In Situ Monitoring of Nitrate Content in Leafy Vegetables Using Mid-Infrared Attenuated Total Reflectance Spectroscopy coupled with Intelligent Algorithm

Fei Ma
Institute of Soil Science Chinese Academy of Sciences
changwen Du (chwdu@issas.ac.cn)
Institute of Soil Science Chinese Academy of Sciences https://orcid.org/0000-0002-9064-3581
Shuailin Zhen
Institute of Soil Science Chinese Academy of Sciences
Yaxiao Du
Institute of Soil Science Chinese Academy of Sciences

Methodology

Keywords: Leafy vegetables, nitrate, mid-infrared attenuated total reflectance, intelligent algorithm, extreme learning machine

Posted Date: April 1st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-20060/v1

In situ monitoring of nitrate content in leafy vegetables using mid-infrared attenuated total reflectance spectroscopy coupled with intelligent algorithm

Fei Ma[a], Changwen Du[a,b*], Shuailin Zheng[a,b], Yaxiao Du[a,b]

*Correspondence to: C Du, Institute of Soil Science Chinese Academy of Sciences, Nanjing 210008.E-mail: chwdu@issas.ac.cn

a *The State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science Chinese Academy of Sciences, Nanjing 210008*

b *University of Chinese Academy of Sciences, Beijing 100049, China*

E-mail address:

Fei Ma: E-mail: fma@issas.ac.cn

Changwen Du: E-mail: chwdu@issas.ac.cn

Zhen Shailin: E-mail: 1605659752@qq.com

Du Yaxiao: E-mail: duyaxiao@foxmail.com

ABSTRACT:

Background: Vegetables are one of the most important nitrate sources of human diary diet. Establishing fast and accurate *in situ* nitrate monitoring approaches that could be used in the plant growth process and vegetable markets is essential.

Results: Incorporating the unique feature of N–O asymmetric stretch absorption in the mid-infrared region (1500-1200 cm^{-1}), portable Fourier-transform infrared attenuated total reflectance (FTIR-ATR) spectroscopic instruments, along with the Euclidean distance-modified intelligent algorithm extreme learning machine (ED-ELM) model, were employed to evaluate the nitrate contents in leafy vegetables. A total of 1224 samples of four popular vegetables (Chinese cabbage, swamp cabbage, celery, and lettuce) were analyzed. The results indicated that the nitrate contents (mean values: Chinese cabbage: 7550 mg/kg; swamp cabbage: 4219 mg/kg; celery: 4164 mg/kg; lettuce: 4322 mg/kg) highly exceeded the World Health Organization (WHO))-specified maximum tolerance limits. The ED-ELM model showed a better performance with the root-mean-square-error of 799.7 mg/kg, the determination coefficients of 0.93, the ratio of performance to deviation of 2.22, the optimized calibration dataset number of 100, and the number of hidden neurons of 30.

Conclusion: The results confirmed that FTIR-ATR, along with the suitable model algorithms, could be used as a potential rapid and accurate method to monitor the nitrate contents in the fields of agriculture and food safety.

Keywords: Leafy vegetables; nitrate, mid-infrared attenuated total reflectance; intelligent algorithm; extreme learning machine

Background

Nitrate is the most important form of nitrogen in the environment and human diet. Vegetables are one of the most important sources of vitamins, minerals, and biologically active compounds with regard to human nutrition [1]. Vegetables, in daily human diets, are also the primary sources of ingested nitrates, and about 80-85% of the daily nitrate intake comes from vegetables [2, 3], along with fruits, water, and additives in meat [4, 5].

Generally, nitrate is not treated as a directly toxic ion, but when it is mixed with food, it can be converted into nitrite by commensal bacteria in the mouth and gastrointestinal tract. Nitrites are further converted into nitrosamines, which are carcinogenic and teratogenic N-nitroso compounds [1, 6]. Previous studies have indicated that nitrate may be beneficial for human health [7, 8], and it was assumed that dietary nitrate could generate nitric oxide, which has antimicrobial effects on gut pathogens, thus providing gastric protection against microbial infections [8, 9]. Whether nitrate can be referred to as an essential nutrient or a food contaminant with potential adverse effects depends on its concentration [10]. Thus, the content of ingested nitrate is critical, and it is necessary to set a maximum limit value for nitrate intake [1].

On one hand, nitrate accumulation in plants, especially in most leafy vegetables, is a major concern [11]. Nitrates in the soil are the primary nutrients required for plant growth [12]. Moreover, nitrate fertilizers that have been used in agriculture result in the accumulation of high levels of nitrate in a variety of vegetables [3]. When nitrate uptake exceeds nitrate assimilation by the plant, accumulation of nitrate in plant tissues occurs. Other factors in the plant growth process that may influence nitrate concentration include the plant species, environmental conditions (e.g. light intensity,

68 temperature, and humidity), harvest time, and storage time [13-15]. For instance, a
69 significant decrease in nitrate level is observed at ambient temperatures, but nitrate
70 level remains constant over time during storage under refrigerated conditions [1, 16].
71 On the other hand, risk assessment of the safety of dietary nitrate intake and
72 exposure from vegetables has been a major health concern in many countries in recent
73 decades. According to the International Agency for Research on Cancer [17],
74 vegetables can be divided into three levels based on their nitrate concentrations: low
75 nitrate (< 100 mg/kg), medium nitrate (100−1000 mg/kg), and high nitrate (> 1000
76 mg/kg) [17]. The European Commission established the maximum level of nitrates in
77 two leaf vegetables: 2000-3500 mg/kg in spinach and 2000-4500 mg/kg in lettuce
78 [18].
79 Therefore, establishing fast and accurate nitrate monitoring approaches,
80 especially nondestructive *in situ* methods that could be used in the plant growth
81 process and vegetable markets, is essential. Over the years, spectrophotometry,
82 photometry, potentiometry, spectrofluorimetry, ion chromatography, gas
83 chromatography and high-performance liquid chromatography have been used widely
84 to monitor the contents of various compounds in vegetables and other foods [14,
85 19-23]. However, these approaches require sample pre-preparation and need to be
86 performed in the laboratory. Portable instruments could be used effectively for *in situ*
87 nitrate monitoring. The use of Fourier-transform infrared (FTIR) spectroscopy
88 techniques has been spreading widely owing to the advantages of this nondestructive,
89 fast, and reliable approach for quality assessment in agro-food industries. Nitrogen
90 distribution in the leaves of Chinese cabbage and nitrogen status in rice were
91 measured based on FTIR photoacoustic spectroscopy [24, 25]. The nutritional and

functional components of leafy vegetables, including Chinese cabbage, beans, and pea seeds, can be successfully determined by the near-infrared and mid-infrared total and diffuse reflectance spectroscopy methods using the absorption feature at various wavelength/wavenumber [26-29]. Nitrate in the soil was reported to be detected by attenuated total reflectance (ATR) spectroscopy [30]. Thus, the use of portable FTIR instruments has extended the application of spectroscopy in the field of agriculture in recent years, making it possible for *in-situ* and *real-time* measurements of nitrate contents in vegetables in vegetable markets and the plant growth process.

Based on the analysis of the collected FTIR spectra data, a rapid, robust, computationally efficient artificial intelligence-based model framework was developed for the *in situ* and real-time monitoring of the nitrate content in leafy vegetables. To address complex data sets with many predictor variables, machine learning techniques are extensively used among researchers [31]. Here, estimating vegetable nitrate content using a proper machine-learning model was the basis for *in situ* measurements. Chinese cabbage (*Brassica rapa subsp. chinensis*), swamp cabbage (*Ipomoea aquatica Forssk*), celery (*Apium graveolens L.*), and lettuce (*Lactuca sativa L.*), the main vegetable species grown and consumed in southeast China, were chosen for this study. Fresh samples of the above-mentioned vegetables were purchased in local markets during harvest time (vegetative stage). The objectives of this research were to: i) investigate the nitrate contents in local leafy vegetables; ii) collect the spectral data of these vegetables using a portable FTIR-ATR device and analyze the features of nitrate absorption in vegetables; and iii) build an accurate *in situ* intelligent algorithm method for nitrate content measurements. The results provided a potential method on the development of a fast and feasible approach to monitor nitrate contents even in markets for the healthy management and

consumption of foodstuffs, making this research a good strategy for food safety assessment.

Methods

Collection of leafy vegetable samples

Four species of leafy vegetables, Chinese cabbage, swamp cabbage, celery, and lettuce, were purchased from four large local supermarkets and traditional wet markets in Nanjing, China, from August 2019 to September 2020. A total of 408 samples of the four species of vegetables were purchased (102 samples of each species). Thirty-two samples of each of the four species of vegetables were purchased from four markets on a single day, and they were analyzed by both spectral and laboratory methods on the same day to ensure that the vegetables were fresh and nitrate contents were relatively stable.

Chemicals

Hydrochloric acid (ρ = 1.19 g/mL, analytical reagent grade, AR), potassium nitrate (AR), ammonia (25% wt.), zinc sulfate heptahydrate (AR), potassium ferrocyanide (AR), octanol (AR), and activated carbon powder were purchased from the China National Pharmaceutical Group Corporation (Beijing, China). Deionized water (15 $M\Omega\ cm^{-1}$) was prepared using a laboratory water system.

Nitrate content analysis

Nitrate content was analyzed following the national standard method (GB/T 5009.33–2016, China: Rapid determination of nitrate in vegetables), established by the Chinese National Institute of Metrology. The vegetable samples were mashed

using a planetary ball mill for 10 min. Then, vegetable homogenate (5 g), deionized water (10 g), ammonia buffer (5 mL) (pH = 9.6–9.7), and activated carbon powder were added to a conical flask, and the mixture was stirred (200 r/min) at 25°C for 30 min. The mixture was then transferred to a volumetric flask (250 mL) and mixed with 150 g/L potassium ferrocyanide solution (2 mL) and 300 g/L zinc sulfate solution (2 mL); deionized water was added to bring the volume of the resulting solution to 250 mL. This mixture was kept standing for 5 min and then filtered. Then, the filtered solution (2-3 mL) and deionized water were placed in a volumetric flask (25 mL), and the absorbance of the samples was measured at 219 nm by ultraviolet spectrophotometry. Meanwhile, nitrate standard solutions of different concentrations were measured to plot the standard curve. Finally, the nitrate content was calculated using the following formula:

$$C_{\text{nitrate}} = \frac{\rho \times V_e \times V_u}{m \times V_a} \quad (1)$$

where C is the nitrate content in mg/kg; ρ is the value of nitrate obtained from the ultraviolet spectrophotometry standard curve; V_e is the constant volume in the volumetric flask used in the extraction process (250 mL); V_u is the constant volume for ultraviolet spectrophotometry measurement (25 mL); m is the mass of the vegetable homogenate (accurate to 0.01 g); and V_a is the volume of the filtered solution.

