Journal of Innovative Optical Health Sciences
Vol. 7, No. 4 (2014) 1350071 (9 pages)

DOI: 10.1142/S1793545813500715

Selection of stable equivalent wavebands for near-infrared spectroscopic analysis of total nitrogen in soil

Jiemei Chen*,†, Tao Pan†,‡, Guisong Liu*,†, Yun Han†
and Dingxing Chen†
*State Key Laboratory of Soil and Sustainable Agriculture
Institute of Soil Science, Chinese Academy of Sciences
Nanjing 210008, P. R. China

†Key Laboratory of Optoelectronic Information and
Sensing Technologies of Guangdong Higher Educational Institute
Jinan University, Guangzhou 510632, P. R. China
‡tpan@jnu.edu.cn

Received 31 July 2013
Accepted 16 October 2013
Published 10 December 2013

The selection of stable wavebands for the near-infrared (NIR) spectroscopic analysis of total nitrogen (TN) in soil was accomplished by using an improved moving window partial least squares (MWPLS) method. A new modeling approach was performed based on randomness, similarity and stability, which produced an objective, stable and practical model. Based on the MWPLS method, a search was in the overall scanning region from 400 to 2498 nm, and the optimal waveband was identified to be 1424 to 2282 nm. A model space that includes 41 wavebands that are equivalent to the optimal waveband was then proposed. The public range of the 41 equivalent optimal wavebands was 1590 to 1870 nm, which contained sufficient TN information. The wavebands of 1424 to 2282 nm, 1590 to 1870 nm, and the long-NIR region 1100 to 2498 nm all achieved satisfactory validation effects. However, the public waveband of 1590 to 1870 nm had only a minimum number of wavelengths, which significantly reduced the method complexity. Various equivalent wavebands serve as guidelines for designing spectroscopic instruments. These wavebands could address the restrictions of position and the number of wavelengths in instrument design.

Keywords: Soil; total nitrogen; near-infrared spectroscopy; improved moving window partial least squares; stability.

1. Introduction

Total nitrogen (TN) is the main component of soil and is a crucial monitoring indicator for an agricultural ecological environment. Standard methods for measuring TN in soil cannot be conducted easily. In addition, these methods consume chemical reagents and cause environmental pollution. Therefore, a rapid, chemical-free measurement method for soil TN is of great significance. Near-infrared (NIR) spectroscopy has been proven to be a powerful analytical tool for use in agriculture,[1] food,[2] environment,[3] medicine,[4] and so on.[5] The establishment of a rapid, chemical-free measurement for soil components based on NIR spectroscopy has been a significant research direction in recent years.[6–8]

Given that soil is a complex system with multiple components, the spectroscopic analysis of major soil components has to mitigate noise disturbance. In addition, appropriate wavelength selection and stability are two key objectives. Model stability was not discussed in detail by previous studies because of the need for numerous experiments. Partial least squares (PLS) was widely used for NIR analysis.[9,10] However, according to results from several experiments, waveband selection for the PLS model is necessary because the prediction effect cannot improve with insufficient signal-to-noise ratio in the waveband. TN refers to the sum of organic nitrogen and inorganic nitrogen in soil. Previous studies reported on the information wavebands in the NIR region of nitrogen-based groups.[9,10] However, the NIR spectra of soil indicate the absorbance information of all components. The absorption band of the above functional group cannot be simply taken as the waveband for analysis of TN in soil because of the interference of other components. Therefore, waveband selection has to be based on the model prediction effect by using an appropriate chemometrics method. In this research, the moving window partial least squares (MWPLS) method[11–13] was improved in terms of stability and equivalence. The appropriate waveband for the chemical-free measurement of TN in soil was selected.

An objective and rational evaluation method is essential to spectroscopy analysis. Several experimental results indicate that differences in the divisions of calibration and prediction sets could result in fluctuations in the prediction and in the model parameters, thereby producing unstable results. On the other hand, the random selection of samples for the validation set is reasonable. However, contingencies attributed to random divisions for the calibration and prediction sets may cause the modeling process to appear distorted. For example, a randomly generated calibration set comprises with low TN measured values, whereas the prediction set comprises samples with high TN measured values. Improving the prediction effect under such conditions is difficult and often produces incorrect models. To avoid the evaluation distortion of the model, the calibration and prediction sets need to be divided based on certain similarities in the modeling optimization process. In this paper, a new modeling approach was performed based on randomness, similarity and stability, which produced an objective, stable and practical model.

