Quantitative determination of additive Chlorantraniliprole in Abamectin preparation: Investigation of bootstrapping soft shrinkage approach by mid-infrared spectroscopy

Hong Yan, Xiangzhong Song, Kuangda Tian, Yilin Chen, Yanmei Xiong, Shungeng Min

PII:	S1386-1425(17)30696-0
DOI:	doi: 10.1016/j.saa.2017.08.067
Reference:	SAA 15424
To appear in:	*Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*
Received date:	9 January 2017
Revised date:	30 August 2017
Accepted date:	31 August 2017

Please cite this article as: Hong Yan, Xiangzhong Song, Kuangda Tian, Yilin Chen, Yanmei Xiong, Shungeng Min , Quantitative determination of additive Chlorantraniliprole in Abamectin preparation: Investigation of bootstrapping soft shrinkage approach by mid-infrared spectroscopy, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* (2017), doi: 10.1016/j.saa.2017.08.067

Quantitative determination of additive Chlorantraniliprole in Abamectin preparation: investigation of bootstrapping soft shrinkage approach by mid-infrared spectroscopy

Hong Yan[1], Xiangzhong Song[1], Kuangda Tian[1], Yilin Chen[1], Yanmei Xiong[1,*], Shungeng Min[1,**]

[1]*College of Science, China Agricultural University, Beijing 100193, P.R. China*

* Corresponding author.

** Corresponding author. Tel: +86 010 62733091; Fax:+86 010 62733091

Email addresses: xiongym@cau.edu.cn (Y. Xiong), minsg@cau.edu.cn (S. Min).

Abstract:

A novel method, mid-infrared (MIR) spectroscopy, which enables the determination of Chlorantraniliprole in Abamectin within minutes, is proposed. We further evaluate the prediction ability of four wavelength selection methods, including bootstrapping soft shrinkage approach(BOSS), Monte Carlo uninformative variable elimination(MCUVE), genetic algorithm partial least squares (GA-PLS) and competitive adaptive reweighted sampling(CARS) respectively. The results showed that BOSS method obtained the lowest root mean squared error of cross validation (RMSECV) (0.0245) and root mean squared error of prediction (RMSEP) (0.0271), as well as the highest coefficient of determination of cross-validation (Q_{cv}^2) (0.9998) and the coefficient of determination of test set (Q^2_{test}) (0.9989), which demonstrated that the mid infrared spectroscopy can be used to detect Chlorantraniliprole in Abamectin conveniently. Meanwhile, a suitable wavelength selection method (BOSS) is essential to conducting a component spectral analysis.

Keywords: MIR, BOSS, Chlorantraniliprole, wavelength selection

1. Introduction

Pesticides play an indispensable role in pest control in agriculture. However, as the usage of traditional pesticides increased rapidly over the last few decades, the issues of environmental pollution and pesticide resistance have become more and more seriously[1, 2]. To solve this problem, many approaches have been tried. One of them is by using the botanical pesticides, which is friendly to the environment and suffers less pesticide resistance. Botanical pesticides, therefore, attracted more attention and were spread fast instead of synthesis pesticide in controlling pest[3]. It is nevertheless true that drawbacks still exist, it may take a longer time to produce efficacy and its active ingredients are easy to decompose. In order to speed up efficacy and profiteer, some manufacturers illegally mixed synthetic pesticides into it, which not only pollutes the environment but also cheats the consumers. Thus, it is imperative to control the quality of botanical pesticides[3-8].

As a kind of botanical pesticides, Abamectin is widely used as insecticide, acaricide and nematicide. In addition, the unique mechanism of Abamectin is caused by the interference with nerve physiological activities of insects. Therefore, it is not only easy to avoid generating pesticide resistance, meanwhile, high biological activity, selectivity and security can be assured simultaneously.

Chlorantraniliprole, a new type broad spectrum insecticide, becomes very popular on the market, even in the face of some insecticide resistant pests, it also reveals an outstanding effect of pests control. Since Chlorantraniliprole has surprising rain erosion resistance and photolysis resistance ability, it is usually added into botanical pesticides by manufacturers.

Methods suggested by CIPAC(Collaborative International Pesticides Analytical Committee) for quality

control of pesticide are GC(Gas Chromatography) and HPLC(High Performance Liquid Chromatography)[9], which are relatively complicated and time-consuming. Therefore, it is necessary to develop a fast and convenient method to detect the illicit additive pesticide. MIR(mid-infrared spectroscopy), known as a rapid, non-destructive analysis method, could be highly beneficial for industry, agriculture[10-12], food chemistry[13], pesticide analysis[3, 7, 8, 14, 15] etc. In addition, the absorption bands of fundamental frequency vibrations and the combination of vibrations make MIR fit for quantitative analysis. Techniques based on MIR spectroscopy indeed offer the advantages of requiring little sample preparation with a fast and easy data acquisition. On top of that, several properties can be predicted according to a single spectrum simultaneously. Thus, MIR spectroscopy holds the reason to be applied to detecting the additive pesticides. However, absorption bands are overlapped and low sensitivity, so aiming to predict indirectly, appropriate chemometric tools should be used for multivariate calibration which is largely responsible for advancing spectroscopic techniques. Computational approaches (e.g., principal component analysis (PCA), linear discriminant analysis (LDA), partial least squares (PLS)) permit the processing of large amounts of spectroscopic data variables that subsequently require data reduction approaches in order to identify the sources of variance across spectra and for inter-class variation to be identified.