FTIR-ATR spectroscopy

Each vegetable was ground, and the vegetable juice obtained was scanned on a hand-held TruDefender FTIR spectrometer with an ATR spectra accessory (Thermo Fisher Scientific, USA). Spectra of the samples were recorded over the original range

166 from 4000 to 400 cm^{-1}, with a spectral resolution of 4 cm^{-1}. Atmospheric and
167 instrumental noise was corrected by subtracting the background noise from each scan.
168 Each ground sample was placed on a diamond reflection probe for three
169 measurements, and a blank reference was scanned before the spectra for each sample
170 were recorded. Spectra from the 1224 samples were collected. The spectra of a serial
171 nitrate standard solution (with concentrations ranging from 0 to 15000 mg/kg) were
172 recorded before each of the four samples was scanned.
173
174 **Pre-processing of the spectra**
175 The FTIR-ATR spectra were pre-processed with a Savitzky–Golay smoothing filter to
176 eliminate baseline float and noise and improve the signal-to-noise ratio [24, 25].
177 Savitzky–Golay smoothing seemed to be superior to adjacent averaging because it
178 reserved spectral features, such as peak height and width. The spectral range from
179 1500 to 1200 cm^{-1} was selected based on the absorption characteristics of nitrate.
180 Moreover, the second derivative spectra of nitrate in the range from 1500 to 1200 cm^{-1}
181 were obtained. Principal component analysis (PCA) was also performed. The
182 MATLAB R2013a software (MathWorks, Natick, MA, USA) and related scripts were
183 used to perform other statistical analyses.
184
185 **Theoretical overview**
186 The pre-processed spectra were divided into calibration and validation datasets. Then,
187 the extreme learning machine (ELM) model, an intelligent algorithm, was employed
188 to predict the nitrate contents. To improve the prediction accuracy, the calibration
189 dataset was modified before being calibrated by the Euclidean distance (ED) method.
190 Meanwhile, the partial least squares (PLS) model was used for comparison.

191 Subsequently, the performance of the models and prediction results were evaluated.

192

193 **Calibration and validation datasets**

194 The 1224 vegetable samples were randomly divided into a calibration dataset
195 (training dataset) containing 74% (900 samples) of the spectra, and a validation
196 dataset, containing the remaining 26% (324 samples) of the spectra.

197

198 **Modified Euclidean distance**

199 Each FTIR-ATR spectral curve represented the spectral features and concentration
200 level of nitrate, and the linear matching of the nitrate concentration and spectra curves
201 showed good correlation coefficients [32]. The spectral curves within a small intensity
202 range were similar, and the nitrate concentrations were closer. Based on this
203 observation, the ED method was employed to recognize similar spectral curves in the
204 calibration data set for further modeling. The ED method was selected in this study
205 based on our previous result, which reported that it was suitable for spectral
206 identification [33]. ED between the calibration and target samples was computed
207 using pairs of curves and their derivatives as a measure of similarity for clustering.

208 $$ED_{ik} = \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{kj})^2} \qquad (2)$$

209 where ED_{ik} is the Euclidean distance between the ith target sample x_i and each kth
210 calibration sample x_k, $k \neq i$; and j is the variable index, $j = 1, 2, \ldots p$. The calibration
211 data set sequence was re-ordered in an ascending manner based on the ED results,
212 which meant that similar spectra were near-neighbors.

213

214 **Extreme learning machine model**

ELM belonged to a single hidden-based layer forward network [34]. For a sample set (x_i, t_i), where $x_i = (x_{i1}, x_{i2}, \ldots x_{in})^T \in R^n$ and $t_i = (t_{i1}, t_{i2}, \ldots t_{in})^T \in R^K$, the standard single hidden-based layer forward network with L hidden nodes and activation function $h(x)$ was mathematically modeled as [35] :

$$\sum_{i=1}^{L} \beta_i h_i(x_j) = \sum_{i=1}^{L} \beta_i h_i(w_i \times x_j + b_i) = o_j \quad (3)$$

where $j = 1, 2, \ldots n$; $w_i = (w_{i1}, w_{i2}, \ldots w_{in})^T$ is the weight vector connecting the ith hidden node to the input nodes, $\beta_i = (\beta_{i1}, \beta_{i2}, \ldots \beta_{iK})^T$ is the weight vector connecting the ith hidden node to the output nodes; and b_i is the threshold of the ith hidden node. Then,

$$H\beta = T \quad (4)$$

$$H = \begin{bmatrix} h_1(w_1 x_1 + b_1) & \cdots & h_L(w_L x_1 + b_L) \\ \vdots & \ddots & \vdots \\ h_1(w_1 x_N + b_1) & \cdots & h_L(w_L x_N + b_L) \end{bmatrix}_{N \times L} \quad (4)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times K} \text{ and } T = \begin{bmatrix} t_1^T \\ \vdots \\ t_L^T \end{bmatrix}_{N \times K} \quad (5)$$

The difference between conventional gradient-based solution methods and the ELM method was that the ELM method determined the function by using the formula:

$$\beta = H^{+}T \quad (6)$$

where H^+ is the Moore-Penrose generalized inverse of matrix H.

In addition, the ELM input contained the training dataset and number of hidden neurons L. The output included pre-processing the training data set (normalization), partitioning the available data set into the training and validation data sets, and computing the hidden layer output values of the ELM model. For the validation dataset, each validation spectrum possessed its re-ordered calibration dataset, which

was obtained using the ED method. Thus, the aim of this step was to obtain the correct number of calibration datasets and the ELM hidden layer.

Partial least squares model

This was a bilinear model where a matrix *X*, containing the variables (spectra wavenumber), and matrix *Y*, a function of the variables of *X* (nitrate contents), were used for the prediction of the smallest number of latent variables. In this study, the optimal number of latent variables for each database model was determined based on the minimal root-mean-square-error (*RMSE*) of cross-validation by leave-one-out cross-calibration [36].

Model performance evaluation

The evaluation indices of predictive capability for the ELM and PLS models were coefficients of determination (R^2), the ratio of performance to deviation (*RPD*), *RMSE*, Willmott's index (*WI*), and the Legates and McCabe index (E_{LM}) [35, 37].

$$R^2 = \frac{\sum_{i=1}^{n}(y'_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y'_i - y_i\right)^2} \quad (8)$$

$$RPD = \frac{SD}{RMSE} \quad (9)$$

$$WI = 1 - \left[\frac{\sum_{i=1}^{n}(y'_i - y_i)^2}{\sum_{i=1}^{n}(\left|y' - \bar{y}\right| + \left|y_i - \bar{y}\right|)^2}\right], 0 \le WI \le 1 \quad (10)$$

$$E_{LM} = 1 - \left[\frac{\sum_{i=1}^{n}\left|y_i - y'_i\right|}{\sum_{i=1}^{n}\left|y_i - \bar{y}\right|}\right], (\infty \le E_{LM} \le 1) \quad (11)$$

where y'_i and y_i are the predicted data and data measured by the chemical analysis method, respectively; n is the number of data sets; and SD is the standard deviation. $RMSE_C$ and $RMSE_P$ represented the root-mean-square-error in the calibration and validation dataset models, respectively. The RPD, which is used for normally distributed data, represented prediction accuracy, and should be higher than 1.8. An RPD value between 2 and 2.5 indicated a good quantitative prediction model, while a value higher than 3 suggested excellent performance of the model. A good performance model should have a WI value close to 1. Hence, insensitivity could be overcome because the ratio of model errors, rather than the square of the model error difference, could be analyzed [38]. E_{LM}, which was a more robust parameter than WI, predicted relatively higher values by squaring the differences [35, 37].

In addition, the ratio of $RMSE_P$ to $RMSE_C$ was used to judge the robustness of the model. A ratio lower than 1.2 was usually considered as a measure of robust performance [39, 40].

Results

Nitrate content analysis

A total of 408 vegetable samples were analyzed; the nitrate contents of Chinese cabbage, swamp cabbage, celery, and lettuce are listed in Table 1. The nitrate contents of Chinese cabbage (4063–14104 mg/kg), with an SD of 1664 mg/kg, was extremely high. The highest value was 14104 mg/kg, which was more than four-fold higher than the level indicating serious contamination. The nitrate content of swamp cabbage (2111-6607 mg/kg), with an SD of 1029 mg/kg, showed the narrowest range among the four species, but the average (4219 mg/kg) and median (4196 mg/kg) values were beyond the levels indicating serious contamination. The nitrate content of celery was

805– 8643 mg/kg, with an *SD* of 1214 mg/kg; the average (4164 mg/kg) and median (1214 mg/kg) values were slightly lower than those of the other species. As reported by Kalaycıoğlu and Erim (2019), the fiber compounds in the vegetables could reduce the possible harmful effects of the high nitrate contents [1]. The nitrate content in lettuce (10485 mg/kg), which was two to three fold higher than the maximum nitrate content specified by the European Commission standard, was the highest. Fig. 1 shows the different nitrate contents in the four leafy vegetable samples. The "+" symbol represents the outliers with extreme nitrate contents. The lower and upper lines of the boxplot represent the first and third quartiles (25^{th} and 75^{th} percentiles), respectively, and the median value (50^{th} percentile) was marked by the central line. Two horizontal lines were drawn out from the first and third quartiles to the smallest and largest non-outliners, respectively [35]. Additionally, all the nitrate contents followed the Gaussian distribution.

Spectral characterization

The characteristic absorption of nitrate, in the range from 1500 to 1200 cm^{-1}, in the Chinese cabbage, swamp cabbage, celery, and lettuce samples is shown in Fig. 2(a), and the main peaks of nitrate were found at around 1401 cm^{-1} and 1350 cm^{-1}. The absorption bands at 1401 cm^{-1} and 1350 cm^{-1} were associated with the N–O asymmetric stretching mode (ν_3). They were generated by the splitting of the *v* generation mode into two bands labeled $\nu_{3,\,high}$ and $\nu_{3,\,low}$. Hudson et al. [41] reported that the peak at around 1400 cm^{-1} was not observed in a pure nitrate environment, but the spectra were similar to those of aqueous $Ca(NO_3)_2$. The peak at 1245 cm^{-1} was associated with nitrite (NO_2^-) [42].

Moreover, the second-order derivative spectra were then calculated and plotted,

306 as shown in Fig. 2(b). The peak at approximately 1460 cm^{-1} was associated with N=O
307 vibration, the peaks at 1375 cm^{-1} and 1363 cm^{-1} were attributed to N=O and N–O,
308 respectively, and the peak at 1300 cm^{-1} was associated with N–O vibration [30].

309 The first three principal components (PCs), PC1 (83.58 %), PC2 (9.42 %), and
310 PC3 (4.05 %), containing 97.05% of the total spectral information, were investigated
311 among the four vegetables, as shown in Fig. 3. These PCs for the four species were
312 clustered, indicating that the main information was similar due to the characteristic
313 absorption peaks of nitrate in the range from 1500 to 1200 cm^{-1}. On the other hand,
314 the clustered PCs also indicated that the nitrate spectra could not be separated by the
315 general discrimination method.

