Near infrared spectroscopic analysis of chemical oxygen demand (COD) in wastewater produced by cane sugar manufacture

Tao Pan*, Zenghai Chen and Yu Shan

Key Laboratory of Optoelectronic Information and Sensing Technologies of Guangdong Higher Educational Institutes, Department of Optoelectronic Engineering, Jinan University, Guangzhou 510632, China,
*Corresponding author: tpan@jnu.edu.cn

Introduction

The wastewater from cane sugar processing contains a large number of organic pollutants, which would cause serious environmental issues if discharged into waterways. Chemical oxygen demand (COD) is an important index for the concentration of organic pollutants, and the rapid determination of COD is a significant research branch in environmental protection. Near infrared (NIR) spectroscopy has recently been used to analyse the COD of wastewater,[1-2] and efforts are ongoing to optimise these calibration models. In this paper, Savitzky-Golay (SG) smoothing and partial least squares (PLS) methods[3-8] were applied to NIR spectroscopic data, in order to optimise calibration models and determine the appropriate wavelength range for assessing COD in wastewater from cane sugar processing.

PLS combines the advantages of multiple linear regression and principal component analysis, particularly regarding issues of spectral colinearity, and is widely used in spectroscopic studies. PLS factor (F) is an important parameter. SG smoothing is an effective spectral preprocessing method and the number of smoothing points is an important parameter. SG smoothing modes and PLS factors were jointly optimised for the development of a COD calibration based on NIR spectroscopic data.

Many experimental results have shown that, if the signal to noise ratio in the spectral region used for a PLS model is too low, then the prediction can be difficult to improve. Selecting the appropriate wavelength region is thus necessary for improving model prediction, reducing model complexity, and for designing specialised NIR spectroscopy instruments. Visible light, short-wavelength NIR and long-wavelength NIR were each collected for the study presented here, with the prediction for each region compared, allowing the wavelength range with the highest signal to noise ratio to be identified. A stable result was attained by selecting the optimal SG-PLS model and corresponding wavelength region using the average of predictions from different divisions of the calibration and prediction sets.

Materials and Methods

Experimental materials, instrument and measurement method

Eighty one cane sugar wastewater samples were collected; samples had been treated with a preliminary discharging method. The COD of each sample was measured using the routine potassium permanganate oxidation method, and had values ranging from 52 to 382 $mg.L^{-1}$ with a mean and standard deviation of 230.4 and 97.1 $mg.L^{-1}$ respectively. NIR spectra were collected using a FOSS XDS rapid liquid grating spectrometer (Foss NIR Systems Inc., Denmark) with a 2 mm transmission accessory. The measured wavelength region was 400–2500 nm, at an interval of 2 nm; the ambient conditions while measuring spectra were 25±1°C and 46% relative humidity. Each sample was measured 3 times, and the average spectrum was used for modeling.

Model evaluation indicators

Model evaluation indicators used in this study were root mean squared error of predication (RMSEP) and correlation coefficient of predication (R_P).

Dividing for calibration set and prediction set

The correlation coefficient between COD and the absorbance of all samples was calculated for each wavelength in the measured spectral region (400–2500 nm), and the wavelength with the highest correlation coefficient was 1528 nm. All 81 samples were divided into calibration (50) and prediction sets (31) based on the distribution of COD and the absorbance at 1528 nm for all samples.

A total of 50 different divisions of the calibration and prediction sets were made to attain a stable prediction result. Calibration models were established for each division, and model predictions (e.g. RMSEP) in 50 different divisions were averaged for each combination of model parameters. Based on the average data, the stable optimal model could be selected.

Reference paper as:
T. Pan, Z. Chen and Y. Shan (2012).Near infrared spectroscopic analysis of chemical oxygen demand (COD) in wastewater produced by cane sugar manufacture, in: Proceedings of the15th International Conference on Near Infrared Spectroscopy,
Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 282-284.

Selected wavelength regions

Wavelengths with absorbances higher than 3 are spurious and may lead to poor calibration models. Acknowledging such wavelengths as potential outliers, this study used 8 spectral regions for developing calibrations: (1) the whole region 400–2500 nm; (2) the visible region 400–780 nm; (3) the NIR region 780–2500 nm; (4) the short-wavelength NIR region 780–1100 nm; (5) the long-wavelength NIR region 1100–2500 nm; (6) the long-wavelength NIR region without absorbances higher than three, 1100–1878 and 2088–2338 nm; (7) the NIR region without absorbances higher than three, 780–1878 and 2088–2338 nm; (8) the whole region without absorbances higher than three, 400–1878 and 2088–2338 nm.

SG smoothing

SG smoothing parameters included the order of the derivative (OD; raw spectrum was considered the 'zeroth' order), degree of polynomial (DP) and number of smoothing points (NSP). As some data benefit from more smoothing points, the number of points used in this paper ranged from 5 to 81 (increments of 2); the polynomial degree ranged from 2 to 6. A specific computer algorithm platform was built to calculate the combination of smoothing coefficients for every SG smoothing mode, and a database was consequently constructed for the optimization of smoothing modes. A total of 540 smoothing modes were obtained (covering the original 117 modes). PLS and SG-PLS models were established for the eight wavelength regions, using all SG smoothing modes and PLS factors (1–30) respectively.