To improve the prediction ability of multivariate models which can be affected by uncorrelated variations (e.g., noise, physical and chemical interferences), wavelength selection is necessary for MIR spectroscopy, an appropriate variable subset selection will generate a desirable calibration model with optimal accuracy and robustness. Wavelength selection is a mathematic and statistic tool in chemometrics field, which was used as a powerful method in spectral quantitative models. However, it takes the risk to lose useful

information, so chemometrics enables the analyst with greater reliability[16-18]. Various wavelength selection methods have been developed for this purpose, some of which are based on PLS models[19-21], such as sub-window permutation analysis(SPA), uninformative variable elimination(UVE), competitive adaptive reweighted sampling(CARS)[17], ant colony optimization(ACO), particle swarm optimization(PSO)[22] etc. Also, Monte Carlo uninformative variable elimination(MCUVE)[5], Monte Carlo uninformative variable elimination-sub-window permutation analysis(GA-PLS)[23], competitive adaptive reweighted sampling-sub-window permutation analysis(CARS-SPA), stability competitive adaptive reweighted sampling(SCARS) are modified methods which are upon traditional ways[18, 24-26]. Recently, a novel method named bootstrapping soft shrinkage approach(BOSS) has been proposed with good performance basing on model population analysis(MPA) and weighted bootstrap sampling (WBS)[27]. Rare research about illicit addition to Chlorantraniliprole role is reported. Moreover, the method of BOSS was applied to analyze liquid pesticide system for the first time.

In this study, MIR spectroscopy combined with chemometrics was tested to be proposed as a useful tool to easily and early detect the illicit addition of Chlorantraniliprole technical in Abamectin preparations. Four wavelength selection methods, including BOSS, CARS, GA-PLS, MCUVE were employed to optimize the variable selection. The prediction abilities were evaluated in the pesticide liquid system, and the correlation between the selected wavelengths and the chemical properties was investigated stepwise. Quantitative data was consequently considered to be the main focus of this approach. Finally, advice about detection of illicit addition on site was proposed.

2. Theory

2.1. PLS Model

PLS is aiming to model the relationship between the data matrix **X** and a response vector **y.** Spectral data matrix X contains *J* variables in columns and *N* samples in rows. Vector y with order *N*×1 denotes the measured property of interest. The purpose of PLS is to select the best optimal number A (A≤N) of PLS factors. Eqs. (1) and (2) illustrate the PLS model.

$$X = TP^T + E_A \quad (1)$$

$$y = Tq^T + f_A \quad (2)$$

While T(N×A) is a score matrix, P(J×A) is the loading matrix with the vectors pa (a=1,2,…… , A) as columns, q(1×A) as the y-loading vector, E_A and f_A are the residual matrix of X and y-vector, respectively. By applying cross-validation(CV), the optimal number of A can be acquired.

2.2. GA-PLS

GA was proposed by Lucasius and Kateman according to the 'Darwin's classical rules'[28], which is mainly ruled by the 'struggle of life'.

The genetic algorithm begins with initiation of the population, and the best selected variables can be formed based on the biological evolution, such as reproduction, mutation, elitism and migration. Here, the maximal cross validation was used as a criterion to find the best individual variables, the combination of their genomes occurred, ultimately, new chromosomes would be formed and take place of the previous one if their performances are better.

2.3. MCUVE

Based on PLS regression coefficients, UVE is a variable selection method with regression coefficient *b*, and the core idea of this method is that the regression coefficient is regarded as the determination indicator of variable importance index, the spectral variables comparing with random noise of regression coefficient and regression coefficient of contribution under random noise level variables.

As already described, MCUVE employ Monte Carlo sampling in the variable space of UVE method. A regression coefficient matrix, $b_i = [b_1, \cdots, b_p], i = 1, \cdots, N$, is calculated after N runs, and the definition of stability is:

$$s_j = \text{mean}(b_j) / \text{std}(b_j),\ j = 1, \cdots, p \tag{3}$$

mean(b_j) and std(b_j) are the mean and standard deviation of the absolute regression coefficients of each variable[5].

2.4. CARS

CARS is based on absolute regression coefficients to evaluate the importance of variables. Monte Carlo is employed for sampling.

The exponentially decreasing function (EDF) is then employed to enforce feature selection, and remove variables with small absolute regression coefficients as well.

$$r_i = ae^{-ki} \tag{4}$$

$$a = (\frac{m}{2})^{1/(N-1)} \tag{5}$$

$$k = \frac{\ln(\text{m}/2)}{N-1} \tag{6}$$

Consecutively, adaptive reweighted sampling (ARS) is performed to realize a competitive feature selection based on the regression coefficients.

Finally, cross validation is adopted to select the subsets according to the lowest RMSECV.

2.5. BOSS

BOSS (The bootstrapping soft shrinkage) was developed by Baichuan Deng in 2016. This method is supposed to select informative variables with the existence of colinearity[27]. The steps are listed here:

(1) K subsets are generated by using BSS, all the variables are assigned with equal weights(w).

(2) build K PLS sub-models with all the subsets and pick out best models with the lowest RMSECV.

(3) sum up the normalized regression vector to obtain new weights for variables.

$$w_i = \sum_{k=1}^{K} b_{i,k} \tag{7}$$

(4) new subsets are generated by WBS according to new weights. This way guarantees that we have better chance to select the variables which have the larger absolute value of regression coefficients.

The subset which has the lowest RMSECV during the iteration is selected as the optimal variable set by repeating step (2)-(4).