316

317 **Optimization of parameters for calibration**

318 Several parameters, including the numbers of calibration data sets, hidden neurons in
319 the ELM model, and latent variables, are vital indices for the process of building a
320 good model, and they should be determined and optimized before modeling. Here, the
321 parameters of ELM, ED-modified ELM (ED-ELM), and PLS methods were
322 investigated separately. The predictive capability of the models was determined by a
323 low *RMSE*, large R^2 (close to 1), and a value of *RPD* that was more than 1.8. For the
324 ELM and PLS methods, there were 900 calibration samples and 324 validation
325 samples. Optimal numbers of hidden neurons and latent variables were optimized, and
326 are listed in Table 2. For the ED-ELM method, each of the 324 validation dataset
327 samples had 900 re-ranked calibration datasets, in which the similarity degree
328 sequence was arranged in a descending order. According to Ma et al. [40], a good
329 calibration dataset should contain large variances and less interference. Thus, the
330 optimal numbers of calibration data sets and hidden neurons were determined based

on the model performance parameters. Table 2 summarizes the performance indices, including the number of calibration data sets (900 samples). The optimal number of hidden layers was forty for the ELM model, while the optimal number of latent variables was seven for the PLS model. The values of $RMSE_C$ were ≈ 1089.91 mg/kg in the ELM model and ≈ 1087.68 mg/kg in the PLS model, indicating that the performance of the two models was similar. For the ED-ELM model, the optimal number of calibration datasets was determined according to the model performance parameters R^2, *RPD*, and $RMSE_P$, as shown in Fig. 4. It was clear that with the increase in the calibration dataset number from 100 to 900, in intervals of 100, the values of R^2 decreased from 0.89 to 0.71, and the values of *RPD* decreased from 3.05 to 1.76. Meanwhile, $RMSE_P$ increased from 636.45 mg/kg to 1035.25 mg/kg. Thus, the optimal number of calibration datasets was in intervals of 100 in the ED-ELM model. This was probably because too large a dataset sample might introduce interference, which would reduce the performance of the model [40].

In any model, the number of neurons in a hidden layer is important to determine the ideal network architecture [38]. A small architecture could short sufficient degrees of freedom to correctly learn the predictor data, while an elaborately large architecture may not converge in a reasonable model execution time, or it may over-fit rather than generalized the data [35]. Thus, the number of hidden neurons was optimized to the range from 10 to 90, and the ratio of $RMSE_P$ to $RMSE_C$, *RPD*, R^2, *WI*, and E_{LM} were calculated as functions of the number of hidden neurons. It can be noted in Fig. 5(a) that the ratio of $RMSE_P$ to $RMSE_C$ increased with an increase in the number of hidden neurons. In contrast, the values of *RPD* and R^2 decreased. Considering that the ratio of $RMSE_P$ to $RMSE_C$ should be lower than 1.2, *RPD* should be higher than 1.8, and R^2 should be close to 1, the optimal number of hidden neurons was found to be 30

356 ($RMSE_P$ / $RMSE_C \approx 1.13$; $RPD \approx 2.2$, and $R^2 \approx 0.93$). Meanwhile, the WI and E_{LM}
357 values in Fig. 5(b) are presented in a parabolic shape, with an increase in the number
358 of hidden neurons, and the values of $WI \approx 0.85$ and $E_{LM} \approx 0.64$ were the highest when
359 the number of hidden neurons was 30. Thus, using the ELM model, the number of
360 calibration datasets was set as 100, and the optimal number of hidden neurons was
361 determined to be 30.
362
363 **Nitrate prediction model**
364 Based on the optimization parameters, the validation data set was predicted using the
365 ELM, PLS, and ED-ELM models and the scatter plots are displayed in Fig. 6. The
366 performance indices were $RMSE_P \approx 995.77$ mg/kg, $R^2 \approx 0.70$, and $RPD \approx 1.76$ in the
367 ELM model and $RMSE_P \approx 1172.01$ mg/kg, $R^2 \approx 0.65$, and $RPD \approx 1.66$ in the PLS
368 model. The RPD values predicted by both the ELM and PLS models were lower than
369 1.8, indicating that the models were not reliable. For the ED-ELM model, the $RMSE_P$
370 was ≈ 799.67 mg/kg, the R^2 was ≈ 0.93, and the RPD was ≈ 2.22, indicating that this
371 model showed the best performance.
372 The Taylor diagram graphically depicted how closely the position of a prediction
373 pattern matched the measured values, and it was used for evaluating the performance
374 of multiple models [43, 44]. The position between the predicted and measured
375 patterns was quantified in terms of the correlation of the (R^2), centered $RMSE_P$ with
376 the SD to evaluate the model that yielded values that were closest to the predicted
377 values. The closer the predicted pattern to the measured pattern, the better the
378 performance of the model is. The Taylor diagrams of the PLS, ELM, and ED-ELM
379 models are shown in Fig. 7, showing the models that yielded data closest to measured
380 data in the validation dataset. The predicted patterns that agree well with the measured

values are placed at the nearest position marked ‘measured’. The organ color contours indicate the centered $RMSE_P$ values. The red square at the bottom of the diagram represents the position of the measured nitrate content. It represented that: $R^2 = 1$, $SD = 1$, and centered $RMSE_P = 0$, when compared to itself. The blue diamond represents the position of the values ($R^2 \approx 0.65$, $SD \approx 1.11$, and centered $RMSE_P \approx 0.67$) predicted by the PLS model; the yellow rounded shape shows the position of the values ($R^2 \approx 0.69$, $SD \approx 0.98$, and centered $RMSE_P \approx 0.57$) predicted by the ELM model; and the green triangle represented the position of the values ($R^2 \approx 0.93$, $SD \approx 0.93$, centered $RMSE_P \approx 0.47$) predicted by the ED-ELM model. Thus, the PLS model-derived values were the farthest from the measured values, while the ED-ELM model-derived values were the nearest to the measured values.

Nitrate status monitoring

According to the WHO-specified tolerance levels of nitrate content in fresh vegetables, four scales were used: the low level (nitrate content ≤432 mg/kg), allowing consumption of raw vegetables; medium level (432 mg/kg <nitrate content ≤785 mg/kg), indicating that consumption of raw vegetables was not allowed; high level (785 mg/kg <nitrate content ≤1234 mg/kg), indicating that consumption of raw or salted vegetables was not allowed; and serious contamination level (1234 mg/kg <nitrate content ≤3100 mg/kg), indicating that consumption of the vegetable in any form was not allowed [45]. Considering the WHO-specified tolerance level of nitrate, nitrate contents in all four species of vegetables exceeded the maximum limit of consumption. The FTIR-ATR spectra could be used for detecting nitrate in various vegetables.

Discussion

To some extents, the prediction accuracy depends on the proper modelling algorithms and when optimal parameters are determined. Though the FTIR-ATR displayed that the nitrate spectral intensity positively coincided with the nitrate content, the PLS model results were not satisfied, which meant the FTIR-ATR spectra and the nitrate contents was non-linear in the range from 1500 to 1200 cm^{-1}. It was probably because other compounds in vegetable included in the range were seemed as interference; another reason might be the nitrate absorption peak split to two peaks at 1401 cm^{-1} and 1350 cm^{-1} at with high concentration were not following linear relationship with nitrate contents [41]. To deal with non-linear modeling, intelligent algorithms have been developed. ELM is a developed algorithm for both classification and regression [41]. The advantages of ELM are extremely fast speed, less human intervenes and great computational scalability [46-47]. To improve the prediction accuracy, the calibration dataset was modified before being calibrated by the Euclidean distance method based on the spectra feature of target sample, which belongs to self-adaptive models, to obtain a similar-sample dataset [33, 40]. Therefore, FTIR-ATR spectra couple with extreme learning machine model with proper parameters showed rapid, accuracy and *in-situ* measurement results, and it could be potentially used in plant growth management and food safety testing.

Conclusions

In this study, the nitrate content levels in four leafy vegetables, Chinese cabbage, swamp cabbage, celery, and lettuce, were investigated. Portable FTIR-ATR spectroscopic instruments, along with the ED-ELM model, were used to predict the nitrate contents. The unique feature of nitrate was its absorption in the mid-infrared region at 1500-1200 cm^{-1}, and the absorption bands at 1401 cm^{-1} and 1350 cm^{-1} were

associated with the ν_3, N–O asymmetric stretch, which was generated by the splitting of the *v* generate mode into two bands labeled $\nu_{3,\ high}$ and $\nu_{3,\ low}$. The peak at 1245 cm^{-1} was associated with nitrite (NO_2^-). The results indicated that the nitrate contents in the vegetables exceeded the corresponding WHO-specified maximum tolerance limits (Chinese cabbage: 4063–14104 mg/kg; swamp cabbage: 2111–6607 mg/kg; celery: 805–8643 mg/kg; lettuce: 2567–10485 mg/kg), and these vegetables could only be consumed after being completely boiled. Moreover, the ED-ELM model (with performance indices of $RMSE_P \approx 799.67$ mg/kg, $R^2 \approx 0.93$, and $RPD \approx 2.22$) showed the best performance, compared to that of the ELM ($RMSE_P \approx 995.77$ mg/kg, $R^2 \approx 0.70$, $RPD \approx 1.76$) and PLS ($RMSE_P \approx 1172.01$ mg/kg, $R^2 \approx 0.65$, $RPD \approx 1.66$) models. The results indicated that FTIR-ATR, along with the ED-ELM model method, was a rapid and accurate *in situ* method to estimate nitrate contents.

Abbreviations

ATR, attenuated total reflectance spectroscopy; ED, Euclidean distance; ED-ELM, Euclidean distance-modified intelligent algorithm extreme learning machine model; ELM, extreme learning machine; E_{LM}, Legates and McCabe index; FTIR, Fourier-transform infrared spectroscopy; FTIR-ATR, Fourier-transform infrared attenuated total reflectance spectroscopy; PCA, principal component analysis; PC1, the first principal component; PC2, the second principal component; PC3, the third principal component; PLS, partial least squares; R^2, coefficients of determination; RMSE, root-mean-square-error; $RMSE_C$, the root-mean-square-error of the calibration dataset; $RMSE_P$**,** root-mean-square-error of the validation dataset; RPD, the ratio of performance to deviation; WI, Willmott's index;

Author Information

Corresponding Author

* Phone: +86-25-86881565. Fax: +86-25-86881000. Email:chwdu@issas.ac.cn.

ORCID

Changwen Du: 0000-0002-9064-3581

Notes

The authors declare no competing financial interest.

Author's contributions

CWD designed and directed the experiment, and made revisions of the manuscript. FM conducted the experiment, processed the data analysis and composed the manuscript. SLZ collected the vegetables, prepared the samples and recorded the spectra data. YXD collected the vegetables, prepared the samples and measured nitrate contents. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Competing interests

All the authors declare that they have no competing interests.

Availability of data and materials

The datasets supporting the conclusion of this article are included within the article (additional files).

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (XDA23030107) and the Innovational Project in Agriculture

484 from Jiangsu Province (CX(17)3010).

