processes

MDPI

Article

Rapid Analysis of Raw Meal Composition Content Based on NIR Spectroscopy for Cement Raw Material Proportioning Control Process

Zhenfa Yang, Qingmei Sui and Lei Jia *

School of Control Science and Engineering, Shandong University, Jingshi Road, Jinan 250061, China
* Correspondence: jialei@sdu.edu.cn; Tel.: +86-13806415319

Abstract: Due to fast analysis speed, analyzing composition content of cement raw meal utilizing near infrared (NIR) spectroscopy, combined with partial least squares regression (PLS), is a reliable alternative method for the cement industry to obtain qualified cement products. However, it has hardly been studied. The raw materials employed in different cement plants differ, and the spectral absorption intensity in the NIR range of the raw meal component is weaker than organic substances, although there are obvious absorption peaks, which place high demands on the generality of modeling and accuracy of the analytical model. An effective modeling procedure is proposed, which optimizes the quantitative analytical model from several modeling stages, and two groups of samples with different raw material types and origins are collected to validate it. For the samples in the prediction set from Qufu, the root mean square error of prediction (RMSEP) of CaO, SiO_2, Al_2O_3, and Fe_2O_3 were 0.1910, 0.2307, 0.0921, and 0.0429, respectively; the average prediction errors for CaO, SiO_2, Al_2O_3, and Fe_2O_3 were 0.171%, 0.193%, 0.069%, and 0.032%, respectively; for the samples in the prediction set from Linyi, the RMSEP of CaO, SiO_2, Al_2O_3, and Fe_2O_3 were 0.1995, 0.1267, 0.0336 and 0.0242, respectively, the average prediction errors for CaO, SiO_2, Al_2O_3, and Fe_2O_3 were 0.154%, 0.100%, 0.022%, and 0.018%, respectively. The standard methods for chemical analysis of cement require that the mean measurement error for CaO, SiO_2, Al_2O_3, and Fe_2O_3 should be within 0.40%, 0.30%, 0.20%, and 0.15%, respectively. It is obvious that the results of both groups of samples fully satisfied the requirements of raw material proportioning control of the production line, demonstrating that the modeling procedure has excellent generality, the models established have high prediction accuracy, and the NIR spectroscopy combined with the proposed modeling procedure is a rapid and accurate alternative approach for the analysis of cement raw meal composition content.

Keywords: near infrared spectroscopy; component analysis; cement raw meal; oxides content; rapid determination

check for updates

Citation: Yang, Z.; Sui, Q.; Jia, L. Rapid Analysis of Raw Meal Composition Content Based on NIR Spectroscopy for Cement Raw Material Proportioning Control Process. *Processes* **2022**, *10*, 2494. https://doi.org/10.3390/pr10122494

Academic Editors: Juan Francisco García Martín and Vincenzo Russo

Received: 25 October 2022
Accepted: 17 November 2022
Published: 24 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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

As an indispensable element of practically all modern structures, cement plays an invaluable role in social and economic growth; there are already many researchers engaged in enhancing the quality of cement products and meaningful results have been achieved [1–3]. Cement manufacturing involves the preparation of raw materials, calcination of clinker, and the creation of cement. Being the initial stage in cement manufacture, the quality of the raw meal heavily determines the quality of the cement product. The preparation process is depicted in Figure 1, the controller transports the calcium, silica, aluminum, and iron raw materials in specific ratios to the raw material mill for grinding, the coarse powder from the mill is selected by the separator and sent back to the raw material mill for grinding, and the fine powder is transported to the raw meal store via the conveyor belt and elevator at the air chute [4,5]. Raw meal samples are collected at the air chute and sent to the laboratory to analyze the content of CaO, SiO_2, Al_2O_3, and Fe_2O_3, and the control room adjusts the proportion of raw materials according to the measured results to achieve feedback control.

The above demonstrates that rapid and accurate determination of the aforementioned four components is a crucial prerequisite for quality assurance of cement raw meals.

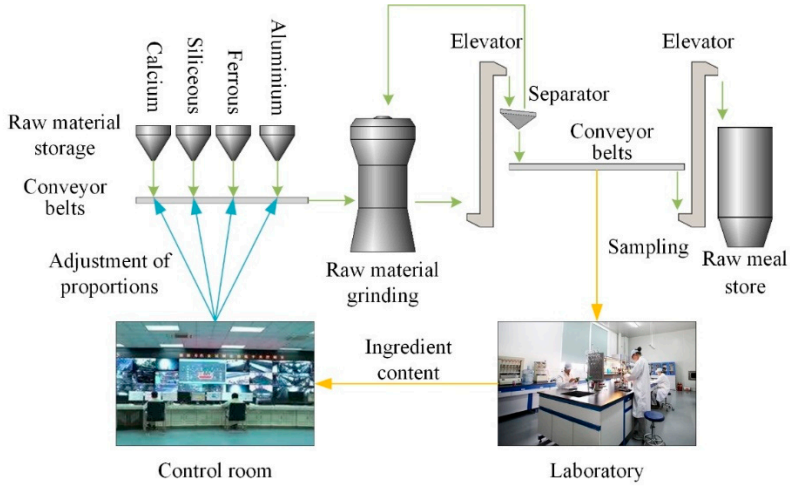


Figure 1. The preparation process of cement raw meal.

Currently, there are four methods for determining the composition content of cement raw meal: the wet chemical method [6], X-ray fluorescence (XRF) [7,8], transient gamma-ray neutron activation [9], and laser induced breakdown spectrometry (LIBS) [10]. However, each of these methods has its own limitations and is unable to obtain the oxide content rapidly and accurately. Wet chemical analysis and XRF are both standard methods of analysis for raw meal. The former is highly accurate and reproducible, but the process is complex, and the time consumed to analyze a sample exceeds 2 h; the latter requires sample preparation and takes about dozens of minutes per sample, which presents a relative high cost and safety issues due to the use of radioactive source [11,12]. Transient gamma-ray neutron activation method is not time-consuming, but the sensing instrument needs to be installed on the conveyor belt for transporting raw materials, which has strict requirements for the running speed, width, and thickness of the conveyor belt, as well as high maintenance costs and the possibility of radioactive source leakage [9]. Although fast, LIBS requires the breakdown of the sample in order to obtain the plasma state spectrum, which is less repeatable and stable due to issues such as laser energy jitter, sample surface inhomogeneity, unstable plasma plume diffusion, and weaker elemental energy leap spectral lines that are difficult to identify [10,12]. As a consequence, the quick and precise analysis of raw meal oxide content is still a significant issue that has to be solved right away for the cement industry.

Due to its rapid detection speed (less than 3 min for each sample), simultaneous detection of many components, no sample pretreatment required, repeatability, non-destructiveness, and environmental friendliness [13–15], near infrared (NIR) spectroscopy has demonstrated significant potential in the field of composition analysis and is widely used in food [16–18], agriculture [19–21], pharmaceutical [22,23], chemical [24,25], and many other fields [26,27]. For example, Tsegay et al. [28] developed non-destructive NIR spectroscopy analytical models to predict oil and major fatty acid contents of Ethiopian sesame and Niemi et al. [29] applied NIR spectroscopy to determine protein content in North Atlantic seaweed. The majority of the detection objects of NIR spectroscopy are organic substances with strong absorption intensity, so direct modeling is able to produce a model with strong predictive power to address the demand for analysis after the NIR spectra acquisition of the samples and the determination of the reference values of the components' content. However, the four main ingredients of cement raw meal are all inorganic, although there are obvious absorption peaks in the near infrared range, the spectral absorption intensity is weaker than organic substances, so direct modeling yields models with relatively large prediction errors. Additionally, raw materials employed in different

cement plants may vary, which places high demands on the generality of the modeling and poses a tremendous challenge for raw meal composition content determination.

Optimizing the model at different stages of modeling is able to improve the accuracy of quantitative analysis models. At present, optimization methods of quantitative analytical model for NIR spectroscopy have received a lot of investigation. In the sample set partitioning session, Galvao et al. [30] proposed an algorithm named sample set partitioning based on joint x–y distance (SPXY) to divide the diesel samples into the calibration set and the prediction set, the random selection (RS) and Kennard–Stone method (KS) were also employed for comparison, the results indicated that the quantitative analysis model established based on the calibration set obtained from the proposed method was able to predict more accurately the specific mass of the diesel; Agus et al. [31] proposed an algorithm based on Mahalanobis distance and Kennard–Stone method that allows for better segmentation of the dataset, and the artificial neural network model built using the sample set acquired by this method performs better than the traditional Kennard–Stone methodology. In the outlier elimination session, Pierna et al. [32] used uncertainty assessment and the convex packet method to identify outliers in several sample sets and compared traditional methods such as the Mahalanobis distance method and the X-residual method, which provides a practical method for outlier elimination. Based on the definition of outliers and the principle of partial least squares regression, Li et al. [33] proposed an outlier elimination method for NIR spectral analysis to identify anomalous samples by comparing the contribution of each latent variable of the partial least squares model, and validated the method using NIR spectral data of orange juice samples, the root mean square error of cross validation (RMSECV) was decreased from 16.870 to 4.809 and the root mean square error of prediction (RMSEP) was lowered from 3.688 to 3.332 following the removal of the 6 anomalous samples. In the spectral preprocessing session, Sampaio et al. [34] increased the meaningful information in the NIR spectra of rice and increased the determining precision of straight-chain starch content by using multiple scattering correction (MSC), standard normal variate transformation (SNV), 1st derivative (1D), 2nd derivative (2D), and Savitzky–Golay smoothing convolution (SG). In the feature variable selection session, L. Nørgaard et al. [35] proposed the interval partial least-Squares regression (iPLS) algorithm, the central idea of which is to divide the entire spectrum into multiple equal-width spectral intervals, build a local piecewise partial least-squares model on each interval separately, and then choose the interval that corresponds to the model that performs the best as the feature interval. He et al. [36] used stepwise regression combined with the regression coefficient method to select optimal wavelengths to optimize the original full band partial least squares (PLS) models. Riccardo et al. [37] made the appropriate modifications to the genetic algorithm by combining it with partial least squares and applying it to feature selection of near-infrared spectra, the effectiveness of the method was validated on the 5 public datasets, the results showed that the predictive power of the model built by the algorithm with the selected variables is much better than that of the model built using the full spectrum, indicating that the genetic algorithm is an effective method for feature selection of spectral datasets.

