



Increasing the accuracy of atmospheric ammonia concentrations calculated from Open-Path Fourier transform infrared spectra using partial least squares model by scanning and removing interference spectral data

Ming-Quan GUO^a, Qian-Wei SONG^b, Li-Min SHAO^{a,*}

^a Department of Chemistry, University of Science and Technology of China, Hefei, Anhui 230026, China

^b State Key Laboratory for Chemo/Biosensing and Chemometrics, Advanced Catalytic Engineering Research Center of the Ministry of Education and College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, China

ARTICLE INFO

Keywords:

PLSR
Atmospheric ammonia
Open-Path Fourier transform infrared
Variable selection

ABSTRACT

Monitoring atmospheric pollutants has raised worldwide concerns in last decades, due to deterioration of air quality. Multivariate regression methods have gained extensive applications in calculating and monitoring concentrations of air pollutants, for example ammonia (NH₃), using Open-Path Fourier Transform Infrared (OP/FT-IR) spectra data. However, the prediction accuracy of multivariate regression models is interfered heavily by the dominant and omnipresent absorption bands of atmospheric H₂O vapor and CO₂ in OP/FT-IR spectra. Hence, a new method of variable selection, referred to as window of scanning and removing interference information (WSRII), is developed to remove interference data of OP/FT-IR spectra. The key of WSRII is to confirm informative variables according to the change of root mean square error of calibration in the partial least squares regression (PLSR) model after removing a spectral window. If the change is greater than 0, the spectral window is reserved as an informative information window. Then, the new matrixes of spectral data are reconstructed by using the informative information windows, which are selected by changing position of the spectral window in full wave number range. Based on this, a PLSR model is rebuilt to predict accurately NH₃ concentrations. The results showed that the proposed method was able to eliminate uninformative information and improved prediction accuracy of PLSR models. Moreover, this process of variable selection is significantly potential to improve prediction accuracy of PLSR model for monitoring atmospheric NH₃ concentrations in real time.

1. Introduction

The pollutants in the atmosphere, such as ammonia (NH₃), methane (CH₄) and fine particulate matter (PM_{2.5}), have raised worldwide concerns in last decades [1–3]. Many researchers reported that the air pollutants were related to increasing of lung cancer, chronic obstructive pulmonary disease and respiratory disease [4–6], which posed a serious threat to human health. NH₃ is the only alkaline molecule in the atmosphere, which is easy to react with acid group for example SO₄²⁻, to generate ammonium salt particles [7]. So, NH₃ in the atmosphere is commonly regarded as a precursor to form non-organic PM_{2.5}, which is able to cause severe hazard for respiratory system of human body [8]. A study showed airborne PM_{2.5} had a positive relationship with increasing respiratory or cardiovascular disease [9]. And then the deposited PM_{2.5} caused a series of environmental problems such as soil acidification and enhances eutrophication [10,11]. Thus, it is of great significance for hu-

man health to monitor and calculate NH₃ emission of atmosphere in real time.

Open-Path Fourier Transform Infrared (OP/FT-IR) spectrometry is widely used to monitor emissions of pollutant molecules, such as NH₃, CH₄ and nitrous oxide (N₂O), in the atmosphere [12–14]. The reports revealed that some atmospheric pollutant molecules, from various sources such as industrial plants and agricultural operations, were monitored successfully via OP/FT-IR spectrometry [15–17]. Due to the potential to extract meaningful chemical information, multivariate regression methods had been used for quantitative analysis of OP/FT-IR spectra data [18,19]. Therein, partial least squares regression (PLSR) is a very popular multivariate regression method, which was applied to multicomponent spectral analyses especially in vibrational spectroscopy, for instance infrared (IR) spectroscopy and Raman spectroscopy [20]. However, the interferences from other atmospheric molecules are always added to OP/FT-IR spectra, which seriously disturbs the stability and accuracy of PLSR models. For example, the dominant and omnipresent absorption bands of H₂O vapor in IR spectra and moment-changing tem-

* Corresponding author.

E-mail address: lshao@ustc.edu.cn (L.-M. SHAO).