2. Materials and Methods

2.1. *Soil sampling*

Leafy vegetables are the main vegetable cultivation species of Pearl River Delta of China. Nitrogen testing of soil in the major vegetable production areas is very important for reasonable fertilization and effective yield. A total of 163 samples of the farmland soil (yellow brown earth) were collected from several leafy vegetables fields of the major vegetable production areas located in Pearl River Delta. Sampling points were arranged uniformly about 5 to 6 in each acre field. Soil samples were collected in representative topsoil (0–20 cm) of each sampling point. The samples were ground after drying, and then were sifted by using a 0.25-mm soil sifter. The TN content of each sample was measured by using the Kjeldahl method, a standard soil analysis method. The measured values were used for the calibration and validation of spectroscopic analysis. The TN in all samples ranged from 0.60 to 1.78 g kg^{-1}, and the mean values and the standard deviations were 1.178 and 0.220 g kg^{-1}, respectively.

2.2. *Experimental instruments and measurement methods*

The spectroscopy instrument used was an XDS Rapid Content™ Grating Spectrometer (FOSS, Danmark) equipped with a diffuse reflection accessory and a round sample cell. The scanning scope of the spectrum spanned 400 to 2498 nm with a 2-nm

wavelength interval, including the overall NIR region and a large part of the visible region. Wavebands of 400 to 1100 nm and 1100 to 2498 nm were adopted for Si and PbS detection, respectively. The same soil sample was repeated to collect spectra thrice, and the mean value of the three measurements was used for modeling. The spectra were measured at $25°C \pm 1°C$ and $46\% \pm 1\%$ RH.

2.3. *Sample set division and model optimization frame*

First, 60 samples were randomly selected from a total of 163 samples as the validation set. The remaining 103 samples were used as the modeling set. The modeling set was divided into similar calibration (60 samples) and prediction (43 samples) sets for a total of 50 times. The mean value and the standard deviation of the root mean square error and the correlation coefficients for all divisions were denoted by $\text{M-SEP}_{\text{Ave}}$, $\text{M-R}_{\text{P,Ave}}$, $\text{M-SEP}_{\text{Std}}$ and $\text{M-R}_{\text{P,Std}}$. These values served as the basis for the discussion on prediction accuracy and stability; $\text{M-SEP}^{+} = \text{M-SEP}_{\text{Ave}} + \text{M-SEP}_{\text{Std}}$ was a comprehensive indicator of prediction accuracy and stability. A smaller M-SEP^{+} indicates higher accuracy and stability. Model parameters (such as waveband and PLS factor) were selected according to the minimum $\text{M-SEP}_{\text{Ave}}$ or M-SEP^{+}. Finally, the selected model was re-validated against the validation set. The randomly selected validation samples, which were not subjected to the modeling optimization process, were regarded as the prediction set while the original modeling set was used as calibration set. The validation root mean square error and validation correlation coefficients of prediction were then calculated and denoted by V-SEP and V-R_{P}, respectively.

The two sets could be considered similar when the TN's mean value and the standard deviation of the calibration set were close to those of the prediction set. All modeling samples were randomly divided into calibration and prediction sets for a sufficient number of times. The mean value and standard deviation of the TN of the calibration, prediction and whole modeling sets were denoted by $\text{TN}_{\text{C,Ave}}$, $\text{TN}_{\text{C,Std}}$, $\text{TN}_{\text{P,Ave}}$, $\text{TN}_{\text{P,Std}}$, TN_{Ave} and TN_{Std}, respectively. The similarity degree was defined by the following:

$$a_0 = \max\left\{\frac{|\text{TN}_{\text{C,Ave}} - \text{TN}_{\text{P,Ave}}|}{\text{TN}_{\text{Ave}}}, \frac{|\text{TN}_{\text{C,Std}} - \text{TN}_{\text{P,Std}}|}{\text{TN}_{\text{Std}}}\right\} \times 100\%. \quad (1)$$

Improved similarity was achieved with smaller a_0. In this paper, 50 divisions that satisfy $a_0 < 10\%$ were retained for modeling.

2.4. *Optimization frame of the MWPLS method*

The parameters of the MWPLS method were as follows: (1) beginning wavelength and its serial number (B), (2) number of wavelengths (N) and (3) PLS factor (F).[14,15] The range of parameters B, N, F were denoted by $\mathcal{B}$, $\mathcal{N}$, $\mathcal{F}$, respectively, which can be set according to the actual situation; the parameter space $\mathcal{B} \otimes \mathcal{N} \otimes \mathcal{F}$ is shown in Fig. 1. The PLS models were established for all combinations (B, N, F). The corresponding $\text{M-SEP}_{\text{Ave}}$, $\text{M-R}_{\text{P,Ave}}$, $\text{M-SEP}_{\text{Std}}$, $\text{M-R}_{\text{P,Std}}$ and M-SEP^{+} were then calculated and matched to a cube of the parameter space.

