

## QUANTITATIVE ANALYSIS OF TOTAL NITROGEN CONTENT IN MONOAMMONIUM PHOSPHATE FERTILIZER USING VISIBLE-NEAR INFRARED SPECTROSCOPY AND LEAST SQUARES SUPPORT VECTOR MACHINE

L. S. Wang<sup>1\*</sup>, R. J. Wang<sup>1</sup>, C. P. Lu<sup>1</sup>, J. Wang<sup>1</sup>, W. Huang<sup>1</sup>, Q. Jian<sup>1</sup>,  
Y. B. Wang<sup>1</sup>, L. Z. Lin<sup>2</sup>, L. T. Song<sup>1</sup>

<sup>1</sup> Institute of Intelligent Machines, Chinese Academy of Sciences,  
Hefei 230031, China; e-mail: lswang@iim.ac.cn

<sup>2</sup> Electronic Engineering Institute, Hefei 230037, China

*A quantitative analysis method to determine the total nitrogen content in monoammonium phosphate (MAP) fertilizer using visible-near infrared (Vis-NIR) spectroscopy and least squares support vector machine (LS-SVM) is proposed. Sample set partitioning based on the joint x-y distance (SPXY) was used to select the calibration set. Fourteen spectral pre-processing methods were then employed to deal with the spectral data including Savitzky–Golay (SG) smoothing, first derivative ( $D^1$ ) and second derivative ( $D^2$ ) with SG smoothing, multiplicative scatter correction (MSC), standard normal variate (SNV), wavelet, and combination thereof. Next, the LS-SVM model with radial basis function kernel was established with the best pre-processing method, and its performance was compared with that of partial least squares (PLS) model. The results revealed LS-SVM calibration with the discrete wavelet transform provided the best prediction for total nitrogen content in MAP fertilizer, yielding  $R^2$ , root mean square error of prediction (RMSEP), and ratio of performance to deviation (RPD) values of 0.91, 0.101, and 3.34, respectively.*

**Keywords:** Vis-NIR spectroscopy, monoammonium phosphate fertilizer, total nitrogen content, pre-processing, least squares support vector machine.

## КОЛИЧЕСТВЕННЫЙ АНАЛИЗ СОДЕРЖАНИЯ АЗОТА В МОНОАММОНИЙФОСФАТНОМ УДОБРЕНИИ С ИСПОЛЬЗОВАНИЕМ ВИДИМОЙ И БЛИЖНЕЙ ИК СПЕКТРОСКОПИИ И МЕТОДА ОПОРНЫХ ВЕКТОРОВ И НАИМЕНЬШИХ КВАДРАТОВ

L. S. Wang<sup>1\*</sup>, R. J. Wang<sup>1</sup>, C. P. Lu<sup>1</sup>, J. Wang<sup>1</sup>, W. Huang<sup>1</sup>, Q. Jian<sup>1</sup>,  
Y. B. Wang<sup>1</sup>, L. Z. Lin<sup>2</sup>, L. T. Song<sup>1</sup>

УДК 543.42.062

<sup>1</sup> Институт интеллектуальных машин Китайской АН,  
Хэфэй, 230031, Китай; e-mail: lswang@iim.ac.cn

<sup>2</sup> Институт электроники, Хэфэй 230037, Китай

(Поступила 15 марта 2018)

*Предложен метод количественного определения содержания азота в моноаммонийфосфатных (MAP) удобрениях с использованием спектроскопии видимого–ближнего ИК диапазона (Vis-NIR) и метода опорных векторов и наименьших квадратов (LS-SVM). Для выделения калибровочного набора использовано разделение выборки по расстоянию между осями x-y (SPXY). Для обработки спектральных данных взяты 14 методов, в том числе сглаживание Савицкого–Голя (SG), нахождение первой ( $D^1$ ) и второй производных ( $D^2$ ) со сглаживанием SG, коррекция мультипликативного рассеяния (MSC), стандартная нормальная переменная (SNV), вейвлет-преобразование, а также и их комбинация. Разработана модель LS-SVM с радиальной базисной функцией, отвечающей наилучшему методу предварительной обработки. Ее производительность сравнивалась с моделью частичных наименьших квадратов (PLS). Калибровка LS-SVM с дискретным вейвлет-преобразованием показала наилучший прогноз для содержания азота в MAP-удобрении с  $R^2 = 0.91$ , среднеквадратичной ошибкой прогноза RMSEP = 0.101 и отношением производительности к отклонению RPD = 3.34.*

**Ключевые слова:** спектроскопия видимой–ближней ИК области, моноаммонийфосфатное удобрение, содержание азота, предварительная обработка, метод наименьших квадратов и опорных векторов.

