool for the rapid, online assessment of soil nitrogen supply across diverse soil types *(Crescini et al., 2025)*.

Existing 3D approaches often struggle with two major limitations: traditional geostatistical methods like Regression Kriging (RK) effectively capture spatial autocorrelation but may underperform in modeling complex non-linear relationships between environmental covariates and soil depth; purely machine learning approaches often lack explicit spatial structure constraints, potentially introducing vertical discontinuities.

The integration of Regression Kriging and BP neural networks offers a new approach for three-dimensional soil prediction. Regression Kriging captures the linear influences of environmental factors through covariate regression, while the BP neural network models the complex nonlinear relationships within the data. Their complementary strengths help overcome the limitations of using either method alone. Specifically, Regression Kriging effectively handles spatially autocorrelated residuals, whereas the BP neural network enhances the mapping between environmental variables and soil properties.

MATERIALS AND METHODS

Overview of the Study Area

The study area is located in Longyan City, Fujian Province, China, spanning 116°42'18"–116°43'27"E and 25°16'30"–25°17'24"N. It is a typical hilly region covering approximately 9.6 km², with an average elevation of 560 meters. The area falls within a subtropical monsoon climate zone, characterized by hot, rainy summers with average temperatures of 22–30°C. Influenced by the southeast monsoon and typhoons, precipitation is concentrated and heavy rainfall events are common. Winters are mild and relatively dry, with average temperatures of 6–13°C, occasional cold waves, and no severe freezing conditions. The region receives abundant sunlight and ample rainfall, with an annual precipitation of approximately 1,400–2,000 mm. The rainy season occurs from May to September, accounting for more than 70% of the annual rainfall. During the typhoon season (July–September), the area may experience intense rainfall and strong winds.

Soil Sample Collection

Soil samples were collected from a peach orchard in Longyan City, Fujian Province, from June 24 to 29, 2025, using an undisturbed soil sampler. The cultivar was "Yingzui" honey peach, with plant spacing of about 2 m. Farmyard manure had been applied 10 months earlier using the trench method.

A Latin hypercube sampling method was applied within a 500 m × 500 m experimental field to randomly select 25 candidate sampling points. Based on on-site conditions, inaccessible or unsuitable points were excluded, resulting in 20 final sampling locations were determined. At each point, soil samples were taken from three depth intervals: 5–20 cm, 20–40 cm, and 40–60 cm. After removing plant residues, gravel, and other impurities, approximately 1000 g of soil was placed into clean sample bags, properly labeled, and stored for later analysis. Upon returning to the laboratory, the samples were air-dried, passed through a 10-mesh sieve, and 500 g of processed soil was sealed in a clean sample bag for subsequent measurements.

Methods for Soil Component Measurement

In this study, soil pH was determined using the potentiometric method. Soil organic matter content was measured using the potassium dichromate oxidation–oil bath heating titration method. Available phosphorus was determined by the ammonium fluoride–hydrochloric acid extraction method followed by molybdenum–antimony colorimetry. Available boron was measured using the hot-water extraction method combined with the azomethine-H colorimetric technique. Alkali-hydrolyzable nitrogen was determined using the semi-micro Kjeldahl method. Available potassium was measured using the neutral ammonium acetate extraction method followed by flame photometry. Available iron was quantified using diethylenetriaminepentaacetic acid (DTPA) extraction coupled with inductively coupled plasma optical emission spectrometry (DTPA–ICP–OES). Total calcium content was determined by digesting soil samples with hydrochloric acid, nitric acid, hydrofluoric acid, and perchloric acid, followed by measurement using atomic absorption spectrophotometry.

Research Methods

BP Neural Network Modeling

A BP neural network (BPNN) is a multilayer feedforward network trained using the error backpropagation algorithm and is well suited for analyzing various nonlinear relationships. It consists of an input layer, an output layer, and one or more hidden layers (*Duan et al., 2014; Panda et al., 2010*). Owing to its supervised learning capability, BPNN can serve as an effective calibration method, providing reliable predictive performance (*Liu et al., 2008*).

The training of a BPNN involves two stages: forward propagation and error backpropagation (*Lu et al., 2013*).

(1) Forward Propagation

In a neural network, forward propagation involves computing the input and output values of each neuron. The output value (z_j) can be expressed as:

$$z_j = f_i(\sum_{i=1}^{m} \omega_{ji} x_i + b_j) \quad (1)$$

where: x_i represents the input-layer information, denoting the spectral variables; z_j represents the hidden-layer output; ω_{ji} is the weight between the input layer and hidden layer; and f_i is the transfer (activation) function from the input layer to the hidden layer. In this study, the *trainlm* function was selected as the training algorithm. The b_j denotes the hidden-layer bias.

The output from the hidden layer s_k is then transmitted to the output layer, and the predicted value (s_k) is expressed as:

$$s_k = g_j(\sum_{j=1}^{n} \omega_{kj} z_j + b_k) \quad (2)$$

where: s_k represents the output-layer information (the content of each soil nutrient); g_j is the transfer (activation) function from the hidden layer to the output layer, for which the *Purelin* function was used in this study; ω_{kj} denotes the weights between the hidden layer and the output layer; and b_k is the bias term of the output layer.

