

Article

Point-to-Interval Prediction Method for Key Soil Property Contents Utilizing Multi-Source Spectral Data

Shuyan Liu ^{2,3}, Dongyan Huang ¹, Lili Fu ^{2,3}, Shengxian Wu ¹, Yanlei Xu ⁴ , Yibing Chen ⁵ and Qinglai Zhao ^{1,*}

¹ College of Engineering and Technology, Jilin Agricultural University, Changchun 130118, China; huangdy@jlu.edu.cn (D.H.); wushengxian623@163.com (S.W.)

² Key Laboratory of Bionics Engineering, Ministry of Education, Jilin University, Changchun 130022, China; shuyan22@mails.jlu.edu.cn (S.L.); full23@mails.jlu.edu.cn (L.F.)

³ College of Biological and Agricultural Engineering, Jilin University, Changchun 130022, China

⁴ College of Information Technology, Jilin Agricultural University, Changchun 130118, China; yanleixu@163.com

⁵ Jilin Province Soil and Fertilizer Station, Changchun 130033, China; jlszrlfzz@163.com

* Correspondence: zhaoqinglai@jlaui.edu.cn; Tel.: +86-180-4361-8811

Abstract: Key soil properties play pivotal roles in shaping crop growth and yield outcomes. Accurate point prediction and interval prediction of soil properties serve as crucial references for making informed decisions regarding fertilizer applications. Traditional soil testing methods often entail laborious and resource-intensive chemical analyses. To address this challenge, this study introduced a novel approach leveraging spectral data fusion techniques to forecast key soil properties. The initial datasets were derived from UV–visible–near-infrared (UV-Vis-NIR) spectral data and mid-infrared (MIR) spectral data, which underwent preprocessing stages involving smoothing denoising and fractional-order derivative[s] (FOD) transform techniques. After extracting the characteristic bands from both types of spectral data, three fusion strategies were developed, which were further enhanced using machine learning techniques. Among these strategies, the outer-product analysis fusion algorithm proved particularly effective in improving prediction accuracy. For point predictions, metrics such as the coefficient of determination (R^2) and error metrics demonstrated significant enhancements compared to predictions based solely on single-source spectral data. Specifically, R^2 values increased by 0.06 to 0.41, underscoring the efficacy of the fusion approach combined with partial least squares regression (PLSR). In addition, based on the coverage width criterion to establish reliable prediction intervals for key soil properties, including soil organic matter (SOM), total nitrogen (TN), hydrolyzed nitrogen (HN), and available potassium (AK). These intervals were developed within the framework of the kernel density estimation (KDE) interval prediction model, which facilitates the quantification of uncertainty in property estimates. For available phosphorus (AP), a preliminary assessment of its concentration was also provided. By integrating advanced spectral data fusion with machine learning, this study paves the way for more informed agricultural decision making and sustainable soil management strategies.

Keywords: spectral data; soil properties; point prediction; interval prediction; machine learning



Citation: Liu, S.; Huang, D.; Fu, L.; Wu, S.; Xu, Y.; Chen, Y.; Zhao, Q. Point-to-Interval Prediction Method for Key Soil Property Contents Utilizing Multi-Source Spectral Data. *Agronomy* **2024**, *14*, 2678. <https://doi.org/10.3390/agronomy14112678>

Academic Editor: Dimitrios D. Alexakis

Received: 10 October 2024

Revised: 6 November 2024

Accepted: 13 November 2024

Published: 14 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The agricultural sector operates at the nexus of soil health dynamics and property availability, where achieving an optimal equilibrium of key soil properties—such as organic matter, N, P, and K—profoundly influences crop productivity and yield outcomes [1–3]. A precise evaluation of these essential properties holds paramount importance for informed fertilizer application strategies and the implementation of effective soil management practices [4,5]. Traditional chemical methodologies have long served as the bedrock of soil science and agronomy, offering meticulous insights into soil property estimates through a series of well-established laboratory procedures. These conventional techniques encompass

soil sampling, chemical extraction processes, and quantitative analyses utilizing methods like colorimetric assays, spectrophotometry, and titration [6]. Renowned for their high precision and reliability, these methods have undergone extensive validation across diverse soil types and property compositions [7–9]. Despite the accuracy of traditional soil testing approaches, their reliance on labor-intensive and resource-intensive chemical analyses presents formidable challenges in terms of operational efficiency and scalability [10]. Given these challenges, there is an urgent need to develop innovative technologies that can enhance the efficiency and effectiveness of soil property assessments.

In recent years, advancements in soil property detection technologies have increasingly highlighted the need for rapid, portable, and non-destructive methods. For instance, spectroscopic techniques, particularly visible–near-infrared (Vis-NIR) and MIR spectroscopy, leverage the unique spectral signatures of soil constituents to non-destructively evaluate the concentrations of soil properties, such as carbon, N, P, and K [11–14]. Among these, MIR spectroscopy, especially when coupled with attenuated total reflectance (ATR) probes, has shown promise in effectively analyzing soil components and contaminants due to its sensitivity to organic and inorganic compounds in soils. Studies have demonstrated the utility of MIR-ATR in capturing correlations between soil properties and the elemental composition, as seen in partial least-squares models for major soil elements, thus enhancing quantitative soil analyses [15]. Additionally, MIR-ATR techniques have been successfully applied in diagnostic screening for urban soil contaminants and in rapidly assessing petroleum-contaminated soils, underscoring their versatility in environmental monitoring [16,17]. These methods have gained popularity due to their ability to provide real-time data with minimal soil disturbance. Nevertheless, significant challenges related to soil heterogeneity and variability in moisture content continue to affect the accuracy and reliability of these assessments [18]. Remote sensing technologies, which utilize satellite and aerial imagery, offer the capability to monitor soil characteristics over extensive geographical areas, thus providing valuable spatial data for large-scale agricultural management [19–21]. Despite their advantages, the accuracy of these remote sensing methods often depends on careful calibration with ground truth data, which may be limited in availability and may not adequately represent local soil conditions [22,23]. Moreover, emerging technologies such as electrochemical sensors and laser-induced breakdown spectroscopy (LIBS) facilitate rapid and direct analysis of soil elements, enabling timely decision making in agricultural practices [24–28]. While these methods present clear advantages, they are not without challenges, issues related to sensor stability, environmental interference, and the need for frequent recalibration pose significant obstacles to their practical implementation [29,30]. Overall, while innovative technologies have made substantial strides in soil properties detection, the reliance on data from individual sensors still imposes limitations on the overall accuracy of soil assessments.

To address these limitations, there is a growing interest in developing innovative methods for assessing soil properties through the fusion of multi-sensor data combined with machine learning algorithms. Numerous studies have demonstrated the effectiveness of using multi-source data within fusion networks to enhance the accuracy of soil property predictions. For instance, combinations of remote sensing data with spectral data, spectral data with image data, and LIBS with MIR spectral data have all shown promise in improving predictive outcomes [31–33]. These advancements provide a new direction for the assessment of soil properties. Consequently, it is essential to investigate the fusion methods of multi-source spectral data to maximize the potential of near-field soil sensing technologies for accurate soil property estimation. On the one hand, many of these approaches primarily focus on point predictions for individual properties, which limits their ability to comprehensively evaluate multiple key properties simultaneously [34–37]. Furthermore, there is a notable lack of research on interval prediction methods in this context. Unlike traditional point estimation, interval prediction generates a range of upper and lower limits for property values, which better reflects the inherent uncertainty and variability of soil properties [38]. This approach leverages established chemometric methods to enhance pre-

diction reliability within a comprehensive interval modeling framework, as demonstrated in prior studies by Rodionova and Pomerantsev [39,40]. By broadening the traditional model application, this method accounts for greater variability, which is particularly beneficial in agricultural and environmental research [39]. Currently, most applications of interval prediction are concentrated in fields such as photovoltaic energy generation and water demand forecasting, with relatively few studies addressing the interval prediction and evaluation of soil property levels [41,42]. This gap presents an opportunity for further research to explore how interval prediction techniques can be effectively adapted and applied to soil property assessments. On the other hand, machine learning represents a pivotal component within the realm of multi-source data fusion technology. Models like partial least square regression (PLSR), random forest (RF), support vector machine (SVM), and deep learning (DL) exemplify the potential of amalgamating diverse data streams to enhance the precision and efficacy of property predictions [31,34,43–45]. These methodologies offer a sophisticated framework for synthesizing heterogeneous data inputs, thereby elevating the granularity and reliability of soil property estimations. Despite the substantial progress made in leveraging machine learning for multi-source data fusion, considerable opportunities for advancement remain, particularly in the area of point-to-interval prediction of soil property [46]. The transition from singular data points to comprehensive prediction intervals remains a frontier where further refinement and innovation are warranted.

In addition, the preprocessing and fusion methods applied to raw spectral data are critical factors influencing estimation accuracy [47]. Various studies have employed techniques such as baseline correction, smoothing, and denoising for spectral preprocessing, with fractional order derivatives (FODs) also emerging as a particularly effective approach for enhancing the accuracy of Vis-NIR estimations [48–50]. Despite these advancements, a standardized method for processing spectral data has yet to be established, and comparative studies assessing the performance of different preprocessing techniques remain scarce. The fusion of spectral data focuses on integrating information during the post-processing phase, and numerous studies have demonstrated its efficacy in enhancing both the accuracy and stability of soil property estimations [51–53]. However, when performing multi-task predictions for soil properties, inconsistencies among the information provided by the different spectral bands can lead to distortions in the fusion results, particularly in scenarios characterized by strong spectral aliasing or noise [54]. Consequently, effectively managing noise and preventing overfitting presents significant challenges in current research efforts. Addressing these issues is essential for optimizing the performance of predictive models in soil science.

In this study, we present a novel approach that integrates diverse spectral data to facilitate precise point-to-interval forecasts of critical soil property contents. Following the collection, preprocessing, and feature extraction of soil spectral data, we systematically designed and compared various fusion strategies to evaluate their impact on prediction accuracy and associated errors. Leveraging a PLSR model, we successfully achieved point predictions for five essential soil properties. Building on this work, we established prediction intervals at varying confidence levels by applying statistical insights derived from the errors associated with the point predictions. This comprehensive analysis not only allowed for the identification of uncertainties in predicting soil properties but also illuminated the underlying causes of these uncertainties.

2. Materials and Methods

2.1. Sample Preparation

The research site is located within a controlled experimental field at the Jilin Academy of Agricultural Sciences (JAAS) in Gongzhuling City, Jilin Province. This geographical zone is defined by a temperate continental climate, marked by an average annual precipitation of approximately 594.8mm and an average annual temperature of 5.6 °C. The pronounced seasonal variations observed in this region significantly contribute to the overall fertility of the soil, thereby fostering a conducive environment for a diverse range of agricultural

practices. The nutrient profile within this locale is subject to the interplay of natural processes and agricultural interventions. Hence, point and interval monitoring of soil property contents is imperative for augmenting agricultural yields and sustainability within this region.

Field surveys and soil sampling were carried out in the study area during the fall of 2021. Considering the soil type, topography, and land-use characteristics, a plum sampling method was employed to collect samples from the 5–20 cm soil layer. In total, 109 soil samples were collected. Following the preparatory steps, each soil sample was partitioned into three distinct sections. The first section was dedicated to the collection of UV-Vis-NIR spectral data, and the second to the acquisition of MIR spectral data. The third component involves the chemical analysis of key soil properties, which is essential for accurately determining the actual concentrations of each property.

The chemical measurement of soil properties is conducted using internationally recognized methods. Specifically, soil organic matter was measured by the Walkley and Black dichromate oxidation method [7]. In this process, the soil sample is oxidized with potassium dichromate and sulfuric acid. The organic matter content is then determined based on the consumption of dichromate, which correlates with the organic carbon content in the sample. For TN, we employed the Kjeldahl method, which involves digesting soil samples with sulfuric acid and a catalyst to convert nitrogen into ammonium [55]. Following digestion, the ammonium was distilled and titrated to quantify the total nitrogen content in the sample. HN was assessed using a hydrochloric acid hydrolysis procedure [56]. In this method, soil samples were treated with acidic solutions to release nitrogen from organic compounds. The resulting ammonium can then be quantified using colorimetric methods or ion-selective electrodes, providing a reliable measure of HN. The determination of AK was achieved through the ammonium acetate extraction-flame photometry method [57]. During this procedure, the soil sample is shaken with an ammonium acetate solution, which displaces potassium ions from the soil matrix. The concentration of potassium in the extraction solution is subsequently measured using atomic absorption spectrophotometry or flame photometry. For AP, extraction was performed using the Olsen method, which employed a sodium bicarbonate solution [58]. The concentration of phosphorus in the extract was determined through colorimetric methods, allowing for an accurate assessment of phosphorus availability in the soil.

