Wavelength selection of near infrared spectroscopy analysis for soil organic matter based on ECMWMLR and MWPLS methods

Huazhou Chen and Tao Pan*

Key Laboratory of Optoelectronic Information and Sensing Technologies of Guangdong Higher Educational Institutes, Department of Optoelectronic Engineering, Jinan University, Guangzhou 510632, China
*Corresponding author: tpan@jnu.edu.cn

Introduction

Organic matter content is an important indicator of soil. The direct rapid analytical method for soil organic matter by near infrared (NIR) spectroscopy is very significant. Selecting the appropriate NIR wavelengths for calibration can improve model prediction, reduce model complexity and provide valuable reference information for the design of a dedicated soil spectrometer. In this paper, equidistant combination moving window multiple linear regression (ECMWMLR)[1-3] and moving window partial least squares (MWPLS)[4] were used to select a combination of discrete NIR wavelengths, and a continuous region of NIR wavelengths, respectively, from spectra of soil. To get stable results, all models were obtained using the average data from 50 different divisions of the calibration and prediction sets.

Materials and Methods

Experimental materials, instrument and measurement method

Ninety one farmland soil samples were collected. The organic matter content of each sample was measured using the routine chemical method of potassium dichromate ($K_2Cr_2O_7$) oxidation. The organic matter chemical values ranged from 0.676 to 2.830%, with a mean and standard deviation of 1.622 and 0.520%, respectively. Spectra were measured using a FOSS XDS rapid content grating spectrometer with a diffuse reflectance accessory. The complete spectral range collected was 400–2500 nm with a spectral resolution of 2 nm.

Dividing for calibration set and prediction set

All samples were divided into calibration and prediction sets at a ratio of about 2:1. Model parameters were changed as the calibration and prediction sets were subdivided, to avoid fluctuations of predictions. In order to establish objective models, a division method for the calibration and prediction sample sets based on the optimal single wavelength prediction bias (OSWPB) was proposed.[5] A total of 50 different divisions of the calibration set and the prediction set were made. Calibration models were established for each division. Model predictions (e.g. root mean squared error of prediction) across 50 different divisions were averaged for each combination of model parameters. Based on the average data, the stable optimal model could be selected.

MWPLS method

Parameters of the MWPLS method include the beginning wavelength (B), the number of adopted wavelengths (N_M) and the PLS factor (F). B was set from 400 to 2498 nm, N_M was set from 1 to 1050, and F was set from 1 to 30. A PLS model for each parameter combination (B, N_M, F) was generated.

ECMWMLR method

Parameters of the ECMWMLR method include the beginning wavelength (B), the number of adopted wavelengths (N_E) and the gap of adopted wavelengths (G). B was set from 400 to 2500 nm, N_E was set from 1 to 60, and G was set from 1 to 1050. A multiple linear regression model for each parameter combination (B, N_E, G) was generated.

Model evaluation indicators

The model evaluation indicators include the root mean squared error of prediction (RMSEP) and correlation coefficient of predication (R_P). The parameters RMSEP and R_P were calculated for all 50 divisions, each parameter was averaged, and the mean RMSEP value was used as the goal of model optimisation and parameter design.

Reference paper as:
H. Chen and T. Pan (2012).Wavelength selection of near infrared spectroscopy analysis for soil organic matter based on ECMWMLR and MWPLS methods, in: Proceedings of the15th International Conference on Near Infrared Spectroscopy,
Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 279-281.

Results and Discussion

The NIR spectra of soil samples

The NIR spectra of 91 samples are shown in Figure 1. Based on the prediction of the optimal single wavelength (1072 nm) model, all samples were divided into calibration (64 samples) and prediction (27 samples) sets.

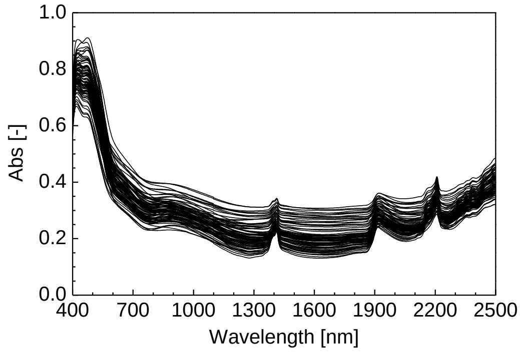

Figure 1. NIR spectra of 91 soil samples

RMSEP of the optimal models based on MWPLS

The optimal model for the fixed B and changed N_M was selected according to RMSEP; RMSEP of the optimal model corresponding to each B is shown in Figure 2. The optimal model for the fixed N_M and changed B was also selected according to RMSEP (Figure 3).

RMSEP of the optimal models based on ECMWMLR

The optimal models for the fixed B, changed N_E and changed G (Figure 4), and fixed N_E, changed B and changed G (Figure 5), were selected according to RMSEP values.

Selected wavelength combination

The global optimal model based on ECMWMLR was selected, and the corresponding B, N_E, G were 1786 nm, 9 and 20, respectively. The mean value and standard deviation of RMSEP and R_P were 0.265 and 0.028%, and 0.871 and 0.027, respectively. The global optimal model based on MWPLS was selected, and the corresponding B, N_M and F were 1692 nm, 95 and 14, respectively. The selected wavelength range was 1692–1880 nm, and the mean value and standard deviation of RMSEP and R_P were 0.275 and 0.033%, and 0.870 and 0.029, respectively.