2.5. Model Validation

To evaluate the performance of four promising variable selection methods, namely GA-PLS, MCUVE, CARS and BOSS. Mean-centered were applied before modeling, and the optimal number of latent variables was determined by 5-fold cross validation. The root-mean-square error of calibration (RMSEC), root-mean-square error of the prediction of test set (RMSEP), Q_{cv}^2 and Q^2_{test}were used to assess model performance. Moreover, the number of optimal latent variables (nLV) and the number of variables selected (nVAR) were also reported.

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{Ncal} (y_i - \hat{y}_i)^2}{Ncal}} \tag{8}$$

$$Q^2_{cv} = 1 - \frac{\sum_{i=1}^{Ncal} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{Ncal} (y_i - \bar{y}_i)^2} \quad (9)$$

While y_i is the experimental of the predicted properties, and $\hat{y}_i$ and $\bar{y}_i$ represent predicted and average respectively. $Ncal$ is the number of calibration samples of the training set. RMSEP and Q^2_{test} hold the equation following the same as RMSEC and Q_{cv}^2.

$$\text{reliability} = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}} \quad (10)$$

Each method was repeated 50 times to assess the stability and reproducibility. The standard derivation (STD) was employed to calculate stability with Eqs(10). Where X_i and $\bar{X}$ are predicted and average value, separately. n is the number of all samples. The smaller the value of the stability, the more stable is the method.

3. Experiments and Software

3.1. Samples

One hundred and fifty samples are prepared with three branches from three different manufactories. Each branch is used to prepare 50 samples, the concentration of Chlorantraniliprole ranges from 0.05 mg/g to 3.0 mg/g of each branch. The three manufactories are Shandong Zouping Pesticide Co., Ltd., North China Pharmaceutical Group Aino Co., Ltd. and Hebei Green Garden Pesticides Co., Ltd.

3.2. Chemicals

Chlorantraniliprole original drug (95%), 1.8% Abamectin EC formulation (three manufacturers are Shandong Zouping Pesticide Co., Ltd., North China Pharmaceutical Group Aino Co., Ltd. and Hebei Green Park Pesticides Co., Ltd.,), carbon tetrachloride (AR, Hebei Green Park Pesticides Co., Ltd.), carbon tetrachloride (AR, Sinopharm Chemical Reagent Co., Ltd.), acetone (AR, Sinopharm Chemical Reagent Co., Ltd.), N- dimethylformamide(AR, Sinopharm Chemical Reagent Co., Ltd.).

3.3. Instruments

Analytical balance (A & D, Japan), 1mm cuvette.

Spectrum 400 Fourier Transform Infrared - Near Infrared Spectrometer (Perkin Elmer, United States).

Spectral acquisition mode: This experiment ATR Annex ZnGe scanning single attenuated total reflection spectroscopy, DTGS detector, scanning 64 times, the number of resolution is 4cm^{-1}, scanning range is 4000 -825 cm^{-1}. Air are used as background reference, background is scanned once every 30 minutes.

3.4. Software

All codes were written in Matlab (Version 2014a, the Mathworks, Inc) on a personal computer with an Intel Core i5-4210U 1.7 GHz CPU and 16 GB RAM.

3.5. Datasets

Each spectrum contains 3351 wavelength points within the range 650~4000cm^{-1}. The content of Chlorantraniliprole is regarded as the interest. The whole dataset was divided into calibration set (120 samples) and independent test set (30 samples) by KS sampling.

4. Results and Discussion

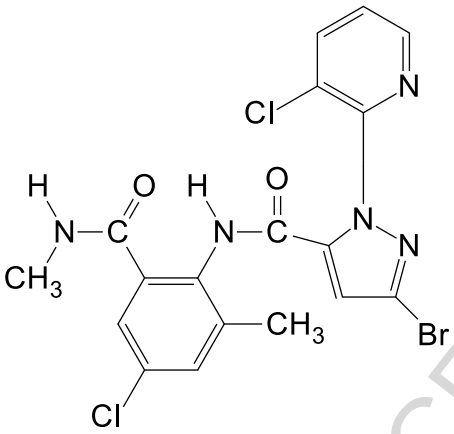


Fig.1 Chemical structure of Chlorantraniliprole

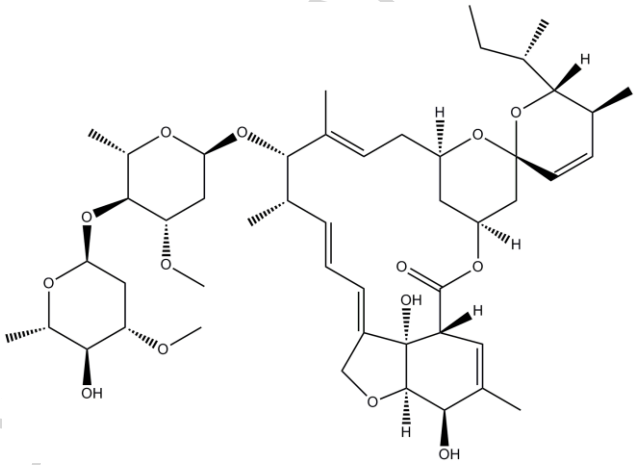


Fig.2 Chemical structure of Abamectin

To evaluate the performance of BOSS, three promising variable selection methods, namely GA-PLS, MCUVE and CARS, were used for comparison. Both the accuracy and robustness were investigated.

The obtained spectra of the prepared samples are shown in Fig. 3. Spectral pre-processing methods including baseline, derivative, smoothing and normalization were all attempted, but they didn't make big

differences and were even worse than the original models. Therefore, no preprocessing methods were used. In the dataset, four wavelength selection methods are compared, including BOSS, CARS, GA-PLS, MCUVE. Several parameters are used to evaluate model performance, RMSECV (Root mean squared error of cross-validation), RMSEP(root mean squared error of prediction of test set),Q^2_{cv},Q^2_{test}, nVAR and nLV.