<https://doi.org/10.1016/j.cjac.2022.100133>

Received 28 February 2022; Received in revised form 4 May 2022; Accepted 25 May 2022

Available online 30 May 2022

1872-2040/© 2022 Changchun Institute of Applied Chemistry, CAS. Published by Elsevier Ltd. All rights reserved.

perature make data analysis rather difficult [21]. Thus, the reducing interferences in OP/FT-IR spectroscopy is of the essence to rapid and accurate monitoring gaseous molecules in the atmosphere.

Variable (or wavelength) selection is a critical step to eliminate interferences in multivariate analysis for datasets with many variables especially when number of samples is much smaller than the number of variables [22]. This technique is always used for the researches of removing uninformative variables in near IR and Raman spectroscopies [23–26]. It always aims on three aspects: (1) to improve the prediction performance of a multivariate analysis model, (2) to build a faster model by reducing dimensionality of data, which can reduce training and utilization time, and provide more cost-effective predictors, and (3) to give a better understanding and interpretation for measured data [27,28]. Up to now, many variable selection methods have been employed to improve analytical performance in model of multivariate regression by reducing uninformative information, such as Monte Carlo uninformative variables elimination (MCUVE) [29], genetic algorithm (GA) [30], successive projections algorithm (SPA) [31], and iteratively retains informative variables (IRIV) [32]. However, these methods of variable selection are always complicated to perform and difficult to implement. Therefore, developing a simple and actualized variable selection method is essential for reducing interference information in OP/FT-IR spectra.

In this work, a new method of variable selection, referred to as window of scanning and removing interference information (WSRII), is proposed for reducing interference data of OP/FT-IR spectra to improve the prediction performance of PLSR model for NH_3 in the atmosphere. A calibration set containing 100 OP/FT-IR spectra was used for training PLSR model and a validation set containing 30 OP/FT-IR spectra was used for testing prediction performance of PLSR model. It was found that the proposed method could improve obviously the prediction accuracy of the PLSR model for NH_3 concentrations of atmosphere. However, the relative errors of prediction concentrations were large in low-concentration range, which was due to that PLSR model in single calibration set couldn't fit the over wide concentration range, ultimately resulting in a deviation of the Lambert-Beer's law. In this case, a data of single calibration set with over wide concentration range was divided into data of two calibration sets in low concentration range and high concentration range. Then, the two calibration sets were further used for evaluating performance of the proposed method and these results showed the WSRII could improve prediction accuracy of PLSR model. Therefore, the proposed method of variable selection is prospective to be used in rapid and accurate monitoring and calculating NH_3 concentrations of atmosphere in real time.

2. Measurement and data processing

2.1. Experiment

OP/FT-IR spectra were measured several years ago in southern Idaho with the Northwest Irrigation and Soil Research Laboratory of the United States Department of Agriculture, which was in a cooperative project for monitoring gaseous emissions around animal farms. The OP/FT-IR spectrometer was manufactured by MDA Corp. (Atlanta, GA) and incorporated a Bomem Michelson 100 interferometer, a 31.5 cm telescope, a cube-corner array retro-reflector, and a Sterling engine-cooled mercury cadmium telluride (MCT) detector. The interferograms were measured with a resolution of 1 cm^{-1} , then corrected for the non-linear response of the MCT detector. The spectra were computed by fast Fourier transform (FFT) from interferograms with a zero-filling factor of 8 and medium Norton-Beer (MNB) apodization function [33]. Spectral data in the region from 1250 to 750 cm^{-1} was used in predicting concentrations of NH_3 . The manipulation of spectra and data processing were done using MATLAB 7.10 (2010) (The Math Works Inc., Natick, MA) on the Windows 7 operating system.

2.2. Calibration set and validation set

The calibration set to train PLSR model contains 100 OP/FT-IR spectra data in the spectral range of $1250\text{--}750 \text{ cm}^{-1}$ and the concentrations of NH_3 are nearly evenly distributed over the range of 0–1400 ppm-m. The data in calibration set firstly was measured as interferograms and then the interferograms were calculated to absorption spectra by using Fourier transform (FT). Further, an iterative procedure was used to calculate accurately the concentrations of NH_3 as standard values (SV), which was described in reference [33]. Finally, the baseline of spectra data in the calibration set was adjusted using wavelet transform. The validation set to test PLSR model contains 30 OP/FT-IR spectra data and the concentrations of NH_3 are in the same range as the calibration set. The validation spectra were processed in the same way as calibration set. The original spectra of calibration and validation were plotted in Fig. S1. The reconstructed spectra of calibration and validation set after WSRII were shown in Fig. S2.