485

486 **References**

487 1. Kalaycıoğlu Z, Erim FB. Nitrate and nitrites in foods: worldwide regional
488 distribution in view of their risks and benefits. J Agric Food Chem. 2019; 67:
489 720–2.
490 2. Sušin J, Kmecl V, Gregorčič A. A survey of nitrate and nitrite content of fruit and
491 vegetables grown in Slovenia during 1996–2002. Food Addit Contam. 2006; 23:
492 385–90.
493 3. Lin YJ, Cheng CJ, Chen JW, Lin Z. Incorporating exogenous and endogenous
494 exposures into dietary risk assessment of nitrates and nitrites in vegetables: A
495 probabilistic integrated toxicokinetic modeling approach. J Agric Food Chem.
496 2020; 68: 1079–90.
497 4. Prasad S, Chetty AA. Nitrate-N determination in leafy vegetables: Study of the
498 effects of cooking and freezing. Food Chem. 2008; 106: 772–80.
499 5. Öztekin N, Nutku MS, Erim FB. Simultaneous determination of nitrite and nitrate
500 in meat products and vegetables by capillary electrophoresis. Food Chem. 2002;
501 76: 103–6.
502 6. Singh N, Bhatla SC. Hemoglobin as a probe for estimation of nitric oxide
503 emission from plant tissues. Plant Methods. 2019; 15: 39.
504 7. Bahadoran Z, Mirmiran P, Jeddi S, Azizi F, Ghasemi A, Hadaegh F. Nitrate and
505 nitrite content of vegetables, fruits, grains, legumes, dairy products, meats and
506 processed meats. J Food Compos Anal. 2016; 51: 93–105.
507 8. Rocha BS, Laranjinha J. Nitrate from diet might fuel gut microbiota metabolism:
508 Minding the gap between redox signaling and inter-kingdom communication. Free

509 Radical Bio Med. 2020; 149: 37–43.
510 9. Gorenjak HA, Cencic A. Nitrate in vegetables and their impact on human health.
511 A review. Acta Alimentaria. 2013; 42: 158–72.
512 10. Pagliano E, Mester Z. Determination of elevated levels of nitrate in vegetable
513 powders by high-precision isotope dilution GC–MS. Food Chem. 2019; 286: 710–
514 4.
515 11. Chen BM, Wang ZH, Li SX, Wang GX, Song HX, Wang XN. Effects of nitrate
516 supply on plant growth, nitrate accumulation, metabolic nitrate concentration and
517 nitrate reductase activity in three leafy vegetables. Plant Sci. 2004; 167: 635–43.
518 12. Colla G, Kim HJ, Kyriacou MC, Rouphael Y. Nitrate in fruits and vegetables. Sci
519 Hortic. 2018; 237: 221–38.
520 13. Nuñez de González MT, Osburn WN, Hardin MD, Longnecker M, Garg HK,
521 Bryan NS, Keeton JT. A survey of nitrate and nitrite concentrations in
522 conventional and organic-labeled raw vegetables at retail. J Food Sci. 2015; 80:
523 C942–49.
524 14. Stachniuk A, Szmagara A, Stefaniak EA. Spectrophotometric assessment of the
525 differences between total nitrate/nitrite contents in peel and flesh of cucumbers.
526 Food Anal Method. 2018: 11: 2969–77.
527 15. Blekkenhorst LC, Prince RL, Ward NC, Croft KD, Lewis JR, Devine A, Shinde S,
528 Woodman RJ, Hodgson JM, Bondonno CP. Development of a reference database
529 for assessing dietary nitrate in vegetables. Mol Nutr Food Res. 2017; 61:1600982.
530 16. Ngo VD, Jang BE, Park SU, Kim SJ, Kim YJ, Chung SO. Estimation of
531 functional components of Chinese cabbage leaves grown in a plant factory using
532 diffuse reflectance spectroscopy. J Agric Food Chem. 2019; 99: 711–8.
533 17. Group IW. IARC Working group on the evaluation of carcinogenic risks to

534 humans. International agency for research on cancer. Ingested nitrate and nitrite,
535 and cyanobacterial peptide toxins. IARC: Lyon, France. International agency for
536 research on cancer 2010.
537 18. Commission E. Commission regulation setting maximum levels for certain
538 contaminants in foodstuffs. Office Journal European Union 2006; 364: 5–24.
539 19. Shariati-Rad M, Irandoust M, Mohammadi S. Determination of nitrite in food
540 samples by kinetic spectrophotometric data and multivariate curve
541 resolution-alternating least squares (MCR-ALS). Food Anal Method. 2017; 10:
542 694–703.
543 20. Tamme T, Reinik M, Roasto M, Juhkam K, Tenno T, Kiis A. Nitrates and nitrites
544 in vegetables and vegetable-based products and their intakes by the Estonian
545 population. Food Addit Contam. 2006; 23: 355–61.
546 21. Wang Q, Huang H, Ning B, Li M, He L. A highly sensitive and selective
547 spectrofluorimetric method for the determination of nitrite in food products. Food
548 Anal Method. 2016; 9: 1293–300.
549 22. Campanella B, Onor M, Pagliano E. Rapid determination of nitrate in vegetables
550 by gas chromatography mass spectrometry. Anal Chim Acta. 2017; 980: 33–40.
551 23. Lin, SL, Hsu JW, Fuh MR. Simultaneous determination of nitrate and nitrite in
552 vegetables by poly (vinylimidazole-co-ethylene dimethacrylate) monolithic
553 capillary liquid chromatography with UV detection. Talanta. 2019; 205: 120082.
554 24. Li CY, Du CW, Zeng Y, Ma F, Shen YZ, Xing Z, Zhou JM. Two-dimensional
555 visualization of nitrogen distribution in leaves of Chinese cabbage (Brassica rapa
556 subsp. chinensis) by the Fourier transform infrared photoacoustic spectroscopy
557 Technique. J Agric Food Chem. 2016; 64: 7696–701.
558 25. Wu K, Du CW, Ma F, Shen YZ, Liang D, Zhou JM. Rapid diagnosis of nitrogen

559 status in rice based on Fourier transform infrared photoacoustic spectroscopy
560 (FTIR-PAS). Plant Methods 2019; 15: 94.
561 26. Carbas B, Machado N, Oppolzer D, Ferreira L, Brites C, Rosa EAS, Barros AIR
562 NA. Comparison of near-infrared (NIR) and mid-infrared (MIR) spectroscopy for
563 the determination of nutritional and antinutritional parameters in common beans.
564 Food Chem. 2020; 306:125509.
565 27. Karunakaran C, Vijayan P, Stobbs J, Bamrah RK, Arganosa G, Warkentin TD.
566 High throughput nutritional profiling of pea seeds using Fourier transform
567 mid-infrared spectroscopy. Food Chem. 2020; 309: 125585.
568 28. Li G, Wang R, Quampah AJ, Rong Z, Shi C, Wu J. Calibration and prediction of
569 amino acids in stevia leaf powder using near infrared reflectance spectroscopy. J
570 Agric Food Chem. 2011; 59: 13065–71.
571 29. Al-Qadiri HM, Al-Holy MA, Lin M, Alami NI, Cavinato AG, Rasco BA. Rapid
572 detection and identification of pseudomonas aeruginosa and escherichia coli as
573 pure and mixed cultures in bottled drinking water using Fourier transform infrared
574 spectroscopy and multivariate analysis. J Agric Food Chem. 2006; 54: 5749–54.
575 30. Shao YQ, Du CW, Shen YZ, Ma F, Zhou JM. Evaluation of net nitrification rates
576 in paddy soil using mid-infrared attenuated total reflectance spectroscopy. Anal
577 Methods. 2017; 9: 748–55.
578 31. Knoll L, Breuer L, Bach M. Large scale prediction of groundwater nitrate
579 concentrations from spatial data using machine learning. Sci Total Environ. 2019;
580 668: 1317–27.
581 32. Shao YQ, Du CW, Shen YZ, Ma F, Zhou JM. Rapid determination of N isotope
582 labeled nitrate using Fourier transform infrared attenuated total reflection
583 spectroscopy. Chinese Anal Chem. 2014; 42: 747–52.

584 33. Ma F, Du CW, Zhou JM, Shen YQ. Optimized self-adaptive model for assessment
585 of soil organic matter using Fourier transform mid-infrared photoacoustic
586 spectroscopy. Chemometr Intell Lab. 2017; 171: 9–15.
587 34. De Clercq D, Wen Z, Fei F, Caicedo L, Yuan K, Shang R. Interpretable machine
588 learning for predicting biomethane production in industrial-scale anaerobic
589 co-digestion. Sci Total Environ. 2020; 712: 134574.
590 35. Sharma E, Deo RC, Prasad R, Parisi AV. A hybrid air quality early-warning
591 framework: An hourly forecasting model with online sequential extreme learning
592 machines and empirical mode decomposition algorithms. Sci Total Environ. 2020;
593 709: 135934.
594 36. Ma F, Du, CW, Zhou JM, Shen YZ. Investigation of soil properties using different
595 techniques of mid-infrared spectroscopy. Eur J Soil Sci. 2019; 70: 96–106.
596 37. Willmott CJ, Robeson SM, Matsuura K. A refined index of model performance.
597 Int J Climatol. 2012; 32: 2088–94.
598 38. Yaseen ZM, Ebtehaj I, Bonakdari H, Deo RC, Danandeh Mehr A, Mohtar WHMW,
599 Diop L, El-shafie A, Singh VP. Novel approach for streamflow forecasting using a
600 hybrid ANFIS-FFA model. J Hydrol. 2017; 554: 263–76.
601 39. Coûteaux MM, Sarmiento L, Hervé D, Acevedo D. Determination of
602 water-soluble and total extractable polyphenolics in biomass, necromass and
603 decomposing plant material using near-infrared reflectance spectroscopy (NIRS).
604 Soil Biol Biochem. 2005; 37: 795–9.
605 40. Ma F, Du CW, Zhou JM. A self-adaptive model for the prediction of soil organic
606 matter using mid-infrared photoacoustic spectroscopy. Soil Sci Soc Am J. 2016;
607 80: 238–46.
608 41. Hudson PK, Schwarz J, Baltrusaitis J, Gibson ER, Grassian VH. A spectroscopic

study of atmospherically relevant concentrated aqueous nitrate solutions. J Phys Chem A. 2007; 111: 544–8.

42. Shaviv A, Kenny A, Shmulevitch I, Singher L, Raichlin Y, Katzir A. Direct monitoring of soil and water nitrate by FTIR Based FEWS or membrane systems. Environ Sci Technol. 2003; 37: 2807–12.

43. Taylor KE. Summarizing multiple aspects of model performance in a single diagram. J Geophys Res-Atmos. 2001; 106: 7183–92.

44. Yearsley JR, Sun N, Baptiste M, Nijssen B. Assessing the impacts of hydrologic and land use alterations on water temperature in the Farmington River basin in Connecticut. Hydrol Earth Syst Sci. 2019; 23: 4491–508.

45. Zhou ZY, Wang MJ, Wang JS. Nitrate and nitrite contamination in vegetables in China. Food Rev Int. 2000; 16: 61–76.

46. Chen H, Tan C, Lin Z. Ensemble of extreme learning machines for multivariate calibration of near-infrared spectroscopy. Spectrochim Acta A. 2020; 229: 117982.

47. Peng J, Li L, Tang YY. Combination of activation functions in extreme learning machines for multivariate calibration. Chemometr Intell Lab. 2013; 120: 53–8.