In this paper, an effective modeling procedure for the analysis of the NIR spectra of the raw cement composition content was proposed, which optimizes the partial least squares model in sample set partitioning, outlier elimination, spectral preprocessing, and feature selection. Even if the raw material types and origins of cement raw meal are different, accurate analytical models can also be obtained through the modeling procedure. To verify the performance of the procedure, two groups of cement raw meal samples with different raw material types and sources were collected, then the PLS models were established and validated, respectively. Finally, the average prediction error of the models was contrasted with the detection requirement of the feedback control of the production line, respectively. The NIR spectroscopy combined with the proposed modeling procedure provides a rapid and accurate alternative method for the analysis of raw meal composition content in the cement raw material proportioning control process, which will contribute

to improving the quality and productivity of cement products and reducing energy and material consumption and CO_2 emissions in the cement industry.

2. Materials and Methods

2.1. Analytical Procedure for NIR Spectroscopy of Cement Raw Meal

The analytical procedure contains 7 steps, as shown in Figure 2:

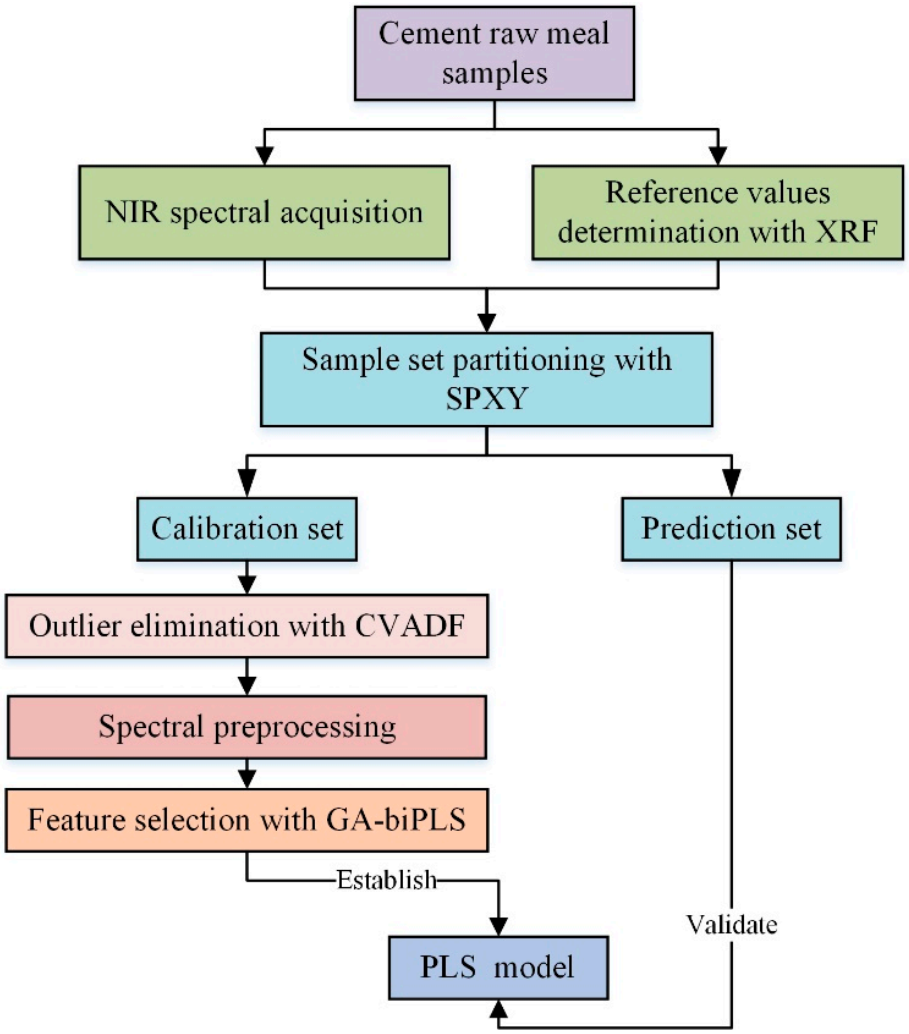


Figure 2. Analytical procedure for NIR spectroscopy of cement raw meal.

(1) Collection of a certain number of samples.

(2) Acquisition of NIR spectra of the samples and determination of the reference values of the four oxides by XRF method.

(3) Division of the samples into calibration and prediction sets using SPXY method.

(4) Outlier elimination with cross-validation-absolute-deviation-F-test (CVADF) algorithm.

(5) Spectral valuable information enhancement with the optimal preprocessing method.

(6) Spectral feature selection by backward interval PLS and genetic algorithm (GA-biPLS).

(7) Modeling based on the calibration set using the PLS method and validation of the model by the prediction set.

2.2. Materials

A group of 202 samples of cement raw meal, composed of limestone, mudstone, and steel slag, was collected from the air ramp at the production line of the Zhonglian Cement Plant in Qufu, Shandong Province and stored in sealed PE compact bags. In addition, another group of 119 samples, consisting of calcareous marl, clay, and coal gangue, were

collected from the Zhonglian Cement Plant in Linyi, Shandong Province in the same preservation manner. As shown in Figure 3, the colors of the samples also show significant discrepancies due to the different composition of the raw material.

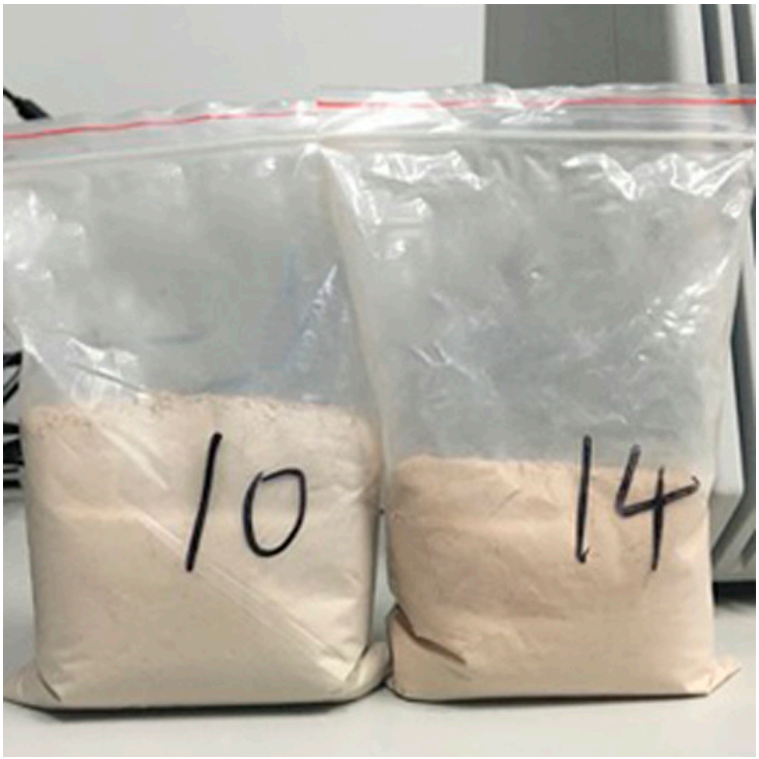


Figure 3. Raw meal samples from two different cement plants.

2.3. NIR Spectra Acquisition

A Fourier NIR spectrometer (MB3600, ABB, Zurich, Switzerland) was applied for the acquisition of raw meal samples. Polytetrafluoroethylene (PTFE) was employed as spectral acquisition background. The number of successive scans and the spectral resolution are 64 and 4 cm^{-1}, respectively, with the wave number range of 10,000–4000 cm^{-1}. During the acquisition process, the room temperature was 24–26 °C and the humidity was 45–55%; moreover, the background spectrum of PTFE was re-acquired every 6 samples to minimize the influence of ambient temperature and humidity.

As described in Figure 4, the spectra of the two groups of samples show a clear separation. The lines above the red dashed line indicate the spectra of the sample from Linyi and the line below indicate the spectra of the sample from Qufu. It is obvious that the diffuse absorbance of the former is significantly higher than that of the latter. However, the positions of the main absorption peaks of the spectra are essentially consistent. The strongest absorption peak appeared at 5200 cm^{-1} is related to the combination of stretching and bending of the O–H group of water, while the peak at 7080 cm^{-1} is related to the combination of the first overtone of O–H anti-symmetric stretching and O–H symmetric stretching of the water molecule. The peak at 4270 cm^{-1} is related to the combination of stretching and bending of the Ca-O bond. The relatively weak absorption peak at 4526 cm^{-1} is related to the combination of stretching and bending of the Si–O, Fe–O, and Al–O bonds [38,39].

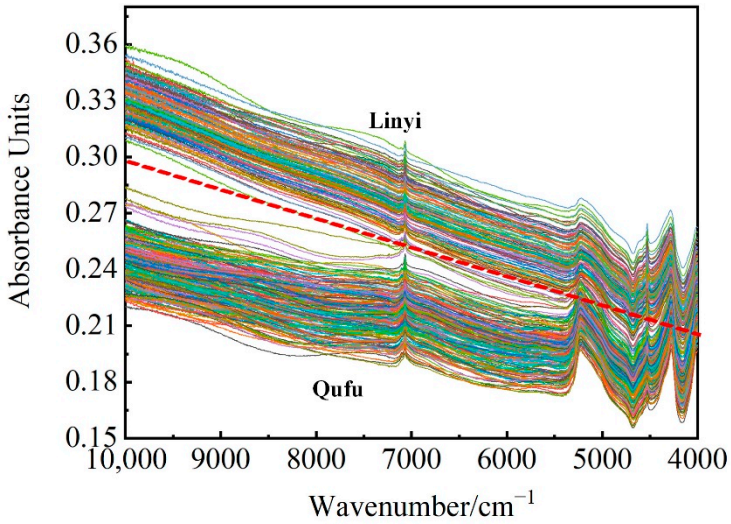


Figure 4. Near infrared spectra of cement raw meal samples from two cement plants.

2.4. Reference Value

X-ray fluorescence (XRF) spectroscopy was conducted to determine the standard values of the four oxides in the samples using an XRF spectrometer (ARL ADVANT'X, Thermo Fisher Scientific, Waltham, MA, USA) according to GB/T176-2017. The samples were prepared by the powder pressing method, using a powder mill to crush them to a suitable particle size, after that the samples were made into smooth and crack-free specimens with a certain strength by pressing them under a certain pressure. Then, the specimens were put into the XRF spectrometer to obtain the oxides content.