**Introduction.** Monoammonium phosphate (MAP) has long been an important phosphate fertilizer for farmers worldwide. This fertilizer is water-soluble and dissolves rapidly in adequately moist soil. Upon dissolution, the two basic components of the fertilizer separate to release ammonium ( $\text{NH}_4^+$ ) and phosphate ( $\text{H}_2\text{PO}_4^-$ ), both of which are required by plants for healthy, sustained growth. One of the most important aspects in the production of MAP fertilizer is quality control, which is traditionally conducted using samples collected in the production line and then analyzed in the laboratory. However, traditional methods of analysis are tedious and time consuming and therefore do not meet the needs for quick analysis in MAP production.

Near infrared spectroscopy (NIRS) is a powerful and versatile tool used in a wide variety of fields because it is nondestructive, requires little or no sample preparation, and is suitable for on/in-line monitoring. Accordingly, it has become one of the most widely applied analytical methods for determination of critical attributes closely related to product quality in the agricultural [1], food production [2], petrochemical [3], and pharmaceutical industries [4]. A few studies have investigated fertilizer quality detection using the NIRS technique. For example, a method for determination of moisture content in potash fertilizer using NIRS was proposed by H. Farajiet et al. [5], in which a predictive model was established using three wavelengths. Additionally, a predictive model of total nitrogen, phosphate, and stone powder that was established by Guo Zheng has since been accepted for *in situ* quality control in the production of compound fertilizer [6]. Song Le et al. [7] proposed an approach to rapid analysis of the content of nutrients in compound fertilizer by NIRS with the partial least squares (PLS) model, in which the PLS model was validated based on six kinds of compound fertilizers available in the market. J. A. Janse Van Vuuren et al. [8] developed a method to determine the composition of bulk blend fertilizers using NIRS for quality control, which was shown to be acceptable for measuring the content of nitrogen, phosphorus, potassium, and sulfur.

The present study was conducted to: (1) investigate the feasibility of using visible-near infrared (Vis-NIR) spectroscopy to predict total nitrogen content in MAP fertilizer, and (2) obtain the optimal calibration model based on comparison of pre-processing methods and calibration methods. To the best of our knowledge, this is the first attempt to develop Vis-NIR calibration for determination of total nitrogen content in MAP fertilizer. The results presented herein will be very useful to MAP quality control.

**Experimental.** A total of MAP fertilizer samples was collected from the production line of the Feidong phosphate fertilizer plant, Anhui, China.

Visible-near infrared spectra were measured using a Vis-NIR reflectance sensor manufactured by Veris Technology Incorporation. The Vis-NIR is based on an optical shoe that has a nitrite-hardened wear plate containing a sapphire window at the bottom. A tungsten halogen bulb in the system illuminates the MAP fertilizer samples through the window, and an optic directs reflected light into a fiber-optic cable for transmission to two separate spectrometers. One spectrometer (Ocean Optics USB4000) uses a silicon CCD to measure light intensity from 350 to 1050 nm at a resolution of 2.3 nm, while the second (Hamamatsu C9914GB) uses an InGaAs detector to measure light intensity from 900 to 2200 nm at a resolution of 8 nm. Internal shutters automatically actuate every 15 min to collect dark and reference spectra, which is critical to ensuring data quality. Each sample was measured three times after being rotated by  $120^\circ$  in between each scan. Following each sample measurement, the obtained spectra were averaged.

Optimal data division into a training dataset and an independent test subset is essential to near infrared qualitative and quantitative modeling for spectral data analysis. There are several existing data division methods for the selection of training and test datasets, including random selection (RS) [9], Kennard Stone (KS) [10], and sample set partitioning based on the joint  $x$ - $y$  distance (SPXY) [11]. Random sampling (RS) is a popular technique because of its simplicity and because a group of data randomly extracted from a larger set follows the statistical distribution of the entire set. However, RS does not guarantee the representativity of the set, nor does it prevent extrapolation problems [12]. An alternative to RS that is often employed is the Kennard–Stone (KS) algorithm. The KS method is designed to cover multidimensional spaces in a uniform manner by maximizing the Euclidean distances between the instrumental response vectors ( $x$ ) of the selected samples. However, the KS method is limited in that the statistics of the dependent variable ( $y$ ) are not taken into account. Sample set partitioning based on the joint  $x$ - $y$  distance (SPXY) is an extension of the KS algorithm that is employed to select samples according to differences in both their  $x$  (instrumental responses) and  $y$  (predicted parameter) spaces. In this study, SPXY was used to divide the sample set into a calibration and validation sets.