2.3. Theory and algorithm

The multivariate regression analysis of OP/FT-IR spectra is based on the Lambert-Beer's law. The indirect linear model in multicomponent spectral analysis is as follows:

$$C = X \cdot b + e \quad (1)$$

Where X is an m -by- n matrix of m spectra. m and n represent the numbers of samples and wavelength or wave number points, respectively. C is an m -by-1 vector of sample concentrations. Vectors e and b represent the error vector and the vector of regression coefficient with m elements, respectively. In order to solve Eq. (1) for vector b , PLSR is used. After vector b is obtained, a PLSR model is ready to calculate concentrations of unknown samples.

In a PLSR model, data matrix X is resolved with the assistance of C to loadings and scores matrices. The scores matrix T is used in regression that is illustrated as the following equations.

$$X = T \cdot P + E \quad (2)$$

$$C = T \cdot Q + F \quad (3)$$

Where P and T denote the loadings and scores matrices for X , respectively. Q refers to the loading vector for C . F and E are the residuals for C and X , respectively.

In the algorithm of WSRII, the process of leave-one-out cross validation in data of calibration set was firstly carried out to determine the factor number for PLSR model. Then, a PLSR model with a selected factor number was trained using data of calibration set (X and C) in the whole wave number range of $1250\text{--}750 \text{ cm}^{-1}$. The root mean square error of calibration (RMSEC) was calculated and denoted as R_1 . Further, a spectral window, which starts at the i -th wave number and ends at the $(i+h)$ -th wave number, is selected as the data cell for variable selection, where h is the size of spectral window. The matrix data of the selected spectral window was removed from matrix X and the residuary data of X was used for another PLSR model. The RMSEC was calculated and denoted as R_2 . On the previously selected factor number, the value of R_2 was compared with R_1 . When the values of R_2 is greater than R_1 , the selected spectral window is reserved as an informative window for PLSR model, otherwise, it is deleted (Fig. S3). In wave number range of $1250\text{--}750 \text{ cm}^{-1}$, the start point of wave number i of selected spectral window moved $h \text{ cm}^{-1}$ at a time and a loop was carried out to collect all informative information windows by comparing values of R_1 and R_2 . The new matrices of calibration and validation sets were reconstructed to use the spectral data of informative information windows. Further, a new PLSR model was built with the reconstructed calibration and used in calculating NH_3 concentrations of unknown samples. It is noteworthy that the

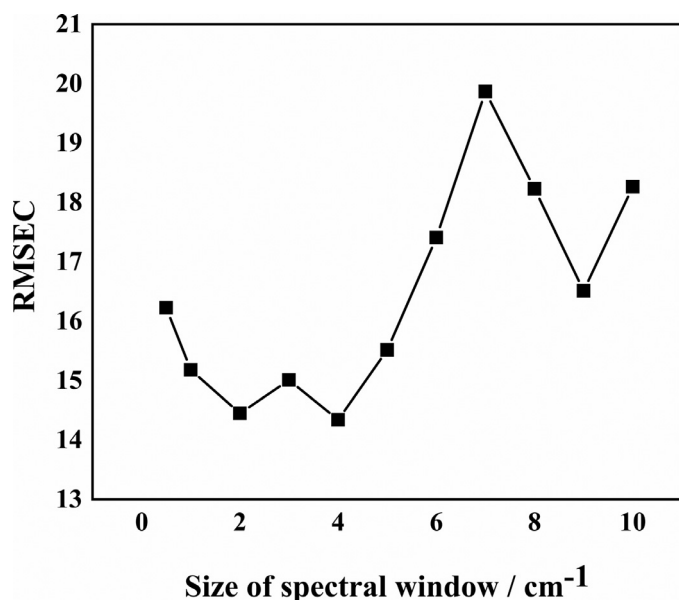


Fig. 1. RMSEC in different size of spectral window.

proposed method does not use extra parameters for variable selection and exclusively based on the calculated values of RMSEC to eliminate interference information. The proposed WSR II is readily comprehensible and easy to implement compared with most methods of variable selection.

