

Determination of soil properties using Fourier transform mid-infrared photoacoustic spectroscopy

Changwen Du ^{*}, Jianmin Zhou, Huoyan Wang, Xiaoqin Chen, Anning Zhu, Jiabao Zhang

National Key Laboratory of Soil and Agricultural Sustainability, Institute of Soil Science Chinese Academy of Sciences, Nanjing 210008, China

ARTICLE INFO

Article history:

Received 17 September 2007

Received in revised form 7 April 2008

Accepted 9 April 2008

Available online 22 April 2008

Keywords:

Soil nutrients

Photoacoustic spectroscopy

Partial least squares

ABSTRACT

Photoacoustic spectroscopy (PAS) is based on the absorption of electromagnetic radiation by analyte molecules, and this technique has emerged as a valuable tool for the study of materials like biological, chemical and geological samples. In this paper, Fourier transform mid-infrared photoacoustic spectroscopy (FTIR-PAS) was used in the prediction of soil properties. Air-dried soil samples ($n = 56$) from Fengqiu Ecology Experimental Station Chinese Academy of Sciences were involved in this experiment, and FTIR-PAS spectra of these soil samples were recorded. These FTIR-PAS spectra indicated abundant soil information, but overlapping of absorption made it difficult to make direct measurement of soil properties. Partial least squares (PLS) models based on soil FTIR-PAS spectra was developed to predict available nitrogen (N), phosphorus (P), potassium (K) and organic matter content of soil. 42 soil samples were firstly used in leave-one-out cross-validation, and calibration error, calibration coefficient, validation error and ratio of standard deviation to prediction error (RPD) were obtained to optimize the PLS factor number; then based on the optimized PLS models the soil properties of the other 14 soil samples were predicted. The calibration statistics showed that the PLS model was suitable to use in the prediction of available N, P, K and organic matter content of soil. This prediction technique was non-destructive, and no sample pre-treatment was needed, which made FTIR-PAS a very promising method for fast evaluation of soil properties as well as soil quality.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Soil is a complex mixture of minerals, oxides, organic matter, microorganisms and other compounds that have been produced during the soil formation process, which provides an extremely large variety of soils. In particular, the topmost layers of agricultural soil are very delicate and dynamic in their behavior. Therefore, soil analysis has been a routine practice for the evaluation of soil fertility. Efficient land-resource assessment and soil management (such as the so-called “precision farming” concept) requires accurate representation of this variability. Standard laboratory techniques for soil analysis are based on chemical analysis, and are time-consuming and not suitable for such applications [1,2].

Spectroscopy, both in the near-infrared (NIR) and mid-infrared (MIR) ranges, is a promising technology for soil analysis, as demonstrated in the recent review paper of Viscarra et al. [3]. Although such techniques are more rapid and generic than

standard procedures [4], time-consuming sample preparation is still required in transmittance measurements (KBr pellet preparation), and it is difficult to use in quantitative measuring [1]; reflectance spectroscopic methods are widely used in quantitative analysis, which are very useful for soil analysis. However, for diffuse reflection IR spectroscopy (DRIFTS) the particle size of soil sample has a heavy influence on quantitative analysis of soil, and it takes time to prepare the soil particles into the same size [5]. Attenuated total reflection (ATR) spectroscopy requires minimal sample preparation, but the soil sample must be close to water saturation in order to ensure good contact between the sample and the ATR crystal; as a result, most bands of interest are hidden by the much stronger water bands, and much information is lost [6].

Fourier transform infrared photoacoustic spectroscopy (FTIR-PAS) is based on the absorption of electromagnetic radiation by analyte molecules. Non-radiative relaxation processes (such as collisions with other molecules) lead to local warming of the sample matrix. Pressure fluctuations are then generated by thermal expansion, which can be detected by a very sensitive microphone (Fig. 1) [7]. The resulting spectrum differs from both of equivalent transmittance or reflectance spectra since the technique detects non-radiative transitions in the sample. The detected

^{*} Corresponding author. Tel.: +86 25 86881565; fax: +86 25 86881000.
E-mail address: chwd@issas.ac.cn (C. Du).