2.2. UV-Vis-NIR and MIR Spectral Data Measurement

The UV-Vis-NIR spectral data of the 109 soil samples were measured in the laboratory using a UV-3600IPLUS, 220C (A12615900129) spectrometer (Shimadzu, Kyoto, Japan). The instrument is equipped with a deuterium lamp (200–330 nm) and a tungsten-halogen lamp (330–2500 nm), enabling it to cover the entire spectral range from UV to NIR. A spectral resolution of 2 nm was employed, generating a total of 1151 spectral data points. The type of photometric value was set to reflectance, which can indicate the presence of specific soil constituents, such as organic matter or minerals. Prior to the measurements, the instrument was calibrated using a barium sulfate (BaSO_4) white reference plate with near 100% reflectance. The reflectance values of the soil samples in each spectral band were recorded by calculating the ratio of the sample's reflectance intensity to the standard white reference, and the scanning was repeated three times for each sample, with the average taken as the final reflectance spectral data.

The MIR spectral data for the soil samples were acquired using an IRAffinity-1S Fourier-transform infrared (FTIR) spectrometer (Shimadzu, Kyoto, Japan). Spectral measurements were conducted across a wavenumber range of $4000\text{--}400\text{ cm}^{-1}$, with the spectral resolution set to 2 cm^{-1} . In this study, transmission MIR spectroscopy was employed to obtain transmittance spectral data for soil samples, which is particularly suitable for capturing the detailed transmittance characteristics of soil components, providing distinct peaks that correlate to molecular vibrations. Similar to the UV-Vis-NIR measurements, a reference sample of potassium bromide (KBr) powder, exhibiting negligible absorption

in the mid-infrared region, was utilized for background scanning. The intensity signals from both the measurement sample and the reference sample were converted to a digital format and subsequently subtracted, yielding the final MIR spectra. We opted to calculate the average of the three scans to minimize the variability caused by environmental factors during measurements and to provide a more stable representation of each sample's spectral signature.

2.3. Spectral Data Preprocessing

The UV-Vis-NIR and MIR spectral profiles underwent smoothing via the Savitzky–Golay (SG) method employing a window size of 15 and a polynomial order of 2 [59]. Subsequently, the raw spectra were subjected to 0-1 order derivative transformations at 0.2 intervals using the Grünwald–Letnikov fractional order derivative (FOD) approach [60]. This fractional order derivative technique, unlike conventional integer order derivatives, such as first or second order, enhances the detection of subtle spectral absorption features, thereby facilitating the extraction of more informative data [61]. To optimize the preprocessing of the spectra, a PLSR model was constructed using spectra data transformed by different FODs. A 10-fold cross-validation approach was employed, where the dataset was divided into ten subsets, with each subset iteratively serving as a testing set while the remaining subsets were used for training [62]. This process involved averaging the results of each iteration and minimizing the prediction error, assessed through the root mean square error (RMSE) metric while maximizing the coefficient of determination (R^2) as much as possible. By systematically evaluating the performance of different FOD transformations, we were able to identify the most effective FOD for spectral enhancement and subsequent model development.

2.4. Feature Selection

Principal component analysis (PCA) is a powerful statistical technique used for dimensionality reduction and feature selection, which is particularly suited for high-dimensional spectral data, such as UV-Vis-NIR and MIR spectra. By transforming the original features into orthogonal principal components (PCs) that capture the maximum variance in the data, which are ordered by the amount of variance they explain in the data. The key steps include data standardization, covariance matrix computation, eigenvalue decomposition, selection of PCs based on variance explained, data transformation, and interpretation of the most informative features identified through the PCs [63,64]. For UV-Vis-NIR and MIR data, PCA can help identify the most informative wavelength regions that contribute the most to the overall variance in the data, which can be useful for tasks like quantitative analysis from points to intervals.

2.5. Data Fusion Strategy

2.5.1. Series Splicing (SS)

To construct the new feature matrix, we connected features selected from the UV-Vis-NIR and MIR spectral data that explain the maximization of the variance of the data. This approach involves initially identifying the most informative features from each spectral range using optimal feature selection methods. Once the features are determined, they are systematically combined in a sequential manner, as illustrated in Figure 1, resulting in a comprehensive feature matrix that integrates the significant spectral characteristics from both UV-Vis-NIR and MIR spectra. This concatenated feature matrix is expected to enhance the overall predictive performance by leveraging the complementary information provided by the distinct spectral regions.

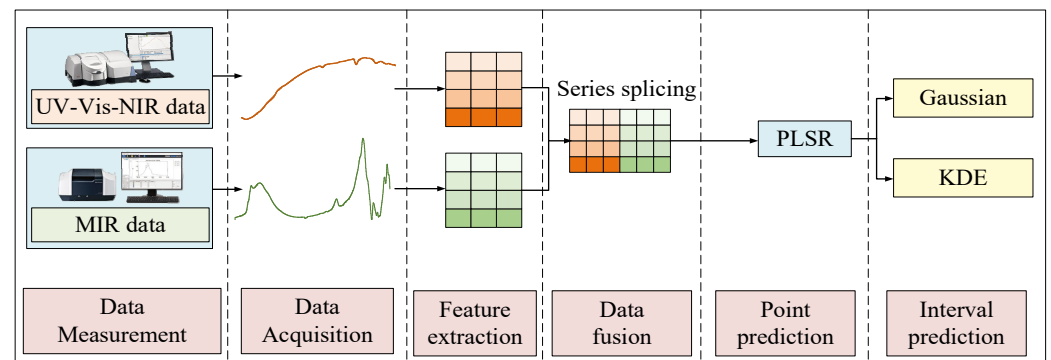


Figure 1. The structure of series splicing based on multi-source spectral data.

2.5.2. Outer-Product Analysis (OPA)

The principle of OPA involves creating a new feature set by calculating the outer product of the feature vectors from two sets of spectral data [65]. The outer product is a matrix that captures the interactions between every pair of features from the two datasets, potentially enhancing the discriminative power of the combined data [66]. The specific implementation methodology is detailed in Figure 2. The band numbers of UV-Vis-NIR and MIR were set as n_1 and n_2 , respectively, with the total number of samples being m . To generate a high-dimensional feature matrix, each sample undergoes OPA, which involves the outer product of the UV-Vis-NIR and MIR spectral matrices. Where $U_{std}[i]$ and $M_{std}[i]$ represent the UV-Vis-NIR and MIR spectral matrices of the i th sample, respectively, and $*$ denotes the outer product operation. The outer product vectors of all samples are then stacked into a matrix to obtain the fused high-dimensional feature matrix Z_i .

$$Z_i = U_{std}^T[i, :] \otimes M_{std}[i, :] \quad (i = 1, 2, \dots, m) \quad (1)$$

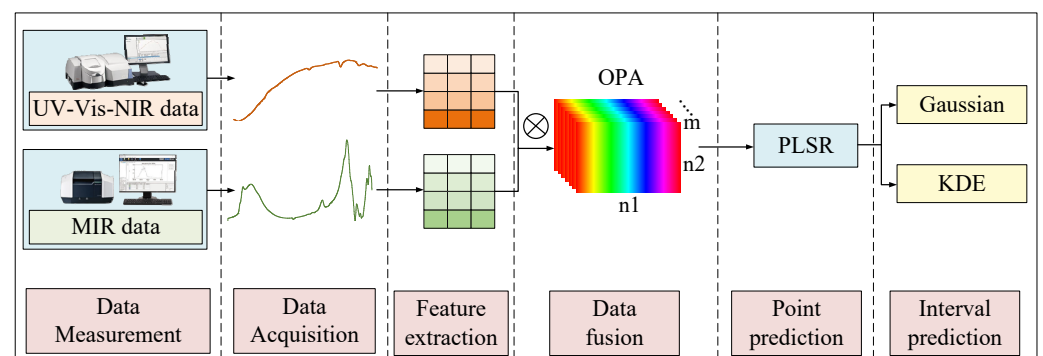


Figure 2. The structure of the OPA based on multi-source spectral data.

2.5.3. Granger–Ramanathan Averaging (GRA)

The GRA algorithm for data fusion relies on predictions derived from single-sensor UV-Vis-NIR and MIR spectral data. This approach involves assigning weights to each spectral band based on the outcomes of the Granger causality test, which is typically executed through regression analysis [67]. The final output is generated by constructing a linear model that correlates the fused prediction results with the individual predictions obtained from the different sensors. The structure of this is shown in Figure 3.

$$Y = W_0 + (W_1 \times X_1) + (W_2 \times X_2) \quad (2)$$

Here, the fused prediction result is denoted as Y . The weights, represented by W_1 and W_2 , are ascertained through the Granger causality test, encapsulating the significance of each band in the predictive process. Concurrently, X_1 and X_2 symbolize the projections

of soil properties predicated on UV-Vis-NIR and MIR data correspondingly, W_0 is the intercept, embodying the distinct informational contributions of these spectral ranges.

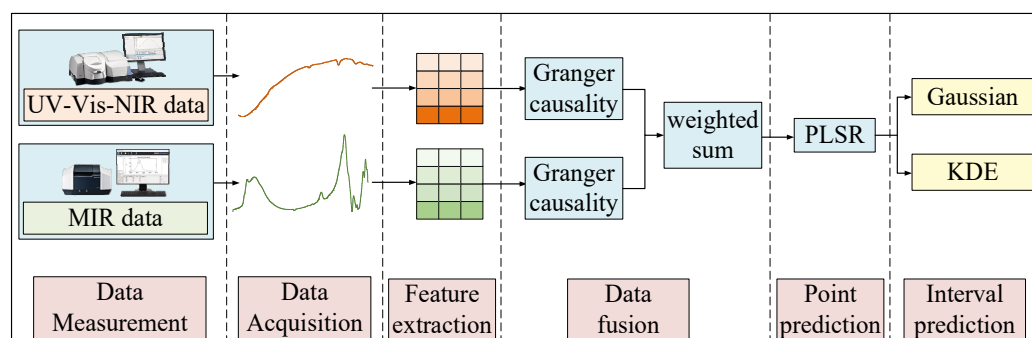


Figure 3. The structure of the GRA based on multi-source spectral data.

2.6. Point Prediction Model

The initial phase of modeling involves point prediction utilizing machine learning techniques. Prior to constructing the model, the Kennard–Stone (K-S) algorithm was employed to divide the dataset of all soil properties into a training set comprising 65 samples (60%), a testing set comprising 22 samples (20%), and a validation set comprising 22 samples (20%) [68]. Specifically, the training set was employed to fit the model, allowing it to learn patterns and relationships within the data. Following this, the validation set was used to fine-tune the model's hyperparameters and provide an initial assessment of its performance. The testing set was reserved for the ultimate stage of assessment, where it serves as the ultimate benchmark for validating the model's robustness and ensuring that it can perform accurately in real-world scenarios. PLSR was then applied to develop a predictive model for estimating soil property contents at the point level. As one of the most prevalent modeling methods in spectral analysis, PLSR effectively addresses the high dimensionality and multicollinearity inherent in spectral data. It emphasizes the relationships among variables, performs well with small sample sizes, and possesses certain nonlinear modeling capabilities. A critical parameter in constructing a PLSR model is determining the optimal number of latent variables (LVs), which significantly influences the model's performance. In this study, we focused on identifying the optimal number of principal component factors (PCFs). To achieve this, we employed a 10-fold cross-validation approach. For each model configuration, we incrementally increased the number of PCFs and evaluated the predictive performance using the test set. The primary goal was to minimize the RMSE, a key indicator of the model's prediction accuracy. Through this iterative process, we were able to identify the optimal number of PCFs that resulted in the lowest prediction error.

2.7. Interval Prediction Model

In the next phase of modeling, probability interval prediction is achieved through the analysis of the absolute errors associated with point predictions. In this study, we aimed to validate the model's effectiveness and the non-parametric estimation approach by employing both Gaussian probability interval prediction and Kernel density estimation (KDE) interval prediction for comparing interval prediction results at various confidence levels. Specifically, the Gaussian interval prediction utilizes statistical analysis of the standard deviation of the absolute errors from the validation set to derive the upper and lower bounds of the confidence intervals [69].

KDE is fundamentally based on inferring the distribution of an entire dataset from a finite number of samples. It is a non-parametric method for estimating the probability density function of a random variable, particularly useful when the underlying distribution is unknown or does not conform to standard parametric distributions [70,71]. KDE is employed to predict probability intervals, which serve to estimate the upper and lower bounds of soil property contents at a specified confidence level. This approach offers

valuable insights into the distribution and variability of property content across different soil samples.