The plot In Figure 5 provides a view of the distribution of the oxides content of the two groups of samples. Although the NIR spectra of the two groups of samples demonstrated significant discrepancies, the mean values for the content of the four components are essentially consistent. Overall, the content of CaO, SiO_2, Al_2O_3, and Fe_2O_3 are distributed in the ranges of 39.5–44.5%, 11–14.5%, 2–4%, and 1.5–3.5%, respectively. Samples from Qufu have a slightly higher overall range of CaO content than Linyi, while the opposite was true for Fe_2O_3, with little difference in SiO_2 and Al_2O_3 content.

2.5. Sample Set Partitioning Based on Joint x–y Distance

Commonly known methods of sample set partitioning, such as Kennard–Stone method and Duplex, select samples based solely on the Euclidean distance between the NIR spectra of cement raw meals to obtain a calibration subset for modeling and a prediction subset for external validation of the model. This implies that only the distribution of the sample in the spectral space is considered in the sample selection process, not the component content space, thus limiting to some extent the representativeness of the calibration subset samples.

Sample set partitioning based on joint x–y distance replaces the Euclidean distance with the joint x–y distance, taking into account the NIR spectrum and the oxide content of the sample for the selection of the calibration subset. The formula for computing the joint x–y distance and the process of implementing the method have already been described in References [30,39], the readers can refer to these papers for more details.

2.6. Outlier Elimination by Cross-Validation-Absolute-Deviation-F-Test

Cross-validation is a method of internal validation of data used in the quantitative calibration model building process. The basic principle is that a quantitative calibration model is built based on some elements of the dataset and predictions are made for the remaining elements. The F-test, a statistical test proposed by Ronald Aylmer Fisher, who is regarded as the father of modern statistics, is also known as the joint hypothesis test, the variance ratio test and the chi-square test, and can be employed to determine whether there is a significant difference between two sets of data.

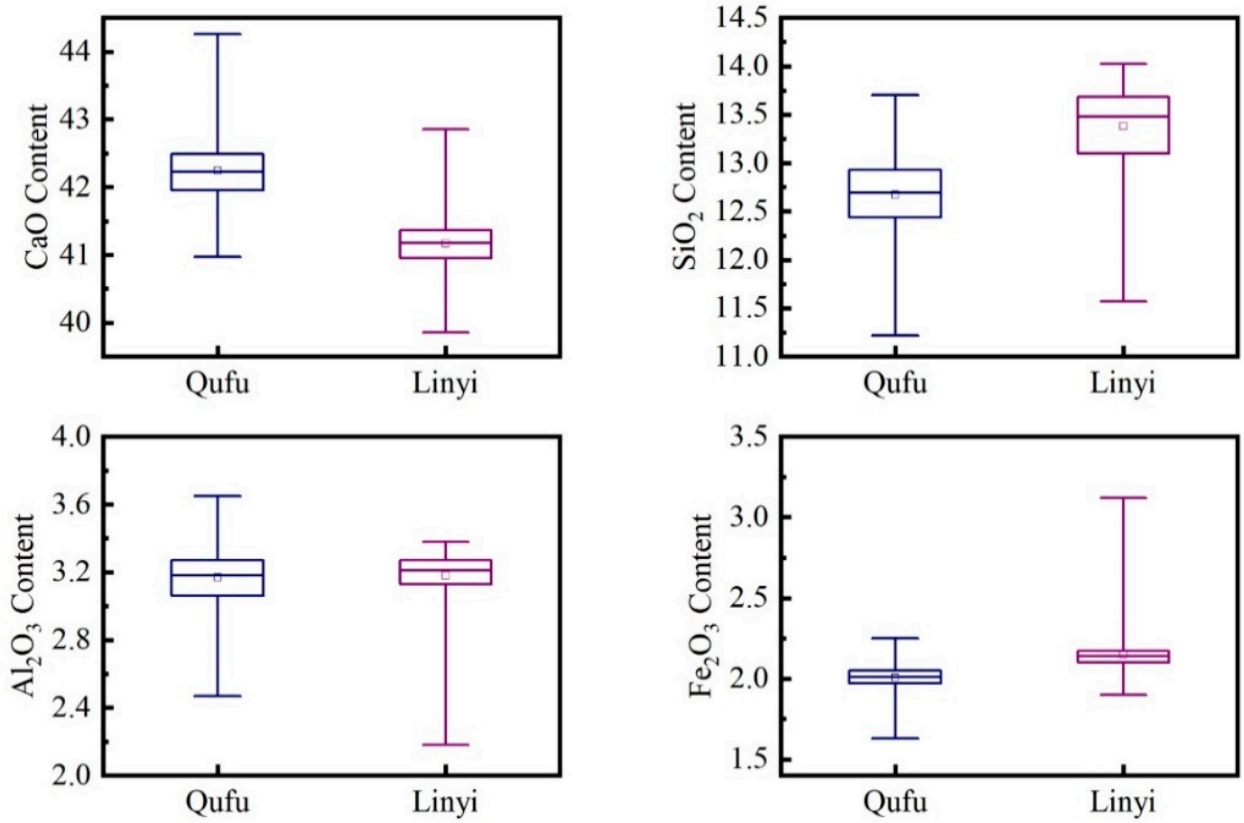


Figure 5. The box and whisker plots of the oxides content distribution of raw meal samples from two cement plants.

In our past work, an outlier elimination method named cross-validation-absolute-deviation-F-test was proposed, which can identify abnormal samples due to operational errors, background noise or extreme component content. The fundamental idea is to obtain the reference and predicted values for each sample in the raw meal sample set by cross-validation, and to identify and reject samples with excessive prediction errors by F-test. Partial least squares method is employed in the cross-validation process to model quantitative calibration models. The implementation of the algorithm has been described in detail in Reference [11].

2.7. Spectral Preprocessing

Spectral preprocessing is undertaken to attenuate the effects of various interfering factors on the NIR spectra of cement raw meal, to enhance the effective information closely related to the oxide content, to reduce the spatial complexity of the quantitative calibration model and to enhance the predictive performance of the model. In the process of analyzing the oxide content of cement raw meal by NIR spectroscopy, the spectra need to be processed repetitively in order to choose an optimal preprocessing method to effectively filter out the noise information in the spectra. Commonly used spectral preprocessing methods include multiple scattering correction (MSC) [40], standard normal variate transformation (SNV) [41], 1st derivative (1D) and Savitzky–Golay convolution smoothing (SG) [42], etc. As can be seen from Figure 6, MSC and SNV are applied to eliminate the effects of inhomogeneous sample particle size, tightness, scattered light from the sample surface, specular reflections and light range variations on the NIR spectra, 1D is able to effectively eliminate the baseline drift of the spectra and overcome background interference, and SG is the most widespread method for removing stochastic noise from NIR spectra.

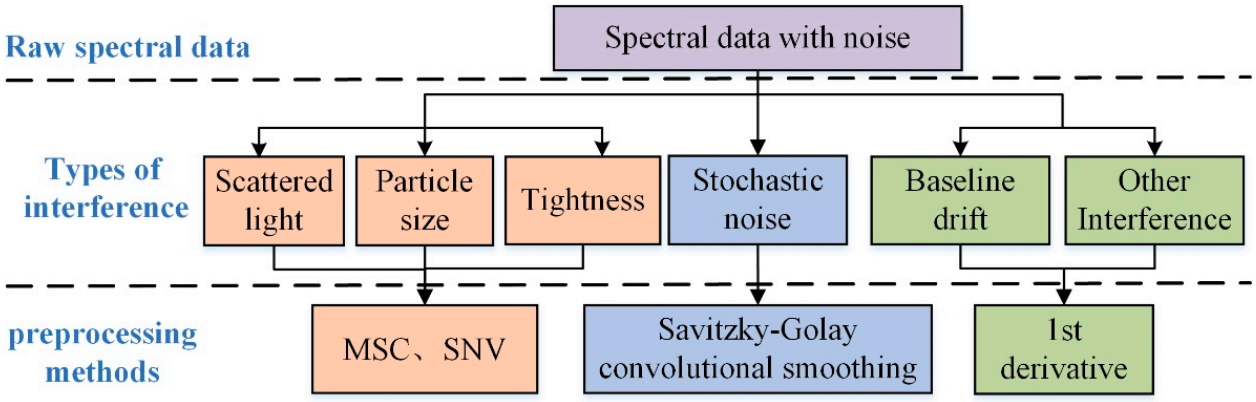


Figure 6. Common spectral preprocessing methods.

2.8. Feature Selection by Backward Interval PLS and Genetic Algorithm

The genetic algorithm (GA) alone has certain limitations when applied to the selection of feature variables for NIR spectra: (1) the local search capability is poor and takes too long when the number of variables is large; (2) the ratio of the number of variables to the number of samples cannot be too high, otherwise it tends to reinforce noise rather than meaningful information. When the backward interval PLS algorithm (biPLS) is employed alone, the optimal combination of spectral subintervals can be obtained, but there are still more irrelevant and redundant variables within each subinterval, which will affect the prediction accuracy of the quantitative calibration model.

The backward interval PLS and genetic algorithm (GA-biPLS) cleverly combines the advantages of both methods and overcomes the limitations of both by first selecting feature intervals that are highly correlated with oxide content, and then performing secondary selection on the selected feature intervals to eliminate irrelevant and redundant variables within the feature wave number interval. The calculation steps of the algorithm can be reviewed in Reference [43].

2.9. Quantitative Calibration Model

Partial least squares (PLS) regression combines the advantages of multiple linear regression, principal component regression and typical correlation analysis, and is able to orthogonally decompose both the spectral matrix X and the oxide content matrix Y of the sample, ensuring that the oxide content information is involved in the decomposition process of the spectral matrix, swapping the score of the spectral matrix with that of the oxide content matrix, and obtaining principal components that are intimately related to the oxide content information of the sample, which is currently the most widely adopted quantitative calibration method.