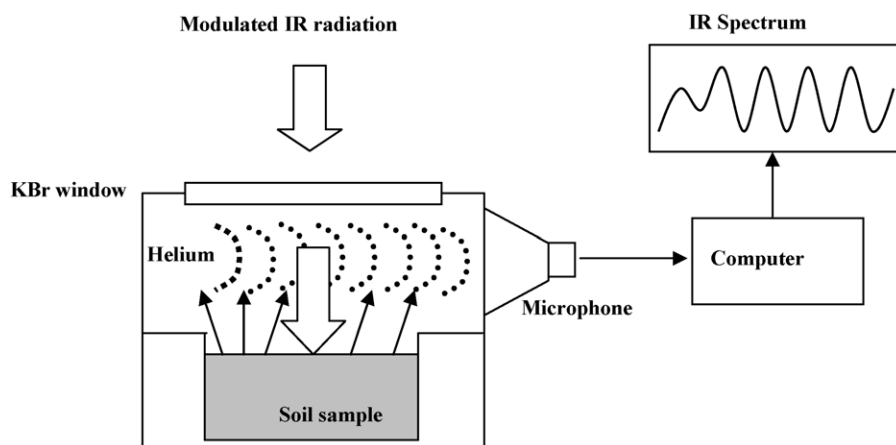


Fig. 1. Schematic description of the photoacoustic spectroscopy setup.

signal in photoacoustic spectroscopy is proportional to the sample concentration, and can be used with highly absorbing samples without any pre-treatment (soil pellet and paste preparation or soil extraction), which will be very useful and suitable for soil testing. Using multivariate statistic analysis, such as partial least squares (PLS), FTIR-PAS made it a wide application with quantitative purpose in solid substance, such as the analysis of depth profiling in coated paper [8,9], chemical composition of woods [10], determination of carboxyl content in high-yield kraft pulps [11], characterization of potato chips [12], beef and pork [13], process of bio-film [14], and measurement of ink in printed paper [15]. However, the application of FTIR-PAS in soil analysis is very limit. Du et al. [6] characterized the FTIR-PAS spectra of agricultural Mediterranean soils, and identification of these soils was successfully conducted, which showed excellent performance in quality analysis [16].

Therefore, in the present study, we applied the FTIR-PAS technique in the quantitative analysis of soil. The spectral data were obtained, and PLS models were used to determine available N, P, K, and organic matter content in the soil. Soil samples were divided into two sets, and one set was used to conduct leave-one-out cross-validation, in which PLS models for each soil property were optimized; then the other set was used to validate the optimized PLS models. It appears that, based on our comprehensive literature review, the work described here is the first attempt ever made to develop the FTIR-PAS technique for quantitative analysis of soil properties.

2. Experimental

2.1. Preparation of the soil samples

56 topsoil samples (Fluvo-aquic soil, 0–20 cm depth) were collected from experiment field with varied fertility in Fengqiu Ecology Experimental Station Chinese Academy of Sciences. The samples were air-dried at room temperature, and passed through a 2 mm sieve.

2.2. Measurement of soil nutrients

Soil properties were determined by conventional chemical method [17]. Soil organic matter was determined by wet oxidation; soil available nitrogen was extracted by KCl; soil available phosphorus was extracted by NaHCO_3 ; and soil available potassium was extracted by NH_4OAc . Table 1 showed available N, P, K and organic matter content of 56 soil samples.

2.3. FTIR-PAS spectra measurements

Photoacoustic spectra were recorded for all soil samples using a Bruker Vector 22 spectrophotometer (Bruker, Germany) equipped with a photoacoustic cell (Model 300, MTEC, USA). After placing the sample (about 200 mg) in the cell holding cup (diameter 5 mm, height 3 mm) purging the cell with dry helium (10 mL/min) for 10 s, the scans were conducted in the wave-number region of $4000\text{--}500\text{ cm}^{-1}$ with a resolution of 4 cm^{-1} and a mirror velocity of 0.6329 cm/s , and 16 successive scans were recorded.