Stability of PLS factor: PLS can comprehensively screen the spectroscopic data and can extract information variables. The PLS factor F is a major parameter that corresponds to the number of spectral integrated variables on behalf of the sample information. The selection of a reasonable F is both necessary and difficult.[16,17] In this paper, F was selected by considering the number of divisions for the calibration and the prediction sets. Thus, the

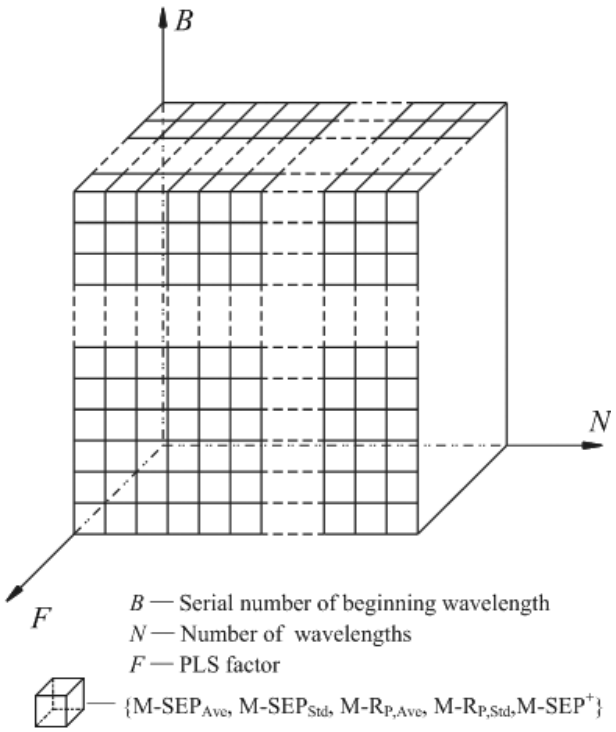


Fig. 1. Schematic of the parameter space $\mathcal{B} \otimes \mathcal{N} \otimes \mathcal{F}$.

optimized PLS factor exhibited stability and practicality. Any waveband corresponded to a unique combination of parameters $(B, N) = (B_0, N_0)$; the optimal PLS model of the waveband was selected according to the following:

$$\text{M-SEP}_{\text{Ave}}(B_0, N_0) = \min_{F \in \mathcal{F}} \text{M-SEP}_{\text{Ave}}(B_0, N_0, F), \quad (2)$$

the corresponding optimal F was denoted by F_{B_0,N_0}; $\text{M-SEP}_{\text{Ave}}(B_0, N_0)$ was a minimum projection of $\text{M-SEP}_{\text{Ave}}(B_0, N_0, F)$ in the plane B-N; and the $\text{M-R}_{\text{P,Ave}}$, $\text{M-SEP}_{\text{Std}}$, $\text{M-R}_{\text{P,Std}}$ and M-SEP^+ of (B_0, N_0, F_{B_0,N_0}) were also determined.

Global optimal model: The global optimal model was selected according to the following:

$$\text{M-SEP}^*_{\text{Ave}} = \min_{\substack{B \in \mathcal{B} \\ N \in \mathcal{N} \\ F \in \mathcal{F}}} \text{M-SEP}_{\text{Ave}}(B, N, F). \quad (3)$$

The corresponding parameters were denoted by B^*, N^* and F^*, respectively, and the corresponding $\text{M-R}_{\text{P,Ave}}$, $\text{M-SEP}_{\text{Std}}$, $\text{M-R}_{\text{P,Std}}$ and M-SEP^+ were determined.

Local optimal model: Instrument design typically involves some restrictions of position and number of wavelengths (such as costs and material properties). In some instances, the demand of actual conditions cannot be met by the global optimal waveband. Therefore, local optimal wavebands that correspond to different positions and number of wavelengths are significant. For any fixed $B = B_0$, the local optimal model was selected according to the following:

$$\text{M-SEP}_{\text{Ave}}(B_0) = \min_{\substack{N \in \mathcal{N} \\ F \in \mathcal{F}}} \text{M-SEP}_{\text{Ave}}(B_0, N, F), \quad (4)$$