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (3)$$

In this context, $f(x)$ represents the density function estimated at a specific point. The variable n denotes the total number of observations in the dataset, while X_i refers to the individual observed values. The parameter h indicates the bandwidth.

2.8. Model Construction and Evaluation

2.8.1. Evaluation Indicators of Point Prediction

The evaluation criteria for assessing the point prediction model include the coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE) [72]. During the modeling process, each dataset—training, validation, and testing—was evaluated using performance metrics. These stages form a coherent sequence. However, the final evaluation of the model is primarily based on its performance on the testing set, as this provides the most reliable measure of its predictive power when applied to unknown samples. It is worth noting that R_C^2 , $RMSE_C$, and MAE_C refer to the R^2 , RMSE, and MAE for the training set. Similarly, R_V^2 , $RMSE_V$, and MAE_V represent the same metrics for the validation set, while R_P^2 , $RMSE_P$, and MAE_P indicate the corresponding values for the testing set. In this context, a higher R^2 value, approaching 1, alongside minimized RMSE and MAE values, signifies that the model is more accurate and performs better. Their mathematical expressions are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - f_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - f_i)^2} \quad (5)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - f_i| \quad (6)$$

where y_i represents the true value, f_i denotes the predicted value, $\bar{y}$ denotes the measurement on average, and n indicates the number of data points.

2.8.2. Evaluation Indicators of Interval Prediction

The prediction interval coverage probability (PICP) and prediction interval normalized average width (PINAW) are widely employed metrics to evaluate the accuracy and quality of probabilistic interval forecasting results [73]. The PICP represents the likelihood that an actual observation falls within the predicted interval, whereas PINAW serves as a measure to assess the normalized average width of the probabilistic interval prediction model. However, as these two metrics tend to exhibit a somewhat contradictory relationship, the coverage width criterion (CWC) is necessary to concurrently consider the degree of coverage and narrowness of the prediction intervals. Ideally, a high-quality prediction interval should achieve a PICP as close to 1 as possible while maintaining minimal values for both PINAW and CWC. The calculation formulae for CWC are as follows:

$$CWC = PINAW \left[1 + \gamma_{PICP} e^{-\eta(PICP - \mu)}, \eta > 0 \right] \quad (7)$$

$$\gamma_{PICP} = \begin{cases} 0 & PICP \geq \mu \\ 1 & PICP < \mu \end{cases} \quad (8)$$

where η is a parameter greater than 0, with this paper using a value of 1, and μ is the given confidence level.

3. Results and Discussion

3.1. Statistical Distribution of Chemical Testing Results for Key Soil Properties

The statistical analysis presented in Table 1 reveals notable variability and distribution patterns in the contents of various soil properties. Among these, the contents of SOM and TN exhibit a moderate range of variability, indicating a relatively consistent presence of these properties across the dataset. This consistency is crucial as it contributes to the stabilization of soil fertility and enhances the inherent control over soil property dynamics. In contrast, the statistical analysis of HN, AK, and AP demonstrates significant variability and pronounced fluctuations. The extensive ranges of these properties suggest they are more susceptible to environmental influences and management practices. Factors such as soil type, microbial activity, and prevailing environmental conditions can significantly impact the availability of these properties, highlighting their dynamic nature. The results indicate notable variability in the distribution of different soil property contents, which reflects geographical representation.

Table 1. Distribution statistics of five soil properties.

Soil Property	Max	Min	Mean	SD
SOM (g/kg)	25.50	11.50	19.27	3.33
TN (g/kg)	3.04	0.78	1.20	0.30
HN (mg/kg)	1389.11	86.38	163.11	177.77
AK (mg/kg)	138.00	573.00	233.66	68.94
AP (mg/kg)	6.90	56.00	22.11	9.46

3.2. UV-Vis-NIR and MIR Spectral Data Acquisition

Figure 4 presents the preprocessed spectral curves of the UV-Vis-NIR and MIR regions, which align well with findings reported in the related literature. It is important to highlight that the transmittance values from the raw MIR data should be transformed into absorbance values in accordance with the Lambert–Beer law [74]. This transformation is crucial, as it more accurately reflects the light absorption characteristics of the sample, thereby enhancing the interpretation of the soil’s chemical composition and properties.

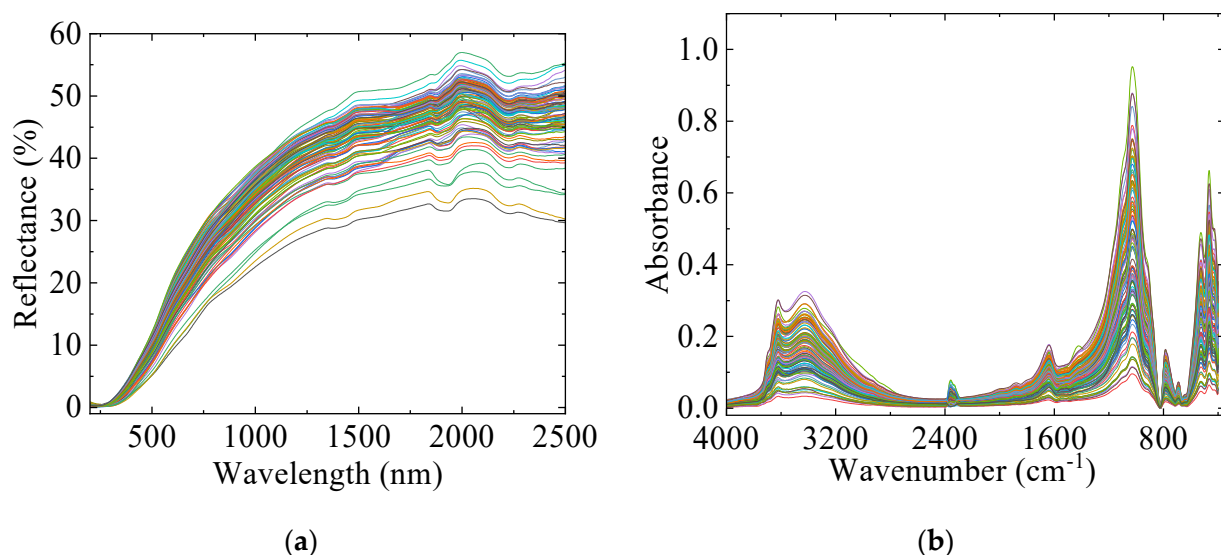


Figure 4. UV-Vis-NIR and MIR reference spectral curves of soil samples: (a) UV-Vis-NIR; (b) MIR.

3.3. Estimation of Soil Property Contents Using Different FODs

An estimation model utilizing PLSR was developed to predict key property contents based on full-spectrum 0–1st order (0.2-order interval) derivatives, with the results outlined

in Tables 2 and 3, which attached the values of PCFs for each PLSR model. The findings demonstrate that all FOD transformations led to varying degrees of enhanced model accuracy compared to the original reflectance (order = 0), indicating the efficacy of FODs in mitigating baseline interferences while amplifying spectral characteristics. One of the most noteworthy is the PLSR model, built on UV-Vis-NIR spectral data transformed by the first-order derivative, which exhibited the highest predictive accuracy, showcasing improved R^2 values and reduced RMSE in the test set when contrasted with the original reflectance spectra. Similarly, the PLSR model utilizing MIR spectral data subjected to 0.6th-order derivative preprocessing showed the highest prediction accuracy, with enhanced R^2 values and reduced RMSE in the test set relative to the original absorbance spectra. To delve deeper into the property estimation mechanisms across different FOD preprocessing methods, the correlation trends between full-band spectra and each property content at various FOD levels were scrutinized. The Pearson correlation coefficient (PCC) was utilized to quantify the strength of the linear relationship between spectral features and soil properties. The closer the PCC's absolute value is to 1, the stronger the correlation. This analysis is detailed in Figure 5, offering insights into the relationships between spectral features and property content variations.

Table 2. Point prediction results of different FODs on key soil property contents in the UV-Vis-NIR band.

Soil Property	FOD	Number of PCFs	Testing Set		
			R_p^2	RMSE _p	MAE _p
SOM	0.0	12	0.71	2.15	1.63
	0.2	21	0.77	2.29	1.62
	0.4	16	0.70	1.92	1.37
	0.6	14	0.77	1.51	1.02
	0.8	14	0.71	1.59	1.07
	1.0	18	0.79	1.29	1.17
TN	0.0	17	0.73	0.10	0.07
	0.2	12	0.70	0.15	0.11
	0.4	25	0.71	0.09	0.07
	0.6	18	0.69	0.11	0.09
	0.8	28	0.74	0.17	0.12
	1.0	22	0.78	0.16	0.14
HN	0.0	19	0.64	50.64	31.93
	0.2	18	0.70	48.29	32.42
	0.4	15	0.69	48.45	33.90
	0.6	27	0.73	191.83	81.34
	0.8	19	0.75	45.35	24.15
	1.0	21	0.72	5.27	4.26
AK	0.0	18	0.63	38.41	30.30
	0.2	19	0.61	39.02	31.25
	0.4	17	0.62	29.44	27.11
	0.6	26	0.65	36.87	29.94
	0.8	22	0.65	28.33	25.13
	1.0	28	0.67	27.34	23.80
AP	0.0	35	0.37	8.75	6.61
	0.2	22	0.41	4.69	3.86
	0.4	18	0.31	5.27	3.90
	0.6	35	0.40	8.38	6.07
	0.8	33	0.43	7.50	6.36
	1.0	24	0.42	4.69	3.85

Table 3. Point prediction results of different FOD derivatives on key soil property contents in the MIR band.

Soil Property	FOD	Number of PCFs	Testing Set		
			R _P ²	RMSE _P	MAE _P
SOM	0.0	18	0.62	1.84	1.37
	0.2	22	0.61	2.16	1.84
	0.4	16	0.65	2.03	1.65
	0.6	22	0.72	1.52	1.25
	0.8	20	0.71	1.35	1.09
	1.0	18	0.64	2.12	1.47
TN	0.0	21	0.58	0.24	0.16
	0.2	14	0.61	0.13	0.10
	0.4	21	0.65	0.24	0.15
	0.6	25	0.67	0.21	0.15
	0.8	19	0.61	0.30	0.16
	1.0	22	0.64	0.09	0.07
HN	0.0	18	0.65	116.50	68.29
	0.2	27	0.53	101.49	48.39
	0.4	25	0.70	109.36	53.51
	0.6	19	0.66	17.95	13.44
	0.8	24	0.62	205.11	125.64
	1.0	30	0.48	262.37	114.34
AK	0.0	31	0.58	37.45	32.44
	0.2	26	0.49	31.57	25.01
	0.4	18	0.51	30.74	19.27
	0.6	19	0.59	32.09	27.61
	0.8	15	0.46	39.57	33.45
	1.0	21	0.57	32.88	27.72
AP	0.0	22	0.28	5.39	4.06
	0.2	28	0.26	6.93	5.84
	0.4	23	0.31	5.17	4.48
	0.6	21	0.40	4.80	4.34
	0.8	25	0.32	5.26	4.53
	1.0	32	0.30	8.75	8.09

The PCC analysis between the UV-Vis-NIR spectral bands and the SOM, TN, HN, AK, and AP contents reveals a gradual improvement in the correlation as the derivative order increases (Figure 5a,c,e,g,i). At a fractional order derivative of 1, the distribution of strongly correlated bands exhibits distinctive patterns compared to other derivative orders. Previous research suggests that SOM, TN, and HN exhibit specific spectral response bands within the UV–visible–near-infrared range. In detail, bands strongly correlated with SOM content tend to cluster around wavelengths such as 825 nm, 853 nm, 1300–1400 nm, 1650 nm, and 2200–2300 nm, reflecting the stretching vibration and bending vibrations of functional groups like C-H, C-O, and C=O [75,76]. Similarly, bands correlated with soil TN and HN show similarities, particularly near 1900–2180 nm and 1500 nm, associated with vibrational absorptions from amino acids, amine compounds, and their N-H and O-H groups present in the soil [77]. The correlation bands for soil AK and AP partly overlap with important SOM bands, demonstrating varying degrees of overlap. When focusing on the MIR spectral data, the strong correlation bands for TN and HN exhibit pronounced distinctions at fractional order derivatives of 0.4 and 0.6, concentrating around wavelengths like 3350 cm^{−1}, 3096 cm^{−1}, 2930 cm^{−1}, 2860 cm^{−1}, and 2520 cm^{−1}, linked to vibrational absorptions of N-H, C-H, and other organic matter components [78,79]. Although the distributions of strong correlation bands for SOM, AK, and AP at fractional order derivatives do not differ significantly and exhibit varying degrees of overlap, specific bands at 1550 cm^{−1}, 784 cm^{−1}, and 768 cm^{−1} show notable correlations at a fractional order

derivative of 0.6. In summary, it is evident that selecting the first-order derivative for UV-Vis-NIR and the 0.6-order derivative for MIR is optimal for derivative[s] transformations, offering enhanced insights into the relationships between spectral features and property contents in soil samples.