2.4. Data processing

Photoacoustic spectra were pre-processed with a smoothing filter (first-order Savitzky–Golay filter with a 25-point window). The Savitzky–Golay filter method essentially performs a local polynomial regression to determine the smoothed value for each data point. This method is superior to adjacent averaging because it tends to preserve features of the data such as peak height and width, which are usually ‘washed out’ by adjacent averaging, and the detail of the smoothing filter was given by Savitzky and Golay [18].

Quantitative analysis of the PAS spectra was performed using PLS, and a general form of PLS model is expressed as:

$$\begin{aligned} X &= TP^T + E \\ Y &= UQ^T + F \end{aligned}$$

where X is the variable predictor matrix (absorbance); Y is the variable response matrix (soil properties); T and U are the X -scores and Y -scores matrices; P and Q are the X -loading and Y -loading matrices; E and F are the X -residual and Y -residual matrices. The coordinates of the sample in a coordinate system defined by the principal components (PCs) are called scores. The loading vectors are the bridge between the variable space and the PC space. The loadings provide the information about how much each variable contributes to each PC. In the case here, T contains information about the samples, and P contains information about the wavenumbers. The detailed PLS algorithm in the PLS analysis is well described by Geladi and Kowalski [19].

2.5. Optimization of PLS factor

It is found that the results of calibration and validation in PLS modeling were heavily decided by the number of PLS factor.

Table 1

Available N, P, K and organic matter content of 56 soil samples used for calibration and validation

No.	Available N (mg/kg)	Available P (mg/kg)	Available K (mg/kg)	Organic matter (g/kg)
s1	13.14	6.58	132.50	8.71
s2	8.84	1.31	59.50	6.49
s3	18.88	13.78	131.50	13.46
s4	10.51	7.00	134.50	9.65
s5	9.56	18.85	241.30	7.90
s6	11.23	6.55	127.50	9.25
s7	11.95	19.91	231.60	7.56
s8	12.35	5.85	46.00	8.12
s9	21.50	15.26	137.50	13.73
s10	12.35	5.85	46.00	8.12
s11	13.14	6.58	132.50	8.71
s12	12.19	6.12	44.30	9.12
s13	16.73	9.28	123.80	11.30
s14	11.23	6.55	127.50	9.25
s15	16.73	9.28	123.80	11.30
s16	18.88	8.56	125.40	11.04
s17	10.51	7.00	134.50	9.65
s18	18.88	8.56	125.40	11.04
s19	10.04	1.33	247.50	6.76
s20	21.50	8.66	119.50	10.26
s21	11.71	7.05	45.00	8.98
s22	21.50	8.66	119.50	10.26
s23	11.71	17.42	225.60	7.58
s24	8.05	1.47	60.50	6.94
s25	12.19	6.12	44.30	9.12
s26	21.50	15.26	137.50	13.73
s27	9.97	1.75	246.30	6.97
s28	18.88	13.78	131.50	13.46
s29	8.36	1.43	243.80	6.72
s30	16.01	12.63	130.40	13.37
s31	10.75	1.37	63.70	6.48
s32	9.56	18.85	241.30	7.90
s33	9.56	1.50	240.80	6.51
s34	16.01	12.63	130.40	13.37
s35	11.71	17.42	225.60	7.58
s36	11.95	19.91	231.60	7.56
s37	8.36	1.43	243.80	6.72
s38	21.50	11.61	131.00	14.85
s39	9.56	1.50	240.80	6.51
s40	10.04	1.33	247.50	6.76
s41	16.72	9.66	128.30	11.42
s42	10.75	1.37	63.70	6.48
s43	7.65	1.22	52.00	6.23
s44	13.58	18.35	227.50	7.86
s45	7.65	1.22	52.00	6.23
s46	8.05	1.47	60.50	6.94
s47	11.95	5.96	130.90	9.64
s48	8.84	1.31	59.50	6.49
s49	11.95	5.96	130.90	9.64
s50	11.71	7.05	45.00	8.98
s51	16.72	9.66	128.30	11.42
s52	21.50	11.61	131.00	14.85
s53	13.58	18.35	227.50	7.86
s54	9.97	1.75	246.30	6.97
s55	10.51	7.00	134.50	9.65
s56	12.19	7.00	47.50	8.55