626

627 **Table 1.** Nitrate contents in vegetables analyzed by the chemical method

Sample	Range (mg/kg)	Mean (mg/kg)	Median (mg/kg)	Standard Deviation
Chinese cabbage	4063–14104	7550	7543	1664
Swamp cabbage	2111–6607	4219	4196	1029
Celery	805–8643	4164	3991	1214
Lettuce	2567–10485	4322	4163	1035

628

629

630

631 **Table 2.** The performance parameters of the calibration model

Models	Calibration data sets	Hidden layers	Latent variables	$RMSE_C$ (mg/kg)
ELM	900	40	–	1089.91
PLS	900	–	7	1087.68
ED-ELM	100	100	–	322.52

632

633

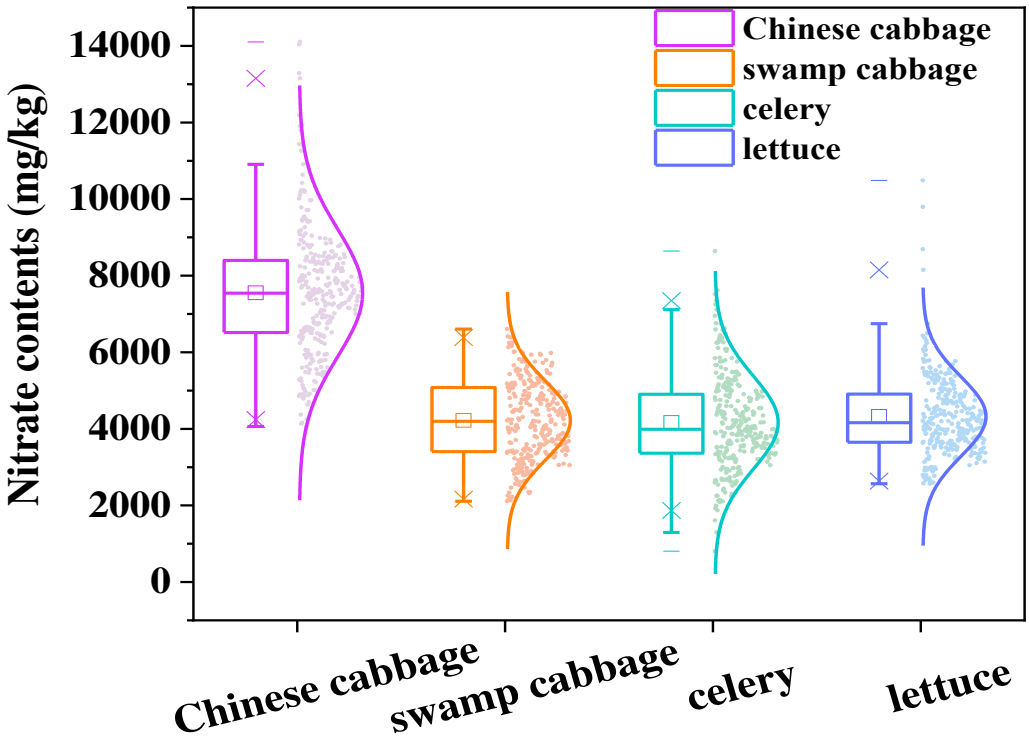


Figure 1. Boxplot of distributions of the nitrate contents of Chinese cabbage, swamp cabbage, celery, and lettuce (306 samples each), following Gaussian distributions.

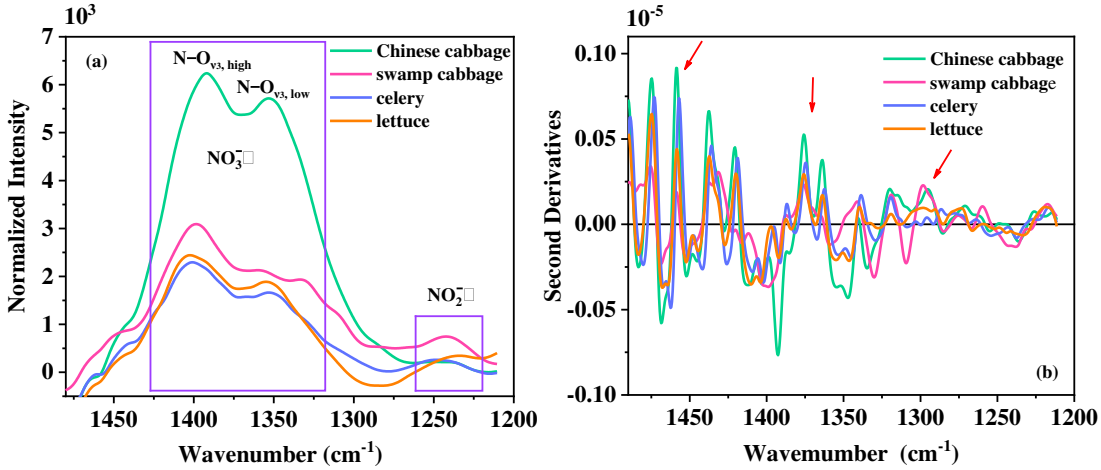


Figure 2. (a) Characteristic FTIR-ATR absorption spectra of nitrate, (b) the second derivative spectra of nitrate, in the range from 1500 to 1200 cm^{-1}, obtained from Chinese cabbage, swamp cabbage, celery, and lettuce.

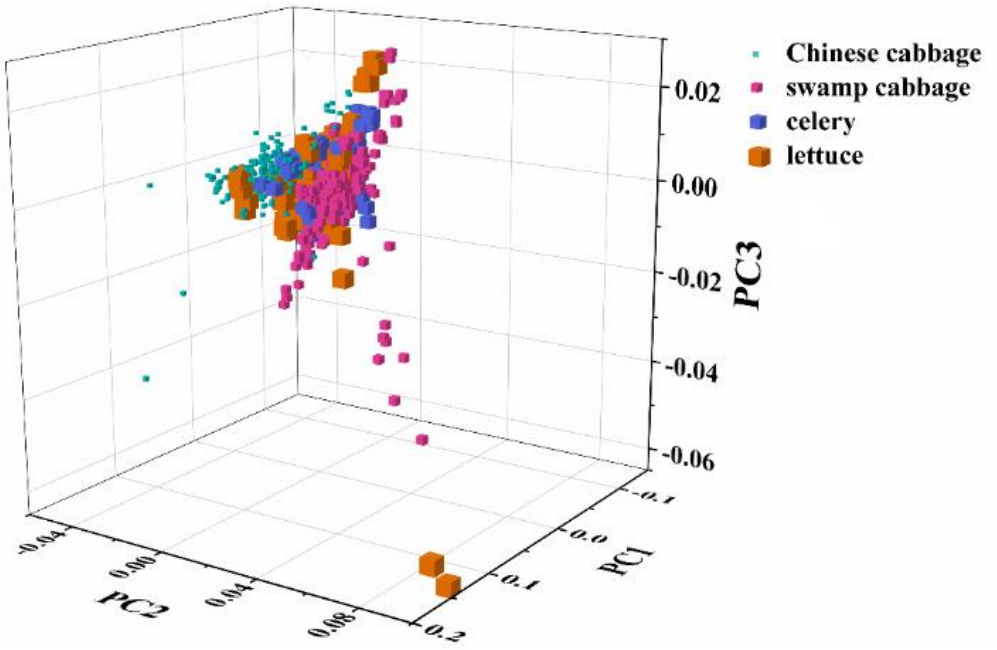


Figure 3. Principal component (PC1, PC2, and PC3) score plots of the vegetable samples; the four types of vegetables were Chinese cabbage, swamp cabbage, celery, and lettuce

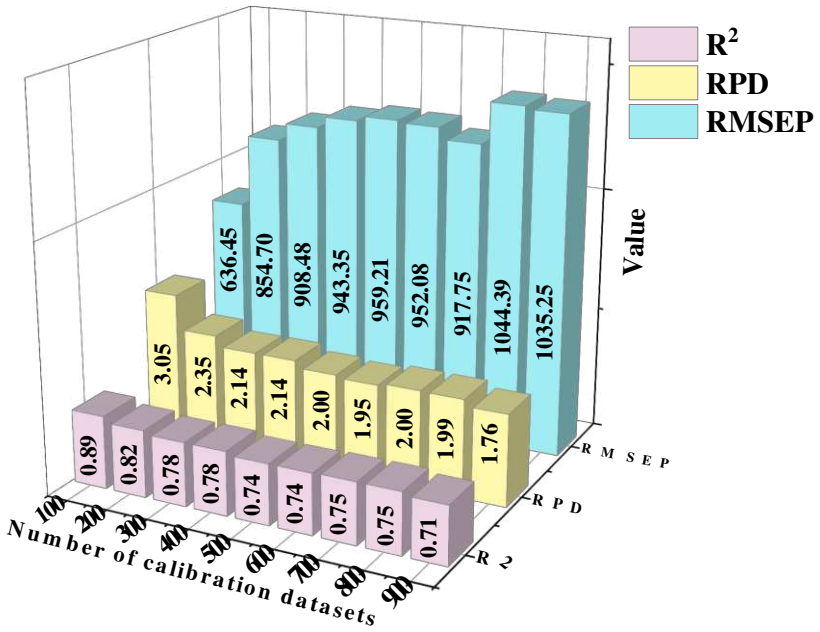


Figure 4. Statistical distribution of the validation parameters R^2, *RPD*, and $RMSE_P$ with various numbers of calibration datasets in the ED-ELM model for the prediction of nitrate content

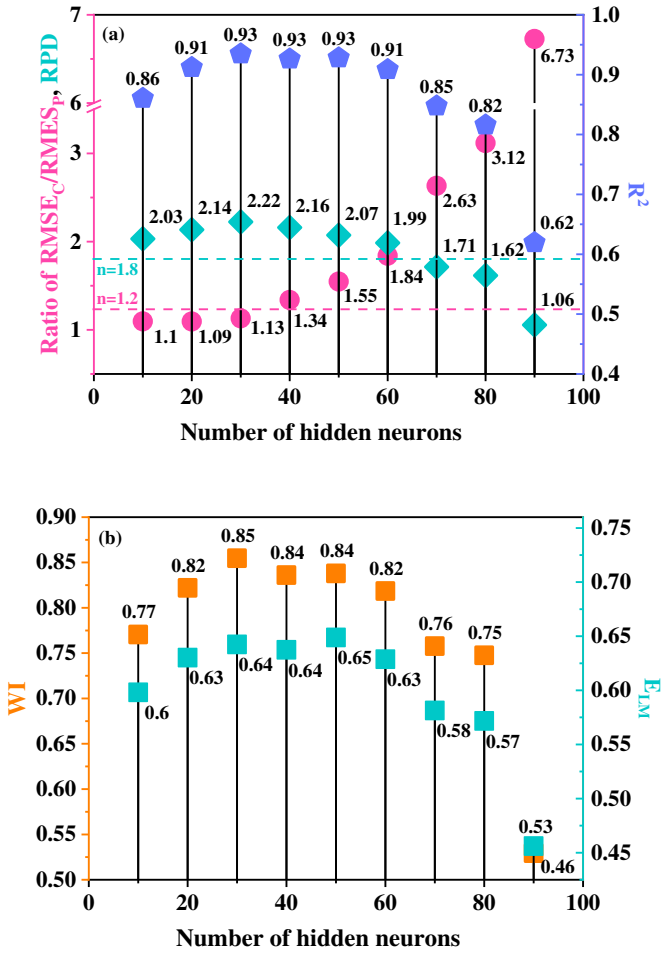
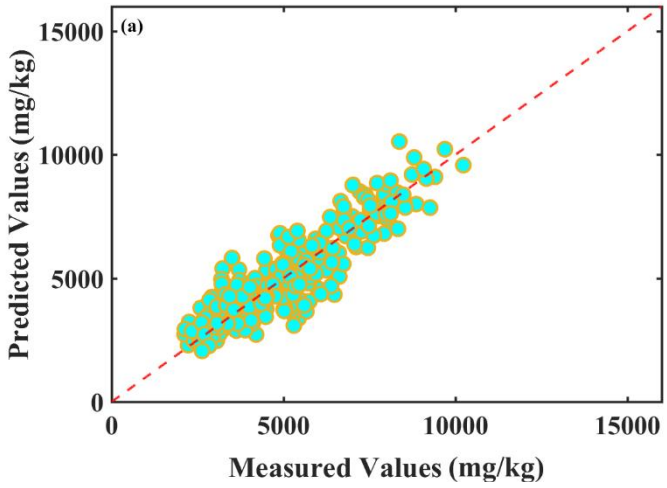
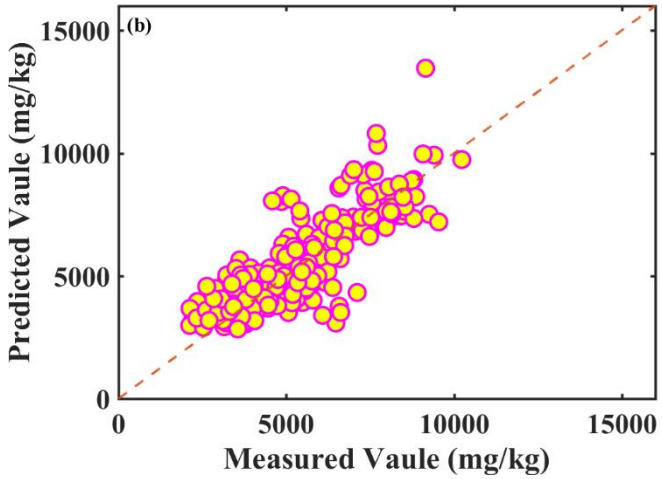