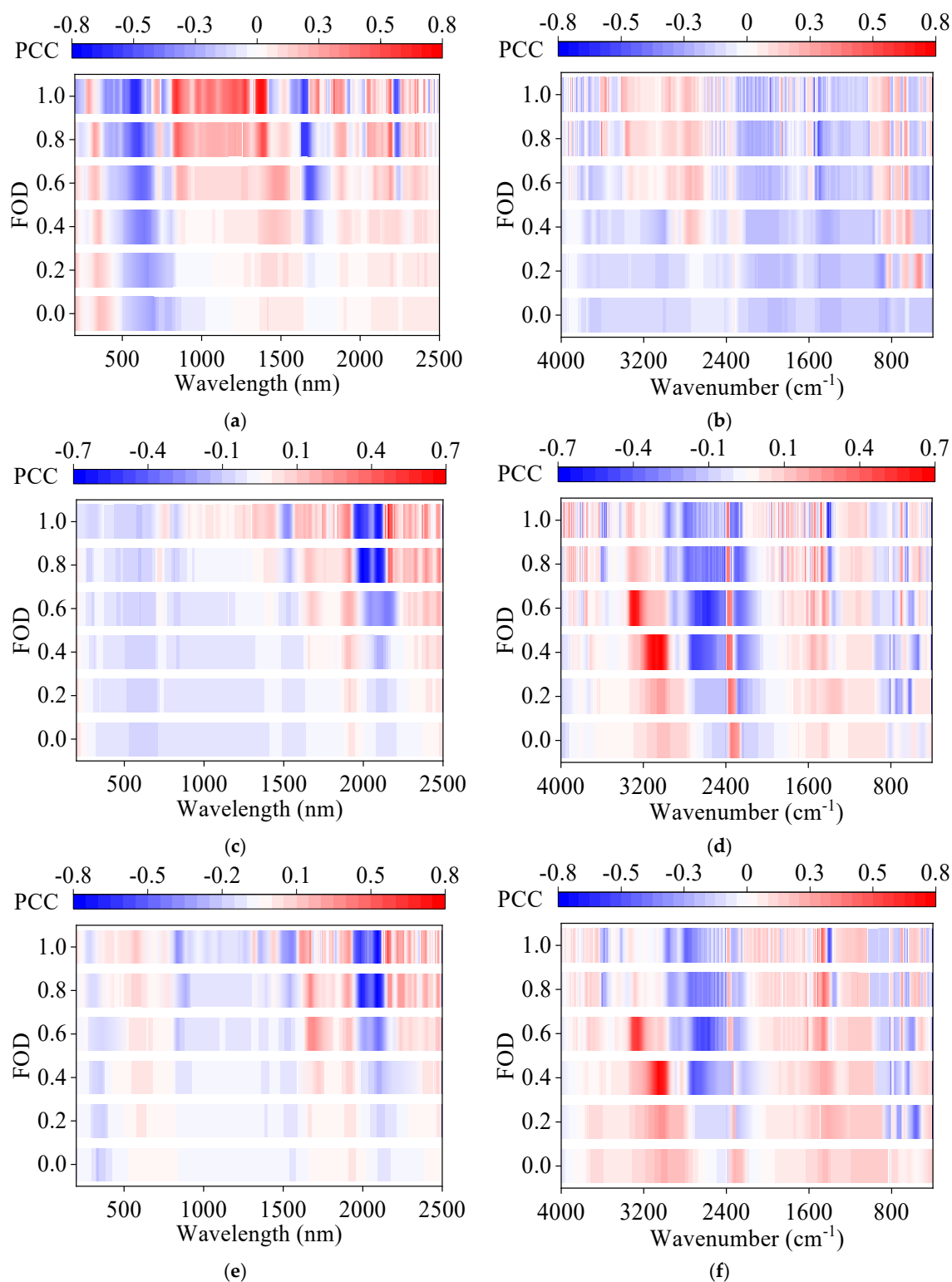


Figure 5. Cont.

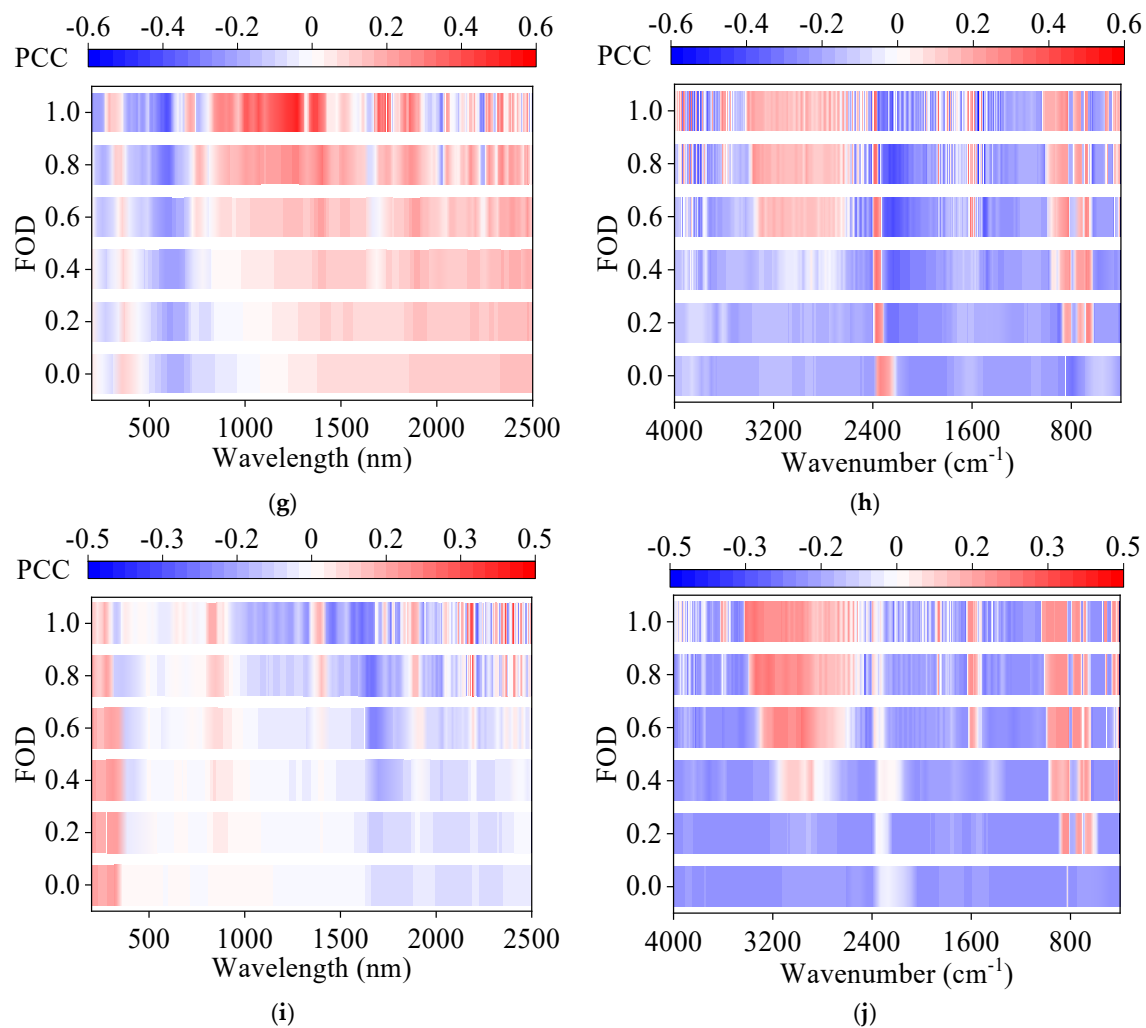


Figure 5. The trend of correlation between the full-band spectrum and the property contents under different FODs: (a) trend of correlation between the UV-Vis-NIR band and SOM; (b) trend of correlation between the MIR band and SOM; (c) trend of correlation between the UV-Vis-NIR band and TN; (d) trend of correlation between the MIR band and TN; (e) trend of correlation between the UV-Vis-NIR band and HN; (f) trend of correlation between the MIR band and HN; (g) trend of correlation between the UV-Vis-NIR band and AK; (h) trend of correlation between the MIR band and AK; (i) trend of correlation between the UV-Vis-NIR band and AP; (j) trend of correlation between the MIR band and AP.

3.4. UV-Vis-NIR and MIR Spectral Feature Matrices Acquisition

The results of the PCA transformation within the selected wavelength range are depicted in Tables 4 and 5, showcasing the number of PCs derived and the contribution rate of each component. In the UV-Vis-NIR spectral domain, the initial 12 PCs extracted post-differentiation encapsulate 90.21% of the total variance. Transitioning to the mid-infrared spectrum, the cumulative contribution of the first five PCs delineates 99.18% of the total variance within the empirically observed spectral data order of 0.6. Figure 6 presents the contributions of the first two principal components alongside the bands of the observed corresponding peaks, following the PCA transformation conducted within the selected wavelength ranges encompassing UV-Vis-NIR and MIR spectra.

Table 4. The results of the PCA transformation in the UV-Vis-NIR spectral band.

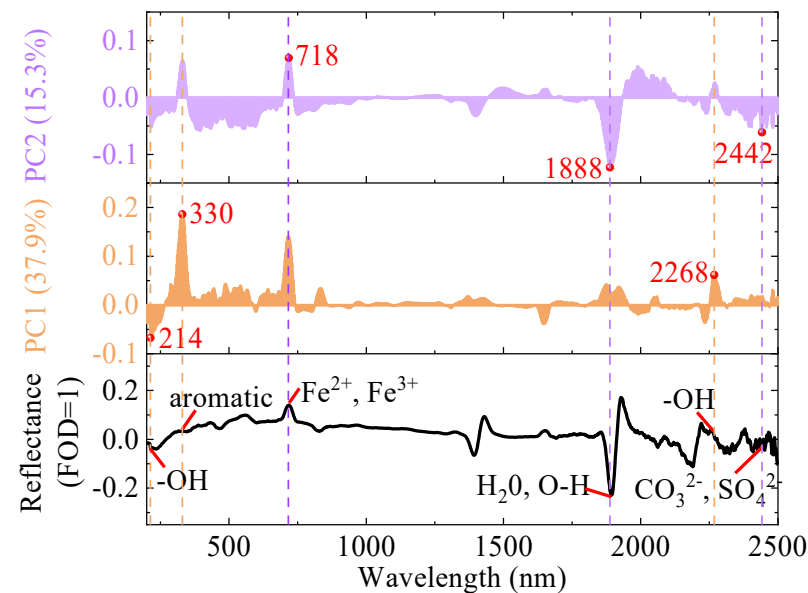
PCs	Contribution Rate (%)	Cumulative Contribution Rate (%)
PC1	37.85	37.85
PC2	15.27	53.12
PC3	9.62	62.74
PC4	9.21	71.95
PC5	6.80	78.76
PC6	3.96	82.72
PC7	1.98	84.70
PC8	1.70	86.40
PC9	1.13	87.83
PC10	0.93	88.46
PC11	0.89	89.35
PC12	0.86	90.21

Table 5. The results of the PCA transformation in the MIR spectral band.