Visible-near infrared spectroscopy is subject to influences of overlapping spectral responses of sample constituents, as well as sources of error including instrumental noise and drift, light scattering, and path length variations that occur during measurements. However, these effects can largely be eliminated by applying suitable preprocessing. Accordingly, various spectral pre-processing algorithms have been developed to accomplish this, such as Savitzky–Golay smoothing (SG) [13], multiplicative signal correction (MSC) [14], standard normal variate (SNV) [15], and wavelet algorithms [16] and derivatives.

Preprocessing was conducted using the SG, first-order derivative ( $D^1$ ), second-order derivatives ( $D^2$ ), MSC, SNV, and wavelet methods and combinations thereof. The SG filter can digitally smooth a given spectrum by approximating it with a specified data window using a specified order polynomial. The SG derivative was applied to reduce baseline offset and noise and increase spectral resolution [17]. The first derivative was applied to remove additive baseline effects, and the second was used to remove sloped additive baselines [18]. The MSC was utilized to remove the baseline drift from spectra caused by scattering and variations in particle sizes and optical length variables. The SNV was used to effectively remove the multiplicative interference caused by scatter and particle sizes. The discrete wavelet transform (DWT) was used for spectral smoothing and noise removal. The most common use of wavelets in spectral analysis is Daubechies (dbN) wavelets, which possess orthogonality and are discrete. Optimal selection of Daubechies wavelets leads to great denoising of spectral signals.

In this study, a filter with a polynomial of order 2 and a window with a width of 31 data points were used. Daubechies wavelets of vanishing moment 6 at decomposition level 8 (db6\_8) was applied. The effects of different preprocessing methods and raw data on PLS models were then compared. All PLS regression models were constructed by leave-one-out cross validation.

Partial least squares (PLS) regression [14] is a highly mature and widespread linear multivariate calibration technique used for quantitative spectroscopic analyses in many fields that is based on a latent variable decomposition of the  $X$  (spectral data) and  $Y$  (references) matrices. A relationship between these two blocks is described by PLS analysis through an inner-relationship of their scores, maximizing the covariance between these inner variables.

The least squares support vector machine (LS-SVM) employs a statistical learning method that has a self-basis of statistical-learning theory and excellent learning performance [19]. When using LS-SVM, there are two crucial problems that must be solved: identification of proper kernel function and the best kernel parameters. For LS-SVM, there are many kernel functions (linear kernel, polynomial kernel, radial basis function (RBF) kernel, spline kernel, bspline kernel, sigmoid kernel, etc.). However, no systematic methodology is available for prior selection of kernel function. In this study, the RBF kernel was used as a training algorithm because it is a nonlinear function and a more compact supported kernel that could reduce the computational complexity of the training procedure while providing good performance.

It is very important to conduct a careful model selection of the tuning parameters to achieve a good LS-SVM regression model. The LS-SVM with RBF kernel algorithm requires two parameters for tuning: the regularization parameter  $\gamma$  and the RBF kernel function parameter  $\sigma^2$ . The  $\gamma$  determines the trade-off between the training error minimization and smoothness [20], while  $\sigma^2$  is the squared bandwidth of the Gaussian curve. To tune these parameters, leave-one-out cross validation was used to select the initial random parameters to be optimized by the standard simplex method [19].

The optimum LS-SVM model was built with the best pre-processing method to predict the total nitrogen content in MAP fertilizer. The performance of the LS-SVM and the PLS model with the best preprocessing method were then compared. Model performance was evaluated by the coefficient of determination ( $R^2$ ), root mean square error of the prediction (RMSEP), and residual predictive deviation (RPD) [21]. Generally, a good model would have high values of  $R^2$  and RPD, and low value of RMSEP for independent validation.

**Results and discussion.** *Calibration and validation sets.* The MAP fertilizer samples were divided into a calibration set and a validation set based on SPXY. The results of the descriptive statistical analysis for the 46 samples after SPXY are shown in Table 1.

