

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)**SciVerse ScienceDirect**

Procedia Engineering 29 (2012) 4124 – 4128

---

---

**Procedia  
Engineering**

---

---

[www.elsevier.com/locate/procedia](http://www.elsevier.com/locate/procedia)

2012 International Workshop on Information and Electronics Engineering (IWIEE)

## Application of Particle Swarm Optimization (PSO) Algorithm to Determine Dichlorvos Residue on the Surface of Navel Orange with Vis-NIR Spectroscopy

Long Xue<sup>a</sup>, Jun Cai<sup>a</sup>, Jing Li<sup>b</sup>, Muhua Liu<sup>b\*,c</sup><sup>a</sup> School of Mechanical and Electronical Engineering, East China JiaoTong University, Nanchang, 330013, China<sup>b</sup> Optics-Electrics Application of Biomaterials Lab, Jiangxi Agricultural University, Nanchang, 330045, China<sup>c</sup> Key lab of nondestructive testing, Ministry of education, Nanchang Hangkong University, Nanchang, 330029, China

---

### Abstract

There is increased interest in the investigation and implementation of rapid, non-destructive methods for detection of pesticide residues. This study reports development of a method for determination the pesticide concentration on the surface of fruit samples by using visible near-Infrared (Vis-NIR) spectroscopy. A total of 330 navel oranges were selected for the calibration (n=220) and prediction (n=110) sets, and sprayed dichlorvos solution diluted with tap water on the surface of fruit samples. After 12 h, the spectra of fruit samples were acquired in the range of 350-1800nm, then pesticide residue was measured using gas chromatography (GC). Chemometrics analysis using Partial Least Squares (PLS) was performed in order to determine the concentration of pesticide residue. The particle swarm optimization (PSO) was used for wavelengths selection of Vis-NIR spectra data in order to develop the PLS prediction model. The model using optimal intervals for predicting the concentration of dichlorvos residue selected by PSO algorithm yielded better results than the PLS model using full wavelengths. This PSO-PLS model could predict dichlorvos residue of 220 samples of prediction set with correlation coefficient of 0.8732. The wavelength selection through a PSO algorithm would enhance the predictive ability when applying PLS model.

© 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of Harbin University of Science and Technology. Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/3.0/).

**Keywords:** Vis-NIR spectroscopy; pesticide residue; particle swarm optimization (PSO);navel orange

---

\* Corresponding author. Tel.: +86-791-83813260; fax: +86-791-83813260.  
E-mail address: [suikelmh@sohu.com](mailto:suikelmh@sohu.com).

## 1. Introduction

Typically, the determination of pesticides in complex matrices, such as fruits and vegetables, involves a sample treatment using different techniques as solid-phase extraction, supercritical fluid extraction, microwave-assisted extraction, and accelerated solvent extraction. Then conventional techniques, such as gas chromatography (GC) and high-performance liquid chromatography (HPLC), were used to measure the chemical concentration. The determination method was time consume and expensive.

Near-infrared spectroscopy (NIRS) is a useful tool due to its flexibility for both qualitative and quantitative analysis, being widely used in chemical industry, agriculture, medicine and other areas[1,2]. There are some researches about determination of pesticide or fungicide by NIR[3,4]. All of this researches show the feasibility of using NIR spectroscopy to detect trace pesticides.

As it is widely accepted as correct that a more predictable, robust and parsimonious NIR calibration model can be obtained in virtue of the effective variable selection and the elimination of spectral range only contributes to the noise. Therefore, besides the spectral data pre-processing, much more attention has been paid to variable selection of one or more spectral ranges to improve partial least squares regression (PLS) models. The main methods for variable selection are stepwise regression analysis (SRA), uninformative variable elimination (UVE)[5,6], interval partial least squares regression (iPLS), Clonal selection feature selection algorithm (CSFS)[7], Particle Swarm Optimization (PSO) [8,9]and genetic algorithm (GA)[10,11].