Figure 5. The ED-ELM model performance indices of (a) the ratio of $RMSE_P$ to $RMSE_C$, RPD, and R^2, (b) WI and E_{LM} values for various numbers of hidden neurons from 10 to 90





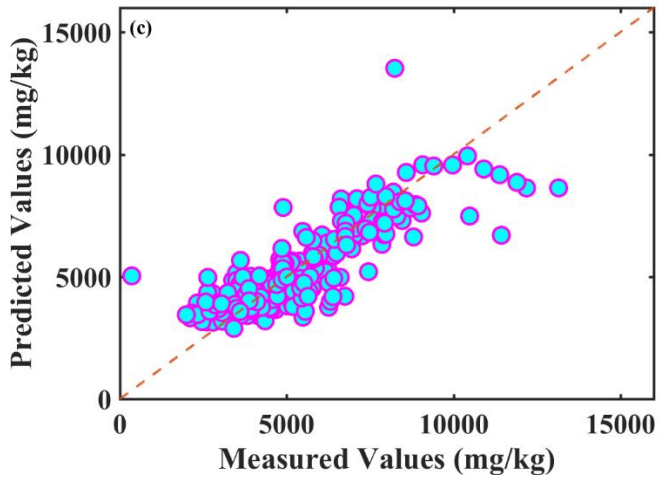


Figure 6. Scatterplots of measured nitrate values and prediction results obtained by (a) the ED-ELM model, (b) the ELM model and (c) the PLS model. The dotted line is the reference line (1:1), and the reference line corresponds to the exact prediction. The samples are distributed along the reference line.

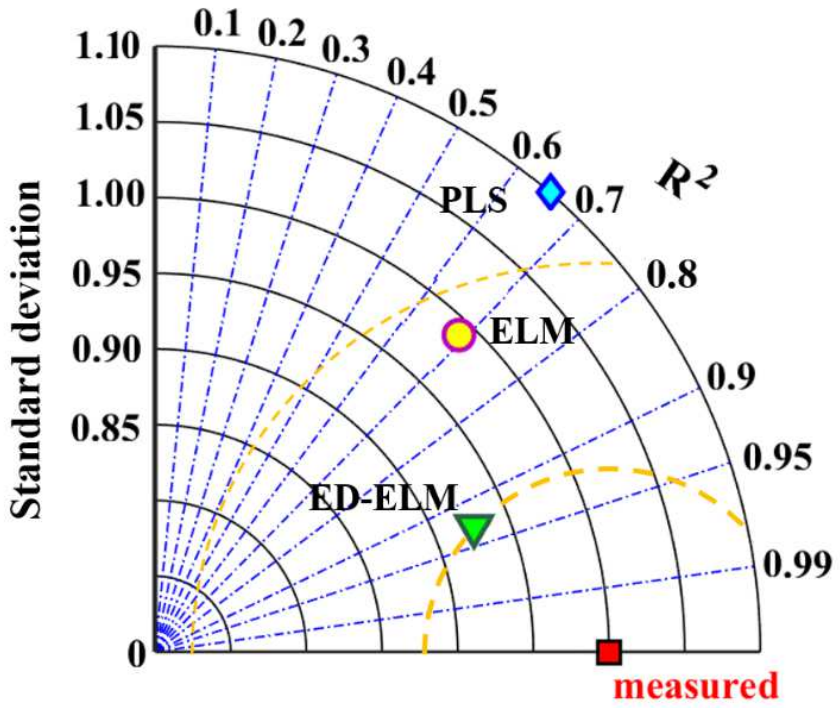
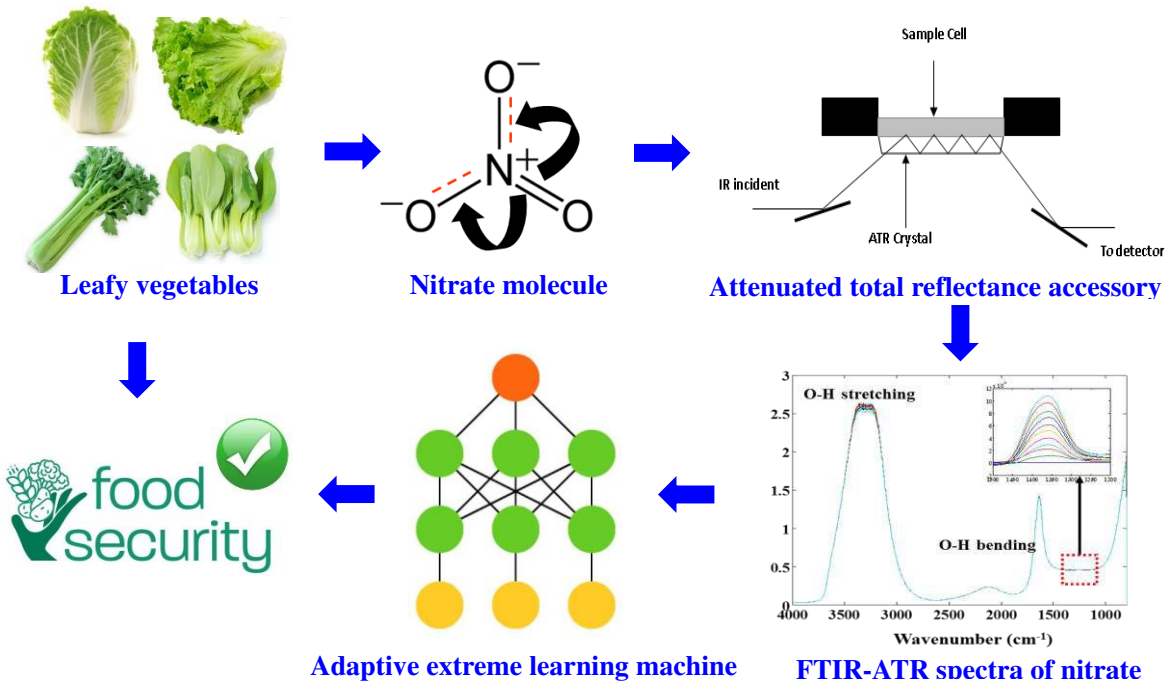


Figure 7. Taylor diagram illustrating the parameters and positions for the measured and predicted obtained by the PLS, ELM, and ED-ELM models.

TOC graphic


Sample Cell
IR incident
ATR Crystal
To detector
Leafy vegetables
Nitrate molecule
Attenuated total reflectance accessory
food security
Adaptive extreme learning machine
O-H stretching
O-H bending
Wavenumber (cm-1)
FTIR-ATR spectra of nitrate

Figures

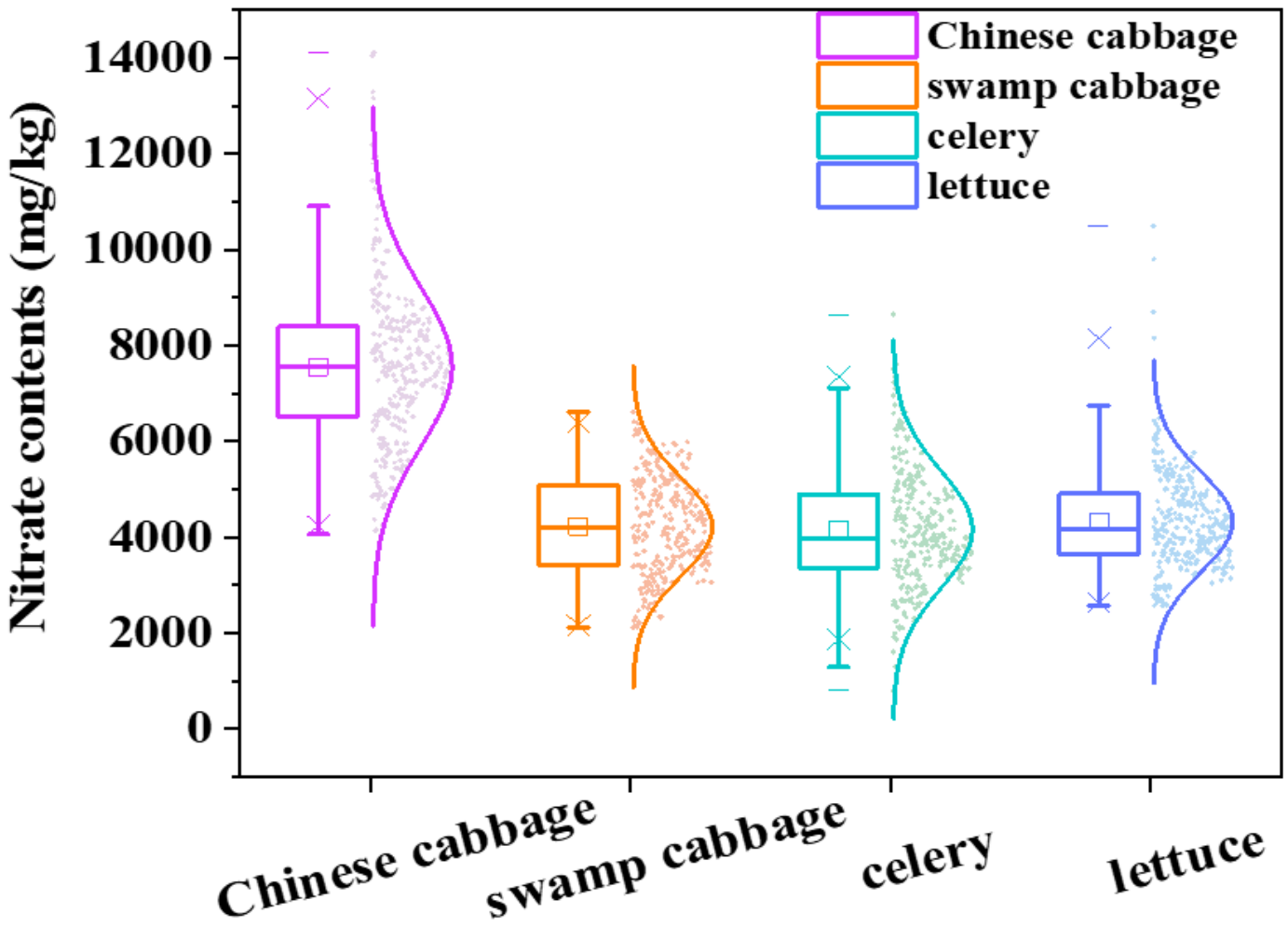


Figure 1

Boxplot of distributions of the nitrate contents of Chinese cabbage, swamp cabbage, celery, and lettuce (306 samples each), following Gaussian distributions.