(2) Error Backpropagation

The number of neurons in the hidden layer is determined according to an empirical formula, as shown in Equation (3):

$$n_h = 2n_i + 1 \tag{3}$$

where: n_h is the number of hidden-layer units, and n_i is the number of input-layer units.

The objective of backpropagation is to compute the gradients of the loss function with respect to all weights ω and biases b, denoted as $\frac{\partial E}{\partial \omega}$ and $\frac{\partial E}{\partial b}$. Using the chain rule, these gradients are calculated layer by layer in the reverse direction. When the difference between the predicted value and the measured value is large, this discrepancy is propagated backward as an error signal. During the backpropagation process, the Levenberg–Marquardt algorithm is used to update the connection weights from the output layer back to the input layer in order to reduce the mean squared error.

$$MSE = \frac{1}{N}\sum_{k=1}^{K}(t_k - y_k)^2 \tag{4}$$

where: t_k is the measured soil nutrient content, y_k is the predicted soil nutrient content, and N is the number of training samples.

Figure 1 illustrates the architecture of the neural network model employed in this study.

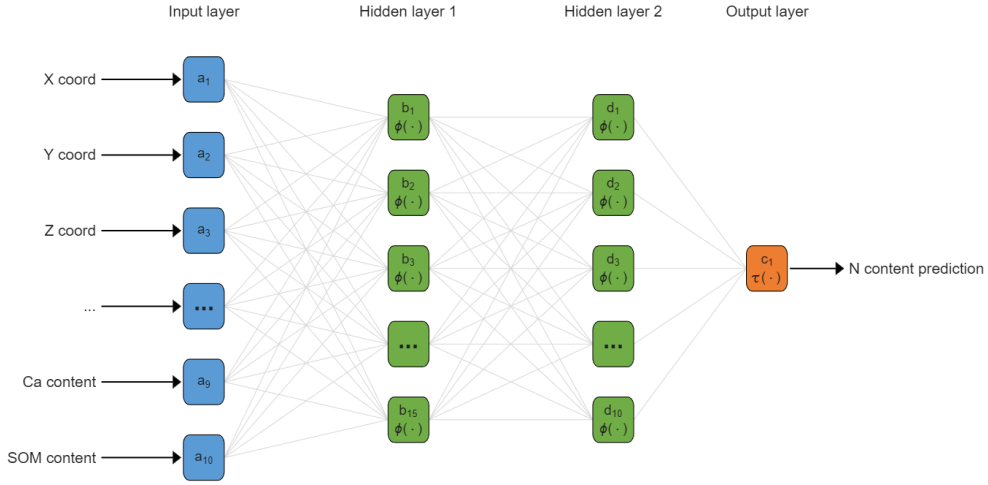

Fig. 1 - BP Neural Network Structure

Note: a_1~ a_{10} represent the 10 input features of the input layer; $\phi(\cdot)$ and $\tau(\cdot)$ respectively denote the activation functions of the hidden layer and output layer; b_1~ b_{15} are the nodes of hidden layer 1; d_1~ d_{10} are the nodes of hidden layer 2; c_1 is the output node of the output layer.

Regression Kriging Residual Correction (Spatial Dependency Modeling)

Regression Kriging (RK) is a method that integrates regression and interpolation techniques into a single framework, enabling simultaneous representation of the relationship between the target soil property and explanatory variables, as well as the spatial structure of the residuals **Error! Reference source not found.**. T he RK model can be expressed as:

$$\hat{Z}(s_0) = \sum_{j=0}^{p}\hat{\beta}_j \cdot X_j(s_0) + \sum_{i=1}^{n}\lambda_i \cdot e(s_i) \tag{5}$$

where: $\sum_{j=0}^{p}\hat{\beta}_j \cdot X_j(s_0)$ is the regression trend component, and $\sum_{i=1}^{n}\lambda_i \cdot e(s_i)$ is the residual term obtained through Ordinary Kriging interpolation. In this study, $\hat{Z}(s_0)$ represents the predicted soil element value at location s_0 obtained using Regression Kriging; $X_j(s_0)$ denotes the known covariates at location s_0; $\hat{\beta}_j$ are the regression coefficients fitted from the sampling data; and p is the number of predictor variables. The λ_i represents the kriging weights determined by solving the Ordinary Kriging system, $e(s_i)$ is the regression residual at sampling location s_i, and n is the number of sampling points.

By incorporating a depth function and depth parameters, the two-dimensional (2D) RK can be extended to three-dimensional (3D) RK *(Hengl et al.,2015)*. The 3D RK model can be expressed as:

$$\hat{Z}(s_0, d_0) = \sum_{j=0}^{p} \hat{\beta}_j \cdot X_j(s_0, d_0) + \hat{g}(d_0) + \sum_{i=1}^{n} \lambda_i (s_0, d_0) \cdot e(s_i, d_i) \quad (6)$$

where: $\sum_{j=0}^{p} \hat{\beta}_j \cdot X_j(s_0, d_0) + \hat{g}(d_0)$ is the three-dimensional regression trend component, and $\sum_{i=1}^{n} \lambda_i (s_0, d_0) \cdot e(s_i, d_i)$ is the residual term obtained via three-dimensional Ordinary Kriging interpolation. Unlike 2D RK, all locations s_i now include the depth dimension (s_i, d_i), where d_i represents the soil depth in s_i .$\hat{g}(d_0)$is the depth function, and $X_j(s_0, d_0)$denotes the value of the covariate at the 3D sampling point. In this study, the equal-area quadratic spline function *(Bishop et al.,1999)* was selected as the depth function.

