Preliminary Study of Determination of *Dinophysis* Concentration Using Hyperspectral Imaging

Ye Shen[1], Hui Huang[1,*], Hong Song[1], Shuyue Zhan[1], Jun Han[1], Ying Chen[1]

1. Ocean College, Zhejiang University, Hangzhou, 310058;

*Corresponding author: huih@zju.edu.cn

Abstract—Conventional measurement methods were based on microscope, which was time-consuming and complex operation. In this paper, the aim of this work was to preliminary study the feasibility of the rapid and precision determination of *Dinophysis* concentration at the wavelength of 495-1030 nm. The transmittance spectra of algae liquid were obtained by hyperspectral transmittance image system. The region of interest (ROI) of each hyperspectral images was segmented and the mean spectra data were extracted from the ROI, then the data was arranged in a matrix and applied to further analysis. Partial least square regression (PLSR) was then used to build model with full spectral wavelengths. Nine wavelengths were chosen as the optimal wavelengths using the successive projections algorithm (SPA) to establish the new simplified model (SPA-PLSR) to correlate the selected spectral feature with *Dinophysis* concentration. By comparing the performance of different models, the best result was by PLSR based on optimal spectral wavelengths with correlation coefficient (R) of 0.9811. The preliminary study revealed the possibility of hyperspectral transmittance imaging as a rapid method to determine *Dinophysis* concentration.

Keywords—Dinophysis concentration; Image processing; Hyperspectral transmittance imaging; Successive projections algorithm; Preliminary study

I. INTRODUCTION

Microalgae, which contains abundant nutritional and bioactive components, is an important biomass resource in ecology and biology [1, 2]. Fast growth, reproduction and high photosynthesis efficiency are representative characteristics of microalgae [3]. Because of the raising problem of energy, food and environment, microalgae biomass industry has gained much more attention recently. However, amongst all the species, some algae are stressed because of its close relationship to red tide, shellfish poisoning, as well as human health. *Dinophysis* is one of these types of toxic algae [4]. Uncontrolled growth of *Dinophysis* may have harmful effect on marine ecosystems by making shellfish poisonous, resulting socioeconomic consequences in the coastal areas, even at low cell densities ($<10^3$ cells·L^{-1}) [5, 6]. Currently the mostly used measurement methods of *Dinophysis* concentration are based on microscopy. Andersen applied conventional microscopy observation to determine the concentration of *Dinophysis* acuminate [7]. Sousa *et al.* used highly sensitive quartz crystal microbalance to detect algae concentration [8]. Although these methods were accurate, complicated sample preparation was involved, which makes them time-consuming and professional technicians required [9]. Therefore, to develop a method for the rapid determination of *Dinophysis* concentration would be beneficial to the tide control and marine environmental protection. Furthermore, it can also prevent food poisoning.

978-0-933957-43-5 ©2015 MTS

Traditional imaging technology was used for algae classification, for its ability of distinguishing morphological information of algae. However, imaging technology only provided limited spectral information [10, 11], which was mainly used for monitoring the external property instead of the internal property [12, 13]. With recent technological progress in spectroscopy and imaging, hyperspectral imaging technology can overcome the limitation of the conventional imaging technique and provide rich spectral and spatial information. Due to the characteristics of spatial information, region of interests (ROIs) of *Dinophysis* can be identified and the spectral data of a sample image was extracted and arranged in a matrix where the rows represented the number of samples and the columns represented the number of variables to developed and applied to agricultural fields [14, 15, 16] and fisheries fields [17, 18, 19]. There were few studies on microalgae using the hyperspectral imaging technology were reported. Mehrubeoglu *et al*. (2012) revealed the capability to differentiate the mixed algae species in liquid media using hyperspectral imaging technology [20]. Davis *et al*. (2014) employed hyperspectral fluorescence microscopy with multivariate curve resolution to track the morphological dynamics and local concentrations of chlorophyll, carotenoids, and lipids in multiple microalgae species in liquid media [21]. These studies implied the possibility of hyperspectral imaging for determination of algae concentration in liquid media. The aim of this work was to evaluate the potentiality of using the hyperspectral transmittance imaging (400-1030 nm) for determining *Dinophysis* concentration. The specific objectives included: sample preparation, imaging system set-up, image acquisition and calibration, wavelength optimization, and data analysis.