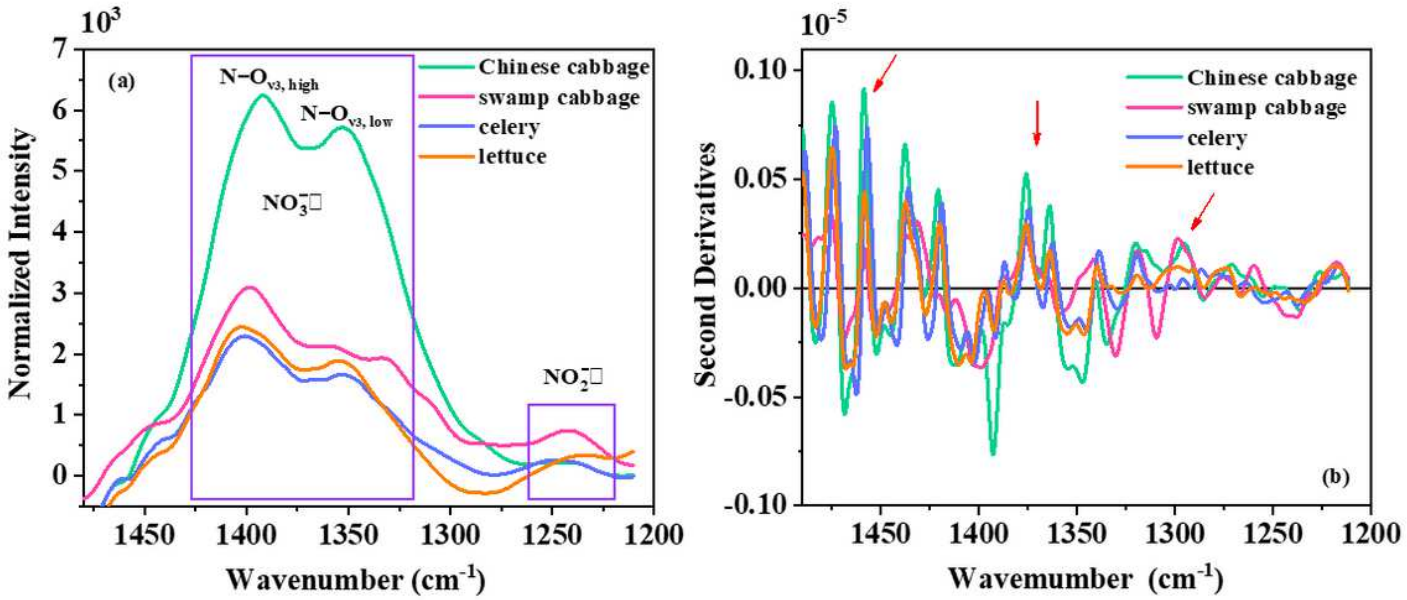


Figure 2

(a) Characteristic FTIR-ATR absorption spectra of nitrate, (b) the second derivative spectra of nitrate, in the range from 1500 to 1200 cm-1, obtained from Chinese cabbage, swamp cabbage, celery, and lettuce.

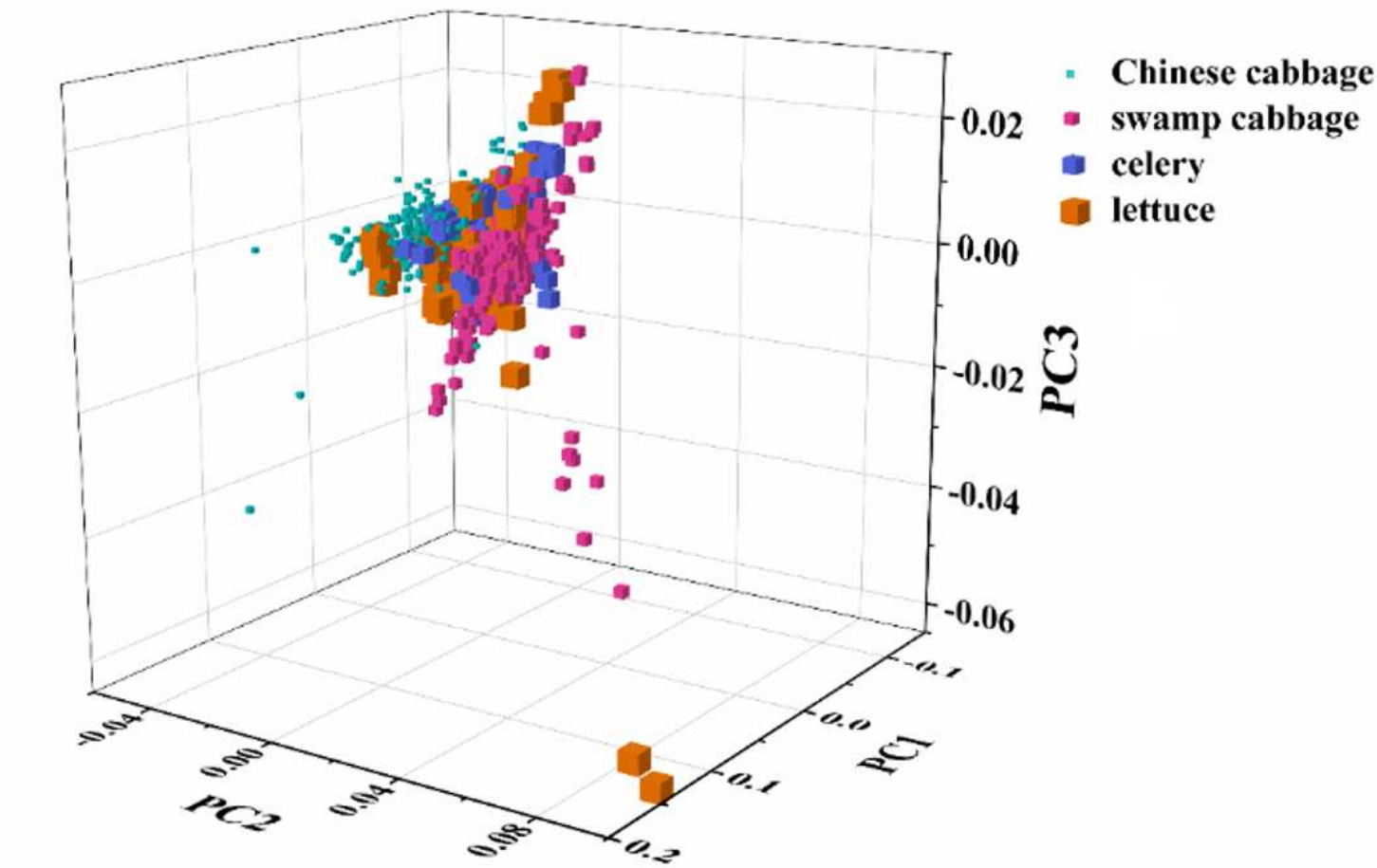


Figure 3

Principal component (PC1, PC2, and PC3) score plots of the vegetable samples; the four types of vegetables were Chinese cabbage, swamp cabbage, celery, and lettuce

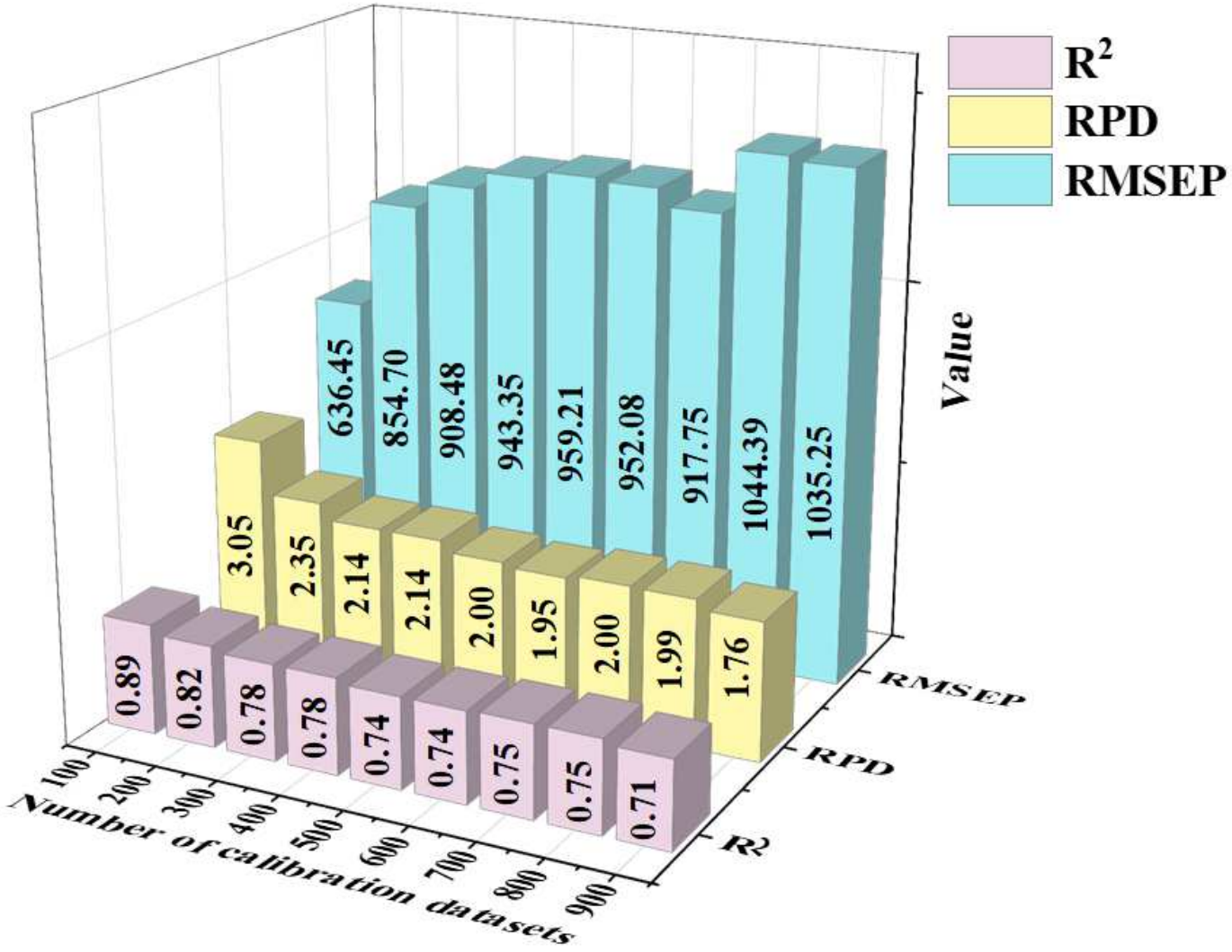


Figure 4

Statistical distribution of the validation parameters R2, RPD, and RMSEP with various numbers of calibration datasets in the ED-ELM model for the prediction of nitrate content