the corresponding optimal N and F were denoted by N_{B_0} and F_{B_0}, respectively; and the $\text{M-SEP}_{\text{Ave}}(B_0)$ was a minimum projection of $\text{M-SEP}_{\text{Ave}}(B_0, N, F)$ on the B-axis. The $\text{M-R}_{\text{P,Ave}}$, $\text{M-SEP}_{\text{Std}}$, $\text{M-R}_{\text{P,Std}}$ and M-SEP^+ of (B_0, N_{B_0}, F_{B_0}) were also determined. Meanwhile, for any fixed $N = N_0$, the local optimal model was selected according to the following:

$$\text{M-SEP}_{\text{Ave}}(N_0) = \min_{\substack{B \in \mathcal{B} \\ F \in \mathcal{F}}} \text{M-SEP}_{\text{Ave}}(B, N_0, F), \quad (5)$$

the corresponding optimal B and F were also denoted by B_{N_0} and F_{N_0}, respectively; and the $\text{M-SEP}_{\text{Ave}}(N_0)$ was a minimum projection of $\text{M-SEP}_{\text{Ave}}(B, N_0, F)$ on the N-axis. $\text{M-SEP}_{\text{Std}}$, $\text{M-R}_{\text{P,Std}}$ and M-SEP^+ of (B_{N_0}, N_0, F_{N_0}) were also determined.

The search range for the MWPLS method in this paper spanned the overall scanning region of 400 to 2498 nm with 1050 wavelengths. To reduce workload and to maintain representativeness, $\mathcal{B}$, $\mathcal{N}$, $\mathcal{F}$ were set as follows:

$$\mathcal{B} = \{1, 2, \ldots, 1050\},$$
$$\mathcal{N} = \{1, 2, \ldots, 100\} \cup \{102, 104, \ldots, 250\} \cup \{260, 270, \ldots, 1050\},$$
$$\mathcal{F} = \{1, 2, \ldots, 20\}. \quad (6)$$

Consequently, the total number of wavebands was 197,355. The computer platform was built by using MATLAB 7.6 software.

3. Results and Discussion

3.1. *Relationship between prediction effect and similarity of sample set*

The VIS-NIR spectra of the 163 soil samples are shown in Fig. 2. To verify the relationship between the prediction effect and the similarity of the sample set, an experiment was conducted. First, the modeling set (103 samples) was divided randomly into calibration (60 samples) and prediction (43 samples) sets for a total of 1500 times; the PLS models were established for all 1500 divisions. The 1500 divisions were arranged according to their similarity degree from small to large and were then divided into seven division groups; the range of the similarity degree of the seven groups were below 1%,

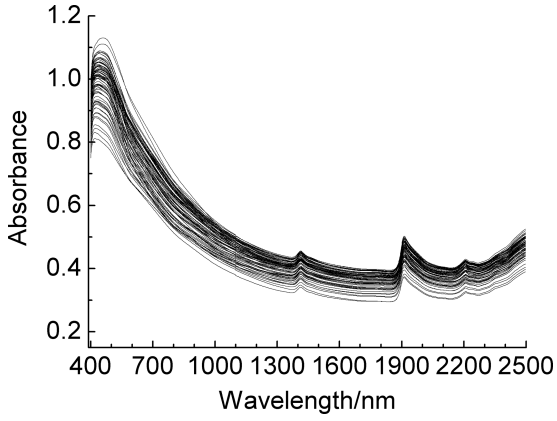


Fig. 2. Visible and NIR spectra of the 163 soil samples.

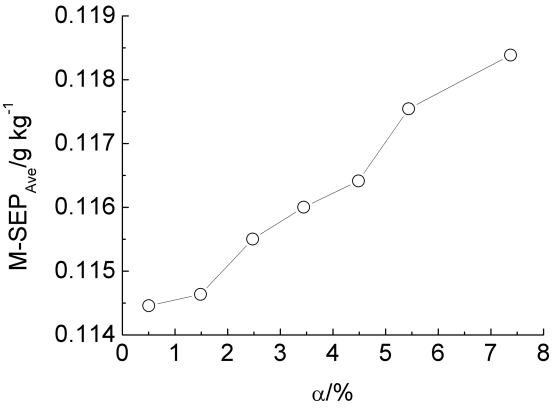


Fig. 3. Relationship between average prediction effects and average similarity degree of each division group.

from 1% to 2%, from 2% to 3%, from 3% to 4%, from 4% to 5%, from 5% to 6% and above 6%. For each division group, the average similarity degree $a^{(j)}$ and the average prediction effects M-SEP$^{(j)}_{\text{Ave}}$ were calculated, $j = 1, 2, \ldots, 7$. As shown in Fig. 3, M-SEP$^{(j)}_{\text{Ave}}$ increases, that is, the prediction effect of models falls, as similarity decreases (and $a^{(j)}$ increases). Therefore, the division for calibration and prediction sets has to be performed based on certain similarities in the modeling optimization process to prevent the evaluation distortion of the model.