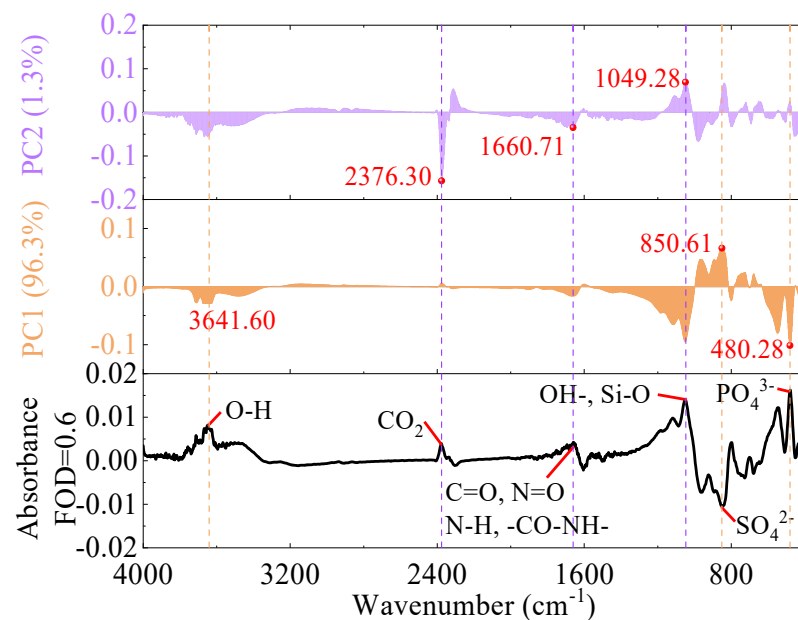
PCs	Contribution Rate (%)	Cumulative Contribution Rate (%)
PC1	96.32	96.32
PC2	1.30	97.62
PC3	0.85	98.47
PC4	0.41	98.88
PC5	0.30	99.18

The position and intensity of the principal component peaks reflect the primary transformation direction of the data and the degree to which the original variables contribute along the principal axes. In Figure 6a, the principal component peaks derived from the first-order spectral data in the UV-Vis-NIR range align closely with the first-order reflectance troughs (or peaks). Among the UV-Vis region, these features are primarily associated with electronic transitions [80]. In contrast, the key features in the NIR region correspond to overtones and combination bands of fundamental vibrational modes typically observed in the MIR region. Rather than capturing the fundamental vibrations directly, the NIR region provides indirect vibrational information by reflecting higher-order harmonics and combinations of these molecular interactions [81]. Specifically, the troughs and peaks observed in the UV-Vis spectra (200–780 nm) likely result from the influence of organic substances and oxygenated compounds present in the soil. Notably, the molecular structures of lignin, humus, or aromatic compounds contain numerous benzene rings and hydroxyl groups, which contribute to these spectral features. Additionally, iron-containing oxides in the soil, such as hematite and magnetite, contribute to these signals through the presence of Fe^{2+} and Fe^{3+} ions. The troughs and peaks primarily arise from $\pi \rightarrow \pi^*$ transitions within the aromatic ring structures or $n \rightarrow \pi$ transitions associated with iron oxides [82]. In the near-infrared spectrum (780–2500 nm), the signals not only reflect the O-H stretching vibrations related to water molecules prevalent in both the atmosphere and soil but also correspond to the vibrational absorption of hydroxyl groups (-OH) in clay minerals, such as montmorillonite and illite [81]. Furthermore, these spectral features may indicate the characteristic absorption of carbonates, sulfates, and other mineral constituents within the near-infrared wavelength range [83]. Compared to UV-Vis-NIR spectral data, MIR spectral data are directly related to the stretching and vibrational characteristics of chemical bonds and functional groups, offering a more nuanced understanding of the molecular structures of both organic and inorganic components. As illustrated in Figure 6b, alongside the characteristic bands for water molecules, several prominent absorption peaks are observed at 1660.71 cm^{-1} and 2376.30 cm^{-1} . These peaks are closely associated with the stretching and deformation modes of various chemical functional groups present in SOM, TN, and HN, including N-H, C=O, -CO-NH-, and N=O groups [84,85]. Furthermore, the absorption peaks associated with O-H, Si-O, PO_4^{3-} , SO_4^{2-} , and other functional groups

found in clay minerals, such as kaolinite and montmorillonite, at 1049.28 cm^{-1} , 850.61 cm^{-1} , and 480.28 cm^{-1} exhibit a strong correlation with the mobile forms of AK and AP [86,87]. Consequently, the final spectral signature matrix comprises a UV-Vis-NIR spectral matrix sized at 109×12 and an MIR spectral matrix sized at 109×5 , encapsulating the essence of the molecular interactions and characteristics across the different spectral regions.



(a)



(b)

Figure 6. Load plots of the first two principal components for the full-band spectral data: (a) UV-Vis-NIR; (b) MIR.

3.5. Comparing Point Prediction Outcomes Across Various Fusion Modes

The process of point prediction modeling encompasses the amalgamation of spectral data originating from diverse sources into three fusion models aimed at prognosticating the soil properties within a specified dataset, in conjunction with PLSR. Within this research

framework, focusing on the minimization of the RMSE through a rigorous 10-fold cross-validation technique. This method was instrumental in pinpointing the optimal number of PCFs, thereby enhancing the efficacy and accuracy of the PLSR model.

Among the three fusion strategies considered, the results of the point prediction modeling are detailed in Table 6. Notably, the OPA fusion method demonstrated superior prediction accuracy compared to the other fusion modes. When assessed against models utilizing single spectral data (as shown in Tables 2 and 3), the prediction accuracy for the five key properties showed varying degrees of improvement across the board. Specifically, the R^2 increased by 0.06 to 0.41. These results underscore the effectiveness of employing multi-source data in predicting key soil property content, highlighting the advantages of integrating diverse data streams to enhance model performance. In the context of UV-Vis-NIR and MIR spectral data, OPA serves to enhance feature interactions through outer product operations, potentially unveiling more intricate and insightful combinations of spectral features [65]. This enhancement allows for the capture of complex nonlinear relationships between the two spectral matrices, a crucial aspect when dealing with highly correlated or interactive spectral datasets. In contrast to direct series fusion methods, the characteristic matrix resulting from OPA fusion better encapsulates the underlying structure of the spectral data, thereby enhancing the overall predictive performance. The GRA method attempts to capture the effects of different spectral features by taking a weighted average approach. However, this linear averaging technique may not be sufficient to fully exploit the potential nonlinear relationships present in complex spectral data.

Table 6. Point prediction results under different fusion strategies.

Fusion Strategy	Soil Property	Testing Set		
		R_p^2	RMSE _p	MAE _p
SS	SOM	0.82	1.05	0.94
	TN	0.79	0.16	0.08
	HN	0.79	16.10	12.58
	AK	0.70	26.06	23.56
	AP	0.41	7.11	5.47
OPA	SOM	0.91	0.37	0.28
	TN	0.90	0.02	0.02
	HN	0.89	2.26	1.74
	AK	0.73	15.53	13.49
	AP	0.53	7.62	5.38
GRA	SOM	0.84	1.12	0.99
	TN	0.84	0.14	0.08
	HN	0.80	11.76	9.20
	AK	0.64	28.71	25.49
	AP	0.50	6.18	4.32

Based on the output of the point prediction model, OPA-PLSR, on the testing set, we present a visual representation of the true and predicted values of soil properties in a fitting graph, as depicted in Figure 7. The fitted line is derived using parameters obtained from the least squares method, a widely used approach in regression analysis that minimizes the sum of the squares of the residuals. The coefficient of determination, R^2 , serves as a statistical measure that quantifies the degree of agreement between the fitted line and the actual data points. A higher R^2 value indicates a stronger correlation, reflecting a closer alignment between the data points and the fitted line.

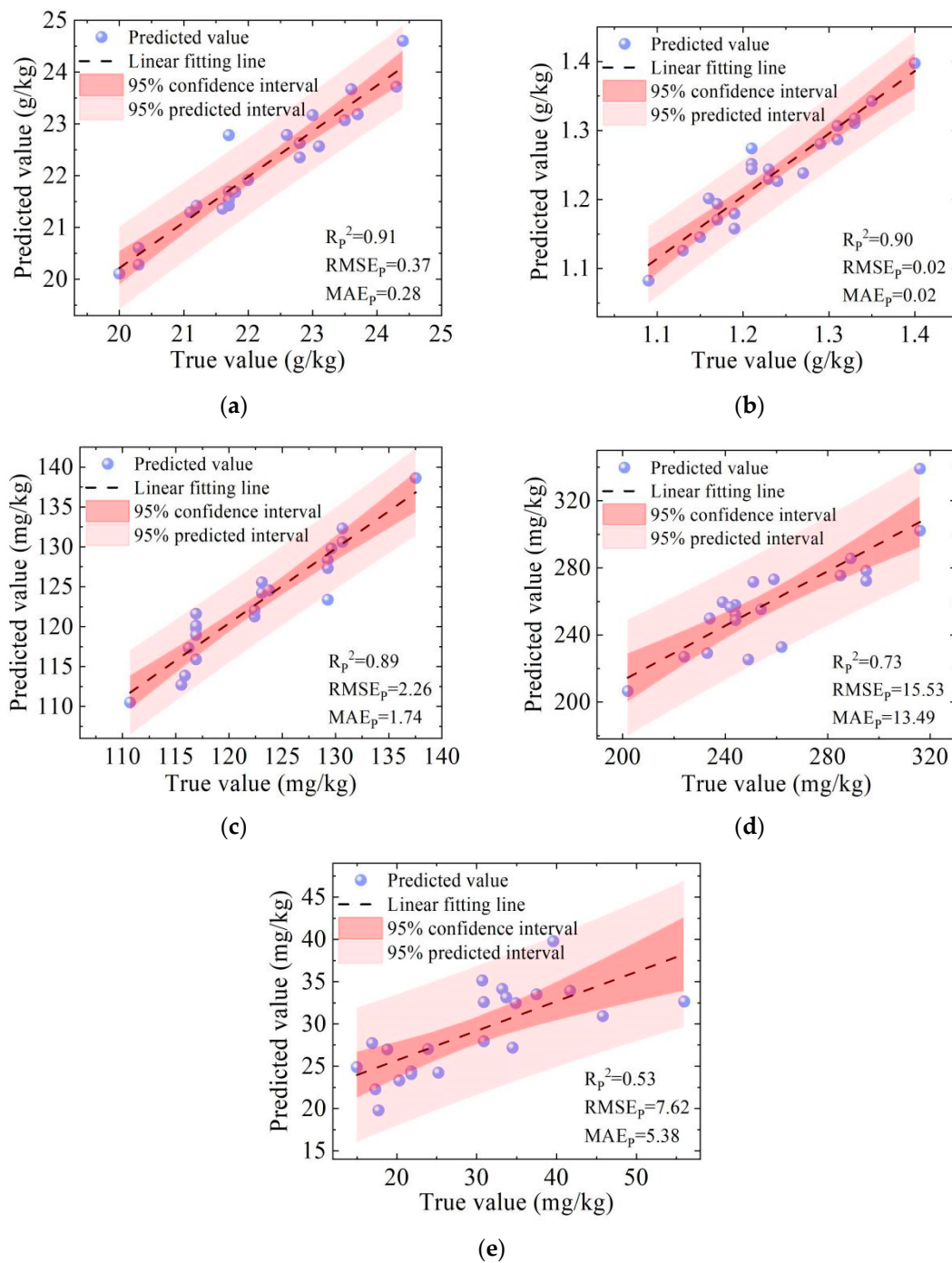


Figure 7. Point prediction results of the testing set based on OPA-PLSR: (a) SOM; (b) TN; (c) HN; (d) AK; (e) AP.

In Figure 7a–c, it is evident that the data points for SOM, TN, and HN are predominantly distributed on either side of the fitted line, demonstrating strong adherence to the predicted trends. The R^2 for these relationships reached impressive values of 0.91, 0.90, and 0.89, respectively. Such results suggest that the model captures the underlying relationships in these soil properties. In contrast, the data points for AK and AP in Figure 7d,e exhibit noticeable deviations from the fitted line. AK in this study can only achieve approximate quantitative detection, with an R^2 of 0.73, while AP can only meet rough accuracy requirements, with an R^2 of 0.53. This discrepancy suggests that while the fitting lines for SOM, TN, and HN effectively capture the underlying trends in the data, the models for AK and AP may require further refinement. This is because AK and AP in

soil do not have inherent spectral response characteristics, and their prediction is mostly based on the relationship with other soil properties that exhibit strong spectral responses, such as water, SOM, and clay minerals [75,88]. The correlation heat map (Figure 8) reveals an absolute correlation value of 0.55 between SOM and AK, suggesting that the spectral response characteristics of SOM may indirectly reflect the distribution of AK in the soil. Comparatively, the weak relationships between AP and both SOM and other soil properties indicate that AP availability is likely influenced by factors beyond just the measured soil characteristics [89]. For example, AP dynamics can be heavily impacted by factors such as soil pH, microbial activity, and interactions with other properties. The complex nature of phosphorus cycling in soils makes it more challenging to accurately predict AP content based on more easily measured soil properties like SOM, leading to weaker prediction accuracy for AP compared to other properties.