2.10. Evaluation of Model Performance

The performance of the quantitative calibration model is evaluated by two indexes: the coefficient of correlation of prediction set (R_p) and the root mean square error of prediction (RMSEP). R_p and RMSEP are defined as Equations (1) and (2), respectively. In these formulas, y_i represents the reference value of sample i obtained by XRF, $\hat{y}_i$ indicates the estimated value of sample i predicted by PLS model, $\hat{y}$ represents the mean of the reference values for all samples in the prediction subset, and n indicates the number of samples in the prediction subset. R_p shows the linear correlation between the predicted results and reference values, and RMSEP represents the efficiency and accuracy of the PLS models. The smaller the RMSEP is, the better the prediction performance of the model will be.

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

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

2.11. Software

NIR Spectra Acquisition was carried out by Horizon MB software (version 3.4.0.3, ABB, Zurich, Switzerland), implementations of algorithms (such as SPXY, CVADF, GA-biPLS, PLS, etc.) were coded by the authors in MATLAB software (version R2016a, MathWorks, Natick, MA, USA).

3. Results and Discussion

3.1. Samples Collected in Qufu

Reference values for the content of CaO, SiO_2, Al_2O_3, and Fe_2O_3 in cement raw meal samples were measured by XRF method and the results are shown in Table 1. The content of CaO is distributed between 40.97 and 44.26% and is concentrated in the range of 42–42.5%; the content of SiO_2 is distributed between 11.22 and 13.70% and is concentrated in the range of 12.25–13%; the content of Al_2O_3 is distributed between 2.47 and 3.65% and is concentrated in the range of 3–3.3%; the content of Fe_2O_3 is distributed between 1.63 and 2.25% and is concentrated in the range of 1.9–2.1%.

Table 1. Distribution of reference values for oxides content.

Oxide	Maximum/%	Minimum/%	Deviation/%	Mean/%
CaO	44.26	40.97	3.29	42.25
SiO_2	13.70	11.22	2.48	12.67
Al_2O_3	3.65	2.47	1.18	3.17
Fe_2O_3	2.25	1.63	0.62	2.01

The 202 cement raw material samples were split into 2 groups at a ratio of approximately 3:1 using SPXY method, with 152 samples in the calibration set and 50 samples in the prediction set, and the statistical results for the 4 ingredients are shown in Table 2. For each of the four oxides, the highest and lowest content of samples were included in the calibration set and the content range of samples in the calibration set covered those in the prediction set, respectively; therefore, the modeling criteria were met.

Table 2. Oxides content range for calibration set and validation set samples.

Oxide	Subsets	Number	Maximum/%	Minimum/%	Deviation/%	Mean/%
CaO	Calibration	152	44.26	40.97	3.29	42.26
	Prediction	50	43.01	41.61	1.40	42.22
SiO_2	Calibration	152	13.70	11.22	2.48	12.67
	Prediction	50	13.22	12.21	1.01	12.68
Al_2O_3	Calibration	152	3.65	2.47	1.18	3.17
	Prediction	50	3.39	2.91	0.48	3.16
Fe_2O_3	Calibration	152	2.25	1.63	0.62	2.00
	Prediction	50	2.11	1.87	0.24	2.01

The cross-validation-absolute-deviation-F-test algorithm was performed to identify and reject outliers for the four ingredients in the calibration set samples. In the cross-validation process, samples in the calibration set were divided into five subsets, and the results are shown in Table 3. Five outliers of CaO were identified, with sample numbers 29,

33, 42, 54, and 69; seven outliers of SiO_2 were determined, with sample numbers 9, 52, 54, 69, 71, 105, and 107; five outliers of Al_2O_3 were detected, with sample numbers 30, 52, 54, 69, and 137; and seven outliers of Fe_2O_3 were ascertained, with numbers 5, 6, 30, 31, 105, 120, and 129.

Table 3. Rejected outliers by cross-validation-absolute-deviation-F-test (CVADF) method.

Oxide	Total Number	Serial Number
CaO	5	29, 33, 42, 54, 69
SiO_2	7	9, 52, 54, 69, 71, 105, 107
Al_2O_3	5	30, 52, 54, 69, 137
Fe_2O_3	7	5, 6, 30, 31, 105, 120, 129

The optimal spectral preprocessing method for each oxide was chosen, as described in the Reference [38], the near infrared spectra of the calibration set samples were enhanced with interesting information using Savitzky–Golay convolution smoothing (SG), multiple scattering correction (MSC) with 1st derivative (1D), 1st derivative (1D), and standard normal variate transformation (SNV), respectively, to obtain the spectra employed for subsequent modeling of CaO, SiO_2, Al_2O_3, and Fe_2O_3, as shown in Figure 7.

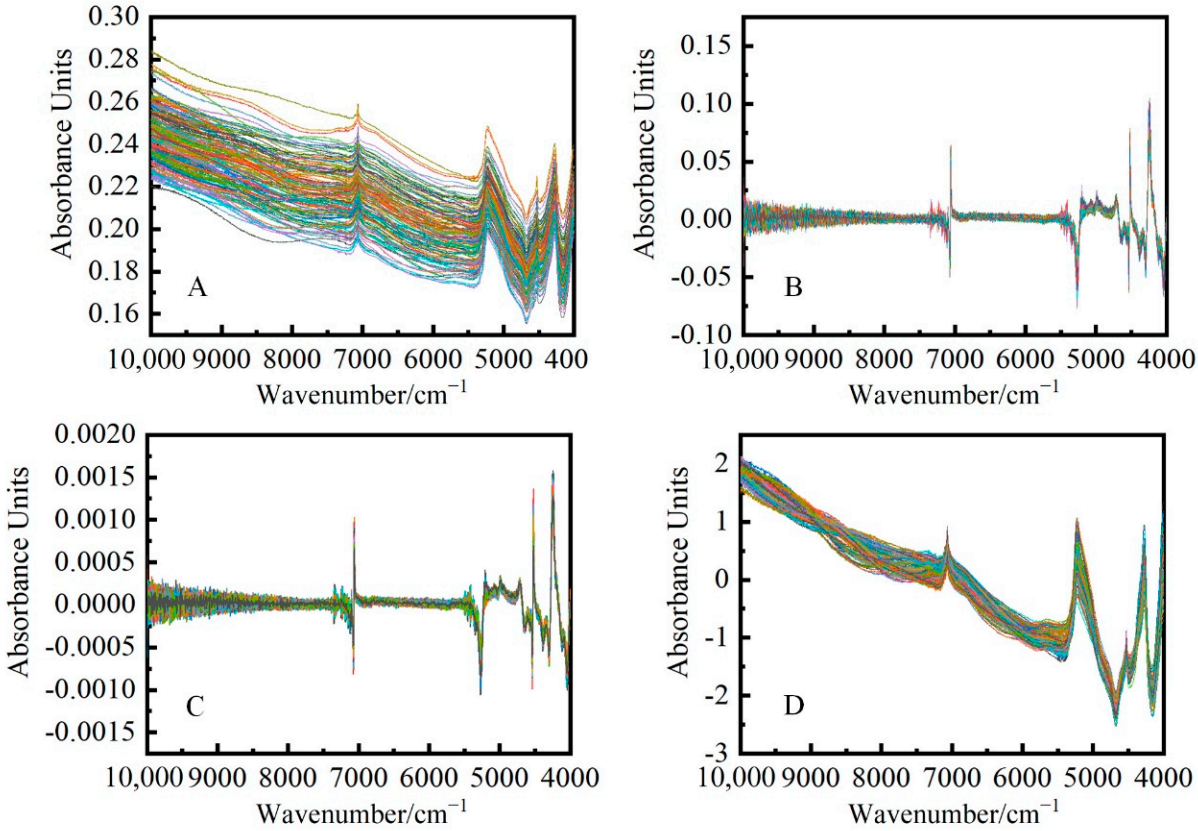


Figure 7. NIR spectra after preprocessing: (**A**) CaO; (**B**) SiO_2; (**C**) Al_2O_3; (**D**) Fe_2O_3.

Subsequently, backward interval partial least squares and genetic algorithm was applied to the four effective components to select the characteristic wavenumber variables. Firstly, feature intervals of each oxide were selected by the biPLS algorithm. The entire spectrum (4000–10,000 cm^{-1}) of cement raw meal was split into 5, 10, 20, 30, 40, and 50 equal intervals, respectively, and 6 biPLS models were obtained, the model with lowest RMSEP value was selected. Then, the range of intervals was narrowed, assuming that the number of intervals of the selected model is m, the full spectrum was divided into $m - 5, \ldots, m - 1, m + 1, \ldots, m + 4$, and $m + 5$ equal intervals, respectively, and 10 biPLS models were established. After that, the intervals corresponding to the lowest RMSEP

and highest R_p models were selected as the characteristic intervals. For CaO, the range of 4000–10,000 cm^{-1} was divided into 30 subintervals, and 8 subintervals (1st, 2nd, 3rd, 4th, 5th, 12th, 14th, and 22nd) were selected; for SiO_2, the range was split into 10 subintervals, and 2 subintervals (1st and 5th) were chosen, and then the genetic algorithm; for Al_2O_3, the range was partitioned into 5 subintervals, and the 1st subinterval was picked; for Fe_2O_3, it was divided into 40 subintervals, a total of 25 subintervals (2nd, 3rd, 4th, 6th, 8th, 9th, 10th, 11th, 13th, 15th, 16th, 17th, 19th, 20th, 21st, 23rd, 24th, 25th, 26th, 27th, 28th, 30th, 31st, 32nd, and 33rd) were selected.

Then genetic algorithm was used to select effective variables from the feature intervals obtained by biPLS method. The GA program parameters employed in this paper were the following: (1) the number of chromosomes was 30; (2) the probability of mutation was 0.01; (3) the probability of cross-over was 0.5; (4) maximum number of variables selected in the chromosome was 30; (5) the number of runs was 100; (6) the iterations selected as termination criteria were 100. The characteristic variables obtained for CaO, SiO_2, Al_2O_3, and Fe_2O_3 are shown in the shaded portion of Figure 8.