II. MATERIAL AND METHODS

A. Sample Preparation

Dinophysis were obtained from the Department of Marine Science, Ocean College, Zhejiang University. The *Dinophysis*, whose original concentration was 4000 cells·ml^{-1}, was diluted five times. Six groups of different concentration, 4000 cells·ml^{-1}, 2000 cells·ml^{-1}, 4000/3 cells·ml^{-1}, 1000 cells·ml^{-1}, 800 cells·ml^{-1} and 2000/3 cells·ml^{-1} were obtained, labeled and then transported to the laboratory of Hyperspectral Imaging, Zhejiang University. Each group was imaged individually using the hyperspectral imaging system described below.

B. Hyperspectral Imaging System

Fig.1 showed a hyperspectral transmittance imaging system which consisted of a zoom lens, controllable conveyor, sample-loading platform, a imaging spectrograph (ImSpectorV10E, Spectral Imaging Ltd., Oulu, Finland, 400-1030 nm), a high performance CCD camera (672 × 512 pixels) (A312f, Basler vision technologies, Ahrensburg, Germany) with a spectral resolution of 2.8 nm, was attached to the spectrograph, and a computer system equipped with the Spectral Image System V10E software which controlled the motor speed, exposure time, image acquisition, and image correction. The light source, which consisted of a 20W tungsten halogen lamp, a regulator and a transformer, was installed below the sample.

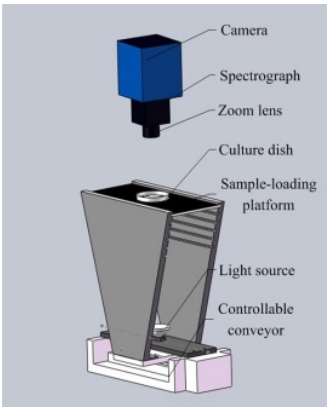

Fig.1 Hyperspectral transmittance imaging system

C. Image Acquisition And Calibration

Using the system described in Fig.1, the sample was placed in the culture dish on the sample-loading platform and was scanned line by line. The speed and exposure time of the conveyor were adjusted prior to image acquisition. The conveyor moved at a speed of 2.82 mm/s and the exposure time was 0.095s. After scanning, for each sample, a three dimensional hypercube (x×y×λ) was obtained and be stored in raw format for further analysis, which contained both spatial (x×y) and spectral (λ) information.

To eliminate the influences of detector sensitivity and illumination, two images recorded as bright and dark references were acquired for image correction before acquiring the hyperspectral image. The dark image was acquired by covering the camera lens completely and the bright image was obtained by transmitting light through the glass. Finally, the corrected transmittance images were obtained by the following equation:

$$R = \frac{A - D}{B - D} \tag{1}$$

Where R was the corrected transmittance image, A was the raw transmittance image, B was the white reference image and D was the dark reference image. All the further analyses were conducted on the corrected transmittance images.

D. ROI Identification

After image acquisition and calibration, the region of interest (ROI) corresponding to sample part were identified (Fig.2). The ROI identification was carried out based on the binary image. Image spectral data of each tested sample within ROIs were extracted and arranged together to form a matrix. The value in each unit of the matrix referred to the transmittance value of a sample at a specific wavelength, and then processed by multivariate data analysis methods. The quantitative models, which were the relationship of the concentration and the spectral data, were established using PLSR algorithm. It aimed to search the best function to correlate spectroscopic data (***X***) with *Dinophysis* concentration (***Y***), where ***X*** (size $\boldsymbol{m}\times\boldsymbol{n}$) was the matrix of calibration spectra, ***m*** was the number of samples, ***n*** was the number of variables.

E. Optimal Wavelength Selection

Optimal wavelength selection played an important role in data analysis. Due to the huge hyperspectral data, the extracted spectral data at whole spectral wavelengths was considered redundant,

which may influence the calibration efficiency and increase the calculate burden. Therefore, successive projections algorithm (SPA) was used to eliminate irrelevant variables to simplify the calibration modeling and improve the accuracy of the results [23, 24]. Then the performances of established models were evaluated and compared to verify the feasibility for measuring the *Dinophysis* concentration. The main steps of hyperspectral imaging technique for evaluating the feasibility for measuring *Dinophysis* concentration were shown in Fig.2.