Involving of more PLS factors in the models resulted in better calibration, however, the validation results would turn poor, which would lead to over-training. Therefore, optimization should be made to obtain low calibration and validation error and high calibration coefficient. The soil samples were split randomly into a calibration set containing 75% of the spectra and a validation set containing the remaining 25%, i.e., 42 soil samples were used as calibration set, and 14 soil samples were used as validation set. Leave-one-out cross-validation was made using the 42 soil samples in calibration set, and the number of PLS factor was optimized by the cross-validation calibration coefficient, cross-validation calibration error and the ratio of standard deviation to prediction error (RPD) [20]. The prediction

error was presented in terms of the root mean squares error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y - y')^2}{N}}$$

where y and y' denote the predicted and actual values, and N is the number of soil samples.

RPD was also calculated to confirm the optimized PLS factors.

$$RPD = \frac{SDC}{RMSEV}$$

where SDC is the standard deviation in cross-validation calibration, and RMSECV is the root mean standard error in cross-validation.

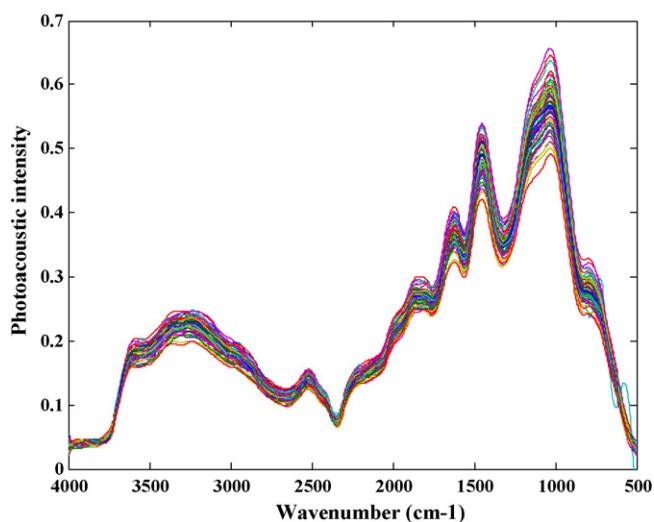


Fig. 2. FTIR-PAS spectra of Fluvo-aquic soil samples.

With the optimized PLS models the soil properties in validation set (14 soil samples) were predicted, and the models' prediction capability was evaluated.

3. Results and discussion

3.1. FTIR-PAS spectra of soils

Fig. 2 showed the FTIR-PAS spectra of typical Fluvo-aquic soil samples, and more absorptions were observed comparing with transmittance and ATR spectra, and visual differences among the

PAS spectra were found, and a detailed comparison has been presented elsewhere [6,21]. Clear absorptions were visible showed in several spectral regions, and in particular around 2800–3700 cm^{-1} , 2200–2600 cm^{-1} , 1800–2100 cm^{-1} , and 900–1600 cm^{-1} . Soil was mainly composed of organic matter, clay minerals and water. Organic matter was complicated material, and had abundant absorptions in mid-infrared region, which made it difficult to have a specific identification. Kaolin (1:1 clay mineral) and Bentonite (2:1 clay mineral) are two popular clay types commonly encountered in the investigated soils. The absorptions of bentonite were demonstrated in the regions of 2800–3700 cm^{-1} (O–H stretching), 1500–1800 cm^{-1} (C=O stretching), 800–1200 cm^{-1} (fingerprint region), in which the absorption in the regions of 2800–3700 cm^{-1} was a wide band, the absorptions of kaolin were indicated in the regions of 3500–3700 cm^{-1} , 1500–2000 cm^{-1} , 800–1200 cm^{-1} (Si–O stretching), in which the absorption in the regions of 800–1200 cm^{-1} was strong; soil calcium carbonate also showed photoacoustic absorption in the regions of 2900–3100 cm^{-1} , 2300–2600 cm^{-1} , 1000–1600 cm^{-1} , 1600–1700 cm^{-1} and 2100–2200 cm^{-1} , and the absorption in the regions 1000–1600 cm^{-1} was very strong, but was heavily interfered by some other absorptions; the absorption in the regions 2300–2600 cm^{-1} was strong enough, and was less interfered, which was useful in quantitative analysis; the absorption in the regions of 2900–3100 cm^{-1} might come from the water attached on the calcium carbonate particle surface. Soil water indicated a strong absorption in the region of 1600–1700 cm^{-1} and 2900–3600 cm^{-1} [16,22] but, the interference was very strong in the region of 2900–3600 cm^{-1} .