Construction of Three-Dimensional Variogram and Fusion Strategy

Traditional 3D RK employs multiple linear regression (MLR) (*Zhang et al. 2020*) to fit the trend surface. Replacing MLR with a BP neural network (BPNN) can greatly enhance the model's ability to capture complex nonlinear relationships. In this study, BPNN was used to determine the trend model. At each sampling point, the residuals of the BPNN predictions were calculated, and a three-dimensional variogram model was constructed and fitted based on these residuals. Finally, 3D kriging interpolation was performed to combine 3D RK with BPNN. The corresponding equation can be expressed as:

$$\hat{Z}_{RK-BPNN}(s_0, d_0) = \sum_{j=0}^{p} \int BPNN\left(X_j(s_0, d_0), w, b\right) + \sum_{i=1}^{n} \lambda_i (s_0, d_0) \cdot e(s_i, d_i) \quad (7)$$

where: $\sum_{j=0}^{p} \int BPNN\left(X_j(s_0, d_0), w, b\right)$ is the deterministic trend value predicted by the BPNN model, and $\sum_{i=1}^{n} \lambda_i (s_0, d_0) \cdot e(s_i, d_i)$ is the residual obtained through three-dimensional Ordinary Kriging interpolation. $\int BPNN$ represents the complex nonlinear transformation defined by the network structure; $X_j(s_0, d_0)$denotes the covariates input to the BPNN; w is the weight matrix learned during BPNN training; b is the bias vector learned during BPNN training; and $e(s_i, d_1)$ is the BPNN prediction residual at location (s_i, d_i).

Figure 2 illustrates the architecture of the hybrid model integrating Regression Kriging and BP Neural Network developed in this study. In this model, features exhibiting weak correlations with the target variable (specifically B and Ca) were excluded, while the remaining architecture matches that of the BPNN model. This strategy aims to reduce model complexity and enhance generalization ability.

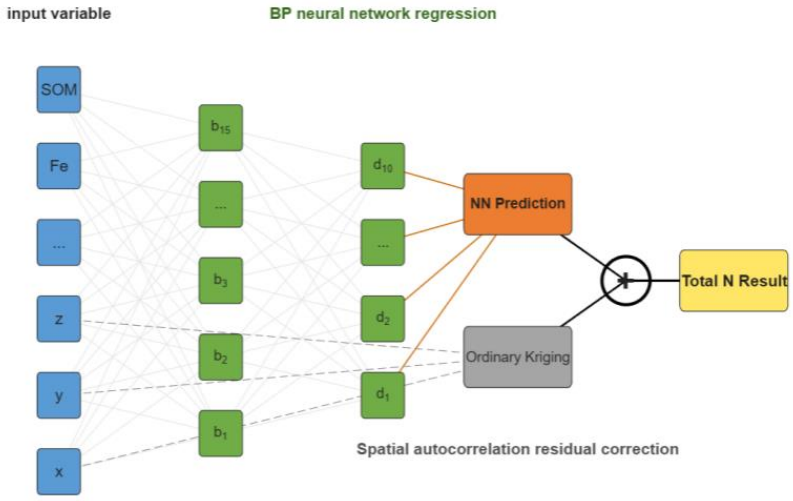

Fig. 2 - RK-BP Structure

Note: b_1~ b_{15} are the nodes of hidden layer 1; d_1~ d_{10} are the nodes of hidden layer 2.

Accuracy Evaluation

The sampling data were randomly divided into two subsets, with 70% of the points assigned to the training set and 30% to the validation set. The accuracy of the models was evaluated using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). R^2 measures the goodness of fit of the model to the data; the closer its value is to 1, the more reliable the model. MAE and RMSE are used to assess the error between predicted and observed values, with smaller values indicating higher model reliability and predictive accuracy. The corresponding formulas are as follows:

$$R^2 = 1 - \frac{\Sigma\left(y - \hat{y}\right)^2}{\Sigma(y - \bar{y})^2} \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \quad (9)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \quad (10)$$

RESULTS

Conventional Statistical Characteristics

As shown in Table 1, the mean nitrogen content in the study area is 117.29 mg/kg, with a minimum of 49.7 mg/kg and a maximum of 285.4 mg/kg. The overall coefficient of variation (CV) is 48.07%, indicating considerable variability. However, the CVs for the 5–20 cm and 20–40 cm layers are around 20%, indicating moderate variability. In contrast, the CV in the 40–60 cm layer exceeds 30%, exhibiting a high degree of variation. Nitrogen content exhibits a decreasing trend with depth, indicating surface-layer accumulation of nitrogen.