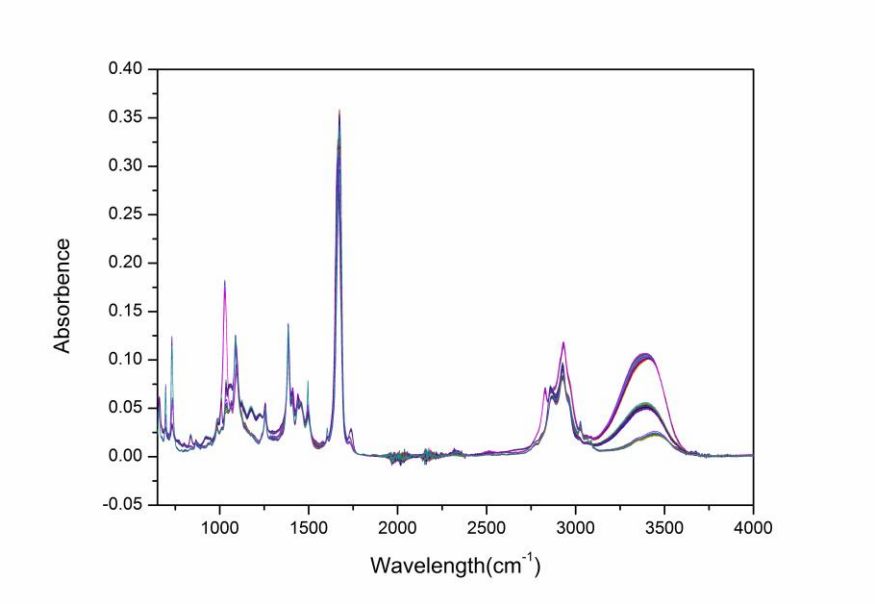


Fig.3 Original MIR spectrum of 150 samples

The details of spectrum are displayed in Fig.4, Fig.5 and Fig.6. Obviously, Chlorantraniliprole has two secondary amide groups, the absorption peaks around 3390cm^{-1} and 3261cm^{-1} with moderate intensity correspond to anti-symmetric and symmetric stretching vibration of N-H. And strong absorption is the characteristic of amide which is known as the "amide region" links to peaks at 1660cm^{-1} to 1637cm^{-1}, due to the C=O stretching vibration. The bands neighboring 1530cm^{-1}representbending vibration of N-H, also, C=C stretching vibration dynamic absorption peaks of benzene ring are around 1500cm^{-1}.

The stretching vibration absorption peak of C=O is around 1665cm^{-1}, as a result, some characteristic peaks of Chlorantraniliprole and Abamectin EC have overlapped, so there is difficulty in quantitative analysis

with single variable composition by choosing the characteristic peak from spectrum, therefore, multivariate calibration model should be set up by chemometrics.

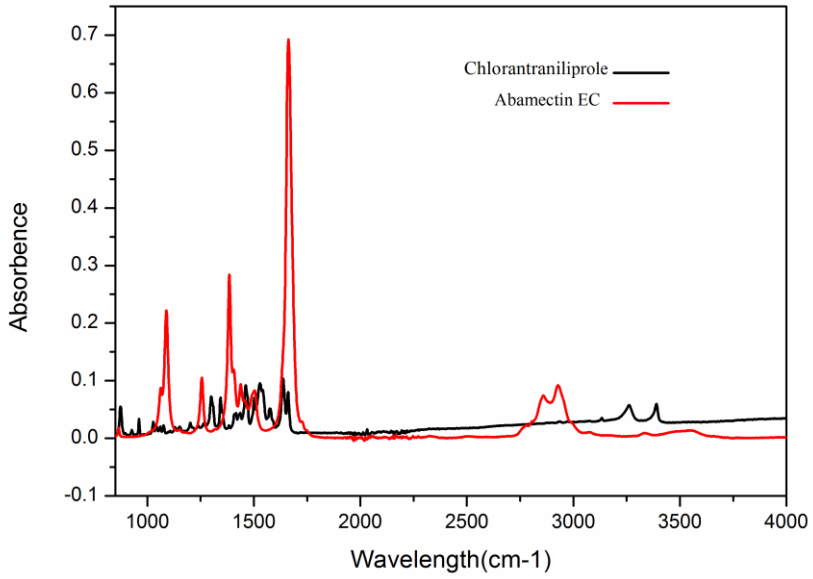


Fig.4 MIR spectrum of Chlorantraniliprole technical and Abamectin EC

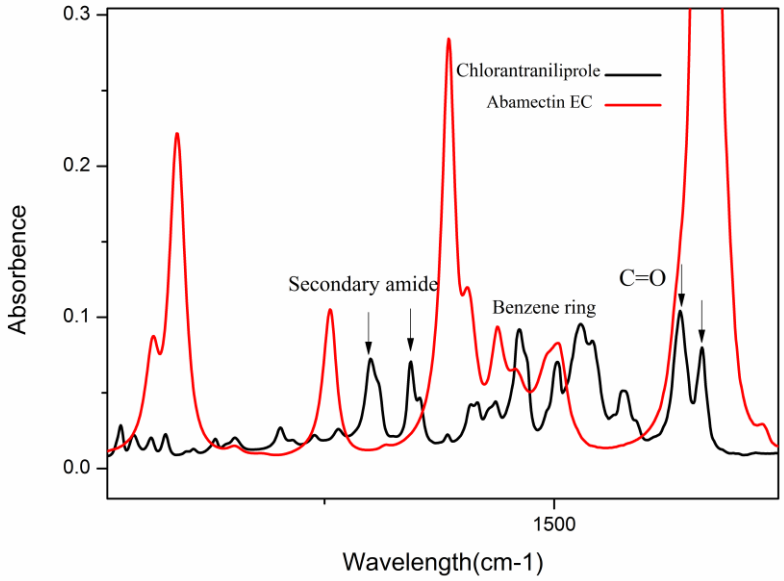


Fig.5 Functional characteristics (secondary amide, C=O and Benzene ring) of Chlorantraniliprole around 1500cm^{-1}