The characteristics of soil photoacoustic absorptions provided a basis in quantitative analysis of soil. However, strong interferences of absorptions made it necessary to use multivariate calibration in quantitative analysis of soil.

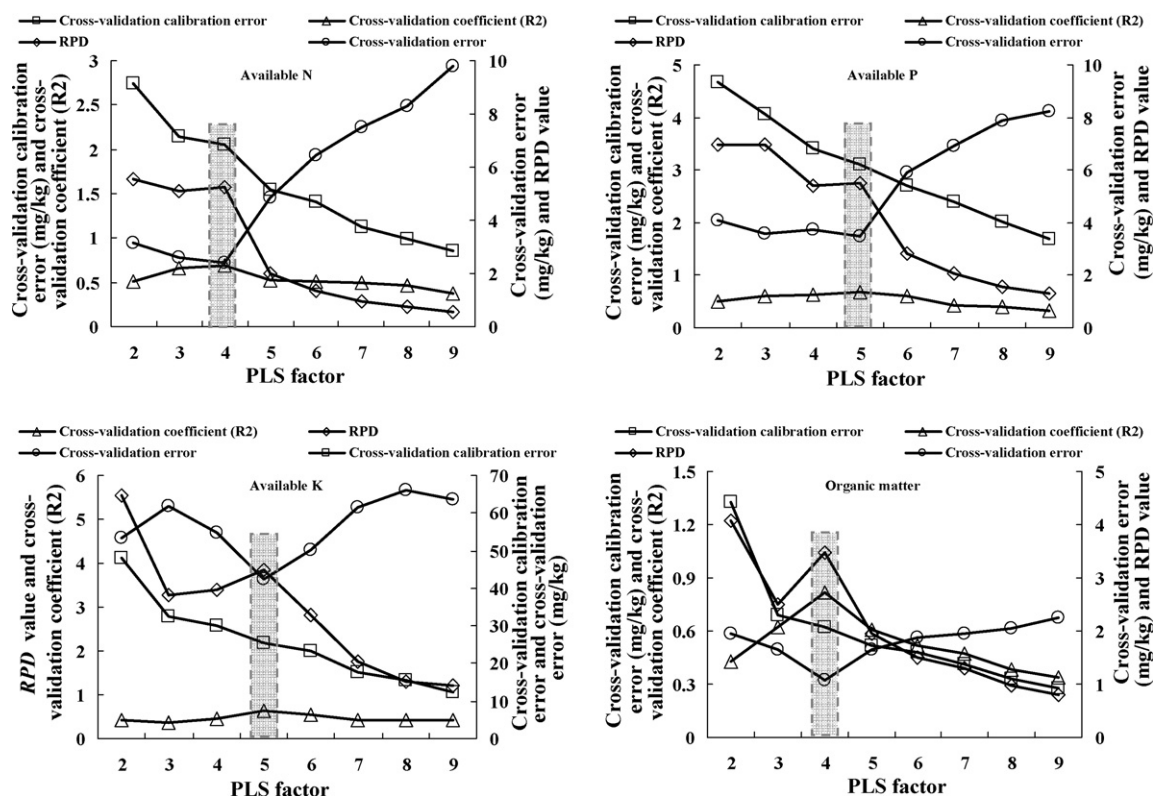


Fig. 3. Optimization of PLS factor number using leave-one-out cross-validation calibration ($n = 42$).