3.2. *Comparison among the visible, short-NIR, long-NIR and overall scanning regions*

The overall scanning region was separated into the visible region (400 to 780 nm), the short-NIR region (780 to 1100 nm) and the long-NIR region (1100 to 2498 nm). PLS models were established in all regions including the overall scanning region, and the results are summarized in Table 1. Table 1 shows that M-SEP$_{\text{Ave}}$ and M-SEP^{+} of the long-NIR region were both significantly better than those of the other three wavebands. Thus, the long-NIR region had the best prediction accuracy and stability among all four wavebands.

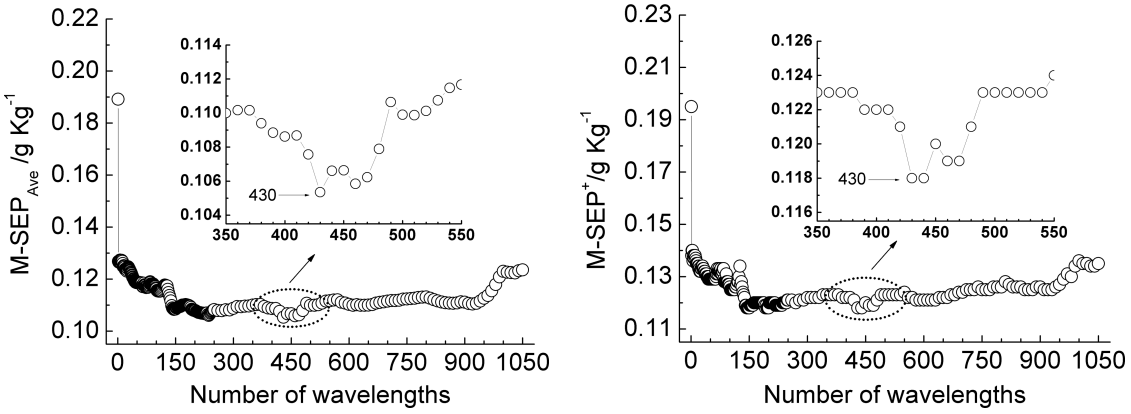
3.3. *Waveband optimization by using the MWPLS method*

The M-SEP$_{\text{Ave}}$ and M-SEP^{+} of the local optimal model for each N and for each beginning wavelength are shown in Figs. 4 and 5. The results show that M-SEP$_{\text{Ave}}$ and M-SEP^{+} for $N = 430$ and for the beginning wavelength 1424 nm had the best values, which indicates that these N and beginning wavelength had the best prediction accuracy and stability. The beginning wavelength and N of the global optimal model were 1424 nm and 430, respectively. The waveband was 1424 to 2282 nm in the long-NIR region. The corresponding prediction accuracy and stability are shown in Table 2. Tables 1 and 2 show that the global optimal model was evidently better than those for long-NIR and the overall scanning regions. Moreover, the number of wavelengths N was reduced significantly, thereby decreasing model complexity.

Actually, Refs. 6–8 have confirmed that simultaneous assessment of various soil properties (such as organic matter and TN of soil) by visible, NIR or mid-infrared diffuse reflectance spectroscopy, respectively are feasible. However, their models were only based on overall regions, and further optimization of wavelength is not performed. This current paper focused on the NIR region for the farmland soil samples (yellow brown earth) located in Pearl River Delta of China. First, the PLS model of overall NIR region was established, and then further optimization of wavelength was accomplished by using the improved MWPLS method with the stability. There is no contradiction between our results and previous findings.

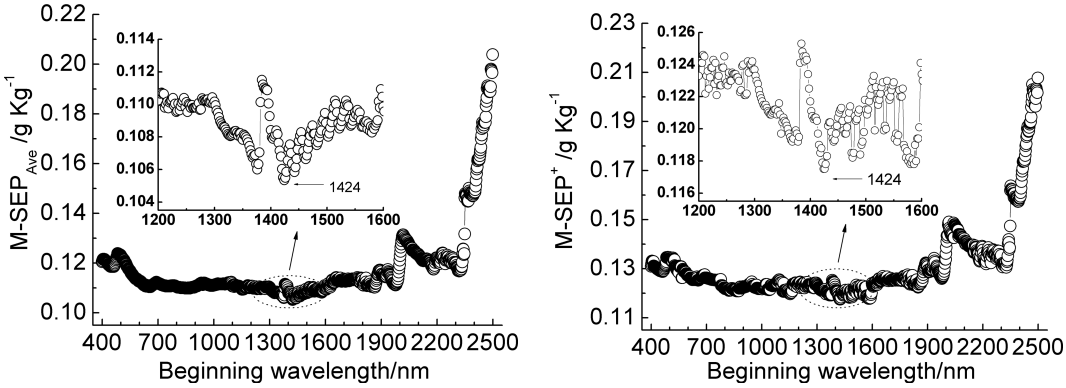
Table 1. Modeling prediction accuracy and stability of PLS models that correspond to the visible, short-NIR, long-NIR, and overall scanning regions.