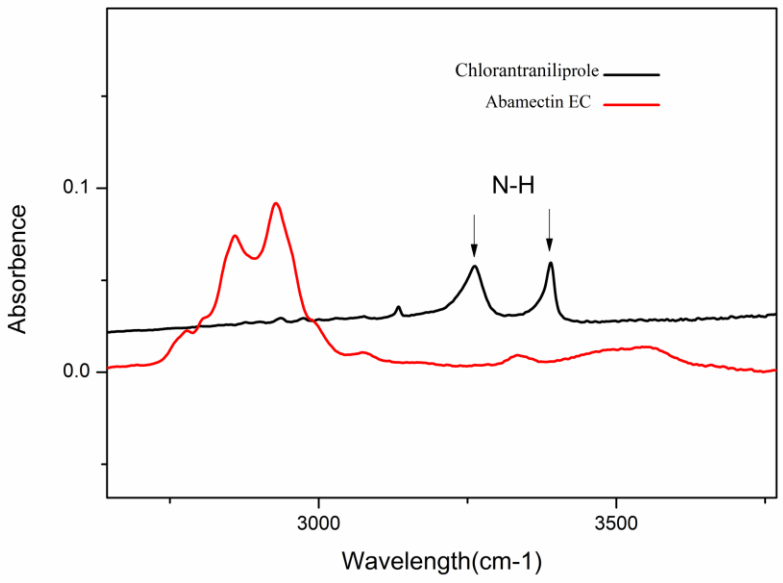


Fig.6 Functional characteristics (N-H) of Chlorantraniliprole around 3000cm^{-1} and 3500 cm^{-1}

The results of PLS, BOSS, CARS, GA-PLS, MCUVE are illustrated in Table 1, Fig.7 and Fig.8. Compared to full spectrum, it was easily noticed that the results of four wavelength selection methods showed great improvements in Q^2_{test}, RMSECV and RMSEP, and the enhancement of prediction abilities of BOSS and MCUVE in this system were most remarkable, it gave the lowest RMSEP(0.0271) and the highest Q^2_{test}(0.9989). The RMSEP of CARS, GA-PLS, MCUVE are 0.0376, 0.0539 and 0.0324 respectively. This phenomenon expressed that with fewer variables, better prediction results can be acquired. As a result, it is essential to carry out variable selection before establishing calibration models. What's more, considering that collinear variables can reduce the stability of calibration models, only choosing the key variables would be a practicable way for modeling.

Table 1 Results of Chlorantraniliprole datasets. nVAR: number of variables; nLV: number of latent variables; RMSECV: root-mean-square error of cross-validation; RMSEP: root-mean-square error of prediction; Q2_CV: coefficient of determination of cross-validation; Q2_test: coefficient of determination of test set.

Characteristics	PLS	GA		MCUVE		CARS		BOSS	
		Mean	STD	Mean	STD	Mean	STD	Mean	STD
nVAR	3351	54	24	878	65	66	16	44	29
nLV	10	10	0	10	1	10	1	10	1
Q2_CV	0.9992	0.9967	0.0012	0.9987	0.0002	0.9980	0.0009	0.9990	0.0006
Q2_test	0.9663	0.9969	0.0010	0.9975	0.0006	0.9969	0.0011	0.9969	0.0022
RMSECV	0.1148	0.0391	0.0062	0.0330	0.0022	0.0403	0.0075	0.0320	0.0054
RMSEP	0.1319	0.0416	0.0075	0.0359	0.0047	0.0402	0.0065	0.0354	0.0084

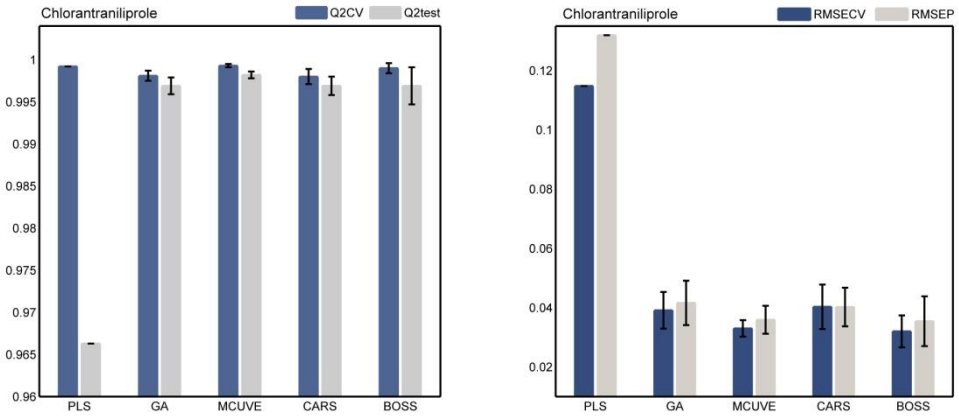


Fig.7 Q^2_{cv} and Q^2_{test} of PLS, BOSS, CARS, GA-PLS, MCUVE models

Fig.8 RMSECV and RMSEP of PLS, BOSS, CARS, GA-PLS, MCUVE models

The selected wavelengths are showed in Fig.9. It was observed that BOSS, CARS, GA-PLS chooses lower number of variables, while MCUVE selects more. GA-PLS has the lowest nVAR. BOSS and CARS have similar selected wavelengths, they both select the region around 1650cm^{-1}, 1500cm^{-1}, 1300cm^{-1}, which is

corresponding to C=O, benzene ring, N-H, respectively. The results verify the reason for variables