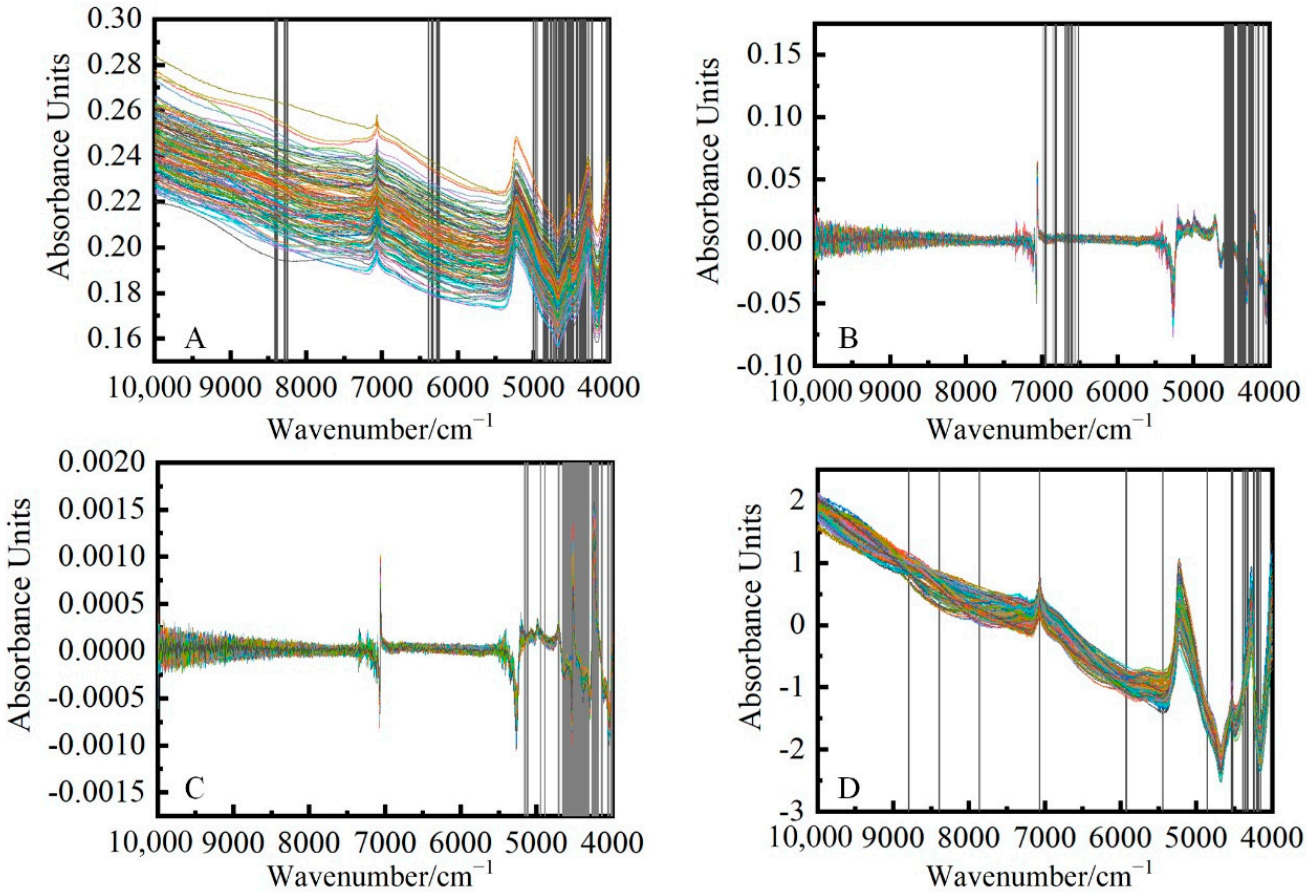


Figure 8. Feature wavenumber variables obtained by backward interval PLS and genetic algorithm (GA-biPLS): (**A**) CaO; (**B**) SiO_2; (**C**) Al_2O_3; (**D**) Fe_2O_3.

Based on the remaining samples in the calibration set and the selected characteristic wavenumber variables, PLS models were developed for CaO, SiO_2, Al_2O_3, and Fe_2O_3, respectively, and the models were externally validated using samples from the prediction set, the parameters are listed in Table 4.

Table 4. The parameters for partial least squares (PLS) models, such as latent variables (LVs), correlation coefficient of calibration set (R_c), correlation coefficient of prediction set (R_p), root mean square error of cross validation (RMSECV), root mean square error of prediction (RMSEP).

Component	LVs	R_c	RMSECV	R_p	RMSEP
CaO	9	0.8727	0.2688	0.8206	0.1910
SiO_2	6	0.7951	0.2854	0.8054	0.2307
Al_2O_3	9	0.8265	0.0981	0.8344	0.0921
Fe_2O_3	8	0.6039	0.0509	0.6584	0.0429

The reference and predicted values for the 50 samples in the prediction set are represented in a line graph as shown in Figure 9, it is obvious that the predicted values of each oxide content obtained by NIR spectroscopy are in general consistent with the reference values obtained by X-ray fluorescence spectroscopy method with small errors.

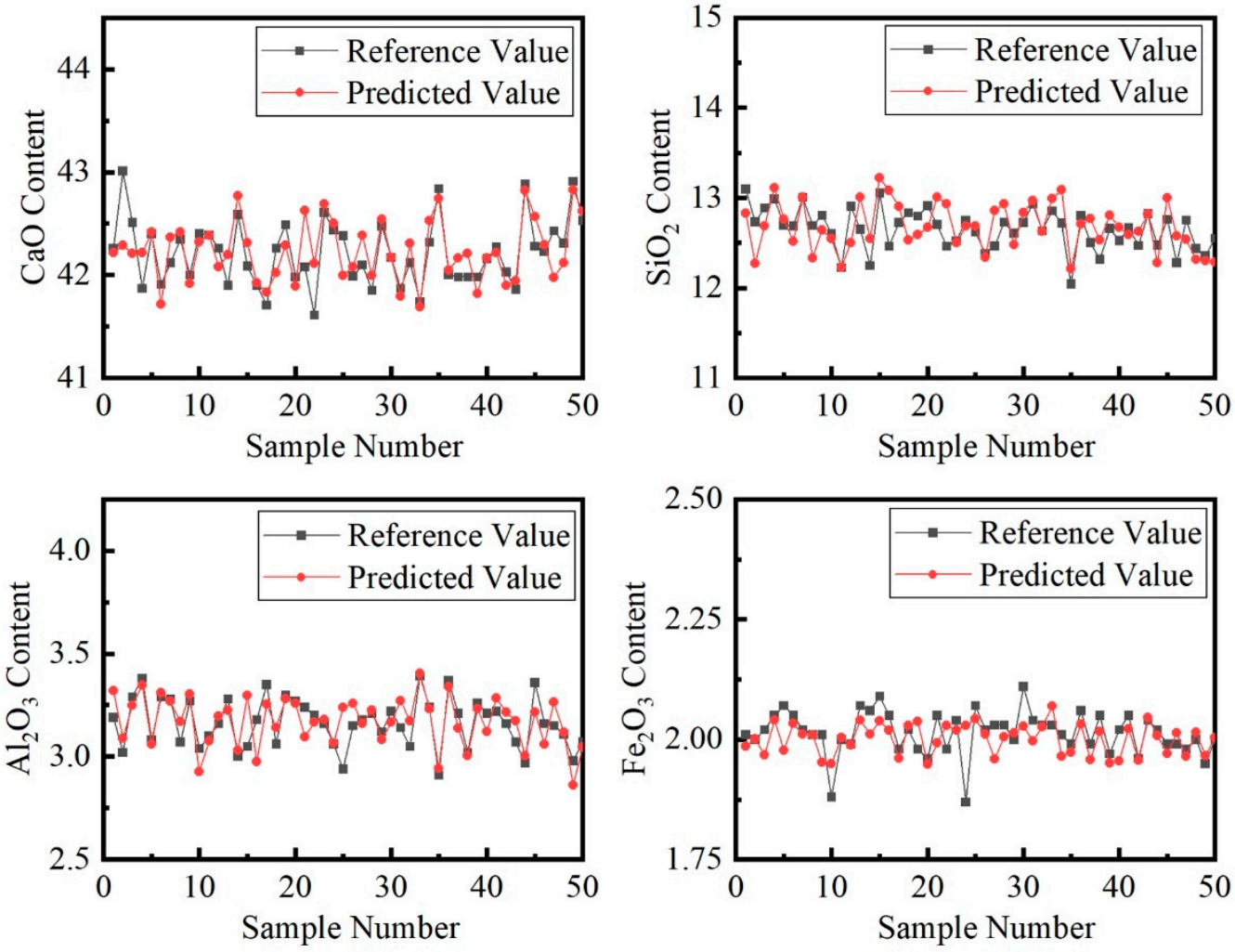


Figure 9. Reference vs. predicted values for the prediction set samples.

The standard methods for chemical analysis of cement (GB/T 176-2017) require that the measurement error for CaO, SiO_2, Al_2O_3, and Fe_2O_3 should be within 0.40%, 0.30%, 0.20%, and 0.15%, respectively. The mean prediction errors of the 50 samples in the prediction set for CaO, SiO_2, Al_2O_3, and Fe_2O_3 were 0.171%, 0.193%, 0.069%, and 0.032%, respectively, which perfectly satisfied the requirements for feedback control in cement production line.

3.2. Samples Collected in Linyi

As listed in Table 5, the reference values for the content of CaO, SiO_2, Al_2O_3 and Fe_2O_3 in the raw meal samples produced in Linyi were determined by XRF method. Although the NIR spectra of the samples in the two groups showed significant differences, the average values of the four ingredients were similar, the content ranges were essentially the same, and the deviation values were not very different.

Table 5. Distribution of reference values for oxides content.

Oxide	Maximum/%	Minimum/%	Deviation/%	Mean/%
CaO	42.85	39.86	2.99	41.17
SiO_2	14.02	11.57	2.45	13.38
Al_2O_3	3.38	2.18	1.20	3.18
Fe_2O_3	3.12	1.90	1.22	2.15

The 119 cement raw meal samples were split into 2 groups at a ratio of approximately 3:1 with the SPXY algorithm, resulting in 89 samples in the calibration set and 30 samples in the prediction set, and the statistical results for CaO, SiO_2, Al_2O_3, and Fe_2O_3 are shown in Table 6. For each of the four oxides, the highest and lowest content samples were included in the calibration set and the content range of the prediction set was contained within that of the calibration set, in accordance with the modeling criteria.

Table 6. Oxides content range for calibration set and validation set samples.