Results and Discussion

The NIR spectra of 81 wastewater samples are shown in Figure 1. Wavelengths centered around 2000 and 2500 nm had strong absorption, low spectral energy, low information quality and much noise, and the 1878–2088 and 2338–2500 nm regions had absorbances higher than 3 (and were considered for elimination).

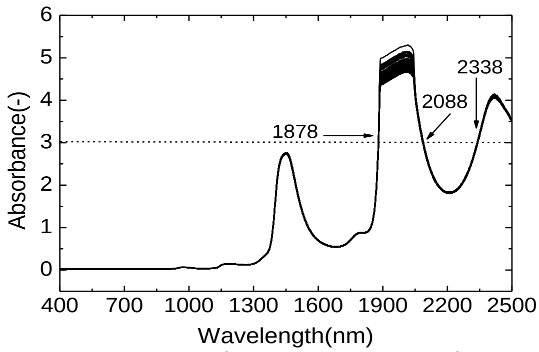

Figure 1. NIR spectra of 81 wastewater samples.

Table 1. Prediction of the optimal PLS and SG-PLS models in the selected wavelength regions.

Waveband (nm)	N	No smoothing			SG smoothing					
		F	RMSEP (mg.l^{-1})	R_P	OD	DP	NSP	F	RMSEP (mg.l^{-1})	R_P
400–780	191	5	63.2	0.784	0	6	55	8	57.8	0.820
780–1100	161	21	56.7	0.847	5	5	7	6	26.2	0.966
780–2500	860	3	84.8	0.526	4	6	57	5	78.5	0.633
1100–2500	700	3	85.1	0.522	4	6	57	5	79.2	0.621
400–2500	1050	3	84.1	0.537	0	6	71	10	68.3	0.739
780–1878, 2088–2338	801	3	81.0	0.624	2	2	81	7	73.8	0.710
1100–1878, 2088–2338	641	2	81.6	0.567	0	2	73	17	74.3	0.763
400–1878, 2088–2338	991	3	77.9	0.650	0	6	69	8	65.9	0.762

N: Number of adopted wavelengths

The predictions of the optimal PLS and SG-PLS models for the eight wavelength regions are shown in Table 1. The model prediction of the SG-PLS model with SG smoothing was substantially better than the PLS model without SG smoothing for each wavelength region, especially in the short-wavelength NIR region (780–1100 nm). Smoothing parameters for the optimum 780–1100 nm SG-PLS model were 5th order derivation, 5th degree polynomial and 7 smoothing points. Six PLS factors were used to construct the model, and the mean and standard deviation RMSEP and R_P values were 26.2 and 3.1 mg.l^{-1}, and 0.966 and 0.009, respectively.

Reference paper as:
T. Pan, Z. Chen and Y. Shan (2012).Near infrared spectroscopic analysis of chemical oxygen demand (COD) in wastewater produced by cane sugar manufacture, in: Proceedings of the15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 282-284.

Conclusion

Short-wavelength NIR (780–1100nm) was the optimal wavelength region for the NIR spectroscopic analysis of COD in sugar cane wastewater. Limiting the available wavelengths down to just 780–1100nm substantially simplified the model. Modelling optimisation and wavelength selection were based on 50 divisions of the calibration and prediction sets, and thus the results were stable and reliable. The results also showed that joint optimisation of SG smoothing modes and PLS factor can substantially improve model predictions.

Acknowledgements

This work was supported by the NSFC (10771087, 61078040), and the science and technology project of Guangdong province (2009B030801239).

References

1. J.C. He, X.L. Yang, L.R. Wang and J.M. Pan, *Acta Scientiate Circumstantiae*, **27**(12), 2105-2108 (2007).
2. J. C. He, X.L. Yang and L.R. Wang, *Infrared Millim. Waves*, **26**(4), 317-320 (2007).
3. A. Savitzky and M.J.E. Golay, *Anal. Chem.* **36**(8), 1627-1637 (1964).
4. K. Nakanishi, A. Hashimoto, M. Kanou, T. Pan and T. Kameoka, *Appl. Spectrosc.* **57** (12), 1510-1516 (2003).
5. T. Pan, A. Hashimoto, M. Kanou, K. Nakanishi and T. Kameoka, *Bioprocess and Biosystems Eng.* **26**(2), 133-139 (2003).
6. J. M. Chen, T. Pan and X.D. Chen, *Opt. Precis. Eng.* **14** (1), 1-7(2006).
7. P. Cao, T. Pan and X.D. Chen, *Opt. Precis. Eng.* **15**(12), 1952-1958 (2007).
8. J. Xie, T. Pan, J.M. Chen, H.Z. Chen and X.H. Ren, *Chinese J. Anal. Chem.* **38**, 342-346 (2010).

Reference paper as:
T. Pan, Z. Chen and Y. Shan (2012).Near infrared spectroscopic analysis of chemical oxygen demand (COD) in wastewater produced by cane sugar manufacture, in: Proceedings of the15th International Conference on Near Infrared Spectroscopy,
Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 282-284.