In this report, the possibility of predicting dichlorvos residue of navel orange using Vis-NIR is studied. And PSO algorithm was used as variable selection for modeling complicated Vis-NIR spectra data combined with PLS.

## 2. Materials and Methods

### 2.1. Sample Collection

Fresh navel orange samples were purchased from local market in Nanchang, Jiangxi, China. The samples were brought to laboratory and screened manually to discard the damaged ones. A total of 330 sound samples were washed and wiped with muslin cloth to remove dirt and water, and then kept in laboratory at 10°C and 60% relative humidity.

### 2.2. Pesticide Solution

The commercial pesticide, dichlorvos (O, O-dimethyl-O-2, 2-dichlorovinyl phosphate), containing 80% dichlorvos was used. Dichlorvos, an organophosphates insecticide, is used as a spray application for pest control. In this study, two concentration levels, 1:800 v/v and 1:1000 v/v were diluted based on the volume of dichlorvos and tap water (the suggested concentration level was 1:500 v/v mentioned in instruction book.). All the samples were divided into 2 sets, one set was sprayed using 1:800 solution and the other was sprayed using 1:1000 solution. In order to obtain different pesticide residue content of navel orange, the amount of solution sprayed on the surface of navel orange was different. Before measuring the actual dichlorvos residue content using GC, the samples were placed in laboratory for 12h.

### 2.3. Spectrum Acquisition

A spectrograph QualitySpec Pro (Analytical Spectral Devices, Inc., USA), with accessory lighting source and optical fiber was used. The fruit sample was placed on the light source accessory. The

reflectance was calculated by comparing the near-infrared energy reflected from the sample with the standard reference. The spectra were acquired in the range of 350nm to 1800nm at 1nm interval. For each sample, six equidistant positions around the equator (approximately  $60^\circ$ ) were selected, and 30 scans were co-added for each scan. Then six spectra were obtained for each sample, and were averaged for further evaluation.

#### 2.4. Pre-processing Method and Data Analysis

All the 330 fruit samples were divided into two groups (calibration set contained 220 samples and prediction set contained 110 samples). Firstly, the samples were arranged in order of their reference dichlorvos residue which measured by GC. Then the first and third samples in every three fruit samples were selected for calibration set and then the second one for prediction set. Partial least squares (PLS) regression was used to develop a prediction model. The software Matlab 7.1 (MathWorks, USA) was used for all calculations.

The Vis-NIR data set was split into 48 equal intervals, each interval had 30 wavelengths except the last one containing 41 wavelengths with the range from 1760nm to 1800nm. Each particle of the swarm represented a possible solution and was coded by a binary string. We used the binary string to represent if the 48 intervals had been selected, so the string length was 48 bits. The bit value “1” represents a select variable; the bit value “0” represents an unselect variable. The “fitness”, which is a measure of adaptation to environment, is calculated for each particle. The fitness function  $F_i$  of each particle is calculated as follows [12]:  $F_i = R_{cal} / (1 + RMSECV_i) \quad i = 1, 2, 3, \dots, n$ .

Where  $RMSECV$  is the root mean square error of cross validation,  $n$  is the number of particle and  $R_{cal}$  is the correlation coefficients between predicted and measured value in calibration set. Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached. In this paper, the population size and generations were 30 and 100, respectively.

### 3. Result and Discussion

#### 3.1. PLS Model in Full Bands

A total of 330 fruit samples' Vis-NIR original spectra are shown in figure 1 (a), and the spectra of fruit samples after pre-processing with  $D^1$  are shown in figure 1 (b).

