Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about *Accepted Manuscripts* in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard <u>Terms & Conditions</u> and the <u>Ethical guidelines</u> still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



www.rsc.org/methods



69x32mm (112 x 112 DPI)

Analytical Methods Accepted Manuscript

Analytical Methods Accepted Manuscript

A rapid determination method for chemical oxygen demand in aquaculture wastewater using ultraviolet absorbance spectrum and chemometrics

Hong Cao¹, Wentai Qu^{2,*}, Xianglong Yang^{1,2}

- College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China. E-mail: cacao1987@zju.edu.cn
- School of Information Science and Engineering, Ningbo Institute of Technology, Zhejiang University, Ningbo 315100, China

Abstract: Ultraviolet (UV) absorbance spectrum was studied for the fast determination of chemical oxygen demand (COD) in aquaculture wastewater. Four calibration methods were investigated and compared in building the model, including multiple linear regression, partial least squares, least squares support vector machine and back-propagation artificial neural network (BP-ANN). In order to eliminate the superfluous information within the UV spectrum, variable selection method of successive projections algorithm (SPA) was taken into consideration to select efficient wavelengths (EWs), which was first time applied in the UV spectroscopic technique for the COD determination in aquaculture wastewater. Six different spectral pretreatment methods were compared. The best prediction result was obtained by the BP-ANN model based on the 11 EWs with the optimum spectral pretreatment method of standard normal variate, and the coefficient of determination of prediction was 0.90, root mean squared error of prediction was 10.96 mg/L and residual predictive deviation was 5.06. The overall results obtained revealed that SPA is a favourable method to select UV spectral EWs, the usage of EWs in model calibration could obtain better prediction results in comparision with full range UV spectrum, and the combination of UV absorbance spectral EWs with chemometric analysis could be applied for the rapid determination of COD in aquaculture wastewater successfully.

Key words: chemical oxygen demand; aquaculture wastewater; chemometrics; ultraviolet absorbance spectrum

Introduction

Aquaculture is an important source of animal food production currently, and has the potential to provide more than half of human's seafood consumption all over the world. Since 1970, the growth of aquaculture production has an annual average increasement of 8.3%, the

^{*} Corresponding author. E-mail: wentaiqu@zju.edu.cn

Analytical Methods

sustained growth is mainly because of intensification of fish ¹. However, intensive aquaculture is often accompanied with the discharged wastewater containing high concentrations of dissolved organic carbon, nitrogenous compounds and phosphorous, which have caused serious environmental problems ². For example, numerous microorganisms take organic carbon as energy substrate, and the consumption of organic carbon will lead to the problem of insufficient dissolved oxygen in water, which is a threat to aquatic organisms. In order to abide by strict government regulations concerning water quality and pollution, aquaculture enterprises have to reduce the pollution degree of the wastewaters before they are discharged into the environment. Therefore, it is important to monitor the quality of aquaculture wastewater rapidly in order to discover problems and to make sure the discharged wastewater have no negative environmental impact.

The quality of aquaculture wastewater can be assessed by many parameters, such as biological oxygen demand (BOD), chemical oxygen demand (COD) and nitrate. Among these parameters, the COD is an important evaluation indicator of organic pollution, it is a more suitable parameter for the evaluation of water pollution in China. Traditionally, the COD value of aquaculture wastewater is measured by manual sampling procedures with the off-line laboratory analysis. The standard COD measurements are proven to give reliable results, however, they have several drawbacks, besides the sampling and sample storage problems, they are time consuming, costly and detrimental to the environment. Furthermore, these methods provide just a snapshot of the water quality, making them not suitable for on-line measurement and control ³.

There have been a lot of new methods that have been applied for monitoring COD rapidly in recent years ⁴⁻⁶, for example, highly organized arrays of TiO2 nanotubes were integrated into a thin-cell photocatalytic reactor for the rapid determination of COD in wastewater samples ⁶. However, considering the high potential techniques for the research and development of rapid COD monitoring, ultraviolet (UV) spectroscopic technique with multiwavelength approach has attracted significant attention and obtained some useful results ⁷⁻¹¹. These studies are on the basis of the knowledge that the interaction between UV light and the sample, and various organic pollutants in wastewater have a strong absorption of UV radiation ¹². For example, multiwavelength approach which used the spectral ranges of 190-350 nm combined with artificial neural network to determine COD value of wastewater samples was studied ³. The multiwavelength approach can obtain favourable results, particularly when monitoring wastewaters that have frequent changes in composition. However, when using this approach, there is a large amount of superfluous information within

Analytical Methods Accepted Manuscript

the UV spectrum ranges of 200-400 nm, reducing the wavelength numbers is required in the model calibration process. Recently, many studies have demonstrated that using effective wavelengths (EWs) can obtain better prediction results than full range spectrum do ^{13, 14}. Therefore, it is necessary to select EWs to develop a favourable model. However, the above studies on the determination of COD using UV spectrum did not take the EWs selection into account. To this end, this research was conducted to select EWs from the UV spectrum range for the COD determination for the first time.