3.2. Optimization of PLS factor

Since soil was a complex mixture, and soil nutrient content was related with many soil components, multivariate calibration techniques were used to extract related information in the FTIR-PAS spectra. PLS regression was used to develop a correlation between the PAS spectra and the soil nutrient content in the soil samples. Fig. 3 demonstrated results of the leave-one-out cross-validation calibration using different PLS factor number varied from 2 to 9. For each soil property, the calibration error kept decreasing to near zero, and the calibration coefficient kept increasing to near 1. However, the validation error became smaller firstly, then turned larger and larger, which meant that modeling involved too many PLS factors would lead to over-fitting. The PLS factor number could be selected with lowest validation error where the calibration error and calibration coefficient were still good enough. 4 PLS factors, which explained 96.98% variance of input vectors and output vectors, were selected for available N and organic matter, and the RPD values were 5.27 and 3.48, respectively; 5 PLS factors which explained 98.58% variance of input vectors and output vectors, were selected for available P and available K, and the RPD values were 5.51 and 3.85, respectively.

RPD is an important statistic parameter used to evaluate the calibration models. In agricultural application, $RPD > 3$ was considered acceptable and $RPD > 5$ excellent [23]. However, there is no critical level of RPD for the infrared analysis in soil science, and acceptable values depend on the intended application of the predicted values. 3 categories based on RPD in the ranges >2 , 1.4–2.0, and <1.4 were used to indicate decreasing reliability of predicting [20]. Dunn et al. [24] and Adam et al. [25] reported the similar results of suitable limits for RPD: <1.6 , poor; 1.6–2.0, acceptable; and >2.0 , excellent. In this study the RPD values in optimized PLS models were acceptable, and comprehensively, the PLS models were excellent for soil organic matter, soil available N and P, and they were relative poor for soil available K. Comparing

with research results of Adam et al. [25] and [20], the predicting ability of soil properties using photoacoustic spectra-based PLS modeling was better enough comparing with reflectance spectra both in near-infrared and mid-infrared region.

3.3. PLS modelization of soil properties

With optimized PLS models a validation was conducted for each soil properties to further verify the model's prediction capability. Fig. 4 showed the relationship between experimental values and predicted values from the optimized PLS models. The best prediction was organic matter, then the available N and P, and last the prediction of available K. Because soil organic matter was sensitive to photoacoustic absorption [6], superior performance of PLS model for soil organic matter was observed. Since soil available N as well as soil available P was closely related with soil organic matter, a good PLS modeling for soil available N and P was then realized, while the prediction of soil available K was not good. Soil available K was not sensitive to photoacoustic absorption, but soil clay minerals were sensitive, and K could be absorbed, fixed or released from the 2:1 soil clay mineral, such as Bentonite, thus the prediction capability of soil available K might because of the relationship of K vs. some 2:1 clay minerals. It was thereof deduced the prediction capability might be better for the soil mainly composed of 2:1 clay mineral, which need further investigation.

Table 2 showed that the results of validation based on cross-validation calibration were better than that of cross-validation, which further confirmed that the FTIR-PAS technique was very suitable in the quantitative analysis of soil.

Soil background was very complicated, and for different soil type the background was completely incomparable. Therefore, different soil types involved prediction should use a huge calibration set which included enough variances as well as less interferences, otherwise the prediction might be very poor. A reasonable method was to select the same soil type both in the