Table 1

Descriptive Statistics of Soil Nitrogen Concentration (mg/kg)

Depth (cm)	n	Min	Max	Mean	SD	CV (%)
5-20	20	146.5	285.4	182.63	34.61	18.95
20-40	20	76.1	197.6	103.55	31.23	30.16
40-60	20	49.7	104.6	65.69	14.22	21.65
5-60	60	49.7	285.4	117.29	56.38	48.07

The content and distribution of nitrogen (N) in soil are governed by complex interactions among various factors. Research by Sun et al. indicates that available N, K, B, Fe, Ca, Mn, and other elements in soil significantly affect peach fruit quality *(Sun et al., 2022)*. Furthermore, pH, phosphorus, organic matter, and other factors exhibit certain correlations with the content of these elements. Therefore, this study selected depth, pH, phosphorus, potassium, boron, iron, calcium content, and organic matter content as auxiliary variables to predict soil nitrogen content. Table 2 demonstrates the distribution patterns of the analyzed soil chemical properties along the vertical profile. pH values exhibited an increasing trend from top to bottom, while the contents of B and Ca remained relatively stable throughout the profile. In contrast, the contents of N, P, K, Fe, and soil organic matter (SOM) generally decreased with depth.

Table 2

Mean values of soil chemical properties by profile layer

Depth (cm)	pH	N (mg/kg)	P (mg/kg)	K (mg/kg)	B (mg/kg)	Fe (mg/kg)	Ca (mg/kg)	SOM (g/kg)
20	4.62	183.98	13.01	70.24	0.241	171.32	28.36	27.32
40	5.01	106.80	6.57	51.91	0.272	70.47	32.68	19.38
60	5.36	66.89	4.21	43.97	0.244	44.86	29.79	13.31
5-60	5.00	117.29	7.69	56.04	0.251	96.05	30.48	19.88

The Pearson correlation coefficients between nitrogen and these auxiliary variables are presented in Table 3. According to the table, there is a strong correlation between pH value, P, K, Fe, SOM and N element content.

Table 3

Correlation analysis between nitrogen and other variables

	pH	P	K	B	Fe	Ca	SOM
Pearson correlation	-0.486**	0.931**	0.529**	0.217	0.763**	0.013	0.889**
Sig. (two-tailed)	< 0.01	< 0.01	< 0.01	0.095	< 0.01	0.921	< 0.01

*Notes: ** significant at 0.01 (two-tailed); * significant at 0.05 (two-tailed).*

Comparison of Prediction Accuracy of Different Models

In this study, a four-layer BP neural network with two hidden layers was used to predict soil nutrient content. The number of neurons in the first hidden layer was set to 15, and in the second hidden layer to 10. The network was trained for 1,000 iterations using the Levenberg-Marquardt (trainlm) algorithm, which automatically adjusts the learning rate. When using Regression Kriging (RK) for prediction, the number of distance lags was set to 20, and the variogram model was specified as exponential. To ensure comparability, the network structure and parameter settings were kept consistent with the BP neural network and RK models described above. Additionally, Ordinary Kriging (OK) and Co-Kriging (CK) were applied as benchmarks to evaluate the validity of the proposed method. The workflow is illustrated in Figure 3.