In this research, we carried out the feasibility evaluation of using UV spectrum for rapid determination of the COD in aquaculture wastewater. The specific targets of this study were: (1) to build the quantitative prediction model between the UV spectrum and COD; (2) to obtain EWs based on successive projections algorithm (SPA); (3) to compare the prediction results of different calibration models built by multiple linear recession (MLR), partial least squares (PLS), least squares support vector machine (LS-SVM), and back-propagation artificial neural network (BP-ANN) based on selected EWs, and (4) to distinguish the optimum model for the determination of COD in aquaculture wastewater.

1 Materials and methods

1.1 Samples collection and preparation

In this study, wastewater samples were collected from three aquaculture farms in Hangzhou (N 30°16′ and E 120°12′), China. A total of 135 samples were obtained between June 2012 and December 2012, which covered temporal and spatial variations. Sampling time was 07:00-11:00 a.m. Wastewater samples were subjected to UV spectrum and COD measurements immediately. The total 135 samples were divided into two sets randomly, 90 samples were used for calibration and the other 45 samples were used for prediction.

1.2 UV spectrum measurement

Figure 1 shows the raw UV spectrum of the 135 aquaculture wastewater samples. For each sample, a UV-Visible spectrophotometer (Cary 60, Agilent, USA) was used to measure the spectrum. the UV spectrum was obtained from 200 to 400 nm by WinUV V5.0 software (Agilent, USA) at room temperature of 20-22°C. The spectrum acquisition was performed with a wavelength resolution of 1 nm. Deionized water was used for baseline measurement. Sample cuvette was made of quartz, its optical path length was 10 mm. The signal-to-noise (SNR) ratio was low in the spectrum between 200 and 220 nm. Therefore, only spectral data within the ranges of 220 to 400 nm were used as input data for analysis.

Page 5 of 12



Fig. 1 Raw UV spectrum of the 135 aquaculture wastewater samples.

1.3 Standard COD analysis

Standard COD analysis was carried out using the EPA approved methods. The sample was homogenized for 30 s in a blender. The acid digest was performed in a dry thermostat reactor (DRB 200, Hach, USA) for 2 hr at 150°C. Hach TNTplus HR 25-20, 1500 mg/L COD agent with potassium dichromate was used as the oxidizing agent. After the sample was cool down to room temperature. A portable spectrophotometer (DR 2800, Hach, USA) was used to perform the measurement of COD value. **Table 1** shows the statistical results of the measured COD value of wastewater samples.

Table 1 Mean, minimum, maximum, standard deviation of COD value of the 135 wastewater samples

Data set	Number of	Mean	Minimum	Maximum	Standard deviation
	samples	(mg/L)	(mg/L)	(mg/L)	(mg/L)
Calibration	90	131.29	46.94	212.65	33.24
Prediction	45	130.00	45.17	217.31	34.59

1.4 Spectral data pretreatment

Spectral data pretreatment was applied to improve the performance of the calibration model. In this study, five pretreatment methods were used, including multiplicative scatter correction (MSC), standard normal variate (SNV), Savitzky-Golay smoothing (SGS), first-derivative (1stDer) and second-derivative (2ndDer). The SGS is a digital filter algorithm that fits a polynomial to the spectral data points by using linear least squares. The MSC is a pretreatment technique used to make compensation for scatter effects of sample's spectrum. The derivatives have the ability to correct the baseline drift and eliminate overlapping peaks in the spectrum. By using the SNV transformation, it's possible to eliminate the slope

variation from each spectrum individually caused by scatter effects. PLS models were developed to evaluate the performance of pretreatment analysis. All related calculations were performed using "The Unscrambler V9.7" (CAMO Software, Norway).

1.5 Variable selection method

 Successive projections algorithm (SPA) was used as the EWs selection method in this study. The SPA is a favourable algorithm for resolving the problem of collinearity among spectral wavelength variables, it selects variables with minimal redundancy. The more details of the SPA procedure can be found in the literature ¹⁵. The SPA calculation was operated in MATLAB V7.11 (The Math Works, USA).