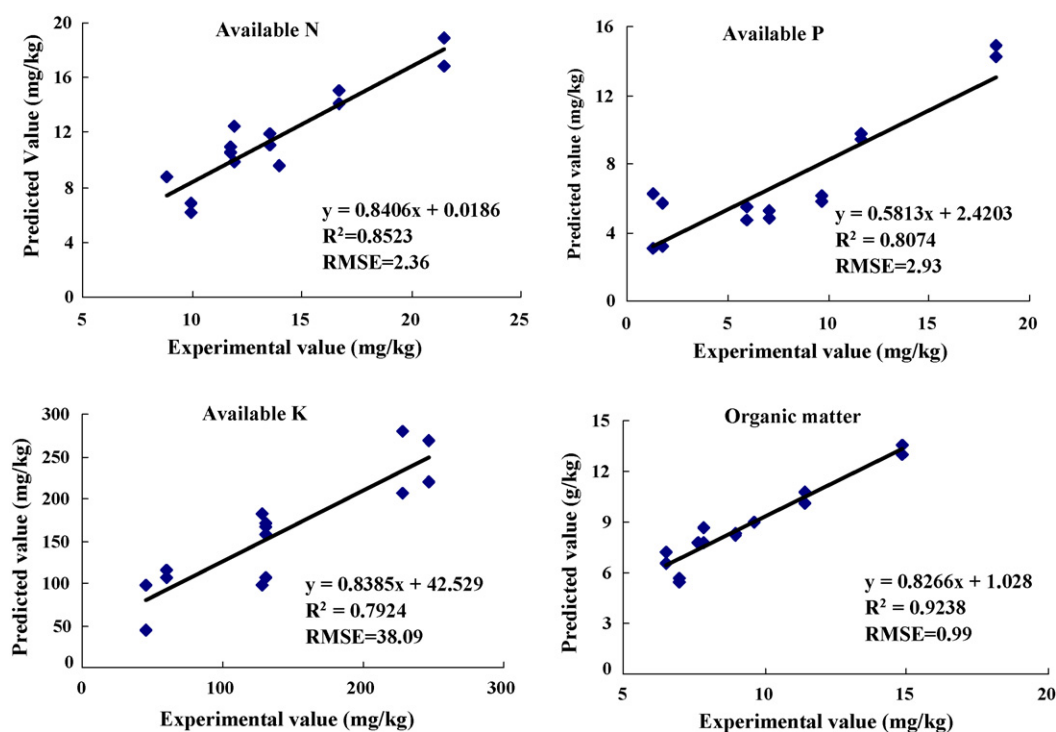


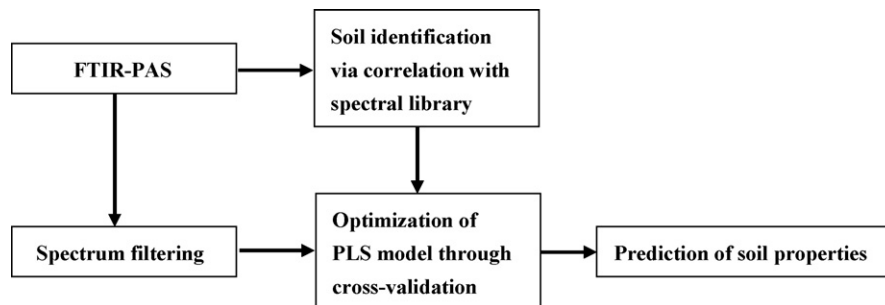
Fig. 4. Validation results using PLS model: plotting of experimental value vs. predicted value ($n = 14$). The optimized PLS factor number: 4 PLS factors for soil organic matter and available N; 5 PLS factors for soil available P and available K (R^2 : correlation coefficient; RMSE: root mean square error).

Table 2

Comparison of leave-one-out cross-validation and validation based on the leave-one-out cross-validation calibration

	Available N	Available P	Available K	Organic matter
Leave-one-out cross-validation ($n = 42$)				
Validation coefficient (R^2)	0.6948	0.6784	0.6455	0.8157
Validation error (RMSE) (mg/kg)	2.41	3.46	42.38	1.07
Validation ($n = 14$)				
Validation coefficient (R^2)	0.8523	0.8074	0.7924	0.9238
Validation error (RMSE) (mg/kg)	2.36	2.93	38.09	0.99

Note: for leave-one-out cross-validation, the optimized PLS factor was 4 for soil organic matter and available N, and 5 for available P and Available K, which was used for the separate validation.

**Fig. 5.** Schematic presentation of the analysis procedure for the prediction of soil properties using FTIR-PAS spectrum.

calibration and prediction, which was why the homogeneous soil set was used in this research. For the prediction of different soil type, the soil identification was recommended [16,26], then the same soil type based models were selected and optimized to do the counterpart prediction, and the procedure was shown in Fig. 5. Before conducting this prediction it was very important to establish a FTIR-PAS spectra library (standard spectra), and the library would contain useful spectroscopic information of soil, which would provide alternative technique in soil analysis.

4. Conclusions

A rapid technique, based on FTIR-PAS spectra and PLS modeling has been developed for the analysis of available N, P, K and organic matter content in Fluvo-aquic soil. The optimized PLS models perform well in the wavenumber region of $4000\text{--}500\text{ cm}^{-1}$, and are especially excellent to the prediction of organic matter, available N and available P in soil. However, they are not good enough for the prediction of available K of soil, and spectra pre-treatment should be made in further research.