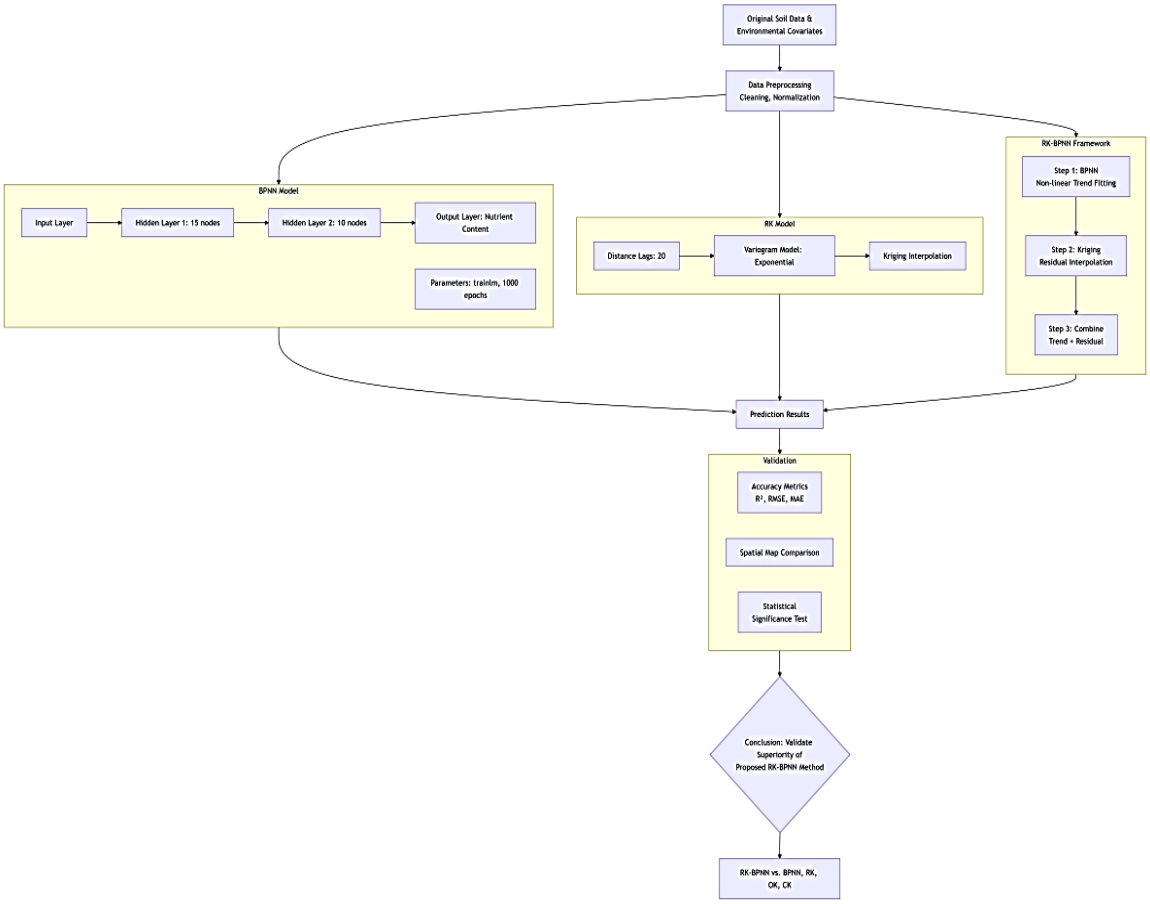

Fig. 3 - The basic logic of the RK–BPNN optimization

Figure 4 shows the spatial distribution of nitrogen (N) predicted by different methods.

Overall, the general spatial pattern of N predicted by all methods is similar: N content decreases with increasing soil depth, showing spatial autocorrelation and stratified heterogeneity, though local details vary significantly. The BP model predictions of N exhibit clear vertical stratification, with uniform values within the same depth layer due to its limited ability to capture horizontal variation.

Both RK and CK produce similar results; although they reflect differences in N concentration within the same layer, the predictions are generally smoother. This smoothing attenuates high N signals in the surface layer, resulting in lower color saturation (representing lower estimated values). This may indicate that the models partially explain surface N enrichment but fail to fully capture localized high values (hotspots), making it difficult to predict areas with large anomalies. CK shows more distinct high-value regions compared to RK, with sharper transitions and richer detail, indicating stronger capability to capture local variability. OK exhibits significant horizontal spatial heterogeneity, with more pronounced high-value areas than CK or RK; however, transitions are smooth, and localized hotspots (e.g., at coordinates (400, 500)) are not accurately captured.

The RK-BP hybrid model successfully identifies and enhances local hotspots, producing higher color saturation, clearer boundaries, broader horizontal coverage of high-value regions, and more pronounced variations. Its predictions closely match observed values, demonstrating strong local fitting ability.