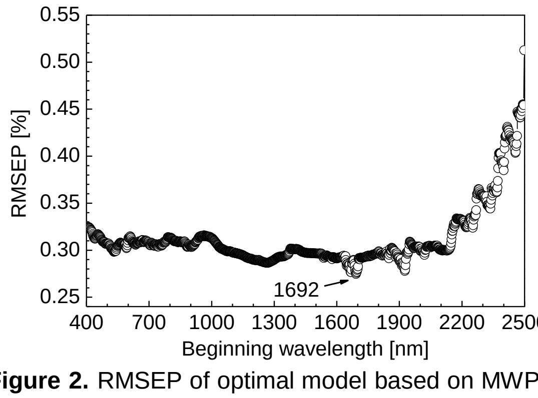

Figure 2. RMSEP of optimal model based on MWPLS corresponding to each beginning wavelength.

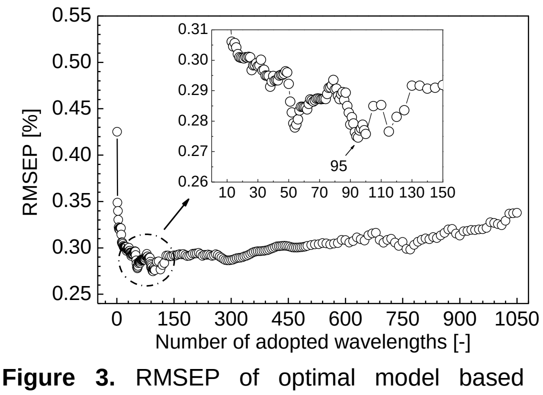

Figure 3. RMSEP of optimal model based on MWPLS corresponding to each number of adopted wavelengths.

Reference paper as:
H. Chen and T. Pan (2012).Wavelength selection of near infrared spectroscopy analysis for soil organic matter based on ECMWMLR and MWPLS methods, in: Proceedings of the15th International Conference on Near Infrared Spectroscopy,
Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 279-281.

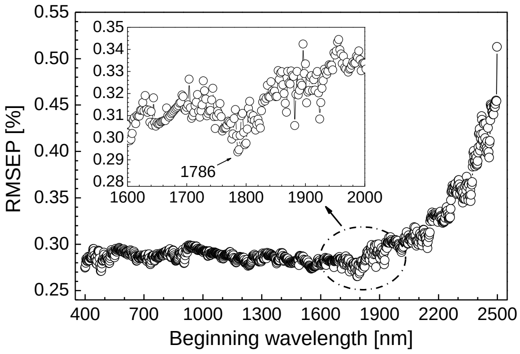

Figure 4. RMSEP of optimal model based on ECMWMLR corresponding to each beginning wavelength.

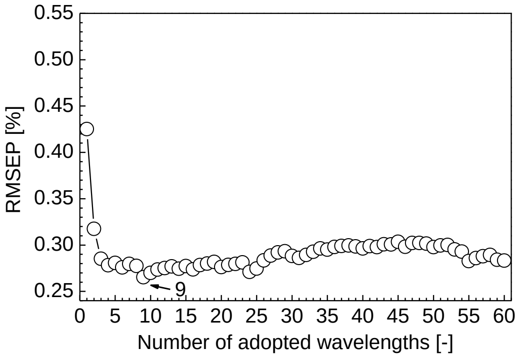

Figure 5. RMSEP of optimal model based on ECMWMLR corresponding to each number of adopted wavelength.

Conclusion

Results presented here were obtained using the average data of prediction from 50 different divisions of the calibration and prediction sets. The optimal ECMWMLR model only used 9 wavelengths to provide a stable model with a slightly better prediction than an optimal MWPLS model. ECMWMLR and MWPLS were both effective methods for selecting the appropriate NIR wavelengths for use in calibrations, , and provided valuable reference information for designing a dedicated soil spectrometer that could collect data from discrete and continuous NIR wavelengths.

Acknowledgements

This work was supported by the NSFC (10771087, 61078040), and the science and technology project of Guangdong province (2009B030801239).

References

1. T. Pan, J. Xie, H.Z. Chen, H. Yin and X.D. Chen, in *Near Infrared Spectroscopy: Proceedings of the 14th International Conference*, Ed by S. Saranwong, S. Kasemsumran, W. Thanapase and P. Williams, IM Publications, Chichester, UK, pp. 867-870 (2010).
2. T. Pan, H.Z. Chen, J. Xie, H. Yin and X.D. Chen, in *Near Infrared Spectroscopy: Proceedings of the 14th International Conference*, Ed by S. Saranwong, S. Kasemsumran, W. Thanapase and P. Williams, IM Publications, Chichester, UK, pp. 159-162 (2010)
3. T. Pan, J. Xie, Y. Shan, H.Z. Chen, *Adv. Mater. Res.* **181-182**, 647-650 (2011).
4. J.H. Jiang, R.J. Berry, H.W. Siesler and Y. Ozaki, *Anal. Chem.* **74**, 3555-3565 (2002).
5. J. Xie, T. Pan, J.M. Chen, H.Z. Chen and X.H. Ren, *Chinese J. Anal. Chem.* **38**, 342-346 (2010).

Reference paper as:
H. Chen and T. Pan (2012).Wavelength selection of near infrared spectroscopy analysis for soil organic matter based on ECMWMLR and MWPLS methods, in: Proceedings of the15th International Conference on Near Infrared Spectroscopy,
Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 279-281.