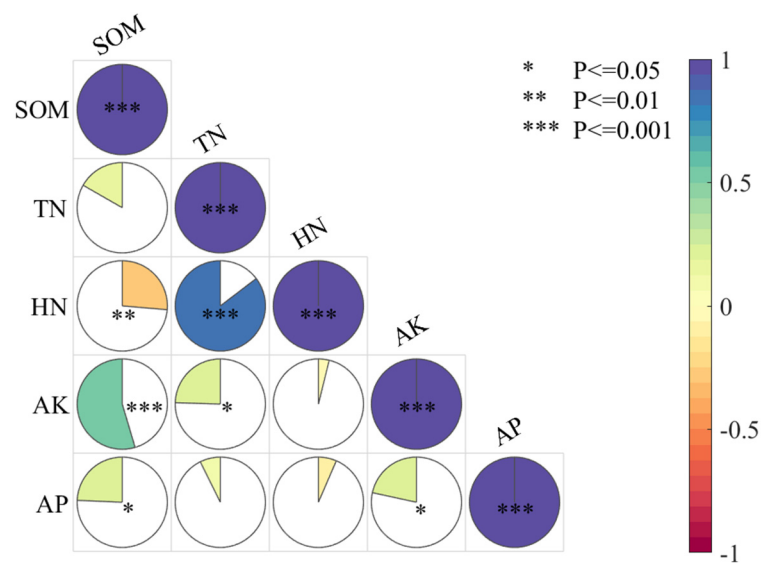


Figure 8. Heat map of the correlations among soil properties. Note: the absolute values of the PCCs are positively associated with the proportion of filled colors within the circle.

3.6. Performance Comparison of Interval Prediction Models

Utilizing the optimal OPA-PLSR model, we used the point errors in the validation set to perform interval predictions. Table 7 presents the outcomes of the Gaussian interval predictions and KDE interval predictions across different confidence levels (95%, 80%, and 65%). The analysis reveals that for SOM, TN, HN, and AK, the PICP obtained through both the Gaussian and KDE methods basically reached the specified confidence levels. This highlights the robustness of the predictions, as it shows that the actual values are highly likely to fall within the predicted intervals. Moreover, the PINAW index indicates that the width of the KDE-generated intervals is generally narrower compared to those produced by the Gaussian method at the same confidence level. For instance, at the 95% confidence level, the PINAW for SOM is 1.7894 for the KDE interval, compared to 1.7591 for the Gaussian interval. A similar pattern is observed for TN, HN, and AK, where the KDE method consistently produces tighter prediction intervals. Focusing on SOM, as a representative case, the KDE method achieved a lower CWC at the 95% confidence level, further demonstrating its advantage in precision. This finding suggested that KDE provided a narrower prediction interval while maintaining prediction accuracy, offering more precise and actionable insights for the management of SOM, TN, HN, and AK contents. This reduction in interval range minimizes the risks associated with decision making and enhances resource utilization efficiency, thereby supporting more effective agricultural practices. In contrast, the PICP for AP fell significantly short of the specified confidence level, indicating that the prediction interval failed to encompass the true values at the expected frequency. In the context of the point prediction results, the accuracy for AP was notably

lower compared to the other properties. This suggests that a higher degree of regression error is often associated with the index of poorer interval prediction in interval modeling. These findings highlighted a notable shortcoming in both the Gaussian probability interval prediction and KDE interval prediction methodologies. It is essential to emphasize that these approaches were insufficient in accurately capturing the inherent variability and fluctuations in the actual values of AP. While the precision of the prediction outcomes fell short of optimal levels, the interval prediction of AP offered an initial evaluation of its concentration, serving as valuable information to guide subsequent soil testing and in-depth analysis.

Table 7. Results of the Gaussian interval prediction and KDE interval prediction at different confidence levels.

Soil Property	Interval Prediction	Confidence Interval	PICP	PINAW	CWC
SOM	Gaussian probability	95%	1.0000	1.7894	1.7894
		80%	0.8182	1.2302	1.2302
		65%	0.7273	0.8947	0.8947
	KDE	95%	1.0000	1.7591	1.7591
		80%	1.0000	1.2007	1.2007
		65%	0.9091	0.8377	0.8377
TN	Gaussian probability	95%	1.0000	2.5094	2.5094
		80%	1.0000	1.7729	1.7729
		65%	1.0000	0.8197	0.8197
	KDE	95%	0.9546	2.2774	2.2774
		80%	0.8636	1.1042	1.1042
		65%	0.7727	0.7591	0.7591
HN	Gaussian probability	95%	0.9091	1.3141	5.7977
		80%	0.9091	0.8593	0.8593
		65%	0.7273	0.6266	0.6266
	KDE	95%	1.0000	2.4107	2.4107
		80%	0.9546	1.7271	1.7271
		65%	0.9546	1.3673	1.3673
AK	Gaussian probability	95%	1.0000	2.0804	2.0804
		80%	1.0000	1.3603	1.3603
		65%	1.0000	0.9920	0.9920
	KDE	95%	0.9546	1.1353	1.1353
		80%	0.8636	0.8773	0.8773
		65%	0.8636	0.7053	0.7053
AP	Gaussian probability	95%	1.0000	1.7305	1.7305
		80%	0.6364	1.1315	154.4700
		65%	0.4546	0.8252	291.2900
	KDE	95%	0.8182	0.7934	42.1890
		80%	0.6818	0.4711	16.7980
		65%	0.5000	0.3223	28.3380

Hence, KDE was employed as an interval prediction model in this study to forecast the concentrations of crucial soil properties, with the prediction outcomes depicted in Figure 9. The graphical representation illustrates that for standard SOM, TN, HN, and AK, the true values predominantly fall within 95% confidence intervals. This alignment signifies robust upper and lower deviation bounds, demonstrating the efficacy of the proposed OPA-PLSR-KDE model in capturing the error distribution.

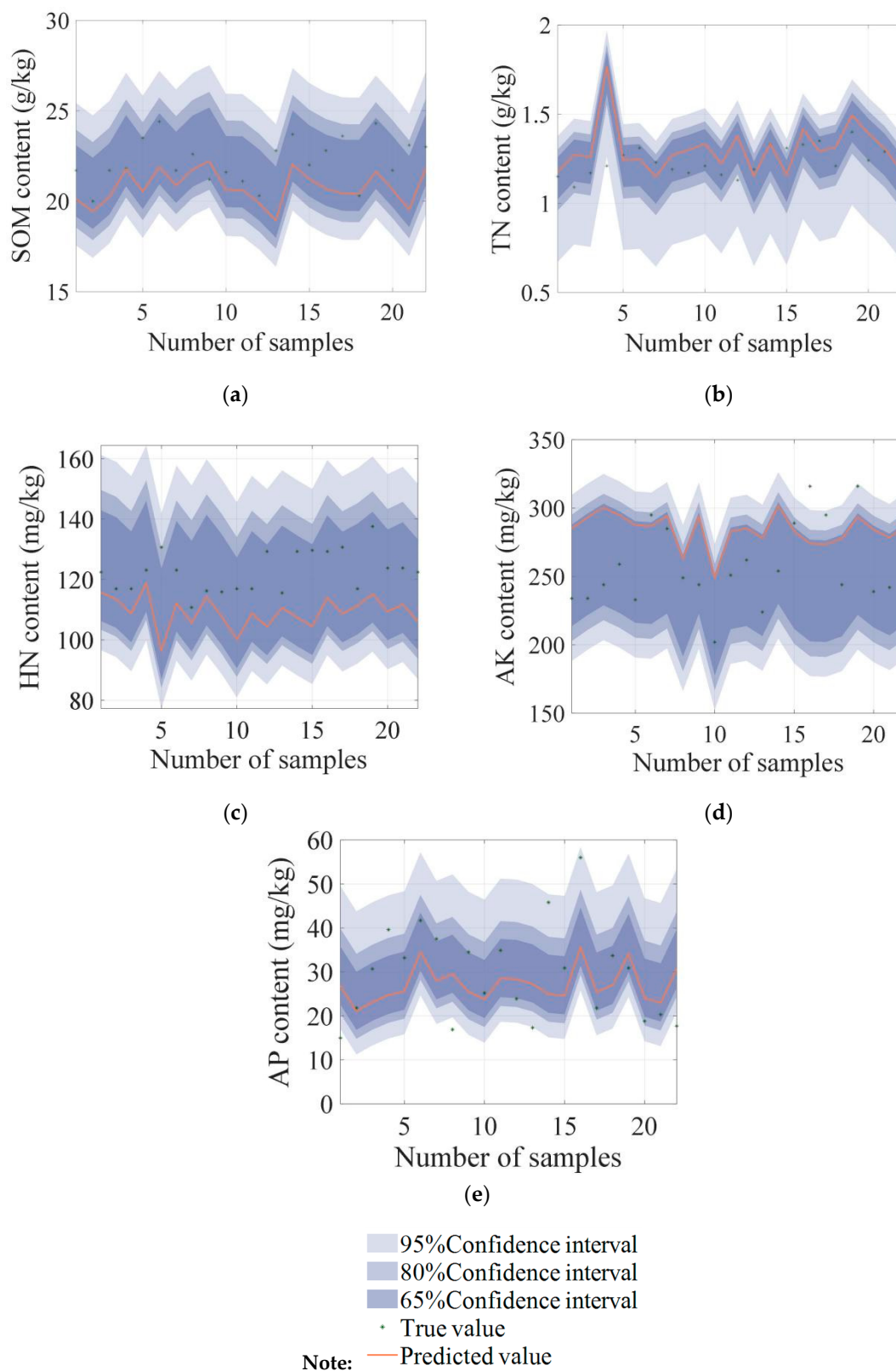


Figure 9. Interval prediction of key soil property contents based on KDE: (a) SOM; (b) TN; (c) HN; (d) AK; (e) AP.

On the one hand, relevant studies have demonstrated that the distribution and behavior of SOM, TN, and HN in soil are relatively stable, primarily influenced by organic matter content and microbial activity [90]. These factors tend to remain consistent over short time

frames, resulting in minimal prediction uncertainty. On the other hand, AK and AP exhibit uneven distributions within the soil and are susceptible to numerous influencing factors, including the soil mineral composition, pH level, and historical fertilization practices [91]. These elements contribute to the complexity of accurately tracking the fluctuations in AK and AP contents, consequently leading to heightened prediction uncertainties.

4. Conclusions

This study aimed to advance the fusion of multi-source spectral data with machine learning algorithms to facilitate precise point-to-interval prediction of unknown soil property concentrations. Initially, the UV-Vis-NIR and MIR datasets were established to organically integrate the relevant feature bands extracted through PCA. This integration provided comprehensive inputs for both point predictions and interval predictions of key soil properties. The prediction models were then validated using the integrated datasets to ensure their robustness and reliability. The results underscore the efficacy of the OPA fusion strategy in significantly enhancing the accuracy of point predictions across the three fusion methods, highlighting the critical role of the integration of diverse spectral data sources in improving prediction performance effectively. Within the OPA-PLSR framework, successful predictions were achieved for SOM, TN, and HN, with correlation coefficients of 0.91, 0.90, and 0.89, respectively. AK also demonstrated approximate quantitative detectability with an R^2 value of 0.73, while AP exhibited a lower R^2 of 0.53, indicating a requirement for further refinement in prediction accuracy. In terms of the interval prediction accuracy, the KDE interval prediction model outperformed the Gaussian probability interval prediction model, demonstrating better CWC at equivalent confidence levels. Reliable interval predictions were attained for SOM, TN, HN, and AK, characterized by narrow average bandwidths and high interval coverage rates. These outcomes effectively quantified the uncertainty surrounding key soil properties, offering a reasonable range of fluctuation in property concentrations. Another thing that should be emphasized is AP, characterized by higher prediction uncertainty. This research not only highlights the potential of advanced spectral fusion techniques in soil property analysis but also advances our understanding of soil property dynamics.

However, a significant limitation of this study is its reliance on a limited number of samples collected in a single year, which may restrict the generalizability of the results. Future research will prioritize expanding soil sample datasets to encompass multiple years, allowing for a more comprehensive understanding of property dynamics over time. For future data processing, priority will be placed on harmonizing the units of spectral data across different spectral bands to ensure consistency and comparability, and effort will be directed toward incorporating comparative analyses of averaged and replicated spectra. Additionally, there will be a focused effort on enhancing the prediction of available nutrients, particularly AK and AP. This will involve refining the related techniques and models to improve the accuracy and reliability of these predictions.

Author Contributions: Conceptualization, S.L., D.H. and Q.Z.; methodology, D.H. and S.W.; software, S.L. and L.F.; validation, Y.X. and Y.C.; investigation, Y.C. and S.W.; resources, D.H. and L.F.; visualization, Y.X. and Q.Z.; writing—original draft preparation, S.L.; funding acquisition, D.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the Key Research and Development Project of the Jilin Provincial Department of Science and Technology (grant number: 20240303041NC).