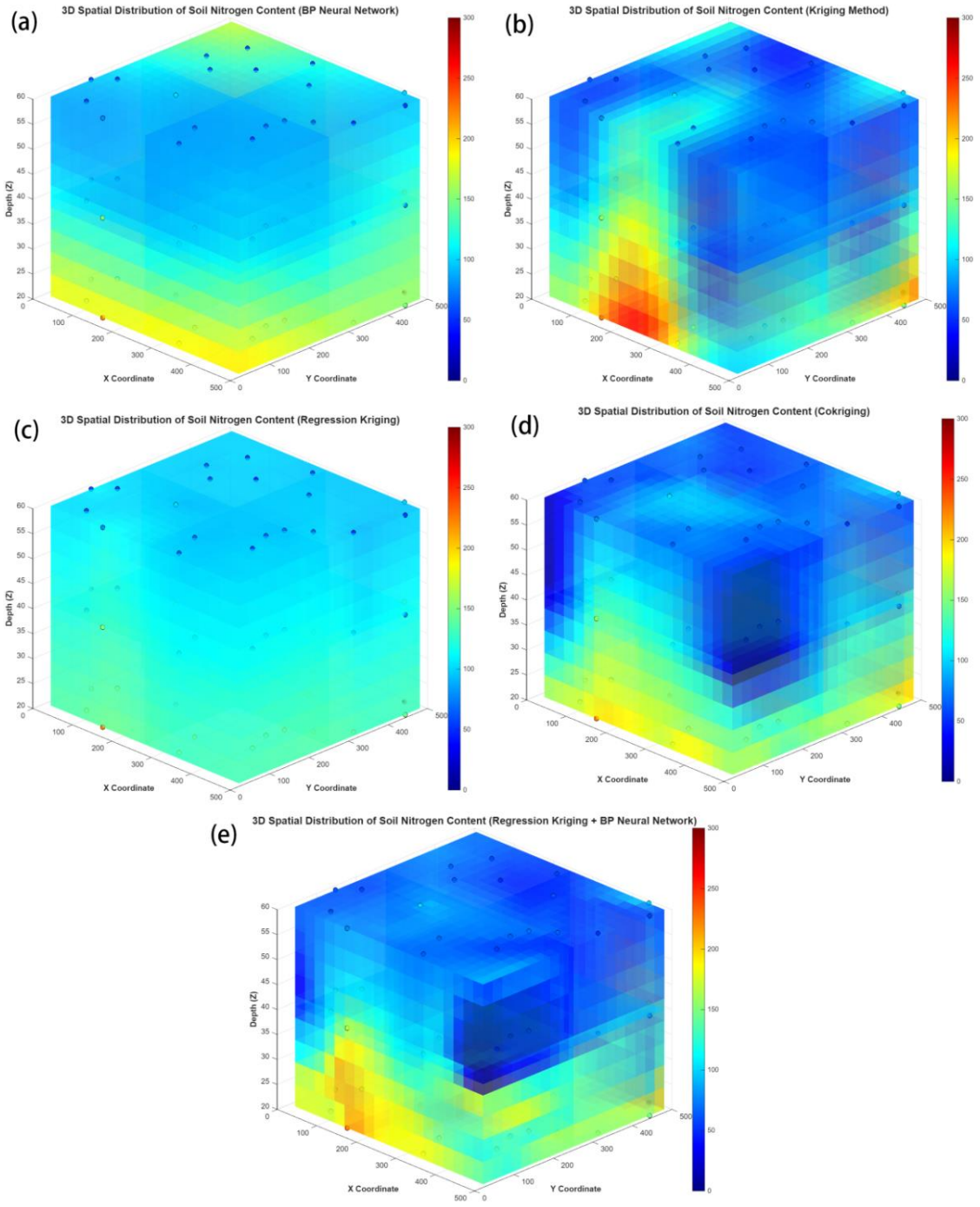

Fig. 4 - Spatial Distribution of Nitrogen Predicted by Different Models

(a) BPNN; (b) OK; (c) RK; (d) CK; (e) RK–BP

Model Performance

In this study, 18 soil samples were used to validate the predictive accuracy of five models. The validation results are shown in Figure 5, where predicted soil nitrogen (N) content is plotted against measured values. The results indicate that the BPNN model performs poorly in predicting soil N content, with the regression line showing the largest deviation from the 1:1 reference line. The three Kriging models exhibit similar predictive ability, with the regression lines moderately deviating from the 1:1 line but showing some differences: in the OK model, the points are closer to the 1:1 line and the regression slope is nearer to 1; in the CK model, a few outliers lie far from the 1:1 line. The RK-BP hybrid model demonstrates superior predictive capability, as its scatter points and regression line are closer to the 1:1 line.