3. Results and discussion

3.1. Investigation of size of spectral window

The effect from size of spectral window (or h) was studied firstly in this work. On the one hand, an overlarge size makes the process of variable selection very rough. It leads to misjudgments where some informative information is included in an uninformative window and eliminated. On the other hand, if the size of spectral window is small, the process of variable selection is time-consuming and complicated for training model. Therefore, an appropriate window size is essential to variable selection in this work. As shown in Fig. 1, the root mean square error of calibration (RMSEC) varies with size of spectral window. It can be found that the accuracy of prediction gets worse, if the size of spectral window is either smaller than 2 cm^{-1} or larger than 4 cm^{-1} . Therefore, the size of spectral window in this work was set as 2 cm^{-1} .

3.2. Criterion of eliminating uninformative variables

The criterion of eliminating uninformative variables is crucial in the process of variable selection. If the criterion is too rigid, some informative information might be removed improperly. However, if the criterion is over tolerant, many uninformative variables might be reserved and subsequently damage the stability and accuracy of the model. In this work, we put forward a criterion with definite physical significance for estimating informative information, which did not need to add an extra parameter in variable selection. As shown in Fig. S3, the R_1 (red line) was regarded as the threshold to variable selection. If value of R_2 was greater than R_1 , the error of prediction in PLSR model increases, which meant the removed window is informative for PLSR model. If value of R_2 was smaller than R_1 , the error of prediction in PLSR model decreased, which illustrated the removed window was uninformative and disturbed the prediction accuracy of PLSR model. So, in this process of variable selection, the proposed criterion is simple, comprehensible and possesses distinct significance.

Table 1

Predicted results of PLSR model and PLSR model after WSR II model of using measured OP/FT-IR Data.

	concentration range (ppm-m)	factor number	RMSEV (ppm-m)
PLSR model	0–1400	5	17.24
	0–600	10	9.64
	600–1400	4	
PLSR model after WSR II	0–1400	4	14.45
	0–600	8	6.43
	600–1400	3	

3.3. Comparison among informative window, IR of spectrum pure NH_3 and atmosphere without NH_3

To further illustrate the significance of the proposed process of variable selection, an OP/FT-IR spectrum of atmosphere without NH_3 , selected informative spectra windows and the IR spectrum of pure NH_3 were shown in the Fig. 2 as a comparison. In Fig. 2, to avoid too many data points affecting the effect of comparison, a continuous informative window was regarded as a dot. So, the x-coordinate of red dots represents the start wavenumber of the selected spectral windows, which contain a number of continuous wave numbers, and the y-coordinate of red dots is the RMSEC. The comparison between the IR spectrum of atmosphere without NH_3 (Fig. 2, black line) and the selected spectra windows (Fig. 2, red dot) demonstrates that some interference information was removed clearly from the spectrum, which also was seen in Fig. S2. The comparison between the selected spectra windows (Fig. 2, red dot) and the IR spectrum of pure NH_3 (Fig. 2, blue line) shows that the spectral features of NH_3 was reserved in informative window.

However, some wave numbers were reserved where water absorbs strongly such as in the range of $750\sim 850 \text{ cm}^{-1}$ and $1150\sim 1250 \text{ cm}^{-1}$. Such wave numbers were probably useful in the PLSR model to deal with interfering water absorbance. In systems of multiple variables, multi-variable analysis is solving several independent equations. Although one unknown variable (NH_3 concentration) is desired, the irrelevant information of NH_3 is necessary for equation set. This reveals that irrelevant information of NH_3 is not necessarily bad for the PLSR model. However, the irrelevant information of NH_3 should not be too much, otherwise it becomes a significant interference. Hence, in this process of variable selection, some irrelevant information was removed from spectrum to improve the prediction accuracy of PLSR model and the other irrelevant information was reserved because these data was useful to solve the equation set.

3.4. Building one model in 0–1400 ppm-m

In order to evaluate the performance of the proposed variable selection method, a calibration set and a validation set were used for building a PLSR model in the concentration range of 0–1400 ppm-m. Then, the root mean square error of validation (RMSEV) and relative errors of each sample were used to quantitatively assess the prediction performance of PLSR regression model.