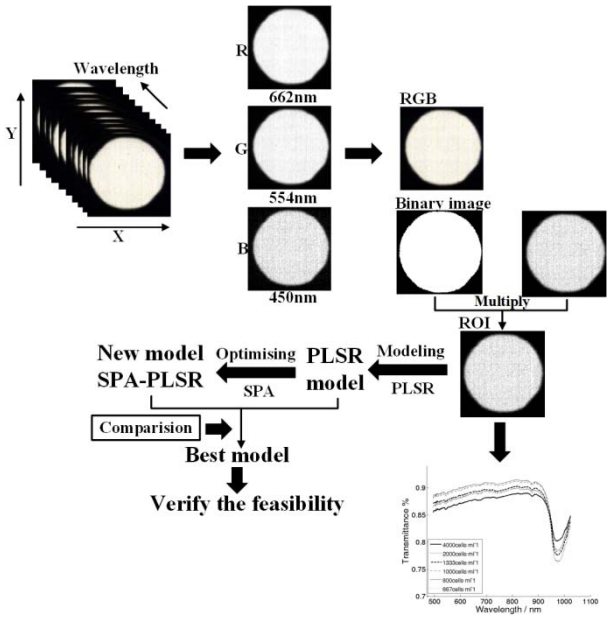

Fig.2. Main step of image segmentation and data analysis.

III. RESULTS AND DISCUSSION

A. Spectral Features Of Dinophysis

The mean spectra (495-1030 nm) of different *Dinophysis* concentration were shown in Fig.3. Due to the presence of noise in the range of 380-494 nm, only spectral wavelengths at the range of 495-1030 nm were used for further analysis (416 wavelengths). The Fig.3 showed that different *Dinophysis* concentration had similar spectral trend, however, there were some difference in the range between 500 nm and 900 nm. It was clear that the transmittance spectra of the *Dinophysis* were varied with concentration. The higher concentration had lower transmittance.

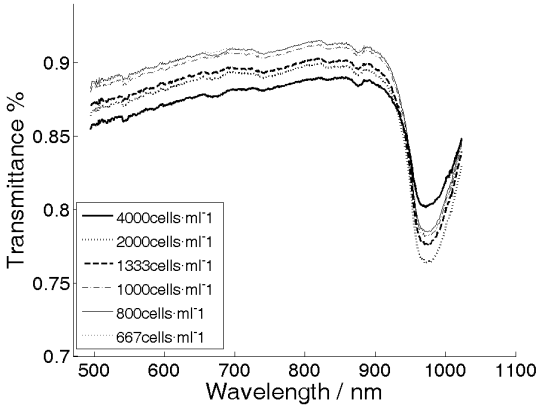

Fig.3 Mean spectra (495-1030 nm) of different *Dinophysis* concentration

B. Data Analysis

When spectral data were extracted and arranged in a matrix (***X, m*** × ***n***), and then the PLSR algorithm was applied to data analysis. The row (***m***) represented the number of samples (***m***=6), the columns (***n***=416) was the number of variables. The PLSR model with full spectral wavelengths was established and the correlation coefficient (R=0.9819) was obtained. It could be noticed that the higher the value of R, the better the model [25]. In Fig.4, it was clear that the measured and predicted values highly linear correlation. Therefore, the result suggested the potentiality of determination of *Dinophysis* concentration.

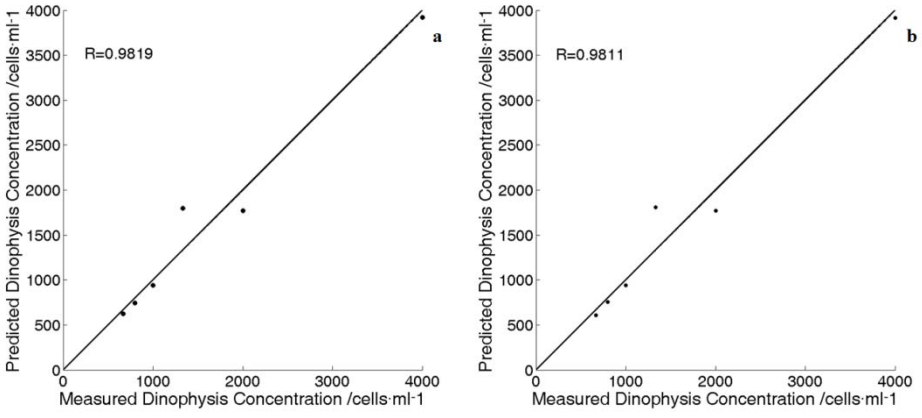

Fig.4 Measured vs. predicted values for the F-PLSR models (a) and the SPA-PLSR models (b)

Due to the extremely large hyperspectral data and some wavelength bands might contain irrelevant information, it was necessary to select the optimal wavelengths to simplify the model and reduce the computational burden. In this study, successive projections algorithm (SPA) was used to remove the redundant variables with less useful information and to select effective wavelength variables from the whole spectra for modeling. In Fig.5, it was obvious that nine variables (533, 584, 619, 747, 817, 909, 951, 1012 and 1022 nm) were selected as the optimal wavelengths and the percentage of spectral variables reduced by 97.84%. The optimal wavelengths were then used to replace the whole spectral wavelengths to develop simplified PLSR model, also named SPA-PLSR, which were adopted to predict the *Dinophysis* concentration. As shown in Table 1, the performance of PLSR and SPA-PLSR model was similar, demonstrating that SPA was useful for simplifying the PLSR model and enhancing the efficiency of the measurement.