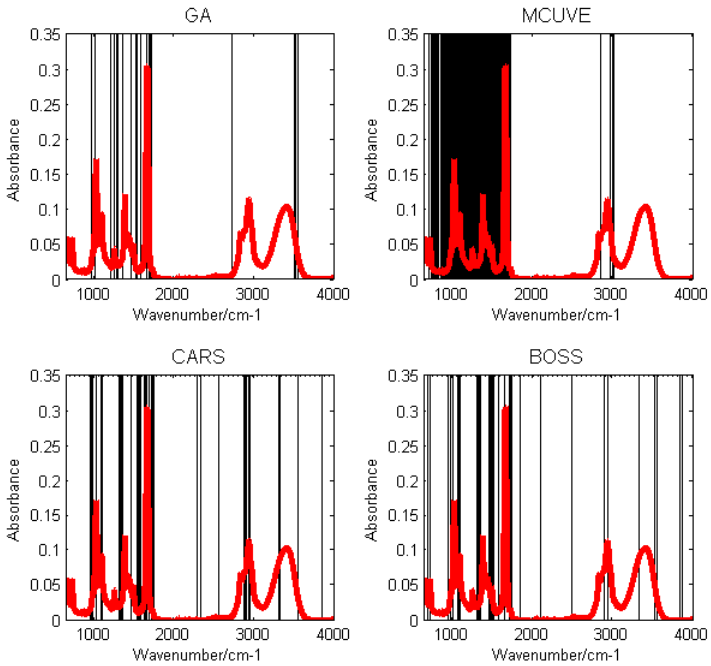


Fig.9 The variables selected by BOSS, CARS, GA-PLS, MCUVE

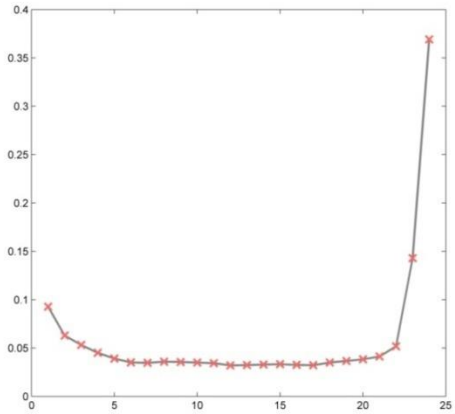


Fig.10. The evolution of RMSEC by BOSS

selection on spectroscopic data, these are different spectral regions which relate to different properties. However, BOSS selected several more variables around 650cm^{-1} than CARS and those variables are more concentrated, which may be the explanation of why BOSS's performance is better. Even though

MCUVE-SPA has the lowest nVAR, but it lost effective information, which leads to the worst performance among four wavelength selection methods. On the contrary, MCUVE retains much more variables than three other methods, despite the result is not bad, it still contains uninformative variables compared with BOSS and CARS. It is clear that the ranking of prediction ability in this system would be as follows: BOSS>CARS>MCUVE>GA-PLS>PLS. Wavelength selection can indeed improve the prediction ability, and BOSS performs the best in this pesticide liquid system.

As we can see, BOSS performs the best among four wavelength selection methods, e.g. BOSS, CARS, GA-PLS, MCUVE. This mainly because that BOSS could make good use of effective information by combining the advantages of both MPA and WBS.

In soft shrinkage strategy, the variables which may be considered as less significant are assigned smaller weights, thus, these variables still have the chance to participate the models. At the same time, they may also prove to be informative. The risk of ignoring essential variables can be reduced by applying soft shrinkage during the whole processes.

WBS also played a role in the good performance of BOSS, different weights assign to different objects, so the objects with larger weights have more chances to be selected. Applying WBS can avoid the influence of colinearity of regression coefficients, also, the combination of BSS and WBS make the variable ranking more reliable.

The key parts of MPA are random sampling and statistical analysis as well as the core feature is that the output of interest not as a single value but a distribution will be considered. In MPA process, the variables with larger weights have more chances to participate. Besides that, MPA also considers the combination among variables, which makes the result more stable and have spectroscopy meaning of selected variables.

In the end, the variables with the highest frequency in top 10% sub-models are retained[29].

As a general framework for statistically extracting interesting information from a large population of sub-models towards better understanding of the chemical data, MPA is promising in developing new chemometrics algorithms, ranging from variable selection and model evaluation to outlier detection and applicability domain definition.

Both CARS and MCUVE are based on large absolute regression coefficients to evaluate the significance of each variable. The larger the absolute regression coefficient is, the more significant the variable is. Monte Carlo strategy for sampling are employed in the variable spaces. The applying of EDF function of CARS causes the impact that some informative variables may be eliminated, on the opposite, MCUVE retained both important and uninformative variables. Thus, the performance of CARS isn't the most powerful one.

BOSS also consider the number of variables (nVAR) in different levels and the weights of variables change during 24 iterations as it is shown in supplementary material. Simultaneously, RMSECV of sub-models decrease with the iterations and achieve the minimum at iteration 16 (Fig.10). At the beginning, the variables with large weight may not turn out to be insignificant for modeling with small weight in the later iterations. The optimal variable set is obtained in iteration 15. The most informative variables are thus obtained at around 1720 nm and 3000 nm.