In this work, the leave-one-out cross validation was firstly carried out in data of calibration set to estimate the factor number and the factor number was always determined at the smallest value of prediction residual error sum of squares (PRESS). Therefore, the factor numbers of the PLSR model and the one after WSR II were selected as 5 and 4 (Fig. 3A and Fig. 3B), respectively. Obviously, the factor number of the PLSR model after WSR II was smaller than the factor number of PLSR model, which was due to the removal of the uninformative information.

In Table 1, the RMSEV value of PLSR model lowered from 17.24 to 14.45 after WSR II. This results showed that the proposed method could reduce prediction errors for atmospheric NH_3 concentrations. As shown in Fig. 3C, the predicted concentrations of the PLSR model were

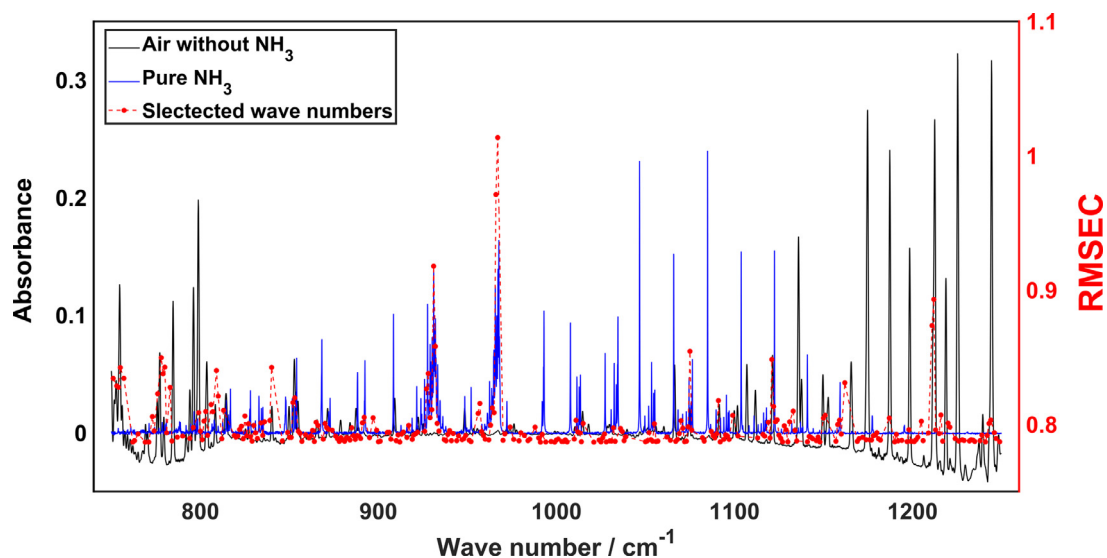


Fig. 2. Comparison of spectrum of atmosphere without NH_3 (black line), IR spectrum of pure NH_3 (blue line) and selected informative wave number (red dot).