TABLE 1. Statistical Properties of Total Nitrogen Content in Monoammonium Phosphate

Sample set	Maximum, %	Minimum, %	Mean, %	Standard deviation
Calibration set (34)	10.96	9.52	10.39	0.36
Validation set (12)	10.81	9.71	10.36	0.34

*Spectral data preprocessing.* Preprocessing is necessary to minimize the effects of unnecessary information and enhance the useful chemical information in model development and improvement [22]. In this study, 14 different preprocessing methods were applied to the spectra. The PLS models developed using different preprocessing methods for predicting total nitrogen content in MAP fertilizers are summarized in Table 2. Comparison of these preprocessing methods revealed that the db6\_8 wavelet produced the best prediction results; therefore, the db6\_8 wavelet was used to develop the predictive model in the following analysis. The number of latent variables ( $LV_s$ ) is a critical parameter influencing the performance of the PLS model that should be optimized by cross validation of the calibration set. The optimal number of latent variables is determined by the lowest RMSECV value. The number of latent variables with different preprocessing methods are listed in Table 2.

TABLE 2. Prediction Results of PLS Models Prepared Using Different Pre-processing Methods

Data	$LV_s$	Calibration set (34)			Validation set (12)		
		$R^2$	RMSECV	RPD	$R^2$	RMSEP	RPD
Original data	8	0.85	0.136	2.65	0.89	0.105	3.24
SG	8	0.84	0.14	2.57	0.68	0.184	1.85
SG D <sup>1</sup>	7	0.88	0.124	2.9	0.70	0.176	1.93
SG D <sup>2</sup>	5	0.830	0.147	2.45	0.54	0.219	1.55
MSC	9	0.90	0.112	3.21	0.84	0.131	2.6
SNV	7	0.85	0.135	2.70	0.88	0.113	3.0
db6_8	8	0.85	0.135	2.70	0.90	0.105	3.24
MSC+SG	9	0.89	0.115	3.13	0.36	0.258	1.32
MSC+db6_8	7	0.90	0.111	3.24	0.84	0.130	2.62
SNV+SG	9	0.89	0.115	3.13	0.38	0.255	1.33
SNV+db6_8	7	0.86	0.133	2.71	0.88	0.111	3.06
MSC+SG D <sup>1</sup>	7	0.88	0.121	2.98	0.70	0.178	1.91
MSC+SG D <sup>2</sup>	6	0.88	0.123	2.93	0.70	0.176	1.93
SNV+SG D <sup>1</sup>	7	0.88	0.122	2.95	0.72	0.173	1.97
SNV+SG D <sup>2</sup>	6	0.88	0.124	2.90	0.73	0.169	2.0

*LS-SVM model.* Overall, 34 samples of MAP fertilizer were selected as a calibration set to establish an LS-SVM model for total nitrogen content, and 12 samples as a validation set were predicted by the LS-SVM model. For tuning, the regularization parameter  $\gamma = 3.23 \times 10^6$  and  $\sigma^2 = 9.11 \times 10^5$  were used to develop LS-SVM calibration model for prediction of the total nitrogen content. The predicted versus reference values for this calibration model in validation set are presented in Fig. 1, and a comparison of the results of the LS-SVM and PLS model with the best pre-processing method is summarized in Table 3. The performance of the LS-SVM model was found to be better than that of the PLS. The reason might be that the LS-SVM method can deal with nonlinear problems in spectra data.

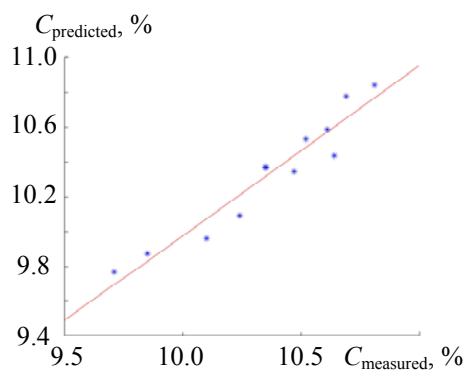


Fig. 1. Scatter plot of measured versus predicted total nitrogen content in validation set. The solid line represents the linear regression line between the predicted and measured values.