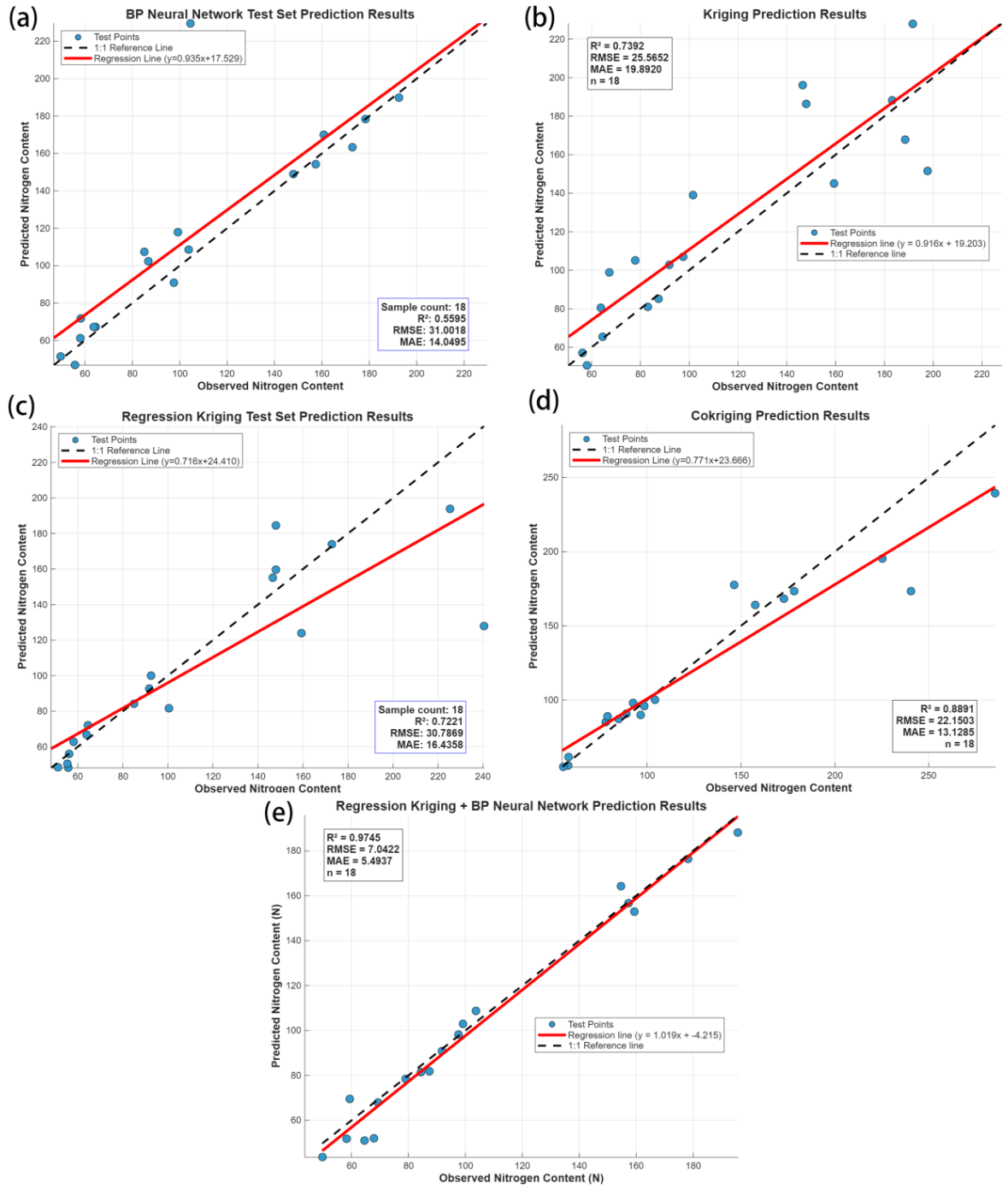

Fig. 5- Scatter Plots of Measured vs. Predicted Soil Nitrogen Content

(a) BPNN; (b) OK; (c) RK; (d) CK; (e) RK–BP

Table 4 summarizes the prediction accuracy of BPNN, OK, RK, CK, and RK-BP for soil nitrogen content. The BPNN model shows the weakest performance, with an R^2 of only 0.5595 and an RMSE of 31.0018, indicating a large average deviation between predicted and observed values; this suggests that a standalone BP neural network has limited ability to predict N content and requires integration with other methods for optimization.

Among the three Kriging models, the CK model demonstrated superior performance with an R2 of 0.8891, an RMSE of 22.1503, and an MAE of 13.1285, outperforming the other two Kriging models. However, its effectiveness relies on the assumption of linear correlation between variables. Similarly, the suboptimal performance of the linear RK model, particularly its higher RMSE, highlights a crucial limitation: soil spatial variability often exhibits strong non-linear characteristics that simple linear regression models are unable to resolve.

To overcome these limitations and further enhance prediction accuracy, a hybrid strategy is required—one that can simultaneously model the complex non-linear relationships between environmental variables and soil properties, while also accounting for spatial autocorrelation in the residuals. Therefore, this study introduces the RK-BP model, which replaces the linear regression component with a BPNN. By combining the powerful non-linear mapping capabilities of BPNN with the spatial interpolation strengths of Kriging, the RK-BP model aims to address the deficiencies of traditional RK and CK, thereby minimizing prediction errors. The RK-BP model achieves the highest R^2 (0.9745) and significantly lower RMSE (7.0422) and MAE (5.4937) compared to the other models, demonstrating superior predictive ability. These results further confirm that the RK-BP model provides the most accurate estimates of soil N content.

Table 4

Comparison of Prediction Accuracy for Soil Nitrogen Content Using the Validation Dataset

Model	R^2	RMSE	MAE
BPNN	0.5595	31.0018	14.0495
OK	0.7392	25.5652	19.8920
RK	0.7221	30.7869	16.4358
CK	0.8891	22.1503	13.1285
RK-BP	0.9745	7.0422	5.4937

(R^2: coefficient of determination; RMSE: root mean square error; MAE: mean absolute error)

As shown in Table 5, the proposed RK-BP hybrid model consistently outperformed all comparative standalone models across various performance metrics. Specifically, the RK-BP model demonstrated significant enhancements in the coefficient of determination (R2), root mean square error (RMSE), and mean absolute error (MAE) compared to BPNN, Ordinary Kriging (OK), Regression Kriging (RK), and Co-Kriging (CK) models. The R2 value of the RK-BP model showed improvements ranging from 9.61% (vs. CK) to 74.17% (vs. BPNN). Concurrently, the error metrics were substantially reduced: RMSE was optimized by 68.21% (vs. CK) to 77.29% (vs. BPNN), and MAE by 58.17% (vs. CK) to 72.39% (vs. OK). These results unequivocally highlight the superior predictive accuracy and robustness of the RK-BP model in our study.

Table 5

Degree of Improvement of the RK-BP Model Compared to BPNN, OK, RK, and CK Models

Comparison model	R^2 optimize	RMSE optimize	MAE optimize
BPNN	+74.17%	+77.29%	+60.90%
OK	+31.83%	+72.46%	+72.39%
RK	+34.95%	+77.13%	+66.59%
CK	+9.61%	+68.21%	+58.17%

CONCLUSIONS

This study used soil sample data from a peach orchard in Fujian Province to investigate the three-dimensional spatial prediction of soil nitrogen as a case study. It systematically evaluated and compared the predictive performance of five models: the BP neural network (BPNN), Ordinary Kriging (OK), Regression Kriging (RK), Co-Kriging (CK), and the proposed RK-BP hybrid model.

The results indicate that, based on the validation dataset, based on the R^2, RMSE, and MAE values from the validation dataset, the RK-BP model provided the most accurate estimates at both the soil sampling point level and the regional scale, achieving an accuracy as high as 97%. Compared with all comparative standalone models, the RK-BP model improved the coefficient of determination (R^2) by 9.61%–74.17% and optimized the root mean square error (RMSE) by 68.21%–77.29%. These results indicate that the hybrid RK-BP model significantly outperforms all comparative single models in terms of prediction accuracy and precision. The model leverages BPNN to capture the complex nonlinear relationships between environmental variables and soil nitrogen, generating a more accurate deterministic trend surface. Subsequently, Ordinary Kriging is applied to the residuals, incorporating essential spatial autocorrelation corrections and addressing the fundamental limitation of BPNN, which ignores spatial location information.

The RK-BP model features an open and modular theoretical framework, and the current results indicate substantial potential for further improvement. The findings provide a reliable reference for optimizing three-dimensional soil nutrient prediction models and enhancing the accuracy of soil nutrient estimates at the regional scale.