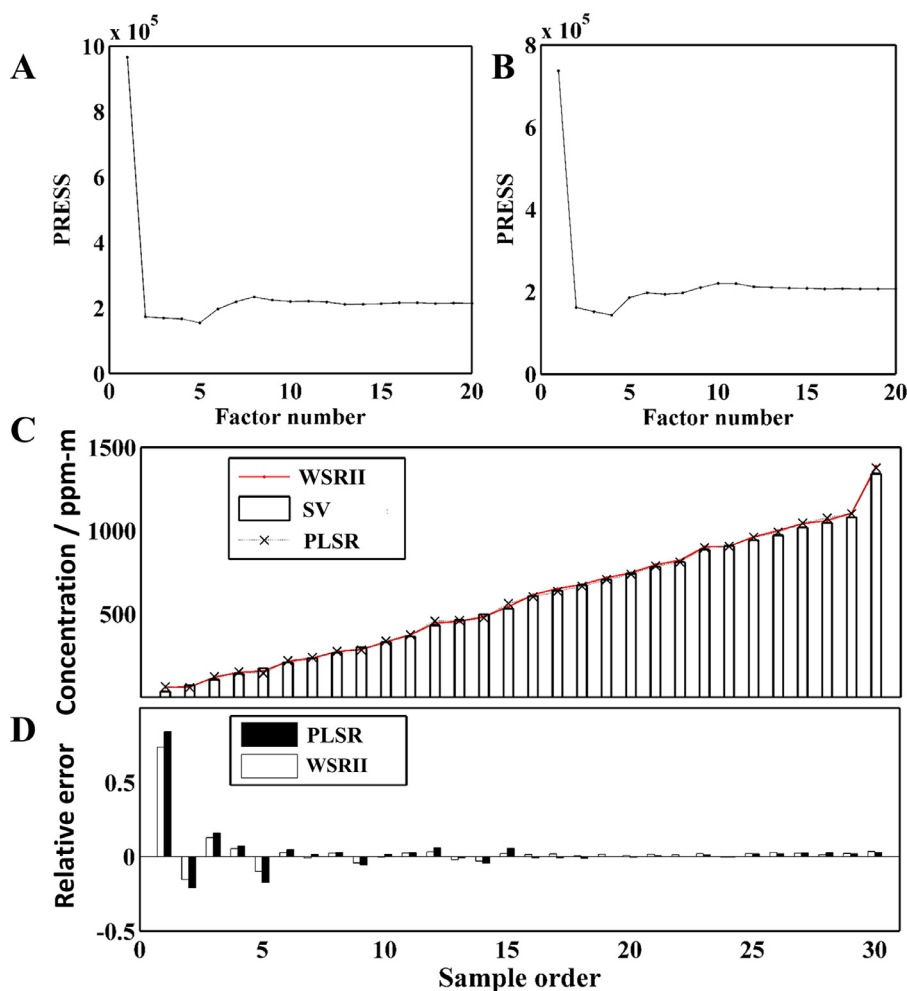


Fig. 3. (A) Results of cross validation of the PLSR model in 0–1400 ppm-m. (B) Results of cross validation of the PLSR model after WSRII in 0–1400 ppm-m. (C) Comparison of prediction results for validation set in PLSR model and the one after WSRII using one calibration set. (D) Relative errors of validation set in 0–1400 ppm-m.

very close to the standard values. The result illustrated that PLSR model possess good accuracy for the prediction of atmospheric NH_3 concentrations. In Fig. 3D, after WSRII, the most samples relative errors of predicted concentrations in the PLSR model were reduced. It illustrated that the PLSR model after WSRII had better prediction results. There-

fore, the proposed variable selection method has good performance to improve the accuracy of prediction.

However, the relative errors in low concentration range (0–600 ppm-m), so far as the maximum value exceeding 0.7, are obviously greater than relative errors in high concentration range (600–1400 ppm-m). It

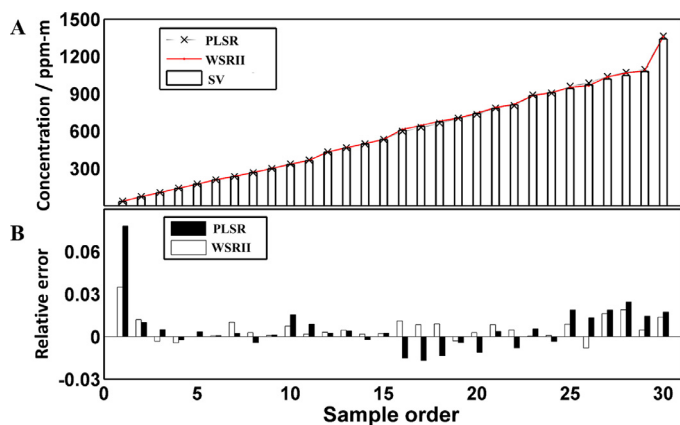


Fig. 4. (A) Comparison of prediction results and (B) relative errors for validation set in PLSR model and PLSR model after WSRII using two calibration sets.

is possible that PLSR model in a single calibration set can't fit over wide concentration range to result in a deviation of the Lambert-Beer's law. The similar problem had been reported in a previous report and it was overcome by multiple calibration sets method [34]. Although the linearity between concentration and absorbance deviates from Lambert-Beer's law in this model, the proposed WSRII can decrease errors by eliminating uninformative information.