TABLE 3. Model Results for the Least Squares Support Vector Machine (LS-SVM) and Partial Least Squares (PLS) Regression Methods Applied for Calibration and Validation Sets for the Prediction of Total Nitrogen in Monoammonium Phosphate Fertilizer

Method	Calibration			Validation		
	$R^2$	RMSEC	RPD	$R^2$	RMSEP	RPD
LS-SVM	0.93	0.095	3.78	0.91	0.101	3.34
PLSR	0.85	0.135	2.70	0.90	0.105	3.24

**Conclusion.** The applicability of the Vis-NIR spectroscopy and LS-SVM to quantify the total nitrogen content in monoammonium phosphate fertilizer was studied. Using 14 different preprocessing methods to deal with the MAP fertilizer spectra, the best performance was achieved by the discrete wavelet transform (db6\_8). Comparison of the LS-SVM and PLS methods revealed that the LS-SVM was more suitable for determination of the total nitrogen content in MAP fertilizer. The results of this study illustrate the feasibility of applying Vis-NIR spectroscopy combined with LS-SVM to quantitatively determine the total nitrogen content in MAP fertilizer.

To further verify the applicability of the proposed method, more samples should be utilized to build a more robust model. Our future work will be directed toward the detection of phosphate and water content in MAP fertilizer.

**Acknowledgment.** This work was supported by the Science and Technology Service Network of the Chinese Academy of Sciences (No. KFJ-SW-ST5-144), the Key R & D project in the Ningxia Hui Autonomous Region (No. XD-XMBW-2016001), and the National Natural Science Foundation of China (No. 31671586).

## REFERENCES

1. *American Association of Cereal Chemists*, Eds. P. C. Williams, K. H. Norris, St. Paul, Minnesota (2001).
2. J. U. Porep, D. R. Kammerer, R. Carle, *Trends Food Sci. Technol.*, **46**, 211–230 (2015).
3. W. I. Friesen, *Appl. Spectrosc.*, **50**, 1535–1540 (1996).
4. Y. Roggo, P. Chaluz, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent, *J. Pharm. Biomed. Anal.*, **44**, 683–700 (2007).
5. H. Faraji, T. Crowe, R. Besant, S. Sokhansanj, H. Wood, *Can. Biosys. Eng.*, **46**, 45–48 (2004).
6. Z. Guo, H. F. Yuan, X. Zhang, C. F. Song, X. Y. Li, J. C. Xie, *Spectrosc. Spectr. Anal.*, **31**, 1521–1523 (2011).
7. L. Song, H. Zhang, X. Y. Ni, L. Wu, B. M. Liu, L. X. Yu, Q. Wang, Y. J. Wu, *Spectrosc. Spectr. Anal.*, **34**, 73–77 (2014).
8. J. A. Janse Van Vuuren, C. A. Groenewald, *Commun. Soil Sci. Plant Anal.*, **44**, 120–135 (2013).
9. M. Daszykowski, B. Walczak, D. L. Massart, *Anal. Chim. Acta*, **468**, 91–103 (2002).
10. W. Wu, B. Walczak, D. L. Massart, S. Heuerding, F. Erni, I. R. Last, K. A. Prebble, *Chemom. Intell. Lab. Syst.*, **33**, 35–46 (1996).
11. R. K. Galvão, M. C. Araujo, G. E. José, M. J. Pontes, *Talanta*, **67**, 736–740 (2005).
12. K. R. Kanduc, J. Zupan, N. Majcen, *Chemom. Intell. Lab. Syst.*, **65**, 221–229 (2003).
13. A. Savitzky, M. J. E. Golay, *Anal. Chem.*, **36**, 1627–1639 (1964).
14. P. Geladi, B. R. Kowalski, *Anal. Chim. Acta*, **185**, 1–17 (1986).
15. R. J. Barnes, M. S. Dhanoa, S. J. Lister, *Appl. Spectrosc.*, **43**, 772–777 (1989).
16. G. A. Blackburn, J. G. Ferwerda, *Remote Sens. Environ.*, **112**, 1614–1632 (2008).
17. H. Cen, Y. He, *Trends Food Sci. Technol.*, **18**, 72–83 (2007).
18. H. Swierenga, A. P. De Weijer, R. J. Van Wijk, L. M. C. Buydens, *Chemom. Intell. Lab. Syst.*, **49**, 1–17 (1999).
19. J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, *Least Squares Support Vector Machines*, World Scientific Publishing Co., Pte, Ltd., Singapore (2002).
20. J. De Brabanter, K. Karsmakers, P. Ojeda, F. Alzate, C. De Brabanter, J. Pelckmans, *ESAT-SISTA Technical Report*, 10–146 (2011).
21. A. M. Mouazen, J. D. Baerdemaeker, H. Ramon, *J. Near Infrared Spectrosc.*, **14**, 189–199 (2006).
22. M. E. Lafargue, M. Feinberg, J. J. Daudin, D. N. Rutledge, *J. Near Infrared Spectrosc.*, **11**, 109–122 (2003).