Waveband (nm)	F	M-SEP$_{\text{Ave}}$	M-SEP$_{\text{Std}}$	M-R$_{\text{P,Ave}}$	M-R$_{\text{P,Std}}$	M-SEP^{+}
400–780	3	0.126	0.010	0.820	0.0314	0.137
780–1100	6	0.133	0.016	0.797	0.0525	0.150
1100–2498	8	0.115	0.014	0.860	0.0336	0.129
400–2498	10	0.124	0.011	0.843	0.0264	0.135



(a) (b)

Fig. 4. M-SEP$_{\mathrm{Ave}}$ and M-SEP^{+} of the local optimal model for each N: (a) M-SEP$_{\mathrm{Ave}}$ and (b) M-SEP^{+}.



(a) (b)

Fig. 5. M-SEP$_{\mathrm{Ave}}$ and M-SEP^{+} of the local optimal model for each beginning wavelength: (a) M-SEP$_{\mathrm{Ave}}$ and (b) M-SEP^{+}.

3.4. *Model space with equivalence*

As previously mentioned, the global optimal waveband 1424 to 2282 nm has a minimum M-SEP$_{\mathrm{Ave}}$ (0.105 g kg^{-1}) and a minimum M-SEP^{+} (0.118 g kg^{-1}); however, statistically speaking, the models with slightly fluctuating prediction accuracy are considered equivalent because of the random and limited modeling samples. Therefore, the optimal accuracy could float upward at an appropriate range (usually 1%). In this paper, minimum M-SEP^{+} floated upward from 0.118 to 0.119 g kg^{-1}, which indicates that the models that satisfy the following inequality can be considered equivalent to the optimal model with stability:

$$\{(B,N,F)|0.118 \leq \text{M-SEP}^{+}(B,N,F) \leq 0.119\}. \quad (7)$$

Table 2. Modeling prediction accuracy and stability that correspond to the optimal MWPLS waveband and to the equivalent optimal waveband.

Waveband (nm)	N	F	M-SEP$_{\mathrm{Ave}}$	M-SEP$_{\mathrm{Std}}$	M-R$_{\mathrm{P,Ave}}$	M-R$_{\mathrm{P,Std}}$	M-SEP^{+}
1424–2282	430	11	0.105	0.012	0.883	0.029	0.118
1590–1870	141	6	0.109	0.009	0.863	0.029	0.118

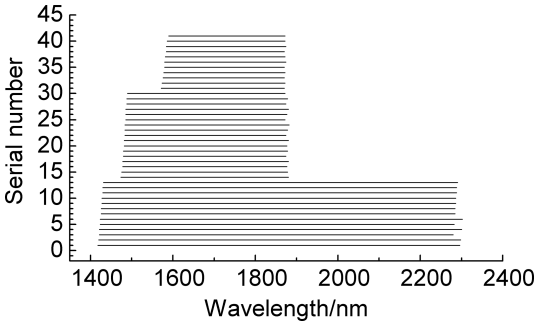


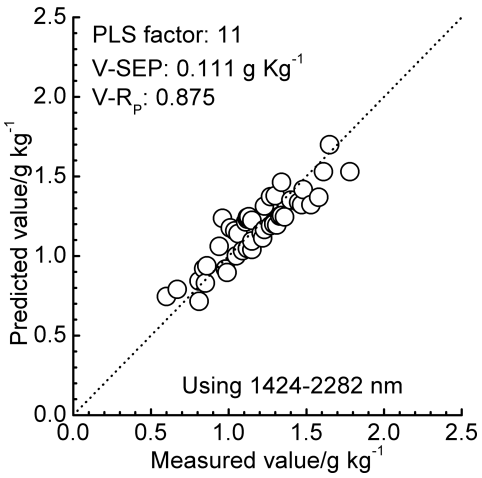
Fig. 6. Position of 41 wavebands in the equivalent model space.