Nevertheless, further details we should notice, the stability of BOSS didn't play well, although BOSS gave the lowest RMSEP, but it has the maximum STD (0.0084) of RMSEP at the same time. One possible reason is that BOSS still considers regression coefficients as a measure of searching the optimal variables. As we know, in PLS1 algorithm, the functions of the parameters, such as loadings, weights and regression coefficients, are dependent with each other. The parameters of PLS algorithm will change

continuously along with the data matrix. In simple terms, the regression coefficients are different in different models. Therefore, this is a problem of BOSS which we need to fix in future study with further investigations and the stability of variable selection methods should be brought to the forefront as well.

5. Conclusion

In this study, we have shown the possibility of applying Mid infrared spectroscopy for fast quantization of prohibited additives of Chlorantraniliprole in Abamectin. Meanwhile, 150 samples were obtained to build the model. The feasibility of mid infrared spectroscopy has been explored, also, four wavelength selection methods, including BOSS, CARS, GA-PLS, MCUVE were compared, the predictive power of BOSS is first assessed in pesticide liquid system. The ranking of performance is BOSS>CARS>MCUVE>GA-PLS>PLS, and these results demonstrate the potential ability of mid infrared spectroscopy to detect Chlorantraniliprole in Abamectin quickly. Mid infrared spectroscopy combined with BOSS-PLS will be suggested as a powerful technique for component testing.

Acknowledgements

This research is financially supported by National Natural Science Foundation of China(Grant No.31301685).

References

[1] Gevao B, Semple K T, Jones K C, Bound pesticide residues in soils: a review. Environmental pollution, 108(2000) 3-14.

[2] Ecobichon, Donald J, Pesticide use in developing countries. Toxicology, 160(2001) 27-33.

[3] K. Qiu, X. Song, Y. Lai, L. Wu, G. Tang, S. Min, Comparison of ATR/transmittance FTIR combined with Beer's law and PLS to determine fipronil in matrine formulation, Analytical Methods, 5 (2013) 4790.

[4] G. Tang, K. Tian, X. Song, Y. Xiong, S. Min, Comparison of several supervised pattern recognition techniques for detecting additive methamidophos in rotenone preparation by near-infrared spectroscopy, Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy, 121 (2014) 678-684.

[5] G. Tang, X. Song, J. Hu, H. Yan, K. Qiu, K. Tian, Y. Xiong, S. Min, Characterization of a Pesticide Formulation by Medium Wave Near-Infrared Spectroscopy with Uninformative Variable Elimination and Successive Projections Algorithm, Analytical Letters, 47 (2014) 2570-2579.

[6] G. Tang, Y. Lai, X. Song, K. Qiu, S. Min, Rapid detection of illicit fipronil in phoxim granules by mid-infrared spectroscopy, Optik - International Journal for Light and Electron Optics, 125 (2014) 2048-2050.

[7] G. Tang, J. Hu, H. Yan, Y. Zhao, Y. Xiong, S. Min, Determination of active ingredients in matrine aqueous solutions by mid-infrared spectroscopy and competitive adaptive reweighted sampling, Optik - International Journal for Light and Electron Optics, 127 (2016) 1405-1407.

[8] K. Qiu, X. Song, G. Tang, L. Wu, S. Min, Determination of Fipronil in Acetamiprid Formulation by Attenuated Total Reflectance-Mid-Infrared Spectroscopy Combined with Partial Least Squares Regression,

Analytical Letters, 46 (2013) 2388-2399.

[9] S. Armenta, S. Garrigues, M. de la Guardia, Partial least squares-near infrared determination of pesticides in commercial formulations, Vibrational Spectroscopy, 44 (2007) 273-278.

[10] S. Armenta, M. de la Guardia, Vibrational spectroscopy in soil and sediment analysis, Trends in Environmental Analytical Chemistry, 2 (2014) 43-52.

[11] M. Janotta, M. Karlowatz, F. Vogt, B. Mizaikoff, Sol–gel based mid-infrared evanescent wave sensors for detection of organophosphate pesticides in aqueous solution, Analytica chimica acta, 496 (2003) 339-348.

[12] M. Forouzangohar, J.A. Baldock, R.J. Smernik, B. Hawke, L.T. Bennett, Mid-infrared spectra predict nuclear magnetic resonance spectra of soil carbon, Geoderma, 247-248 (2015) 65-72.

[13] P. Wang, J. Sun, T. Zhang, W. Liu, Vibrational spectroscopic approaches for the quality evaluation and authentication of virgin olive oil, Applied Spectroscopy Reviews, 51 (2016) 763-790.

[14] S. Armenta, S. Garrigues, M. de la Guardia, Green Analytical Chemistry, TrAC Trends in Analytical Chemistry, 27 (2008) 497-511.

[15] N.K. Niazi, B. Singh, B. Minasny, Mid-infrared spectroscopy and partial least-squares regression to estimate soil arsenic at a highly variable arsenic-contaminated site, International Journal of Environmental Science and Technology, 12 (2014) 1965-1974.

[16] Y.-H. Yun, W.-T. Wang, B.-C. Deng, G.-B. Lai, X.-b. Liu, D.-B. Ren, Y.-Z. Liang, W. Fan, Q.-S. Xu, Using variable combination population analysis for variable selection in multivariate calibration, Analytica chimica acta, 862 (2015) 14-23.