1.6 Multivariate calibration methods

Four calibration methods were applied to compare the prediction performance, namely MLR, PLS, BP-ANN and LS-SVM. PLS regression is widely applied in chemometric calibration (Gerlach, Kowalski, & Wold, 1979). As a multivariate statistical technique, PLS has been effectively used for establishing prediction models in spectral data analysis. PLS does spectral matrix and response matrix (the property to be predicted) decomposition simultaneously, eliminates the useless information. The more detailed PLS procedure can be found in the literature ¹⁶. "The Unscrambler V9.7" (CAMO Software, Norway) was used to develop the PLS models.

The MLR is a common calibration method which is simple and intuitive. However, it fails to work when the number of variables exceeds that of samples. In this study, the wavelength number of spectrum was more than that of the wastewater samples. Therefore, MLR method can not be used directly, it is necessary to select EWs before modeling with MLR. The MLR models were also developed by the software "The Unscrambler V9.7" (CAMO Software, Norway).

Support vector machine (SVM) is a statistical learning method based on structural risk minimization, it has the capability of nonlinear multivariate calibration ¹⁷. SVM has been successfully applied in pattern recognition and function regression ^{18, 19}. LS-SVM is an improved method of SVM which was proposed by Suykens ²⁰. When modeling with LS-SVM, kernel function, and model parameters are two key parameters to optimize the model. In this study, the kernel function of LS-SVM was the radial basis function (RBF). Leave one out cross-validation and grid search technique were applied to obtain the optimal model parameter sigma²

Analytical Methods

 (σ^2) . The details of the LS-SVM method can be found in the literature ²⁰. The LS-SVM models were developed with the free LS-SVM V1.8 toolbox using MATLAB V7.11 (The Math Works, USA).

Artificial neural network (ANN) is a multivariate calibration method that simulates the principle of the human brain. BP-ANN is a common type of the ANN model because of its powerful nonlinear mapping capability. The BP-ANN modeling process is carried out with the adjustment of each node's weight by using transfer function f(t). In this research, the sigmoidal squashing function was used as the transfer function. The function is represented as Eq. (1):

$$f(t) = \frac{1}{1 + e^{-at}}$$
(1)

where t is the slope parameter of the sigmoid squashing function. Continuous adjustment of weights are made until the validation errors are acceptable.

1.7 Performance indices

To evaluate the performance of the proposed calibration methods, the following indices are used: root mean squared error of prediction (RMSEP), residual predictive deviation (RPD) and coefficient of determination of prediction (R^2_{pre}). RPD is defined as the standard deviation of prediction samples' reference values divided by the standard deviation of prediction error. Generally, an accurate and robust model should be with lower RMSEP value and higher R^2_{pre} and RPD value.

2 Results and discussion

2.1 Selecting the best pretreatment method

To get the evaluation results of six different pretreatment methods, PLS models were developed with spectrum within the ranges of 220 to 400 nm after pretreatment. The prediction results of the PLS models were shown in **Table 2**. After the pretreatment of SGS, the prediction result did not get improved obviously compared to that of the PLS model with raw spectrum. This is because that the raw spectrum has little noise, which shows that the Cary 60 UV-Vis spectrometer can acquire the spectrum of aquaculture wastewater with high quality. The best prediction result was obtained on the basis of the pretreatment method of SNV, the value of R^2_{pre} was 0.74, RMSEP was 17.60 mg/L and RPD was 2.64. Compare the values of RPD, MSC-PLS model, 1stDer-PLS model and 2stDer-PLS model were 96.21%,

Analytical Methods Accepted Manuscript

95.83% and 94.70% that of SNV-PLS model. Therefore, SNV was selected as the pretreatment method for further study.