Oxide	Subsets	Number	Maximum/%	Minimum/%	Deviation/%	Mean/%
CaO	Calibration	89	42.85	39.86	2.99	41.20
	Prediction	30	41.47	40.75	0.72	41.09
SiO_2	Calibration	89	14.02	11.57	2.45	13.32
	Prediction	30	14.01	12.88	1.13	13.56
Al_2O_3	Calibration	89	3.38	2.18	1.20	3.16
	Prediction	30	3.35	3.03	0.32	3.24
Fe_2O_3	Calibration	89	3.12	1.90	1.22	2.15
	Prediction	30	2.21	2.06	0.15	2.14

Outliers were identified and eliminated by CVADF method for the four oxides in the calibration set samples. In the cross-validation process, samples in the calibration set were divided into 5 subsets, and the results are as shown in Table 7. Five outliers were found for CaO (Nos. 1, 12, 20, 62, and 81); three outliers were detected for SiO_2 (Nos. 1, 7, and 9); three outliers were identified for Al_2O_3 (Nos. 2, 11, and 33); and three outliers were discovered for Fe_2O_3 (Nos. 2, 33, and 36).

Table 7. Rejected outliers by CVADF method.

Oxide	Total Number	Serial Number
CaO	5	1, 12, 20, 62, 81
SiO_2	3	1, 7, 9
Al_2O_3	3	2, 11, 33
Fe_2O_3	3	2, 33, 36

The optimal spectral preprocessing method for each oxide was chosen, as described in Reference [38], valuable information was enhanced utilizing Savitzky–Golay convolution smoothing (SG), Savitzky–Golay convolution smoothing with multiple scattering correction (SG with MSC), multiple scattering correction (MSC) and Savitzky–Golay convolution smoothing (SG) for the NIR spectra of the calibration set samples, respectively, to obtain

the spectra for the subsequent modeling of CaO, SiO_2, Al_2O_3, and Fe_2O_3, as shown in Figure 10.

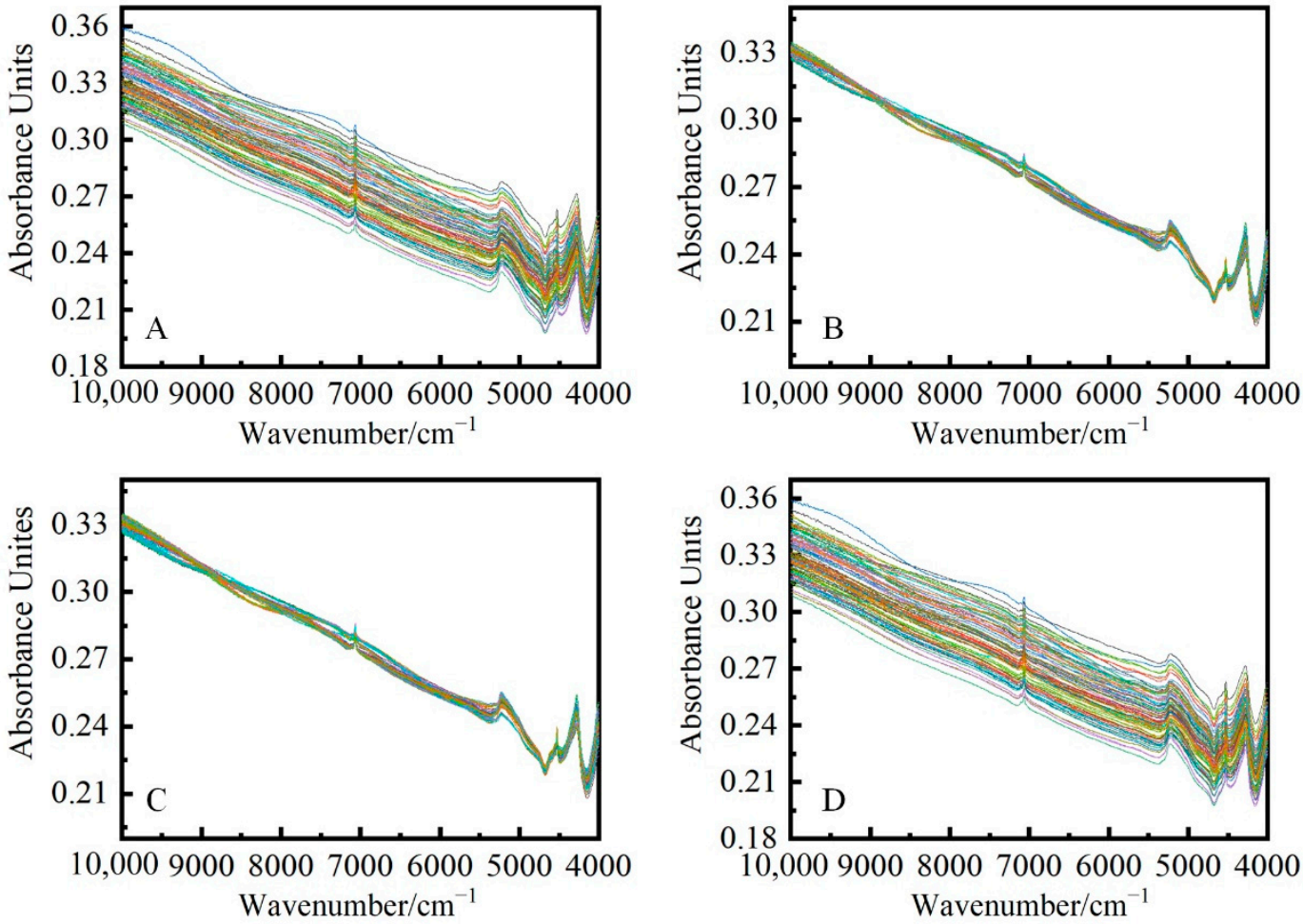


Figure 10. NIR spectra after preprocessing: (**A**) CaO; (**B**) SiO_2; (**C**) Al_2O_3; (**D**) Fe_2O_3.

Afterwards, the GA-biPLS method was employed to pick the feature wavenumber variables for each of the four oxides, the feature intervals of each oxide were selected by the biPLS algorithm in the same way as the samples from Qufu, for CaO, the range of 4000–10,000 cm^{-1} was split into 8 subintervals, and a total of 3 subintervals (2nd, 4th, and 5th) were selected; for SiO_2, the range was divided into 4 subintervals, and the first subinterval was selected; for Al_2O_3, the range was partitioned into 8 subintervals, and a total of 5 subintervals (1st, 3rd, 4th, 5th, and 8th) were chosen; for Fe_2O_3, 30 subintervals were classified, and 10 subintervals (1st, 6th, 10th, 11th, 12th, 13th, 15th, 21st, 25th, and 26th) were picked. Then, the genetic algorithm was employed to select effective variables from the feature intervals obtained by biPLS method with the same parameters as the samples from Qufu, the characteristic variables obtained for CaO, SiO_2, Al_2O_3, and Fe_2O_3 are shown in the shaded portion of Figure 11.

Based on the remaining samples in the calibration set and the determined characteristic wavenumber variables, PLS models are established for CaO, SiO_2, Al_2O_3, and Fe_2O_3 respectively, the parameters are shown in Table 8. Then, the models were externally validated using samples in the predicted set.

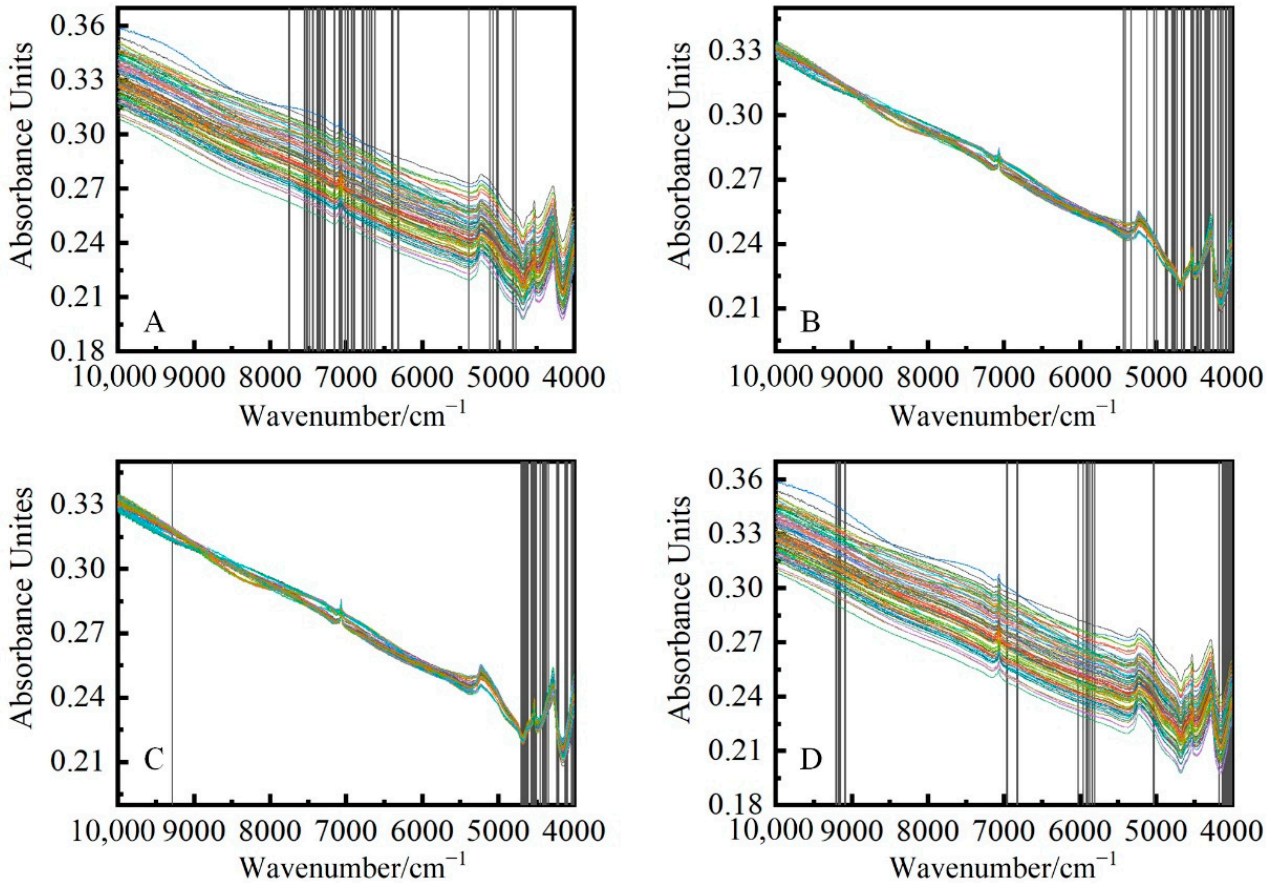


Figure 11. Feature wavenumber variables obtained by GA-biPLS: (**A**) CaO; (**B**) SiO_2; (**C**) Al_2O_3; (**D**) Fe_2O_3.