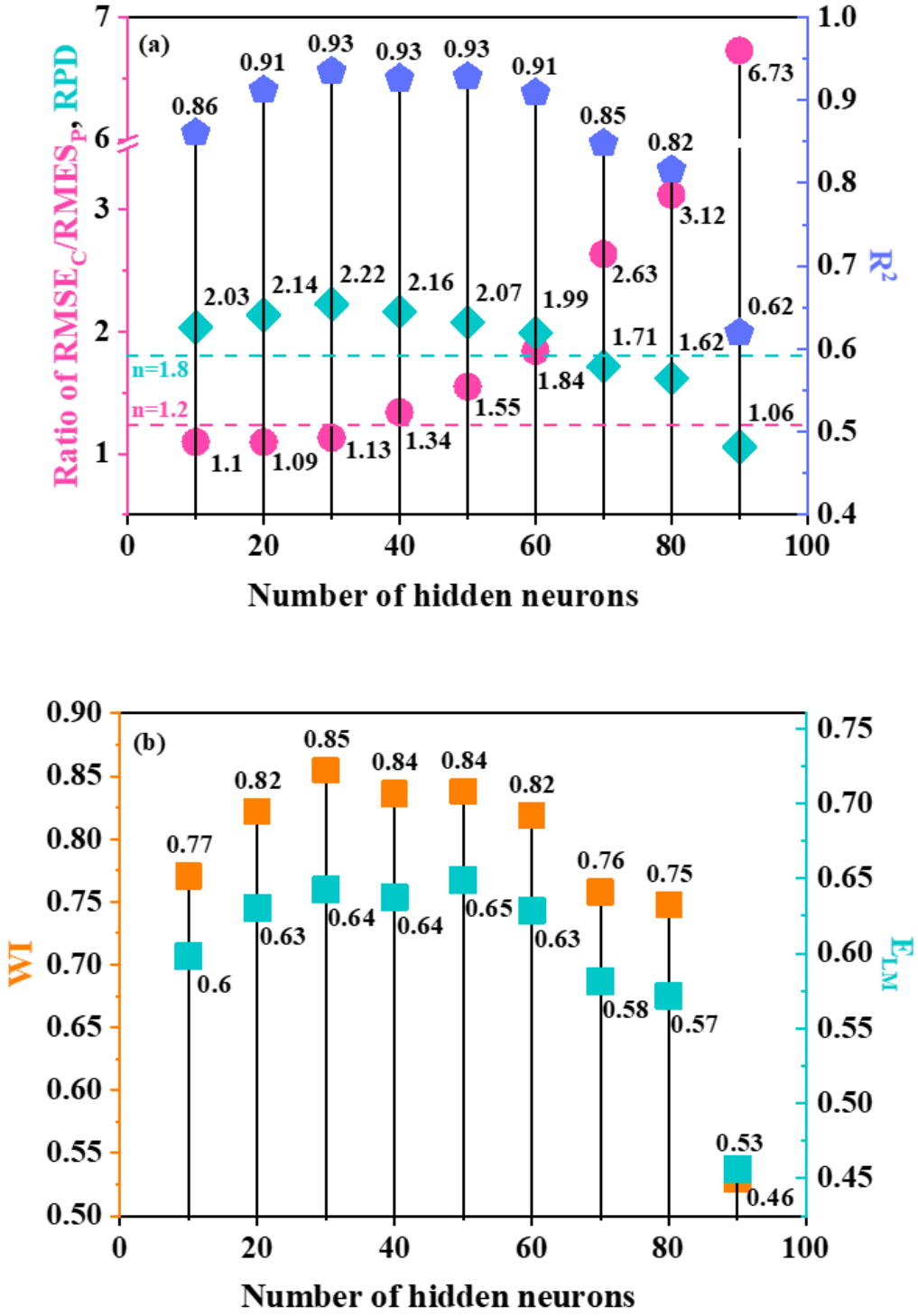


Figure 5

The ED-ELM model performance indices of (a) the ratio of RMSEP to RMSEC, RPD, and R2, (b) WI and ELM values for various numbers of hidden neurons from 10 to 90

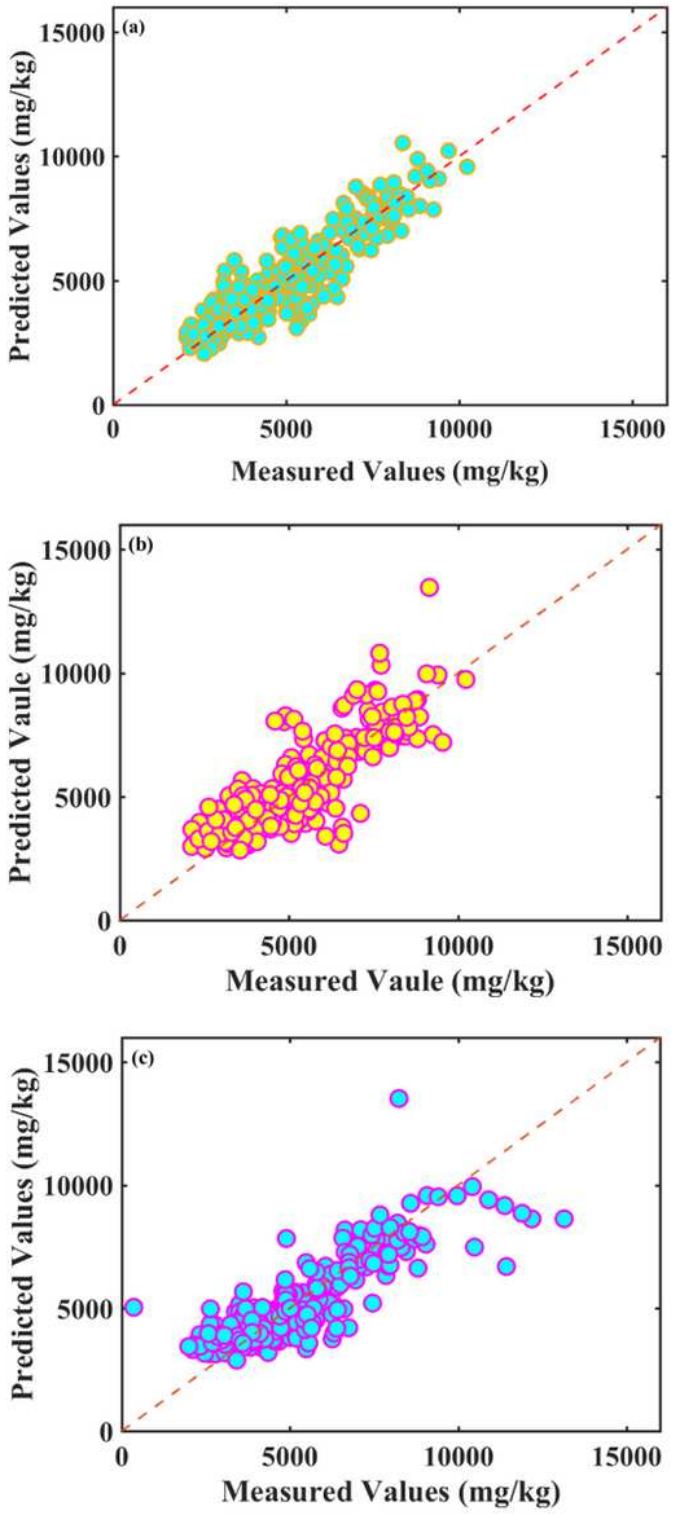


Figure 6

Scatterplots of measured nitrate values and prediction results obtained by (a) the ED-ELM model, (b) the ELM model and (c) the PLS model. The dotted line is the reference line (1:1), and the reference line corresponds to the exact prediction. The samples are distributed along the reference line.

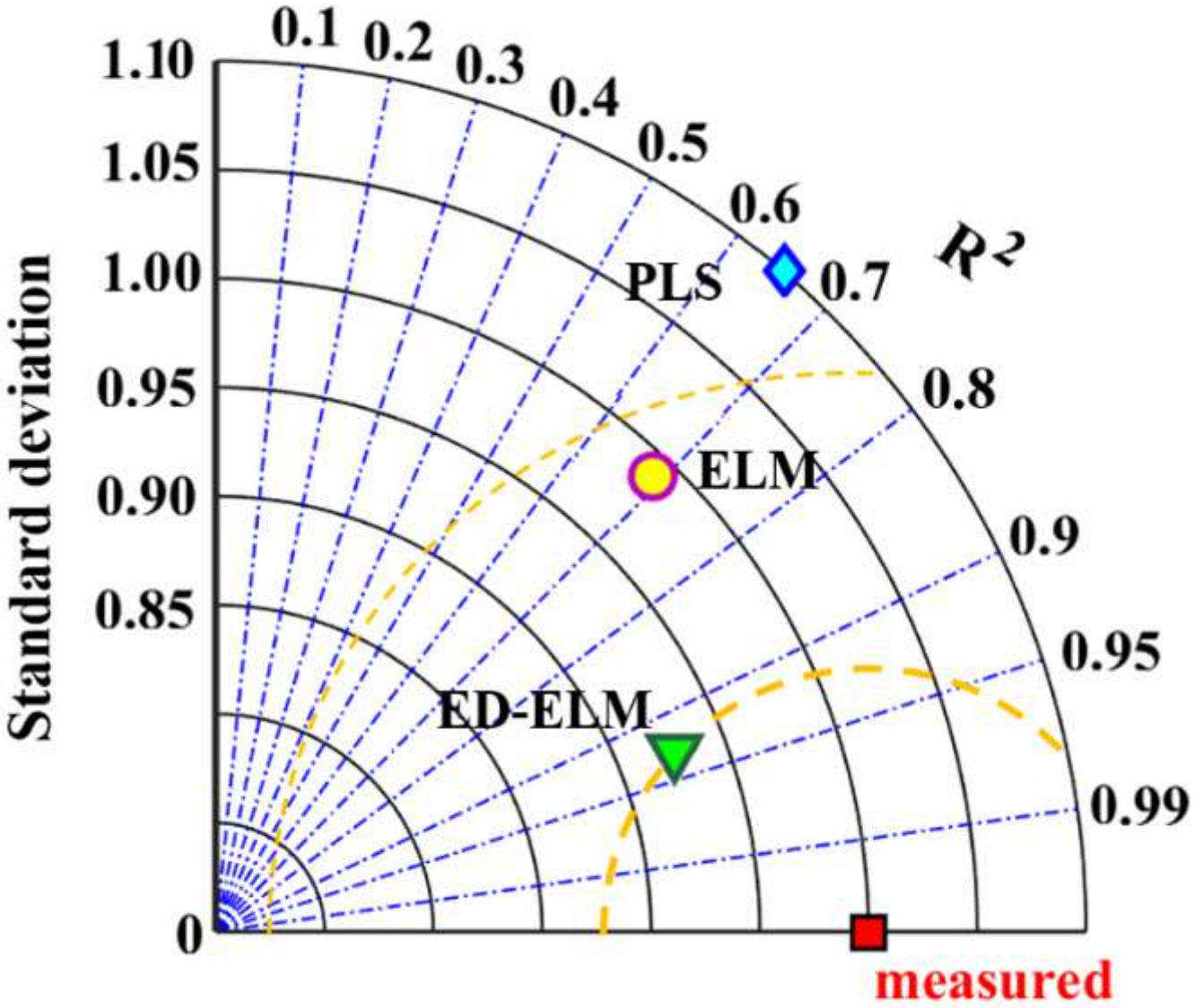


Figure 7

Taylor diagram illustrating the parameters and positions for the measured and predicted obtained by the PLS, ELM, and ED-ELM models.