Pretreat-	Variable	Variable	Calibration	Latent	R ² _{pre}	RMSEP	RPD
ment	selection	number		Varaibles		(mg/L)	
Raw	/	181	PLS	3	0.70	18.59	2.31
SG	/	181	PLS	3	0.71	18.56	2.31
SNV	/	181	PLS	8	0.74	17.60	2.64
MSC	/	181	PLS	8	0.72	18.19	2.54
1 st Der	/	181	PLS	2	0.73	17.71	2.53
2 nd Der	/	181	PLS	3	0.71	18.27	2.50
SNV	SPA	11	MLR	/	0.79	15.85	3.11
SNV	SPA	11	PLS	8	0.78	15.91	3.06
SNV	SPA	11	BP-ANN	/	0.90	10.96	5.06
SNV	SPA	11	LS-SVM	/	0.83	13.93	3.39

2.2 EWs selection based on SPA

SPA was applied to the EWs selection from spectrum within the ranges of 220 to 400 nm. **Figure 2** shows the scree plot of root mean squared error (RMSE) which was generated by SPA. It was used to explain the selection procedures. As shown in **Fig. 2**, the curve trend of RMSE was gentle in the beginning when the selected variable number was from 1 to 2, then a severely fall was followed when the selected variable number was from 2 to 3. After the selected variable number was determined by F-test with significance level $\alpha = 0.25$, the curve flattened out. The black solid circle means that 11 (RMSE = 15.85) variables were selected as EWs. **Figure 3** shows the locations of the selected EWs. In this study, just 6.08% number of spectrum wavelengths within the ranges of 220 to 400 nm were selected as EWs for developing models, It was meaningful for instrument development.



 $\begin{array}{c} 1.6 \\ \begin{array}{c} 225 \\ 243 \\ 248 \\ 248 \\ 0.8 \\ 0.4 \\ 240 \\ 240 \\ 240 \\ 240 \\ 280 \\ 320 \\ 360 \\ 400 \\ \end{array}$

Fig.3 Selected EWs with the SNV spectrum by SPA.

2.3 Comparison of calibration models with EWs

The selected EWs were taken as the input variables of the PLS model, and **Table 2** shows the prediction result. For the SNV-SPA-PLS model, the value of R^2_{pre} was 0.78, the value of RMSEP was 15.91 mg/L and the value of RPD was 3.06. Compared with the SNV-PLS model which was built based on spectrum within the ranges of 220 to 400 nm, the SNV-SPA-PLS model obtained the value of R^2_{pre} increased by 5.41%, the value of RMSEP decreased by 9.60% and the value of RPD increased by 15.91%. Moreover, it should be pointed out that, the SNV-SPA-PLS prediction model had only 11 input variables, which were only 6.08% that of the SNV-PLS prediction model with spectrum within the ranges of 220 to 400 nm. The obtained results indicated that SPA is a good method to select EWs in COD analysis.

Besides the PLS method, we bulit COD determination models based on the selected EWs using MLR, LS-SVM and BP-ANN methods, respectively. **Table 2** shows the prediction results of the three different models. Comparing the results of the four above models, the best prediction result was obtained by the SPA-BP-ANN model with SNV spectrum, the value of R^2_{pre} , RMSEP and RPD was 0.90, 10.96 mg/L and 5.06. RPD value up

Analytical Methods Accepted Manuscript

to 5 shows an accurate prediction ability for COD determination ²¹. Although in many instances, the LS-SVM has indicated a great prediction capability for spectral analysis techniques ²². In this research, in comparison with the SNV-SPA-BP-ANN model, the SNV-SPA-LS-SVM model obtained the value of R^2_{pre} decreased by 7.78%, the value of RPD decreased by 28.78% and the value of RMSEP increased by 27.10%. The above results demonstrated that BP-ANN is an efficient method to establish the prediction model in COD determination with EWs. Better prediction results were obtained by the LS-SVM and BP-ANN models than the MLR and PLS models, which revealed that nonlinear calibration method was more appropriate than linear calibration method for predicting the COD value of aquaculture wastewater. **Figure 4** shows the scatter plot for prediction set by SPA-BP-ANN according to SNV spectrum.



Fig. 4 Relation between reference and predicted COD value by SPA-BP-ANN according to the SNV spectrum in prediction set.

Conclusions

UV spectroscopic technique combined with multivariate analysis was applied for the rapid determination of COD in aquaculture wastewater successfully. Six different spectral pretreatment methods were compared, SNV was selected as the optimized pretreatment method by PLS. Variable selection method of SPA was applied to select EWs, out of 181 wavelengths, only 11 relevant EWs was selected. Four calibration methods were investigated in building the model, the SPA-BP-ANN model with SNV spectrum gave the best performance in this study, the corresponding R^2_{pre} was 0.90, RMSEP was 10.96 mg/L and RPD was 5.06, the good RPD value revealed an accurate prediction ability for COD determination. On the basis of these EWs, a simple spectroscopic system can replace the off-line manual analysis of COD.