3.5. Building two models in 0–600 and 600–1400 ppm-m respectively

In order to further improve prediction accuracy, the method of multiple calibration sets was used in training PLSR model. As shown in Fig. 3C and Fig. 3D, the relative errors were fairly small when the order of samples was greater than 15. It was found that the concentration of the 15th and 16th sample were 531.46 ppm-m and 606.42 ppm-m, respectively. Therefore, the calibration set containing 100 spectra in concentration range of 0–1400 ppm-m was divided into two parts, i.e. a low concentration set with 50 spectra in range of 0–600 ppm-m and a high concentration set with 50 spectra in range of 601–1400 ppm-m, for further research. Based on this, prediction performance of the PLSR model after WSRII was also evaluated using two calibration sets in two different concentration ranges.

As shown in Table 1, the value of RMSEV in the PLSR model decreased from 17.24 to 9.64 and the relative errors were decreased obviously (Figs. 3D and 4B, black bar) after using the method of multiple calibration sets. This result showed that overlarge prediction errors, discussed in the previous section, was due to an over wide concentration range that results in a severe deviation from the linearity of Lambert-Beer's law. In the model of two calibration sets, the factor number was reduced after WSRII, which indicated the proposed method could eliminate interference information. The value of RMSEV were lowered from 9.64 in PLSR model to 6.43 in the PLSR model after WSRII, which illustrated that the proposed WSRII could improve further prediction accuracy of PLSR model.

To further assess the property of WSRII in PLSR models of two calibration sets, we compared the NH_3 concentrations of validation set samples calculated by both the PLSR model and the one after WSRII. As shown in Fig. 4A, the predicted concentrations of PLSR model and the model after WSRII were more close to the standard values. This result further supported the previous assumption, which suggested over wide concentration range led to large errors in the calculated NH_3 concentrations. In Fig. 4B, the relative errors of most samples were lowered obviously in the low concentration range and the relative errors of validation set samples were below 5%. These results showed that the proposed method of variable selection can remove the interference data of OP/FT-IR spectra and improve the prediction accuracy of PLSR model

for NH_3 concentrations in the atmosphere. It is obvious that the proposed variable selection method possesses important potential in rapid and accurate monitoring of the atmospheric NH_3 concentrations.

4. Conclusion

In this work, a new variable selection method, WSRII, is proposed to improve the prediction accuracy of PLSR model for atmospheric NH_3 concentrations using OP/FT-IR spectra. The prediction performance of the proposed method was evaluated by one calibration set and two calibration sets of OP/FT-IR spectra, respectively. The value of RMSEV decreased significantly from 17.24 of the PLSR model to 6.43 of the model after WSRII. Meanwhile, relative errors of validation set samples were reduced below 5%. Therefore, the proposed WSRII can eliminate obviously interference data in OP/FT-IR spectra to improve prediction accuracy of PLSR model. It is noteworthy that the proposed method does not introduce extra parameters in PLSR model for variable selection and is readily comprehensible compared with other ways. Thus, this process of variable selection is significantly potential to monitor and calculate NH_3 emission of atmosphere in real time.

Acknowledgements

We thank Dr. Wanping Wang of West Anhui University for discussing pretreatment of spectral data.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cjac.2022.100133.