Data Availability Statement: The data presented in this study are available from the corresponding author upon reasonable request.

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

References

1. Vanlauwe, B.; Bationo, A.; Chianu, J.; Giller, K.E.; Merckx, R.; Mokwunye, U.; Ohiokpehai, O.; Pypers, P.; Tabo, R.; Shepherd, K.D.; et al. Integrated Soil Fertility Management: Operational Definition and Consequences for Implementation and Dissemination. *Outlook Agric.* **2010**, *39*, 17–24. [\[CrossRef\]](#)
2. Shah, F.; Wu, W. Soil and Crop Management Strategies to Ensure Higher Crop Productivity within Sustainable Environments. *Sustainability* **2019**, *11*, 1485. [\[CrossRef\]](#)
3. Pratt, C.; Kingston, K.; Laycock, B.; Levett, I.; Pratt, S. Geo-Agriculture: Reviewing Opportunities through Which the Geosphere Can Help Address Emerging Crop Production Challenges. *Agronomy* **2020**, *10*, 971. [\[CrossRef\]](#)
4. Kirchmann, H.; Thorvaldsson, G. Challenging Targets for Future Agriculture. *Eur. J. Agron.* **2000**, *12*, 145–161. [\[CrossRef\]](#)
5. Lehman, R.; Osborne, S.; McGraw, K. Stacking Agricultural Management Tactics to Promote Improvements in Soil Structure and Microbial Activities. *Agronomy* **2019**, *9*, 539. [\[CrossRef\]](#)
6. Smith, J.; Johnson, A.; Doe, R. Advances in Soil Assessment Techniques. *Soil Sci. Soc. Am. J.* **2015**, *79*, 714–725.
7. Walkley, A.; Black, I.A. An Examination of the Degtjareff Method for Determining Soil Organic Matter and a Proposed Modification of the Chromic Acid Titration Method. *Soil Sci.* **1934**, *37*, 29–38. [\[CrossRef\]](#)
8. Bremner, J.M. Determination of Nitrogen in Soil by the Kjeldahl Method. *J. Agric. Sci.* **1960**, *55*, 11–33. [\[CrossRef\]](#)
9. Wansu, S.; Yijun, Z.; Gang, Y. Determination of Available Phosphorus in Soil. *Instrum. Anal. Monit.* **2010**, *26*, 331–336.
10. Hartemink, A.E.; Minasny, B. Towards Digital Soil Morphometrics. *Geoderma* **2014**, *230*, 305–317. [\[CrossRef\]](#)
11. Li, H.; Ju, W.L.; Song, Y.M.; Cao, Y.Y.; Yang, W.; Li, M.Z. Soil Organic Matter Content Prediction Based on Two-Branch Convolutional Neural Network Combining Image and Spectral Features. *Comput. Electron. Agric.* **2024**, *217*, 108561. [\[CrossRef\]](#)
12. Soriano-Disla, J.M.; Janik, L.J.; Viscarra Rossel, R.A.; Macdonald, L.M.; McLaughlin, M.J. The Performance of Visible, Near-, and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical, Chemical, and Biological Properties. *Appl. Spectrosc. Rev.* **2014**, *49*, 139–186. [\[CrossRef\]](#)
13. Tan, B.H.; You, W.H.; Tian, S.H.; Xiao, T.F.; Wang, M.C.; Zheng, B.T.; Luo, L.N. Soil Nitrogen Content Detection Based on Near-Infrared Spectroscopy. *Sensors* **2022**, *20*, 8013. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Yu, S.Y.; Bu, H.R.; Dong, W.C.; Jiang, Z.; Zhang, L.X.; Xia, Y.Q. Construction and Evaluation of Prediction Model of Main Soil Nutrients Based on Spectral Information. *Appl. Sci.* **2022**, *12*, 6298. [\[CrossRef\]](#)
15. Janik, L.J.; Skjemstad, J.O.; Raven, M.D. Characterization and Analysis of Soils Using Mid-Infrared Partial Least-Squares. 1. Correlations with XRF-Determined Major-Element Composition. *Soil Res.* **1995**, *33*, 637–650. [\[CrossRef\]](#)
16. Bray, J.G.P.; Rossel, R.V.; McBratney, A.B. Diagnostic Screening of Urban Soil Contaminants Using Diffuse Reflectance Spectroscopy. *Aust. J. Soil Res.* **2009**, *47*, 433–442. [\[CrossRef\]](#)
17. Ng, W.; Malone, B.P.; Minasny, B. Rapid Assessment of Petroleum-Contaminated Soils with Infrared Spectroscopy. *Geoderma* **2017**, *289*, 150–160. [\[CrossRef\]](#)
18. Reijneveld, J.A.; van Oostrum, M.J.; Broelsma, K.M.; Fletcher, D.; Oenema, O. Empower Innovations in Routine Soil Testing. *Agronomy* **2022**, *12*, 191. [\[CrossRef\]](#)
19. Lv, Z.Y.; Wang, F.J.; Cui, G.Q.; Benediktsson, J.A.; Lei, T.; Sun, W.W. Spatial-Spectral Attention Network Guided With Change Magnitude Image for Land Cover Change Detection Using Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4412712. [\[CrossRef\]](#)
20. Xu, Z.Y.; Chen, S.B.; Zhu, B.X.; Chen, L.W.; Ye, Y.H.; Lu, P. Evaluating the Capability of Satellite Hyperspectral Imager, the ZY1-02D, for Topsoil Nitrogen Content Estimation and Mapping of Farmlands in Black Soil Area, China. *Remote Sens.* **2022**, *14*, 1008. [\[CrossRef\]](#)
21. Xu, Y.; Li, B.; Shen, X.; Li, K.; Cao, X.; Cui, G. Digital Soil Mapping of Soil Total Nitrogen Based on Landsat 8, Sentinel 2, and Worldview-2 Images in Smallholder Farms in Yellow River Basin, China. *Environ. Monit. Assess.* **2022**, *194*, 282. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Mehedi, I.M.; Hanif, M.S.; Bilal, M.; Vellingiri, M.T.; Palaniswamy, T. Remote Sensing and Decision Support System Applications in Precision Agriculture: Challenges and Possibilities. *IEEE Access* **2024**, *12*, 44786–44798. [\[CrossRef\]](#)
23. Sharma, R.; Mishra, D.R.; Levi, M.R.; Sutter, L.A. Remote Sensing of Surface and Subsurface Soil Organic Carbon in Tidal Wetlands: A Review and Ideas for Future Research. *Remote Sens.* **2022**, *14*, 2940. [\[CrossRef\]](#)
24. De Cesare, F.; Di Mattia, E.; Pantalei, S.; Zampetti, E.; Vinciguerra, V.; Canganella, F.; Macagnano, A. Use of Electronic Nose Technology to Measure Soil Microbial Activity Through Biogenic Volatile Organic Compounds and Gases Release. *Soil Biol. Biochem.* **2011**, *43*, 2094–2107. [\[CrossRef\]](#)
25. Liu, S.Y.; Chen, X.G.; Xia, X.M.; Jin, Y.; Wang, G.; Jia, H.L.; Huang, D.Y. Electronic Sensing Combined With Machine Learning Models for Predicting Soil Nutrient Content. *Comput. Electron. Agric.* **2024**, *221*, 108947. [\[CrossRef\]](#)
26. Han, P.C.; Yang, K.; Jiao, L.Z.; Li, H.C. Rapid Quantitative Analysis of Potassium in Soil Based on Direct-Focused Laser Ablation-Laser Induced Breakdown Spectroscopy. *Front. Chem.* **2022**, *10*, 967158. [\[CrossRef\]](#)
27. Gazeli, O.; Stefas, D.; Couris, S. Sulfur Detection in Soil by Laser Induced Breakdown Spectroscopy Assisted by Multivariate Analysis. *Materials* **2021**, *14*, 541. [\[CrossRef\]](#)
28. Khoso, M.A.; Shaikh, N.M.; Kalhor, M.S.; Jamali, S.; Ujan, Z.A.; Ali, R. Comparative Elemental Analysis of Soil of Wheat, Corn, Rice, and Okra Cropped Field Using CF-LIBS. *Optik* **2022**, *261*, 169247. [\[CrossRef\]](#)

29. Chen, X.Y.; Zhou, G.H.; Mao, S.; Chen, J.H. Rapid Detection of Nutrients With Electronic Sensors: A Review. *Environ. Sci. Nano* **2018**, *5*, 837–862. [\[CrossRef\]](#)
30. Al-Najjar, O.A.; Wudil, Y.S.; Ahmad, U.F.; Al-Amoudi, O.S.B.; Al-Osta, M.A.; Gondal, M.A. Applications of Laser Induced Breakdown Spectroscopy in Geotechnical Engineering: A Critical Review of Recent Developments, Perspectives and Challenges. *Appl. Spectrosc. Rev.* **2023**, *58*, 687–723. [\[CrossRef\]](#)
31. Li, X.; Li, Z.; Qiu, H.; Chen, G.; Fan, P. Soil Carbon Content Prediction Using Multi-Source Data Feature Fusion of Deep Learning Based on Spectral and Hyperspectral Images. *Chemosphere* **2023**, *336*, 139161. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Zhang, Z.X.; Niu, B.B.; Li, X.J.; Kang, X.J.; Hu, Z.Q. Estimation and Dynamic Analysis of Soil Salinity Based on UAV and Sentinel-2A Multispectral Imagery in the Coastal Area, China. *Land* **2022**, *11*, 2307. [\[CrossRef\]](#)
33. Xu, X.B.; Du, C.W.; Ma, F.; Shen, Y.Z.; Wu, K.; Liang, D.; Zhou, J.M. Detection of Soil Organic Matter From Laser-Induced Breakdown Spectroscopy (LIBS) and Mid-Infrared Spectroscopy (FTIR-ATR) Coupled With Multivariate Techniques. *Geoderma* **2019**, *355*, 113905. [\[CrossRef\]](#)
34. Peng, L.; Wu, X.B.; Feng, C.C.; Gao, L.L.; Li, Q.Q.; Xu, J.W.; Li, B. Assessing the Potential of Multi-Source Remote Sensing Data for Cropland Soil Organic Matter Mapping in Hilly and Mountainous Areas. *Catena* **2024**, *245*, 108312. [\[CrossRef\]](#)
35. Wang, Z.Y.; Wu, W.; Liu, H.B. Spatial Estimation of Soil Organic Carbon Content Utilizing PlanetScope, Sentinel-2, and Sentinel-1 Data. *Remote Sens.* **2024**, *16*, 3268. [\[CrossRef\]](#)
36. Zhou, T.; Geng, Y.J.; Chen, J.; Sun, C.L.; Haase, D.; Lausch, A. Mapping of Soil Total Nitrogen Content in the Middle Reaches of the Heihe River Basin in China Using Multi-Source Remote Sensing-Derived Variables. *Remote Sens.* **2019**, *11*, 2934. [\[CrossRef\]](#)
37. Xu, Y.M.; Wang, X.X.; Bai, J.H.; Wang, D.W.; Wang, W.; Guan, Y.N. Estimating the Spatial Distribution of Soil Total Nitrogen and Available Potassium in Coastal Wetland Soils in the Yellow River Delta by Incorporating Multi-Source Data. *Ecol. Indic.* **2020**, *111*, 106002. [\[CrossRef\]](#)
38. Bates, D.M.; Watts, D.G. *Nonlinear Regression Analysis and Its Applications*; Wiley: New York, NY, USA, 1981.
39. Bystritskaya, E.V.; Pomerantsev, A.L.; Rodionova, O.Y. Non-Linear Regression Analysis: New Approach to Traditional Implementations. *J. Chemom.* **2010**, *14*, 667–692. [\[CrossRef\]](#)
40. Rodionova, O.Y.; Pomerantsev, A.L. Chemometrics: Achievements and Prospects. *Russ. Chem. Rev.* **2006**, *75*, 271. [\[CrossRef\]](#)
41. Xia, X.Y.; Liu, B.; Tian, R.; He, Z.L.; Han, S.Y.; Pan, K.; Yang, J.J.; Zhang, Y.T. An Interval Water Demand Prediction Method to Reduce Uncertainty: A Case Study of Sichuan Province, China. *Environ. Res.* **2023**, *238*, 117143. [\[CrossRef\]](#)
42. Yu, M.; Niu, D.X.; Wang, K.K.; Du, R.Y.; Yu, X.Y.; Sun, L.J.; Wang, F.R. Short-Term Photovoltaic Power Point-Interval Forecasting Based on Double-Layer Decomposition and WOA-BiLSTM-Attention and Considering Weather Classification. *Energy* **2023**, *275*, 127348. [\[CrossRef\]](#)
43. Hong, Y.S.; Chen, S.C.; Hu, B.F.; Wang, N.; Xue, J.; Zhuo, Z.Q.; Yang, Y.Y.; Chen, Y.Y.; Peng, J.; Liu, Y.L.; et al. Spectral Fusion Modeling for Soil Organic Carbon by a Parallel Input-Convolutional Neural Network. *Geoderma* **2023**, *437*, 116584. [\[CrossRef\]](#)
44. Jain, S.; Sethia, D.; Tiwari, K.C. A Critical Systematic Review on Spectral-Based Soil Nutrient Prediction Using Machine Learning. *Environ. Monit. Assess.* **2024**, *196*, 699. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Zhai, W.G.; Li, C.C.; Fei, S.P.; Liu, Y.H.; Ding, F.; Cheng, Q.; Chen, Z. CatBoost Algorithm for Estimating Maize Above-Ground Biomass Using Unmanned Aerial Vehicle-Based Multi-Source Sensor Data and SPAD Values. *Comput. Electron. Agric.* **2023**, *214*, 108306. [\[CrossRef\]](#)
46. Gneiting, T.; Raftery, A.E. Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 477. [\[CrossRef\]](#)
47. Wadoux, A.M.J.-C.; Malone, B.; Minasny, B.; Fajardo, M.; McBratney, A.B. *Soil Spectral Inference with R: Analysing Digital Soil Spectra Using the R Programming Environment*; Springer International Publishing: Cham, Switzerland, 2021.
48. Nawar, S.; Delbecq, N.; Declercq, Y.; De Smedt, P.; Finke, P.; Verdoodt, A.; Van Meirvenne, M.; Mouazen, A.M. Can Spectral Analyses Improve Measurement of Key Soil Fertility Parameters with X-Ray Fluorescence Spectrometry? *Geoderma* **2019**, *350*, 29–39. [\[CrossRef\]](#)
49. Stockmann, U.; Holden, N.M.; McBratney, A.B.; McBratney, A.B.; Minasny, B. An Assessment of Model Averaging to Improve Predictive Power of Portable Vis-NIR and XRF for the Determination of Agronomic Soil Properties. *Geoderma* **2016**, *279*, 31–44.
50. Wang, X.; Zhang, F.; Kung, H.T.; Johnson, V.C. New Methods for Improving the Remote Sensing Estimation of Soil Organic Matter Content (SOMC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR) in Northwest China. *Remote Sens. Environ.* **2018**, *218*, 104–118. [\[CrossRef\]](#)
51. Bates, J.M.; Granger, C.W.J. The Combination of Forecasts. *J. Oper. Res. Soc.* **1969**, *20*, 451–468. [\[CrossRef\]](#)
52. Wang, D.; Chakraborty, S.; Weindorf, D.C.; Li, B.; Sharma, A.; Paul, S.; Ali, M.N. Synthesized Use of VisNIR DRS and PXRF for Soil Characterization: Total Carbon and Total Nitrogen. *Geoderma* **2015**, *243–244*, 157–167. [\[CrossRef\]](#)
53. Benedet, L.; Faria, W.M.; Silva, S.H.G.; Mancini, M.; Dematte, J.A.M.; Guilherme, L.R.G.; Curi, N. Soil Texture Prediction Using Portable X-Ray Fluorescence Spectrometry and Visible Near-Infrared Diffuse Reflectance Spectroscopy. *Geoderma* **2020**, *376*, 114553. [\[CrossRef\]](#)
54. Azcarate, S.M.; Ríos-Reina, R.; Amigo, J.M.; Goicoechea, H.C. Data Handling in Data Fusion: Methodologies and Applications. *TrAC Trends Anal. Chem.* **2021**, *143*, 116355. [\[CrossRef\]](#)
55. Kjeldahl, J.Z. A New Method for the Determination of Nitrogen in Organic Matter. *Z. Anal. Chem.* **1883**, *22*, 366. [\[CrossRef\]](#)
56. Bremner, J.M.; Black, C.A. *Methods of Soils Analysis: Part 2. Chemical and Microbiological Properties*; American Society of Agronomy: New York, NY, USA, 1965.