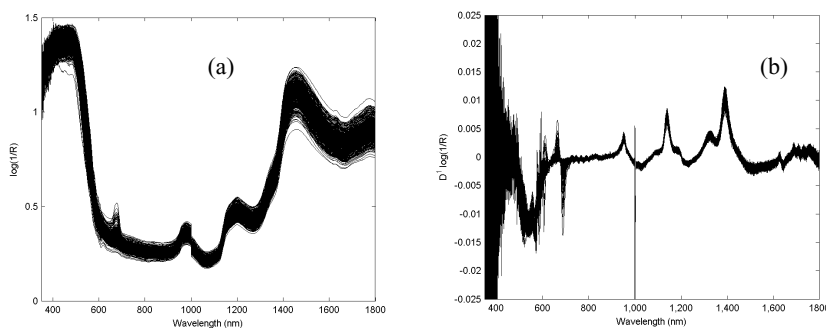


Fig 1. Vis-NIR spectra of navel orange, (a) raw spectra and (b) first derivative

The  $R_{cal}$  and  $RMSECV$  value were 0.8783 and 0.8569,  $R_{pre}$  (Correlation Coefficient between measured and predicted value in the prediction set) and  $RMSEP$  (Root Mean Square Error of Prediction) were 0.7853 and 1.1670, respectively.

### 3.2. PSO-PLS model

By comparing the PLS models of each particle at each generation, there were 26 intervals which included 780 wavelengths. The best 26 intervals were selected in the 84<sup>th</sup> generation, and an overview of the selected intervals is given in figure 2(a). These 26 intervals were used as the input variables of PLS model. Scatter diagram between the measured and predicted dichlorvos on the surface of fruit samples for the predicted set are shown in figure 2(b).

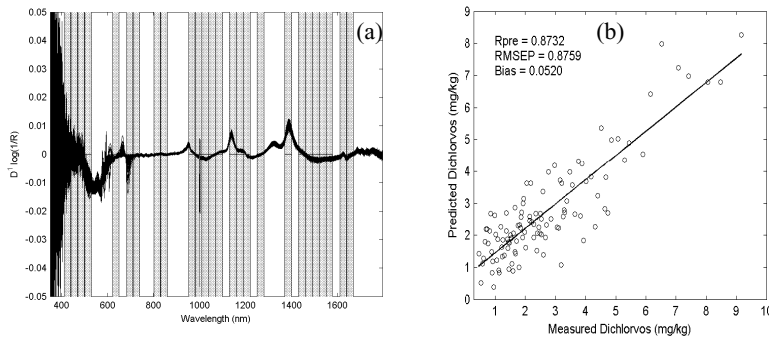


Fig. 2. a. the 26 Selected intervals by PSO and b. Predicted vs. measured dichlorvos values in predicted set with 5 PLS components

### 3.3. Analysis of the Models

In the PSO algorithm, the best score was different each time, so the results would be different in every prediction model. In order to prove the stability of PSO-PLS model, another 3 PSO-PLS models were built. It shows the arithmetic has a good repeatability. The differences between the 4 PSO-PLS models are small. The number of wavelengths is ranging from 720 to 780, and the prediction results are almost identical each other. For the 4 models based on the  $D^1$  pre-processing method, the  $R_{cal}$  ranges from 0.8628 to 0.8692 and  $R_{pre}$  ranges from 0.8654 to 0.8743. It could be concluded that the difference of each PSO-PLS model affected the result of the models weakly, and the number of wavelengths used in PSO-PLS model could be declined by more than 45%.

By comparison the 4 PSO-PLS models, according to the selected times of each interval after 90 generations, the selected 17 optimal intervals were the 3<sup>rd</sup>, 4<sup>th</sup>, 5<sup>th</sup>, 6<sup>th</sup>, 12<sup>th</sup>, 13<sup>th</sup>, 17<sup>th</sup>, 21<sup>st</sup>, 24<sup>th</sup>, 25<sup>th</sup>, 27<sup>th</sup>, 29<sup>th</sup>, 35<sup>th</sup>, 37<sup>th</sup>, 40<sup>th</sup>, 43<sup>rd</sup> and 48<sup>th</sup> interval. The corresponding  $R_{cal}$ ,  $RMSECV$ ,  $R_{pre}$ , and  $RMSEP$  of the PSO-PLS model with 6 PLS component are 0.8641, 0.9023, 0.8735 and 0.8755, respectively.