The pronounced advantages of the PAS approach above the conventional chemical methods are its unique ability to directly predict the soil properties; no pre-treatment of soil sample is needed, therefore this technique is non-destructive to the soil sample, and the total time required for analysis is less than 2 min per sample, which shows a potential in the fast evaluation of soil fertility and soil quality.

Acknowledgements

This work was funded by National “973” project (No. 2005CB121102), and National “863” project (No. 2006AA10A301).

We thank Dr. Raphael Linker and Prof. Avi Shaviv for their kind help in the measurements of FTIR-PAS spectra, and we are also grateful to the revising suggestions from anonymous reviewers.

References

- [1] R.A. Viscarra, A.B. McBratney, *Aust. J. Exp. Agric.* 38 (1998) 765.
- [2] A.B. McBratney, B. Minasny, R.A. Viscarra, *Geoderma* 136 (2006) 272.
- [3] R.A. Viscarra, D.J. Walvoort, A.B. McBratney, L.J. Janick, J.O. Skjemstad, *Geoderma* 131 (2006) 59.
- [4] L.J. Janik, R.H. Merry, J.O. Skjemstad, *Aust. J. Soil Res.* 38 (1998) 681.
- [5] T.T. Nguyen, L.J. Janik, M. Raupach, *J. Soil Res.* 29 (1991) 29.
- [6] C. Du, R. Linker, A. Shaviv, *Appl. Spectrosc.* 61 (2007) 1063.
- [7] D.H. McQueen, R. Wilson, T. Kinnunen, *Trend Anal. Chem.* 14 (1995) 482.
- [8] W.C. Wahls, E. Ketta, J.C. Leyte, *Appl. Spectrosc.* 54 (2000) 214.
- [9] V.G. Gregoriou, E.R. Sheila, *Anal. Chem.* 74 (2002) 2361.
- [10] S. Bjarnestad, O. Dahlman, *Anal. Chem.* 74 (2002) 5851.
- [11] N.K. Bhardwaj, V.Q. Dang, K.L. Nguyen, *Anal. Chem.* 78 (2006) 6818.
- [12] S. Sivakesava, J. Irudayaraj, *J. Sci. Food Agric.* 80 (2000) 1805.
- [13] H. Yang, J. Irudayaraj, *Lebensm.-Wiss. u-Technol.* 34 (2001) 402.
- [14] T. Schmid, C. Helmbrecht, U. Panne, C. Haisch, R. Niessner, *Anal. Bioanal. Chem.* 375 (2003) 1124.
- [15] J. Pan, K.L. Nguyen, *Anal. Chem.* 79 (2007) 2259.
- [16] C. Du, R. Linker, A. Shaviv, *Geoderma* 143 (2008) 85.
- [17] R. Lu, *Agrochemical Analysis Methods in Soil Science*, China Agricultural Publisher, Beijing, 1999.
- [18] A. Savitzky, M.J.E. Golay, *Anal. Chem.* 36 (1964) 1627.
- [19] P. Geladi, B.R. Kowalski, *Anal. Chim. Acta* 185 (1986) 1.
- [20] C. Chang, D.A. Laird, M.J. Mausbach, C.R. Hurburgh, *Soil Sci. Soc. Am. J.* 65 (2001) 480.
- [21] C. Du, J. Zhou, *Chin. J. Anal. Chem.* 35 (2007) 119.
- [22] C. Du, J. Zhou, H. Wang, J. Zhang, A. Zhu, *Spectroscopy and Spectral Analysis*, 28 (2008), in press.
- [23] D.F. Malley, L. Yesmin, D. Wray, S. Edwards, *Commun. Soil Sci. Plant Anal.* 30 (1999) 999.
- [24] B.W. Dunn, H.G. Beecher, G.D. Batten, S. Ciavarella, *Aust. J. Exp. Agric.* 42 (2002) 607.
- [25] P. Adam, S. Balwant, I. Kamrunnahar, *Aust. J. Soil Res.* 43 (2005) 713.
- [26] R. Linker, I. Shmulevich, A. Kenny, A. Shaviv, *Chemosphere* 61 (2005) 652.