ACKNOWLEDGEMENT

This work was supported by the Science and Technology Innovation Special Fund of Fujian Agriculture and Forestry University (KFB24039).

REFERENCES

[1] Bishop, T. F. A., Mcbratney, A. B., & Laslett, G. M. (1999). Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma*, 91(1), 27-45.

[2] Crescini, D., Mascialino, G., Moggia, N., Piubeni, G., Serpelloni, M., & Sardini, E. (2025). PCA- and PLSR-based machine learning model for prediction of urea-n content in heterogeneous soils using near-infrared spectroscopy. *Sensors*, 25(13), 4176.

[3] Duan, & Min, S. (2014). Design and development of detection node in wireless sensor network based on neural network. *Advanced Materials Research*, 1022, 292-295.

[4] Hengl, T., De Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B., Ribeiro, E., ... & Gonzalez, M. R. (2014). Soilgrids1km — global soil information based on automated mapping. *PLoS ONE*, 9(8), e105992.

[5] Hengl, T., Heuvelink, G. B. M., & Rossiter, D. G. (2007). About regression-kriging: from equations to case studies. *Computers & Geosciences*, 33(10), 1301-1315.

[6] Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J. G., Walsh, M. G., Shepherd, K. D., & Tondoh, J. E. (2015). Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. *Plos One*, 10(6), e0125814.

[7] Liu, F., Zhang, F., Jin, Z., He, Y., Fang, H., Ye, Q., & Zhou, W. (2008). Determination of acetolactate synthase activity and protein content of oilseed rape (Brassica napus L.) leaves using visible/near-infrared spectroscopy. *Analytica Chimica Acta*, 629(1-2), 56-65.

[8] Lu, P., Wang, L., Niu, Z., Li, L., & Zhang, W. (2013). Prediction of soil properties using laboratory VIS–NIR spectroscopy and hyperion imagery. *Journal of Geochemical Exploration*, 132, 26-33.

[9] Mervin, St., Luce, Noura, Ziadi, Bernard, Gagnon, Vicky, Lévesque. (2018). Prediction of total carbon, total nitrogen, and pH of organic materials using visible near-infrared reflectance spectroscopy [J]. *Canadian Journal of Soil Science*, 98(1), 175-179.

[10] Minasny, B., Bandai, T., Ghezzehei, T. A., Huang, Y. C., Ma, Y., Mcbratney, A. B., & Widyastuti, M. (2024). Soil science-informed machine learning. *Geoderma*, 452, 117094.

[11] Omondiagbe, O. P., Roudier, P., Lilburne, L., Ma, Y., & Mcneill, S. (2024). Quantifying uncertainty in the prediction of soil properties using mid-infrared spectra. *Geoderma*, 448, 116954.

[12] Orton, T. G., Pringle, M. J., Bishop, T. F. A., Menzies, N. W., & Dang, Y. P. (2020). Increment-averaged kriging for 3-d modelling and mapping soil properties: combining machine learning and geostatistical methods. *Geoderma*, 361, 114094.

[13] Panda, S. S., Ames, D. P., & Suranjan, P. (2010). Application of vegetation indices for agricultural crop yield prediction using neural network techniques. *Remote Sensing*, 2(3), 673-696.

[14] Shurui, L. I. N., Qing, Z. H. U., Beini, Y. I. N., Guishan, Y. A. N. G., Kaihua, L. I. A. O., Xiaoming, L. A. I., & Changqiang, G. U. O. (2010). Generating three-dimensional soil organic carbon density dataset by soil depth function and correction methods in Yangtze river delta, China. *Environmental Modelling & Software*, 192, 106582.

[15] Smith, E. M., Vargas, R., Guevara, M., Tarin, T., & Pouyat, R. V. (2022). Spatial variability and uncertainty of soil nitrogen across the conterminous united states at different depths. *Ecosphere*, 13(7), e4170.

[16] Sun, H., Huang, X., Chen, T., Zhou, P., Huang, X., Jin, W., & Gao, Z. (2022). Fruit quality prediction based on soil mineral element content in peach orchard. *Food Science & Nutrition*, 10(6), 1756-1767.

[17] Yu, S., Bu, H., Dong, W., Jiang, Z., Zhang, L., & Xia, Y. Y. (2022). Construction and evaluation of prediction model of main soil nutrients based on spectral information. *Applied Sciences*, 12(13), 6298.

[18] Yu, X., Liu, Q., Wang, Y., Liu, X., & Liu, X. (2016). Evaluation of MLSR and PLSR for estimating soil element contents using visible/near-infrared spectroscopy in apple orchards on the Jiaodong peninsula. *Catena*, 137, 340-349.

[19] Zhang, Y., Ji, W., Saurette, D. D., Easher, T. H., Li, H., Shi, Z., & Biswas, A. (2020). Three-dimensional digital soil mapping of multiple soil properties at a field-scale using regression kriging. *Geoderma*, 366, 114253.

[20] Zhang, X.Q., Ma, S.X., Li, Z.W., Zheng, D.C., & Song, H.Y. (2024). Prediction of Total Nitrogen Content in Brown Soil Based on Hyperspectral and Combined Prediction Model. *Spectroscopy and Spectral Analysis*, 44(8), 2310-2317.