## 4. Conclusion

Vis-NIR spectra combined with chemometric method were used to predict the dichlorvos residue on the surface of navel orange in the wavelength range of 350-1800 nm. The PSO algorithm was proposed for the wavelengths selection. The model using 17 optimal intervals selected by PSO gave better predictions of dichlorvos residue ( $R_{pre}$  = 0.8735) than the PLS model with full spectra ( $R_{pre}$  = 0.7853). This preliminary research showed that the PSO algorithm could be suitable for selecting wavelengths and

improve the prediction of dichlorvos residue compared with the PLS model with full spectra. However, further studies are needed to test this method with more experimental data collected under different conditions.

## Acknowledgements

The author would like to acknowledge the funding of the Nature Science Foundation of China (NO.30760101), Program for New Century Excellent Talents in University (NCET-09-0168), Jiangxi Provincial Department of Science and Technology (2009BNB05705).

## References

- [1] F. Cámara-Martos, G. Zurera-Cosano, R. Moreno-Rojas, R. García-Gimeno, F. Pérez-Rodríguez, Identification and Quantification of Lactic Acid Bacteria in a Water-Based Matrix with Near-Infrared Spectroscopy and Multivariate Regression Modeling, *Food Analytical Methods* (2011) 1-10.
- [2] D. Cozzolino, W. Cynkar, N. Shah, P. Smith, Varietal Differentiation of Grape Juice Based on the Analysis of Near- and Mid-infrared Spectral Data, *Food Analytical Methods* (2011) 1-7.
- [3] S. Saranwong, S. Kawano, Rapid determination of fungicide contaminated on tomato surfaces using the DESIR-NIR: a system for ppm-order concentration, *J. Near Infrared Spectrosc* 13 (2005) 169-175.
- [4] J. Li, L. Xue, M. Liu, x. Wang, C. Luo, Determination of Dichlorvos Contamination on Nave Orange Surface Based on Least Squares Support Vector Machines, 2010 Sixth International Conference on Natural Computation 6 (2010) 3301-3304.
- [5] C. Jingjing, P. Yankun, L. Yongyu, W. Jianhu, s. Jiajia, Application of Uninformative Variable Elimination Algorithm to Determine Organophosphorus Pesticide Concentration with Near-infrared Spectroscopy, ASABE Paper NO.1008570.St.Joseph, Mich.:ASABE (2010).
- [6] J. Kuligowski, G. Quintás, S. Garrigues, M. de la Guardia, Direct determination of polymerized triglycerides in deep-frying olive oil by attenuated total reflectance–Fourier transform infrared spectroscopy using partial least squares regression, *Analytical and Bioanalytical Chemistry* 397 (2010) 861-869.
- [7] Y. Zhong, L. Zhang, A fast clonal selection algorithm for feature selection in hyperspectral imagery, *Geo-Spatial Information Science* 12 (2009) 172-181.
- [8] H. Shinzawa, J.H. Jiang, M. Iwahashi, I. Noda, Y. Ozaki, Self-modeling curve resolution (SMCR) by particle swarm optimization (PSO), *Anal Chim Acta* 595 (2007) 275-281.
- [9] L.Y. Chuang, J.H. Tsai, C.H. Yang, Binary particle swarm optimization for operon prediction, *Nucleic Acids Res* 38 (2010) e128.
- [10] Q. Fei, M. Li, B. Wang, Y. Huan, G. Feng, Y. Ren, Analysis of cefalexin with NIR spectrometry coupled to artificial neural networks with modified genetic algorithm for wavelength selection, *Chemometrics and Intelligent Laboratory Systems* 97 (2009) 127-131.
- [11] X. Long, L. Jing, L. Muhua, Nondestructive Detection of Soluble Solids Content on Navel Orange with Vis/NIR Based on Genetic Algorithm, *Laser & Optoelectronics Progress* 47 (2010) 123001.
- [12] P. Lu, W. Jia-hua, L. Peng-fei, Region Optimization of SSC Model for Pyrus Pyrifolia by Genetic Algorithm, *Spectroscopy and Spectral Analysis* 29 (2009) 1246-1250.