57. Page, A.L. *Methods of Soil Analysis. Part 2. Chemical and Microbiological Properties*; Wiley: New York, NY, USA, 1992.
58. Olsen, S.R. *Estimation of Available Phosphorus in Soils by Extraction with Sodium Bicarbonate*; Miscellaneous Paper Institute for Agricultural Research Samaru; US Department of Agriculture: Washington DC, USA, 1954.
59. Steinier, J.; Termonia, Y.; Deltour, J. Smoothing and Differentiation of Data by Simplified Least Square Procedure. *Anal. Chem.* **1972**, *44*, 1906–1909. [[CrossRef](#)] [[PubMed](#)]
60. Kharintsev, S.S.; Salakhov, M.K. A Simple Method to Extract Spectral Parameters Using Fractional Derivative Spectrometry. *Spectrochim. Acta Part A* **2004**, *60*, 2125–2133. [[CrossRef](#)]
61. Zhang, Z.P.; Ding, J.L.; Wang, J.Z.; Ge, X.Y. Prediction of Soil Organic Matter in Northwestern China Using Fractional-Order Derivative Spectroscopy and Modified Normalized Difference Indices. *Catena* **2020**, *185*, 104257. [[CrossRef](#)]
62. Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence; Morgan Kaufman Publishing: San Francisco, CA, USA, 1995; Volume 2, pp. 1137–1143.
63. Jolliffe, I.T.; Cadima, J. Principal Component Analysis: A Review and Recent Developments. *Philos. Trans. R. Soc. A* **2016**, *374*, 20150202. [[CrossRef](#)] [[PubMed](#)]
64. Abdi, H.; Williams, L.J. Principal Component Analysis. *WIREs Comput. Stat.* **2010**, *2*, 433–459. [[CrossRef](#)]
65. Jaillais, B.; Pinto, R.; Barros, A.S.; Rutledge, D.N. Outer-Product Analysis (OPA) Using PICA to Study the Influence of Temperature on NIR Spectra of Water. *Vib. Spectrosc.* **2005**, *39*, 50–58. [[CrossRef](#)]
66. Terra, F.S.; Rossel, R.A.V.; Demattê, J.A.M. Spectral Fusion by Outer Product Analysis (OPA) to Improve Predictions of Soil Organic C. *Geoderma* **2019**, *335*, 35–46. [[CrossRef](#)]
67. Granger, C.W.J.; Ramanathan, R. Improved Methods of Combining Forecasts. *J. Forecast.* **1984**, *3*, 197–204. [[CrossRef](#)]
68. Kennard, R.W.; Stone, L.A. Computer Aided Design of Experiments. *Technometrics* **1969**, *11*, 137–148. [[CrossRef](#)]
69. Berger, J.O. *Statistical Decision Theory and Bayesian Analysis*; Springer: New York, NY, USA, 1985.
70. Hemson, G.; Johnson, P.; South, A.; Kenward, R.; Ripley, R.; Macdonald, D. Are Kernels the Mustard? Data from GPS Collars Suggest Problems for Kernel Home-Range Analyses with Least-Squares Cross-Validation. *J. Anim. Ecol.* **2005**, *74*, 455–463. [[CrossRef](#)]
71. Ji, Z.; Niu, D.; Li, M.; Li, W.; Sun, L.; Zhu, Y. A Three-Stage Framework for Vertical Carbon Price Interval Forecast Based on Decomposition–Integration Method. *Appl. Soft Comput.* **2022**, *116*, 108204. [[CrossRef](#)]
72. Ma, J.; Wang, Y.; Niu, X.; Jiang, S.; Liu, Z. A Comparative Study of Mutual Information-Based Input Variable Selection Strategies for the Displacement Prediction of Seepage-Driven Landslides Using Optimized Support Vector Regression. *Stoch. Environ. Res. Risk Assess.* **2022**, *36*, 3109–3129. [[CrossRef](#)]
73. Wang, Y.; Tang, H.; Wen, T.; Ma, J. Direct Interval Prediction of Landslide Displacements Using Least Squares Support Vector Machines. *Complexity* **2020**, *2020*, 7082594. [[CrossRef](#)]
74. Beer, A. Determination of the Absorption of Red Light in Colored Liquids. *Ann. der Phys. und Chem.* **1952**, *86*, 78–88.
75. Rossel, R.A.V.; Walvoort, D.J.J.; McBratney, A.B.; Janik, L.J.; Skjemstad, J.O. Visible, Near Infrared, Mid Infrared or Combined Diffuse Reflectance Spectroscopy for Simultaneous Assessment of Various Soil Properties. *Geoderma* **2006**, *131*, 59–75. [[CrossRef](#)]
76. Summers, D.; Lewis, M.; Ostendorf, B.; Chittleborough, D. Visible Near-Infrared Reflectance Spectroscopy as a Predictive Indicator of Soil Properties. *Ecol. Indic.* **2011**, *11*, 123–131. [[CrossRef](#)]
77. Stenberg, B.; Viscarra Rossel, R.A.; Mouazen, A.M.; Wetterlind, J. Visible and Near Infrared Spectroscopy in Soil Science. *Adv. Agron.* **2010**, *107*, 163–215.
78. Stone, M.L.; Solie, J.B.; Raun, W.R.; Whitney, R.W.; Taylor, S.L.; Ringer, J.D. Use of Spectral Radiance for Correcting In-Season Fertilizer Nitrogen Deficiencies in Winter Wheat. *Trans. ASAE* **1996**, *39*, 1623–1631. [[CrossRef](#)]
79. Janik, L.; Skjemstad, J.O. Can Mid Infrared Diffuse Reflectance Analysis Replace Soil Extractions? *Aust. J. Exp. Agric.* **1998**, *38*, 681–696. [[CrossRef](#)]
80. Stuart, B.H. *Infrared Spectroscopy: Fundamentals and Applications*; Wiley: Chichester, UK, 2004.
81. Rossel, R.A.V.; McGlynn, R.N.; McBratney, A.B. Determining the Composition of Mineral–Organic Mixes Using UV–Vis–NIR Diffuse Reflectance Spectroscopy. *Geoderma* **2006**, *137*, 70–82. [[CrossRef](#)]
82. Norrish, K.; Hutton, J.T. An Accurate X-ray Spectrographic Method for the Analysis of a Wide Range of Geological Samples. *Geochim. Cosmochim. Acta* **1969**, *33*, 431–453. [[CrossRef](#)]
83. Clark, R.N.; Roush, T.L. Reflectance Spectroscopy: Quantitative Analysis Techniques for Remote Sensing Applications. *J. Geophys. Res.* **1984**, *89*, 6329–6340. [[CrossRef](#)]
84. Nguyen, T.T.; Janik, L.J.; Raupach, M. Diffuse Reflectance Infrared Fourier Transform (DRIFT) Spectroscopy in Soil Studies. *Aust. J. Soil Res.* **1991**, *29*, 49–67. [[CrossRef](#)]
85. Baes, A.U.; Bloom, P.R. Diffuse Reflectance Fourier Transform Infrared (DRIFT) Spectroscopy of Humic and Fulvic Acids. *Soil Sci. Soc. Am. J.* **1989**, *53*, 695. [[CrossRef](#)]
86. Farmer, V.C. *The Infrared Spectra of Minerals; Mineralogical Society Monograph*, 4; The Mineralogical Society: London, UK, 1974.
87. Madejová, J. FTIR Techniques in Clay Mineral Studies. *Vib. Spectrosc.* **2003**, *31*, 1–10. [[CrossRef](#)]
88. Shepherd, K.D.; Walsh, M.G. Development of Reflectance Spectral Libraries for Characterization of Soil Properties. *Soil Sci. Soc. Am. J.* **2002**, *66*, 988–998. [[CrossRef](#)]

89. Guppy, C.N.; Menzies, N.W.; Moody, P.W.; Blamey, F.P.C. Competitive Sorption Reactions Between Phosphorus and Organic Matter in Soil: A Review. *Aust. J. Soil Res.* **2005**, *43*, 189–202. [[CrossRef](#)]
90. Schmidt, M.W.I.; Torn, M.S.; Abiven, S.; Dittmar, T.; Guggenberger, G.; Janssens, I.A.; Trumbore, S.E. Persistence of Soil Organic Matter as an Ecosystem Property. *Nature* **2011**, *478*, 49–56. [[CrossRef](#)] [[PubMed](#)]
91. Barrow, N.J. The Effects of pH on Phosphate Uptake from the Soil. *Plant Soil* **2017**, *410*, 401–410. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.