An equivalent model space that contained 41 wavebands was obtained; all wavebands were equivalent optimal wavebands, the beginning wavelength ranged from 1418 to 1590 nm, the ending wavelength ranged from 1870 to 2302 nm, N ranged from 141 to 443, the wavebands spanned 1418 to 2302 nm and the public range was 1590 to 1870 nm with 141 wavelengths. The public range was just one of the equivalent optimal wavebands and had the lowest N; the corresponding prediction accuracy and stability are shown in Table 2. Figure 6 shows the position of 41 equivalent optimal wavebands.

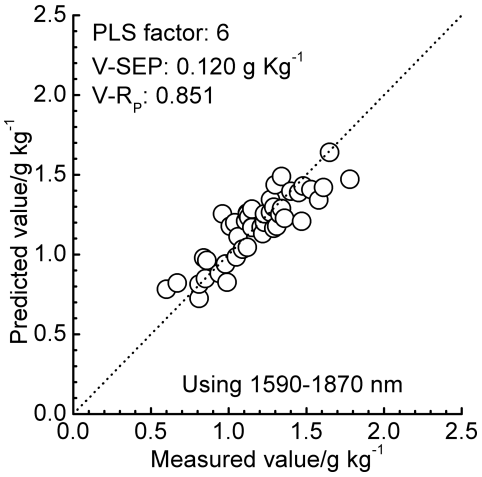
The equivalent model space provided various waveband selections and could solve the restrictions of position and the number of wavelengths in instrument design. The upward range could be adjusted correspondingly.

3.5. *Model validation*

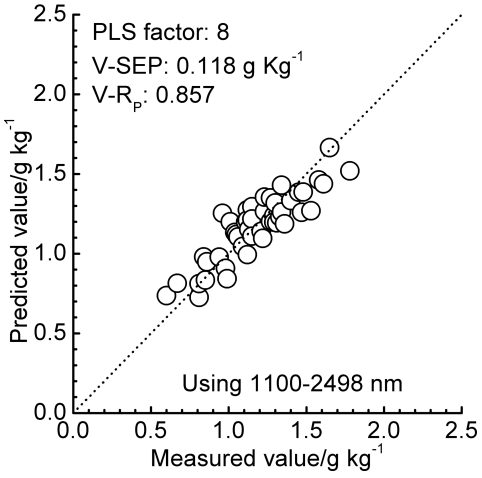
With the PLS models for 1424 to 2282 nm, 1590 to 1870 nm and 1100 to 2498 nm as examples, three models were validated with the validation set. Figure 7 shows the relationship between the predicted and the measured values of TN for 60 validation samples, whereas Table 3 shows the validation effects that correspond to the three wavebands. The results exhibit the high prediction accuracy of the three models and indicate that the predicted TN values of the samples are close to the measured values. However, the public waveband of 1590 to 1870 nm had the lowest N, which significantly reduced method complexity. Satisfactory prediction effects can be achieved for random validation samples because stability was considered in the modeling optimization process.



(a)



(b)



(c)

Fig. 7. Relationship between the predicted values and the measured values of TN for 60 validation samples: (a) using 1424 to 2282 nm; (b) using 1590 to 1870 nm; and (c) using 1100 to 2498 nm.

Table 3. Validation effects that correspond to the optimal MWPLS waveband, the equivalent optimal waveband, and to the long-NIR region.

Waveband (nm)	N	F	V-SEP	V-R_P
1424–2282	430	11	0.111	0.875
1590–1870	141	6	0.120	0.851
1100–2498	700	8	0.118	0.857

4. Conclusions

The selection of stable wavebands for the NIR spectroscopic analysis of TN in soil was accomplished by using the improved MWPLS method in terms of stability and equivalence. A new modeling approach was performed based on randomness, similarity and stability, which produced an objective, stable and practical model. The global optimal waveband and the local optimal wavebands of the MWPLS method were determined. In addition, a model space equivalent to the optimal MWPLS model was proposed. The optimal MWPLS waveband was found to be 1424 to 2282 nm. The model space contained 41 equivalent optimal wavebands; the public waveband was 1590 to 1870 nm and contained sufficient TN information. The wavebands of 1424 to 2282 nm, 1590 to 1870 nm, and the long-NIR region 1100 to 2498 nm all achieved satisfactory validation effects. However, the method complexity was substantially reduced because the public waveband of 1590 to 1870 nm had a minimum number of wavelengths. Stability was considered in the modeling optimization process. Therefore, a satisfactory prediction effect can be achieved for random validation samples. Equivalent wavebands provide valuable guidelines for designing spectroscopic instruments. The proposed methodological framework and computer algorithm are universal and could be applied to other fields.