Table.1 The performance comparison of determination of *Dinophysis* concentration using PLSR model and SPA-PLSR model

Model	Variable number	Correlation coefficient (R)
PLSR	416	0.9819
SPA-PLSR	9	0.9811

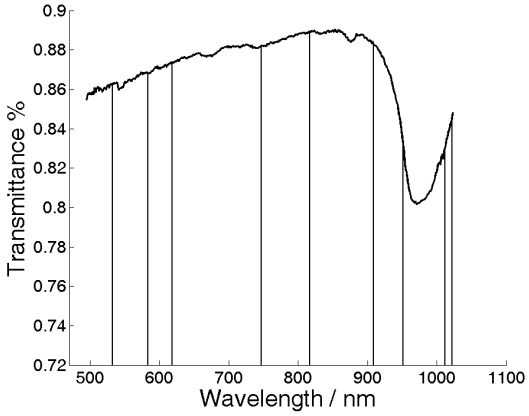

Fig.5 Plot of nine selected wavelength variables by SPA. Columns represent selected wavelength variables.

When the optimal wavelengths were selected, the spectral data were rearranged and formed a new matrix ($\boldsymbol{X_1}$, $\boldsymbol{m_1} \times \boldsymbol{n_1}$) where $\boldsymbol{m_1}$ (6) was the number of samples and the variables $\boldsymbol{n_1}$ was reduced from 416 to 9 wavelengths. PLSR algorithm was then carried out by inputting the simplified spectral dataset to develop the new model and the result were shown in Table 1. It could be found from Fig.4 that the results of SPA-PLSR model and PLSR model based on whole spectral wavelengths (F-PLSR) were similar. In addition, the SPA-PLSR model was less over-fitted than the F-PLSR model and the value of R was 0.9811, which was considered as a well model. Therefore, the SPA-PLSR model was suggested as the optimum calibration model for the determination of *Dinophysis* concentration. The obtained SPA-PLSR function was shown as follows:

$$\boldsymbol{Y}=6547.5-4.3023\boldsymbol{X}_{533nm}-4.3948\boldsymbol{X}_{584nm}-4.5529\boldsymbol{X}_{619nm}-4.8168\boldsymbol{X}_{747nm}-4.9827\boldsymbol{X}_{817nm}-5.1334\boldsymbol{X}_{909nm}+3.9650\boldsymbol{X}_{951nm}+4.4317\boldsymbol{X}_{1012nm}+3.2022\boldsymbol{X}_{1022nm} \quad (2)$$

where $\boldsymbol{Y}$ was the estimated *Dinophysis* concentration, $\boldsymbol{X_i}$ was the value of spectra at the wavelength of i nm.

IV. CONCLUSIONS

The results showed that hyperspectral transmittance imaging technology together with PLSR and SPA could be used for determination of *Dinophysis* concentration. Although the performance of SPA-PLSR model and F-PLSR model was similar, the SPA-PLSR model was less over-fitted and line correlation, therefore, SPA could be applied to simplify the PLSR model. By comparing the two types of model, the relative better model (R=0.9811) obtained by PLSR based on optimal selected wavelengths demonstrated that the potential of using hyperspectral transmittance imaging technology to measure *Dinophysis* concentration. Because of this work was preliminary study, it was necessary to do further improvement. The number of sample was pretty small and therefore lacked of convincing. In the future, wide concentration ranges should be studied to establish more accurate and convincing model. Furthermore, other calibration algorithms and variable selection algorithms should be considered to select the best wavelength variables to establish the better model.

ACKNOWLEDGEMENT

The authors sincerely thank the support of the lab of hyperspectral imaging and many researchers who friendly offered their time and resources for this project.

REFERENCES

[1] Duffy, J.E., *et al*. Biofuels: Algae. *Science*, 2009. 326(5958): p. 1345.

[2] Haag, A.L., Algae bloom again. *Nature*, 2007. 447(7144): p. 520.

[3] Spolaore P., *et al*., Commercial applications of microalgae, *Journal of Bioscience and Bioengineering*, 2006. 101(2): p. 87-96.