References

- [1] Kinney PL. *Curr Environ Health Rep* 2018;5(1):179–86.
- [2] Karanasiou A, Moreno N, Moreno T, Viana M, Leeuw de F, Querol X. *Environ Int* 2012;47:107–14.
- [3] Phillips FA, Naylor T, Forehead H, Griffith DWT, Kirkwood J, Paton-Walsh C. *Atmosphere* 2019;10(4):1–29.
- [4] Ebi KL, Mills DM, Smith JB, Grambsch A. *Environ Health Perspect* 2006;114(9):1318–24.
- [5] Wyzga RE, Rohr AC. *J Air Waste Manag Assoc* 2015;65(5):523–43.
- [6] Stanek LW, Brown JS, Stanek J, Gift J, Costa DL. *Toxicol Sci* 2011;120(S1):S8–27.
- [7] Toro RA, Canales M, Flocchini RG, Morales RGE, Leiva GMA. *Aerosol Air Qual Res* 2014;14(1):33–44.
- [8] Nowak JB, Neuman JA, Bahreini R, Middlebrook AM, Holloway JS, McKeen SA, Parrish DD, Ryerson TB, Trainer M. *Geophys Res Lett* 2012;39(7):L07804.
- [9] Broome RA, Fann N, Cristina TJ, Fulcher C, Duc H, Morgan GG. *Environ Res* 2015;143:19–25.
- [10] Dammers E, Schaap M, Haaime M, Palm M, Wichink KRJ, Volten H, Hensen A, Swart D, Erisman JW. *Atmos Environ* 2017;169:97–112.
- [11] Fowler D, Pilegaard K, Sutton MA, Ambus P, Raivonen M, Duyzer J, Simpson D, Fagerli H, Puzzi S, Schjoerring JK, Granier C, Neftel A, Isaksen ISA, Laj P, Maione M, Monks PS, Burkhardt J, Daemmgen U, Neiryneck J, Personne E, Wichink-Kruit R, Butcherbach-Bahl K, Flechard C, Tuovinen JP, Coyle M, Gerosa G, Loubet B, Altimir N, Gruenhage L, Ammann C, Cieslik S, Paoletti E, Mikkelsen TN, Ro-Poulsen H, Cellier P, Cape JN, Horváth L, Loreto F, Niinemets Ü, Palmer PI, Rinne J, Misztal P, Nemitz E, Nilsson D, Pryor S, Gallagher MW, Vesala T, Skiba U, Brüggemann N, Zechmeister-Boltenstern S, Williams J, O'Dowd C, Facchini MC, de Leeuw G, Flossman A, Chamerliac N, Erisman JW. *Atmos Environ* 2009;43(33):5193–267.
- [12] Russwurm GM, Phillips B. *Appl Opt* 1999;38(30):6398–407.
- [13] Yu L, Shao L. *Chin J Anal Chem* 2015;43(2):226–32.
- [14] Newman AR. *Anal Chem* 1997;69(1):43A–47A.
- [15] Bai M, Suter H, Lam SK, Davies R, Flesch TK, Chen D. *Agr Forest Meteorol* 2018;258:50–5.
- [16] Flesch TK, V Baron S, Wilson JD, Griffith DWT, Basarab JA, Carlson PJ. *Agr Forest Meteorol* 2016;221:111–21.
- [17] Sommer SG, Hutchings N. *Eur J Agron* 2001;15(1):1–15.
- [18] Naguib IA, E Abdelaleem A, Hassan ES, Ali NW, Gamal M. *Spectrochim Acta A Mol Biomol Spectrosc* 2020;239:118513.
- [19] Badaro AT, Garcia-Martin JF, Lopez-Barrera MDC, Barbin DF, Alvarez-Mateos P. *Food Chem* 2020;323:126861.
- [20] Du Y, Liang Y, Jiang J, Berry RJ, Ozaki Y. *Anal Chim Acta* 2004;501(2):183–91.
- [21] Yun Y, Wang W, Tan M, Liang Y, Li H, Cao D, Lu H, Xu Q. *Anal Chim Acta* 2014;807:36–43.
- [22] Yang O, Zhang X, Yan H. *Pattern Recognit* 2020;107.
- [23] Wang X, Feng X, Song X. *Comput Stat Data An* 2020:151.
- [24] Zhang J, Cui X, Cai W, Shao X. *J Chemometr* 2018;32:e2971.

- [25] Xu H, Liu Z, Cai W, Shao X. *Chemometr Intell Lab Syst* 2009;97(2):189–93.
- [26] Shan R, Cai W, Shao X. *Chemometr Intell Lab Syst* 2014;131(15):31–6.
- [27] Guyon I, Elisseeff A. *J Mach Learn Res* 2003;3(Mar):1157–82.
- [28] Lorber A, Kowalski BR. *J Chemometr* 1988;2(1):67–79.
- [29] Cai W, Li Y, Shao X. *Chemometr Intell Lab Syst* 2008;15:188–94.
- [30] Yun Y, Cao D, Tan M, Yan J, Ren D, Xu Q, Yu L, Liang Y. *Chemometr Intell Lab Syst* 2014;130:76–83.
- [31] Paiva HM, S Soares FC, Galvão RKH, Araújo MCU. *Chemometr Intell Lab Syst* 2012;118:260–6.
- [32] Shao L, Griffiths PR. *Anal Chem* 2008;80(13):5219–24.
- [33] Shao L, Wang W, Griffiths PR, Leytem AB. *Appl Spectrosc* 2013;67(3):335–41.
- [34] Shao L, Liu B, Griffiths PR, Leytem AB. *Appl Spectrosc* 2011;65(7):820–4.