Acknowledgments

This work was supported by National Natural Science Foundation of China (No. 61078040), Open Foundation of State Key Laboratory of Soil and Sustainable Agriculture (Institute of Soil Science, Chinese Academy of Sciences, No. 0812201201) and the Science and Technology Project of Guangdong Province (No. 2012B031800917).

References

1. R. Welle, W. Greten, B. Rietmann, S. Alley, G. Sinnaeve, P. Dardenne, "Near-infrared spectroscopy on chopper to measure maize forage quality parameters online," *Crop Sci.* **43**(4), 1407–1413 (2003).
2. J. Y. Chen, H. Zhang, R. Matsunaga, "Rapid determination of the main organic acid composition of raw Japanese Apricot fruit juices using near-infrared spectroscopy," *J. Agr. Food Chem.* **54**(26), 9652–9657 (2006).
3. A. C. Sousa, M. M. L. M. Lucio, O. F. Bezerra, G. P. S. Marcone, A. F. C. Pereira, E. O. Dantas, W. D. Fragoso, M. C. U. Araujo, R. K. H. Galvao, "A method for determination of COD in a domestic wastewater treatment plant by using near-infrared reflectance spectrometry of seston," *Anal. Chim. Acta* **588**(2), 231–236 (2007).
4. K. H. Hazen, M. A. Arnold, G. W. Small, "Measurement of glucose and other analytes in undiluted human serum with near-infrared transmission spectroscopy," *Anal. Chim. Acta* **371**(2–3), 255–267 (1998).
5. Y. Ozaki, "Near-infrared spectroscopy-its versatility in analytical chemistry," *Anal. Sci.* **28**(6), 545–563 (2012).
6. A. Moron, D. Cozzolino, "Application of near infrared reflectance spectroscopy for the analysis of organic C, total N and pH in soils of Uruguay," *J. Near Infrared Specstroc.* **10**(3), 215–221 (2002).
7. D. Cozzolino, A. Moron, "Potential of near-infrared reflectance spectroscopy and chemometrics to predict soil organic carbon fractions," *Soil Tillage Res.* **85**(1–2), 78–85 (2006).
8. R. A. Viscarra Rossel, D. J. J. Walvoort, A. B. McBratney, L. J. Janik, J. O. Skjemstad, "Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties," *Geoderma* **131**(1–2), 59–75 (2006).
9. D. A. Burns, E. W. Ciurczak, *Handbook of Near-Infrared Analysis 2nd Edition*, Marcel Dekker Inc, New York (2001).
10. W. Z. Lu, *Modern Near-Infrared Spectroscopy Analytical Technology*, China Petrochemical Press, Beijing (2007).
11. J. H. Jiang, R. J. Berry, H. W. Siesler, Y. Ozaki, "Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data," *Anal. Chem.* **74**(14), 3555–3565 (2002).
12. Y. P. Du, Y. Z. Liang, J. H. Jiang, R. J. Berry, Y. Ozaki, "Spectral regions selection to improve prediction ability of PLS models by changeable size

moving window partial least squares and searching combination moving window partial least squares," *Anal. Chim. Acta* **501**(2), 183–191 (2004).

13. H. Z. Chen, T. Pan, J. M. Chen, Q. P. Lu, "Waveband selection for NIR spectroscopy analysis of soil organic matter based on SG smoothing and MWPLS methods," *Chemometr. Intell. Lab. Syst.* **107**(1), 139–146 (2011).
14. T. Pan, Z. H. Chen, J. M. Chen, Z. Y. Liu, "Near-Infrared spectroscopy with waveband selection stability for the determination of COD in sugar refinery wastewater," *Anal Methods* **4**(4), 1046–1052 (2012).
15. Z. Y. Liu, B. Liu, T. Pan, J. D. Yang, "Determination of amino acid nitrogen in tuber mustard using near-infrared spectroscopy with waveband selection stability," *Spectrochim. Acta A* **102**, 269–274 (2013).
16. J. Xie, T. Pan, J. M. Chen, H. Z. Chen, X. H. Ren, "Joint optimization of Savitzky-Golay smoothing models and partial least squares factors for near-infrared spectroscopic analysis of serum glucose," *Chin. J. Anal. Chem.* **38**(3), 342–346 (2010).
17. T. Pan, J. M. Liu, J. M. Chen, G. P. Zhang, Y. Zhao, "Rapid determination of preliminary thalassaemia screening indicators based on near-infrared spectroscopy with wavelength selection stability," *Anal Methods* **5**(17), 4355–4362 (2013).