Table 8. The parameters for partial least squares (PLS) models, namely latent variables (LVs), correlation coefficient of calibration set (R_c), correlation coefficient of prediction set (R_p), root mean square error of cross validation (RMSECV), root mean square error of prediction (RMSEP).

Component	LVs	R_c	RMSECV	R_p	RMSEP
CaO	9	0.8727	0.2688	0.7606	0.1910
SiO_2	6	0.7951	0.2854	0.7654	0.2307
Al_2O_3	9	0.8265	0.0981	0.8344	0.0921
Fe_2O_3	8	0.6039	0.0509	0.6584	0.0429

The prediction errors for the 30 samples in the prediction set are represented in a line graph, as shown in Figure 12. It is clear that the quantitative calibration models have high predictive power and low prediction errors. For CaO, the prediction error fluctuates up and down around 0.15%, with a maximum error of 0.38%; for SiO_2, the prediction error varies up and down around 0.10%, with a maximum error of 0.28%; for Al_2O_3, the prediction error fluctuates up and down around 0.02%, with a maximum error of 0.10%; and for Fe_2O_3, the prediction error varies up and down around 0.02%, with a maximum error of 0.06%.

The average prediction errors for the 30 samples In the prediction set are 0.154%, 0.100%, 0.022%, and 0.018% for CaO, SiO_2, Al_2O_3, and Fe_2O_3, respectively, which are fully adequate for the feedback control of cement production line.

Although the types and sources of raw materials used in the Linyi and Qufu samples differ, the models obtained through the proposed modeling process both showed good predictive accuracy, demonstrating that the modeling procedure proposed in this paper has good generality and it is completely capable of meeting the detection requirements of different production lines.

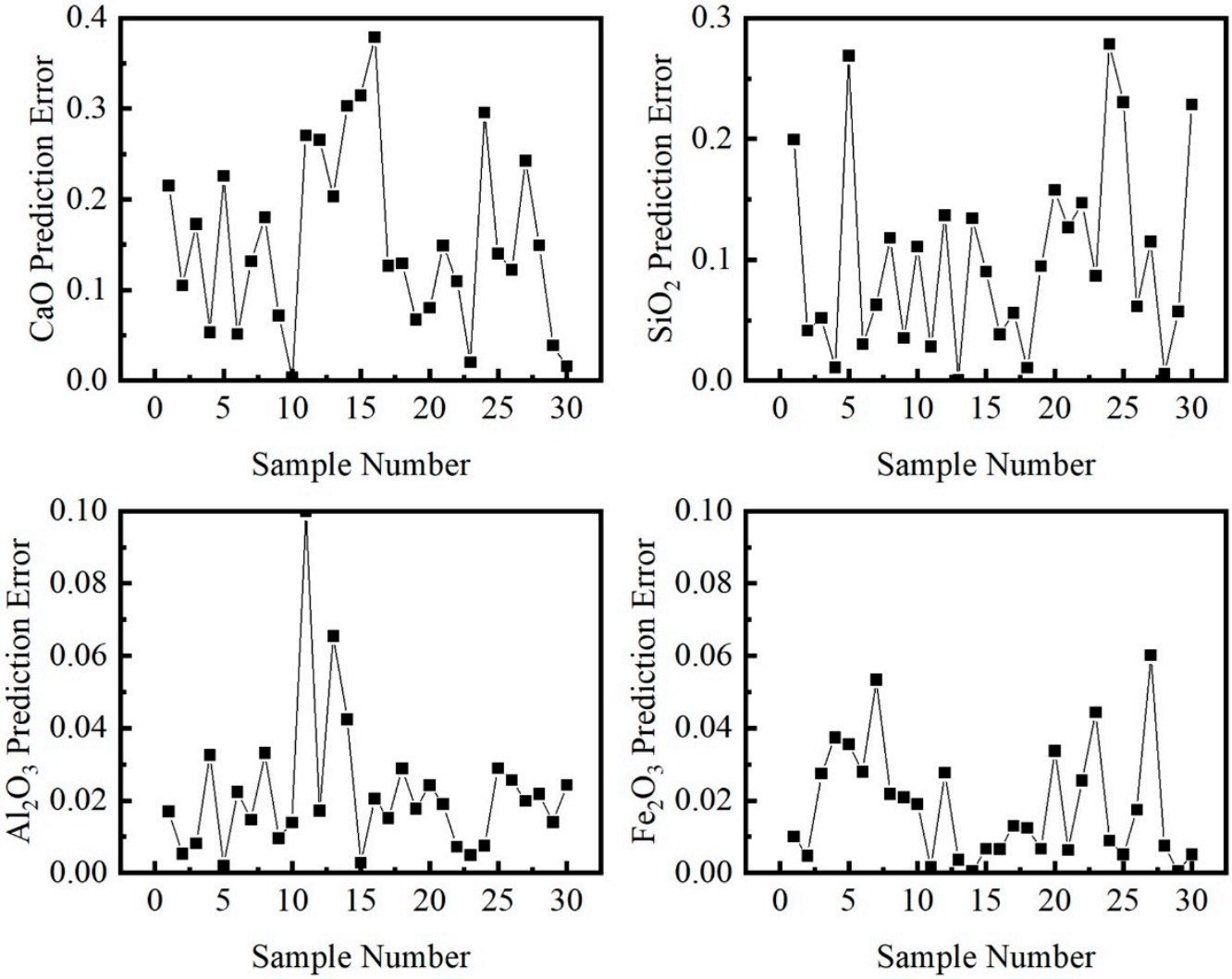


Figure 12. Prediction error for the prediction set samples.

4. Conclusions

In order to obtain accurate composition content of cement raw meal rapidly, a quantitative analytical model modeling process for near infrared spectroscopy is proposed. The modeling procedure is as follows:

- Firstly, the cement raw meal sample set is divided into calibration set and prediction set by SPXY algorithm.
- Secondly, the outliers in the calibration set are eliminated by CVADF approach.
- Thirdly, the optimal spectral preprocessing method is selected to enhance the valuable information in the NIR spectra.
- Fourthly, the spectral feature variables intimately allied to the component content are obtained using GA-biPLS method.
- Fifthly, the PLS model is established and externally validated.

The parameters of the analytical models obtained from the two groups of samples with different raw material types and origins are as follows:

- For the PLS models developed for the samples from Qufu, the R_p of CaO, SiO_2, Al_2O_3, and Fe_2O_3 are 0.8206, 0.8054, 0.8344, and 0.6584, respectively, the RMSEP of CaO, SiO_2, Al_2O_3, and Fe_2O_3 are 0.1910, 0.2307, 0.0921, and 0.0429, respectively.

- For the PLS models built for the samples from Linyi, the R_p of CaO, SiO_2, Al_2O_3, and Fe_2O_3 are 0.8284, 0.8959, 0.8846, and 0.6823, respectively, the RMSEP of CaO, SiO_2, Al_2O_3, and Fe_2O_3 are 0.1995, 0.1267, 0.0336, and 0.0242, respectively.

Despite the fact that the raw materials of cement raw meal samples differ in type and origin, the PLS models obtained through the proposed modeling procedure has excellent predictive performance, indicating that the modeling procedure is highly versatile, and the NIR spectroscopy in conjunction with the modeling procedure provides a rapid alternative method for determining the composition of cement raw meal to satisfy the requirements of different cement production lines for raw material proportioning feedback control.

Author Contributions: Conceptualization, Z.Y. and L.J.; methodology, Z.Y. and Q.S.; software, Z.Y.; validation, Z.Y., Q.S. and L.J.; formal analysis, Q.S.; investigation, Z.Y.; data curation, Z.Y.; writing—original draft preparation, Z.Y.; writing—review and editing, Q.S. and L.J.; visualization, Z.Y.; supervision, L.J.; project administration, Q.S.; funding acquisition, Z.Y., Q.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by National Natural Science Foundation of China (No. 61903224, No. 61873333 and No. 62073193), Shandong Provincial Keypoint Research and Development Program (No. 2019TSLH0301, No. 2019GHZ004 and No. 2021CXGC010204), Shandong Provincial Natural Science Foundation (No. ZR2022QF038).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lan, Y.; Zheng, B.; Shi, T.; Ma, C.; Liu, Y.; Zhao, Z. Crack resistance properties of carbon nanotube-modified concrete. *Mag. Concr. Res.* **2022**, *74*, 1165–1175. [CrossRef]
2. Zhang, W.; Huang, Y. Three-dimensional numerical investigation of mixed-mode debonding of FRP-concrete interface using a cohesive zone model. *Constr. Build. Mater.* **2022**, *350*, 128818. [CrossRef]
3. Shi, T.; Liu, Y.; Zhao, X.; Wang, J.; Zhao, Z.; Corr, D.J.; Shah, S.P. Study on mechanical properties of the interfacial transition zone in carbon nanofiber-reinforced cement mortar based on the PeakForce tapping mode of atomic force microscope. *J. Build. Eng.* **2022**, *61*, 105248. [CrossRef]
4. Kural, A.; Özsoy, C. Identification and control of the raw material blending process in cement industry. *Int. J. Adapt. Control Signal Process.* **2004**, *18*, 427–442. [CrossRef]
5. Tsamatsoulis, D. Modelling and simulation of raw material blending process in cement raw mix milling installations. *Can. J. Chem. Eng.* **2015**, *92*, 1882–1894. [CrossRef]
6. Law, O.W.; Lam, L.; Luk, S.F. Analysis of cement and cement raw meal by atomic absorption spectrophotometry using a new fusion agent. *Talanta* **1995**, *42*, 1265. [CrossRef] [PubMed]
7. Elbagermia, M.A.; Alajtala, A.I.; Alkerzab, M. Chemical analysis of available Portland cement in Libyan market using X-ray fluorescence. *Int. J. Chem. Mol. Nucl. Mater. Metall. Eng.* **2014**, *8*, 73–75.
8. Stutzman, P.; Heckert, A. Performance criteria for chemical analysis of hydraulic cements by X-ray fluorescence analysis. *Adv. Civ. Eng. Mater.* **2014**, *3*, 434–453. [CrossRef]
9. Rad, S.Z.I.; Peyvandi, R.G. An optimal experimental-Monte Carlo combined model for analyzing of the cement raw by prompt gamma neutron activation. *Russ. J. Nondestruct. Test.* **2018**, *54*, 448–454. [CrossRef]
10. Millar, S.; Kruschwitz, S.; Wilsch, G. Determination of total chloride content in cement pastes with laser-induced breakdown spectroscopy (LIBS). *Cem. Concr. Res.* **2019**, *117*, 16–22. [CrossRef]
11. Yang, Z.; Xiao, H.; Sui, Q.; Zhang, L.; Jia, L.; Jiang, M.; Zhang, F. Novel Methodology to Improve the Accuracy of Oxide Determination in Cement Raw Meal by near Infrared Spectroscopy (NIRS) and Cross-Validation-Absolute-deviation-F-Test (CVADF). *Anal. Lett.* **2020**, *53*, 2734–2747. [CrossRef]
12. Fan, J.; Zhang, L.; Wang, X.; Li, Y.; Gong, Y.; Dong, L.; Ma, W.; Yin, W.; Wang, Z.; Li, Z.; et al. Development of a Laboratory Cement Quality Analysis Apparatus Based on Laser-Induced Breakdown Spectroscopy. *Plasma Sci. Technol.* **2015**, *17*, 897. [CrossRef]
13. Shang, H.; Shang, L.; Wu, J.; Xu, Z.; Zhou, S.; Wang, Z.; Wang, H.; Yin, J. NIR spectroscopy combined with 1D-convolutional neural network for breast cancerization analysis and diagnosis. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2023**, *287*, 121990. [CrossRef] [PubMed]
14. Yu, D.-X.; Guo, S.; Zhang, X.; Yan, H.; Zhang, Z.-Y.; Chen, X.; Chen, J.-Y.; Jin, S.-J.; Yang, J.; Duan, J.-A. Rapid detection of adulteration in powder of ginger (*Zingiber officinale* Roscoe) by FT-NIR spectroscopy combined with chemometrics. *Food Chem. X* **2022**, *15*, 100450. [CrossRef]