[4] Heisler J., *et al*., Eutrophication and harmful algal blooms: A scientific consensus. *Harmful Algae*, 2008. 8(1):p. 3-13.

[5] Singh, A., *et al*., An assessment of Dinophysis blooms in the coastal Arabian Sea. *Harmful Algae*, 2014. 34: p. 29-35.

[6] Beatriz R., *et al*., *Dinophysis* Toxins: Causative Organisms, Distribution and Fate in Shellfish. *Marine Drugs*, 2014. 12: p. 394-461.

[7] Andersen, P., Relation between the concentration of *Dinophysis* acuminata and Diarrhetic shellfish poisoning toxins in blue mussels during a toxic episode in the Limfjord(Denmark),2006. *Journal of Shellfish Research*, 2007. 26(4): p. 1081-1087.

[8] Sousa, C., *et al*., Direct and fast detection of Alexandrium minutum algae by using high frequency microbalance. *Journal of Microbiological Methods*, 2014. 104: p. 49-54.

[9] Reguera, B., *et al*., Harmful Dinophysis species: A review. *Harmful Algae*, 2012. 14: p. 87-106.

[10] He, H., *et al*., Rapid and non-destructive determination of drip loss and pH distribution in farmed Atlantic salmon (Salmo salar) fillets using visible and near-infrared (Vis–NIR) hyperspectral imaging. *Food Chemistry*, 2014. 156: p. 394-401.

[11] Dowlati, M., *et al*., Application of machine-vision techniques to fish-quality assessment. *Trends in Analytical Chemistry*, 2012. 40: p. 168-179.

[12] Bakker, W.H., *et al*., Mapping the wavelength position of deepest absorption features to explore mineral diversity in hyperspectral images. *Planetary and Space Science*, 2014. 101: p. 108-117.

[13] Zhang, W., *et al*., Non-destructive internal quality assessment of eggs using a synthesis of hyperspectral imaging and multivariate analysis. *Journal of Food Engineering*, 2015. 157: p. 41-48.

[14] Huang, M., *et al*., Detection of insect-damaged vegetable soybeans using hyperspectral transmittance image. *Journal of Food Engineering*, 2013. 116(1): p. 45-49.

[15] Mahesh, S., *et al*., Hyperspectral imaging to classify and monitor quality of agricultural materials. *Journal of Stored Products Research*, 2015. 61: p. 17-26.

[16] Siripatrawan, U., *et al*., Monitoring fungal growth on brown rice grains using rapid and non-destructive hyperspectral imaging. *International Journal of Food Microbiology*, 2015. 199: p. 93-100.

[17] Dai, Q., *et al*., Potential of visible/near-infrared hyperspectral imaging for rapid detection of freshness in unfrozen and frozen prawns. *Journal of Food Engineering*, 2015. 149: p. 97-104.

[18] Cheng, J., *et al*., Rapid and non-invasive detection of fish microbial spoilage by visible and near infrared hyperspectral imaging and multivariate analysis. *LWT - Food Science and Technology*, 2015. 62(2): p. 1060-1068.

[19] Wu, D., *et al*., Rapid prediction of moisture content of dehydrated prawns using online hyperspectral imaging system. *Analytica Chimica Acta*, 2012. 726: p. 57-66.

[20] Mehrubeoglu, M., *et al*., Hyperspectral Imaging and Analysis of Mixed Algae Species in Liquid Media. 2012.

[21] Davis, R.W., *et al*., Label-free measurement of algal triacylglyceride production using fluorescence hyperspectral imaging. *Algal Research*, 2014. 5: p. 181-189.

[22] Macho, S., *et al*., Monitoring ethylene content in heterophasic copolymers by near-infrared spectroscopy standardization of the calibration model. *Analytica Chimica Acta*, 2001. 445(2): p. 213-220.
[23] He, H., *et al*., Non-destructive and rapid analysis of moisture distribution in farmed Atlantic salmon (Salmo salar) fillets using visible and near-infrared hyperspectral imaging. *Innovative Food Science & Emerging Technologies*, 2013. 18: p. 237-245.
[24] Liu, K., *et al*., A consensus successive projections algorithm – multiple linear regression method for analyzing near infrared spectra. *Analytica Chimica Acta*, 2015. 858: p. 16-23.
[25] Huang, M., *et al*., Prediction of color and moisture content for vegetable soybean during drying using hyperspectral imaging technology. *Journal of Food Engineering*, 2014. 128: p. 24-30.