As far as we know, this is the first research in which EWs selection was conducted in

Analytical Methods

employing the UV spectroscopic technology for the determination of COD of aquaculture wastewater. The selected EWs possess key information related to the COD, they are beneficial for improving the model prediction precision and calibration speed, and interpreting which wavelengths are useful for COD determination.

In this article, a preliminary study on COD determination was presented. Further research would be focused on expanding the spectral region, for example in the NIR spectra region, increasing the number of wastewater samples and optimizing variable selection methods. Moreover, it is also important to compare prediction results under different aquaculture environments for further study.

Acknowledgements

This study was supported by the National Science and Technology Major Project of the Ministry of Science and Technology of China (2008ZX07420-004) and the Natural Science Foundation program of Ningbo (No. 2011A610135).

References

- 1. D. Klinger and R. L. Naylor, *Annu Rev Env Resour*, 2012, **37**, 247-276.
- 2. W. T. Mook, M. H. Chakrabarti, M. K. Aroua, G. M. A. Khan, B. S. Ali, M. S. Islam and M. A. Abu Hassan, *Desalination*, 2012, **285**, 1-13.
- 3. S. Fogelman, H. J. Zhao and M. Blumenstein, *Analytical and Bioanalytical Chemistry*, 2006, **386**, 1773-1779.
- 4. K. H. Lee, T. Ishikawa, S. J. McNiven, Y. Nomura, A. Hiratsuka, S. Sasaki, Y. Arikawa and I. Karube, *Analytica Chimica Acta*, 1999, **398**, 161-171.
- 5. Y. C. Kim, S. Sasaki, K. Yano, K. Ikebukuro, K. Hashimoto and I. Karube, *Analytica Chimica Acta*, 2001, **432**, 59-66.
- 6. Q. Zheng, B. X. Zhou, J. Bai, L. H. Li, Z. J. Jin, J. L. Zhang, J. H. Li, Y. B. Liu, W. M. Cai and X. Y. Zhu, *Advanced Materials*, 2008, **20**, 1044-1049.
- 7. A. Torres and J. L. Bertrand-Krajewski, *Water Science and Technology*, 2008, **57**, 581-588.
- 8. M. C. Sarraguca, A. Paulo, M. M. Alves, A. M. A. Dias, J. A. Lopes and E. C. Ferreira, *Analytical and Bioanalytical Chemistry*, 2009, **395**, 1159-1166.
- 9. X. S. Qin, F. R. Gao and G. H. Chen, *Water Research*, 2012, 46, 1133-1144.
- 10. Y. Q. Wu, S. X. Du and Y. Yan, *Spectrosc Spect Anal*, 2011, **31**, 233-237.
- 11. G. Wu, W. Bi, J. Lv and G. Fu, *Chin Opt Lett*, 2011, 9(s1), s10705.
- 12. A. Charef, A. Ghauch, P. Baussand and M. Martin-Bouyer, *Measurement*, 2000, 28, 219-224.
- 13. R. M. Balabin and S. V. Smirnov, *Analytica Chimica Acta*, 2011, **692**, 63-72.
- 14. D. Wu, P. C. Nie, Y. He and Y. D. Bao, *Food Bioprocess Tech*, 2012, 5, 1402-1410.
- 15. R. K. H. Galvao, M. C. U. Araujo, W. D. Fragoso, E. C. Silva, G. E. Jose, S. F. C. Soares and H. M. Paiva, *Chemometr Intell Lab*, 2008, **92**, 83-91.
- 16. S. Wold, M. Sjostrom and L. Eriksson, *Chemometr Intell Lab*, 2001, **58**, 109-130.

- 17. V. Vapnik, *The Support Vector method of function estimation*, Kluwer Academic Publishers, Norwell, 1998, 55-85.
- 18. C. H. Zhang, Y. J. Tian and N. Y. Deng, *Sci China Math*, 2010, **53**, 151-164.
- 19. S. X. Du, X. L. Wu and T. J. Wu, *Chinese J Anal Chem*, 2004, **32**, 1227-1230.
- J. A. K. Suykens, T. van Gestel and J. de Brabanter, *Least Squares Support Vector Machines*, World Scientific Publishing Company Incorporated, Singapore, 2002, 71-90.
- 21. P. Williams and K. H. Norris, *Near-infrared technology in the agricultural and food industries*, American Association of Cereal Chemists, Saint Paul, 1987, 125-130.
- 22. W. W. Kong, Y. Zhao, F. Liu, Y. He, T. Tian and W. J. Zhou, *Sensors-Basel*, 2012, **12**, 10871-10880.