15. Phanomsophon, T.; Jaisue, N.; Worphet, A.; Tawinteung, N.; Shrestha, B.; Posom, J.; Khurnpoon, L.; Sirisomboon, P. Rapid measurement of classification levels of primary macronutrients in durian (*Durio zibethinus* Murray CV. Mon Thong) leaves using FT-NIR spectrometer and comparing the effect of imbalanced and balanced data for modelling. *Measurement* **2022**, *203*, 111975. [CrossRef]
16. Yang, Y.; Zhuang, H.; Yoon, S.-C.; Wang, W.; Jiang, H.; Jia, B.; Li, C. Quality assessment of intact chicken breast fillets using factor analysis with Vis/NIR spectroscopy. *Food Anal. Methods* **2018**, *11*, 1356–1366. [CrossRef]
17. Sun, X.; Zhu, K.; Jiang, X.; Liu, Y. Non-destructive detection of blackheart and soluble solids content of intact pear by online NIR spectroscopy. *J. Supercomput.* **2020**, *76*, 3173–3187. [CrossRef]
18. Sohn, S.-I.; Pandian, S.; Oh, Y.-J.; Zaukuu, J.-L.Z.; Na, C.-S.; Lee, Y.-H.; Shin, E.-K.; Kang, H.-J.; Ryu, T.-H.; Cho, W.-S.; et al. Vis-NIR Spectroscopy and Machine Learning Methods for the Discrimination of Transgenic *Brassica napus* L. and Their Hybrids with *B. juncea*. *Processes* **2022**, *10*, 240. [CrossRef]
19. Le, B.T. Application of deep learning and near infrared spectroscopy in cereal analysis. *Vib. Spectrosc.* **2020**, *106*, 103009. [CrossRef]
20. Bi, Y.; Li, S.; Zhang, L.; Li, Y.; He, W.; Tie, J.; Liao, F.; Hao, X.; Tian, Y.; Tang, L.; et al. Quality evaluation of flue-cured tobacco by near infrared spectroscopy and spectral similarity method. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2019**, *215*, 398–404. [CrossRef]
21. Nazarloo, A.S.; Sharabiani, V.R.; Gilandeh, Y.A.; Taghinezhad, E.; Szymanek, M.; Sprawka, M. Feasibility of Using VIS/NIR Spectroscopy and Multivariate Analysis for Pesticide Residue Detection in Tomatoes. *Processes* **2021**, *9*, 196. [CrossRef]
22. Ma, H.; Pan, H.; Pan, D.; Ni, H.; Feng, X.; Liu, X.; Chen, Y.; Wu, Y.; Luo, N. Rapid monitoring approaches for concentration process of lanqin oral solution by near-infrared spectroscopy and chemometric models. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *242*, 118792. [CrossRef]
23. Gao, L.; Zhong, L.; Zhang, J.; Zhang, M.; Zeng, Y.; Li, L.; Zang, H. Water as a probe to understand the traditional Chinese medicine extraction process with near infrared spectroscopy: A case of Danshen (*Salvia miltiorrhiza Bge*) extraction process. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *244*, 118854. [CrossRef] [PubMed]
24. Wang, S.; Liu, S.; Zhang, J.; Che, X.; Wang, Z.; Kong, D. Feasibility study on prediction of gasoline octane number using NIR spectroscopy combined with manifold learning and neural network. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *228*, 117836. [CrossRef]
25. Velvarská, R.; Vráblík, A.; Hidalgo-Herrador, J.M.; Černý, R. Near-infrared spectroscopy to determine cold-flow improver concentrations in diesel fuel. *Infrared Phys. Technol.* **2020**, *110*, 103445. [CrossRef]
26. Dennis, J.J.; Wiggins, C.C.; Smith, J.R.; Isautier, J.M.J.; Johnson, B.D.; Joyner, M.J.; Cross, T.J. Measurement of muscle blood flow and O_2 uptake via near-infrared spectroscopy using a novel occlusion protocol. *Sci. Rep.* **2021**, *11*, 918. [CrossRef]
27. Kurata, Y. A comparison of the loading direction for bending strength with different wood measurement surfaces using Near-Infrared spectroscopy. *Forests* **2020**, *11*, 644. [CrossRef]
28. Tsegay, G.; Ammare, Y.; Mesfin, S. Development of non-destructive NIRS models to predict oil and major fatty acid contents of Ethiopian sesame. *J. Food Compos. Anal.* **2023**, *115*, 104908. [CrossRef]
29. Niemi, C.; Mortensen, A.M.; Rautenberger, R.; Matsson, S.; Gorzsás, A.; Gentili, F.G. Rapid and accurate determination of protein content in North Atlantic seaweed by NIR and FTIR spectroscopies. *Food Chem.* **2023**, *404*, 134700. [CrossRef]
30. Galvao, R.K.H.; Araujo, M.C.U.; Jose, G.E.; Pontes, M.J.C.; Silva, E.C.; Saldanha, T.C.B. A method for calibration and validation subset partitioning. *Talanta* **2005**, *67*, 736–740. [CrossRef]
31. Saptoro, A.; Tadé, M.O.; Vuthaluru, H. A modified Kennard-Stone algorithm for optimal division of data for developing artificial neural network models. *Chem. Prod. Process Model.* **2012**, *7*, 13. [CrossRef]
32. Pierna, J.A.F.; Wahl, F.; de Noord, O.E.; Massart, D.L. Methods for outlier detection in prediction. *Chemom. Intell. Lab. Syst.* **2002**, *63*, 27–39. [CrossRef]
33. Li, Z.; Xu, G.; Wang, J.; Du, G.; Cai, W.; Shao, X. Outlier detection for multivariate calibration in Near Infrared spectroscopic analysis by model diagnostics. *Chin. J. Anal. Chem.* **2016**, *44*, 305–309. [CrossRef]
34. Sampaio, P.S.; Soares, A.; Castanho, A.; Almeida, A.S.; Oliveira, J.; Brites, C. Optimization of rice amylose determination by NIR-spectroscopy using PLS chemometrics algorithms. *Food Chem.* **2018**, *242*, 196–204. [CrossRef]
35. Leardi, R.; Nørgaard, L. Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. *J. Chemom.* **2004**, *18*, 486–497. [CrossRef]
36. He, H.-J.; Wang, Y.; Zhang, M.; Wang, Y.; Ou, X.; Guo, J. Rapid determination of reducing sugar content in sweet potatoes using NIR spectra. *J. Food Compos. Anal.* **2022**, *111*, 104641. [CrossRef]
37. Leardi, R. Application of genetic algorithm-PLS for feature selection in spectral data sets. *J. Chemom.* **2000**, *14*, 643–655. [CrossRef]
38. Yang, Z.; Xiao, H.; Zhang, L.; Feng, D.; Zhang, F.; Jiang, M.; Sui, Q.; Jia, L. Fast determination of oxides content in cement raw meal using NIR spectroscopy combined with synergy interval partial least square and different preprocessing methods. *Measurement* **2020**, *149*, 106990. [CrossRef]
39. Yang, Z.; Xiao, H.; Zhang, L.; Feng, D.; Zhang, F.; Jiang, M.; Sui, Q.; Jia, L. Fast determination of oxide content in cement raw meal using NIR spectroscopy with the SPXY algorithm. *Anal. Methods* **2019**, *11*, 3936–3942. [CrossRef]
40. Ren, D.; Qu, F.F.; Lu, A.X. *Near Infrared Spectroscopy Analysis Technology and Application*; Science Press: Beijing, China, 2017.
41. Dhanoa, M.S.; Barnes, R.J.; Lister, S.J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Appl. Spectrosc.* **2016**, *43*, 772–777.

42. Delwiche, S.R.; Rd, R.J. A graphical method to evaluate spectral preprocessing in multivariate regression calibrations: Example with Savitzky-Golay filters and partial least squares regression. *Appl. Spectrosc.* **2010**, *64*, 73–82. [CrossRef] [PubMed]
43. Yang, Z.; Xiao, H.; Zhang, L.; Feng, D.; Zhang, F.; Jiang, M.; Sui, Q.; Jia, L. Fast determination of oxides content in cement raw meal using NIR-spectroscopy and backward interval PLS with genetic algorithm. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2019**, *223*, 117327. [CrossRef] [PubMed]