[17] H. Li, Y. Liang, Q. Xu, D. Cao, Key wavelengths screening using competitive adaptive reweighted

sampling method for multivariate calibration, Analytica chimica acta, 648 (2009) 77-84.

[18] R. Shan, W. Cai, X. Shao, Variable selection based on locally linear embedding mapping for near-infrared spectral analysis, Chemometrics and Intelligent Laboratory Systems, 131 (2014) 31-36.

[19] H.D. Li, Q.S. Xu, Y.Z. Liang, Random frog: an efficient reversible jump Markov Chain Monte Carlo-like approach for variable selection with applications to gene selection and disease classification, Analytica chimica acta, 740 (2012) 20-26.

[20] B.C. Deng, Y.H. Yun, P. Ma, C.C. Lin, D.B. Ren, Y.Z. Liang, A new method for wavelength interval selection that intelligently optimizes the locations, widths and combinations of the intervals, The Analyst, 140 (2015) 1876-1885.

[21] W. Cai, Y. Li, X. Shao, A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra, Chemometrics and intelligent laboratory systems, 90 (2008) 188-194.

[22] Y.H. Yun, W.T. Wang, B.C. Deng, G.B. Lai, X.B. Liu, D.B. Ren, Y.Z. Liang, W. Fan, Q.S. Xu, Using variable combination population analysis for variable selection in multivariate calibration, Analytica chimica acta, 862 (2015) 14-23.

[23] Q.S. Xu, Y.Z. Liang, Monte Carlo cross validation, Chemometrics & Intelligent Laboratory Systems, 56 (2001) 1-11.

[24] T. Mehmood, K.H. Liland, L. Snipen, S. Sæbø, A review of variable selection methods in partial least squares regression, Chemometrics and Intelligent Laboratory Systems, 118 (2012) 62-69.

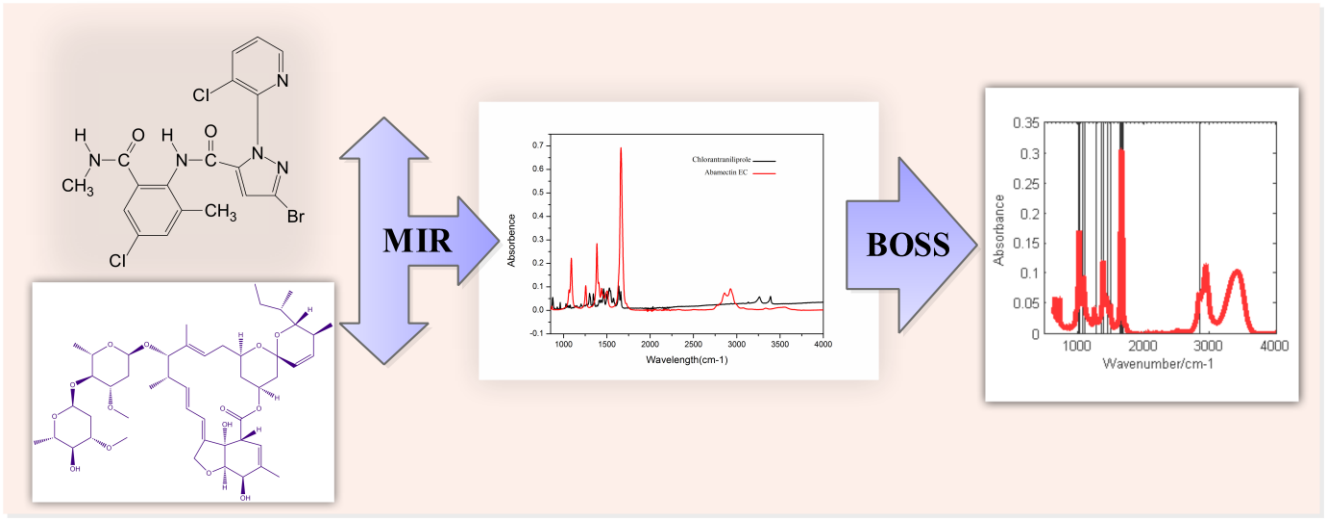
[25] M. Yuan, Y. Lin, Efficient Empirical Bayes Variable Selection and Estimation in Linear Models, Journal of the American Statistical Association, 100 (2005) 1215-1225.

[26] K. Zheng, Q. Li, J. Wang, J. Geng, P. Cao, T. Sui, X. Wang, Y. Du, Stability competitive adaptive reweighted sampling (SCARS) and its applications to multivariate calibration of NIR spectra, Chemometrics and Intelligent Laboratory Systems, 112 (2012) 48-54.

[27] B.C. Deng, Y.H. Yun, D.S. Cao, Y.L. Yin, W.T. Wang, H.M. Lu, Q.Y. Luo, Y.Z. Liang, A bootstrapping soft shrinkage approach for variable selection in chemical modeling, Analytica chimica acta, 908 (2016) 63-74.

[28] M.F.S.L.M.C.C. Oliveros, C.P. Millan, Selection of useful predictors in multivariate calibration, Analytical and bioanalytical chemistry, 380 (2004) 12.

[29] B.-C. Deng, Y.-H. Yun, Y.-Z. Liang, Model population analysis in chemometrics, Chemometrics and Intelligent Laboratory Systems, 149 (2015) 166-176.



Graphical abstract

Highlights:

- Illicit addition of Chlorantraniliprole in Abamectin was first investigated.
- The predict ability of BOSS in liquid system especially in pesticide system is first assessed.
- Possibility of on-site detection of illicit addition was discussed.
- The stability of BOSS needs to be modified