PAPER • **OPEN ACCESS**

Recognition of Heavy Metal Contamination Using Tegillarca Granosa LIBS Data

To cite this article: Fengwen Zhou *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **237** 022030

Recognition of Heavy Metal Contamination Using Tegillarca Granosa LIBS Data

Fengwen Zhou, Pengchao Ye, Hongjuan Fu and Guoli Ji*

Department of Automation, Xiamen University, Xiamen, Fujian Province, 361005, China

*Corresponding author's e-mail: glji@xmu.edu.cn

Abstract. The problem of heavy metal pollution in food and the environment has drawn increasing public concern, because of its irreversible damage. Traditional methods of detecting heavy metals require complicated laboratory conditions and skilled experimenters, and cannot meet cost-effectiveness, rapidity and ease of use. In this paper, we propose a principal component analysis and linear discriminant analysis (PCA-LDA) pattern recognition algorithm for rapid detection of heavy metals in tegillarca granosa by using laser-induced breakdown spectroscopy (LIBS) data. First, the principal component analysis (PCA) algorithm is used to reduce the dimension of spectral data from the previous 2811 dimension to 16 dimensions. Second, the linear discriminant analysis (LDA) algorithm is used for classification. Results show that PCA-LDA has the best classification accuracy compared with the partial least squares discriminant analysis (PLS-DA), soft independent modeling of class analogy (SIMCA) and support vector machine (SVM) algorithms, and achieves a discrimination accuracy of 87%. This work indicates that by combining pattern recognition with LIBS technology, healthy tegillarca granosa can be quickly distinguished from heavy metal polluted samples, which offers possibility of environmental monitoring and in field food detection.

1. Introduction

As the industrialization of the world continues to grow, many industrial wastewaters are shown to containing a large amount of toxic heavy metals that are discharged directly into rivers, lakes and oceans[1]. They not only destroy the natural environment but also seriously affect the living conditions of aquatic animals. The consumption of such long-term aquatic products contaminated by heavy metals over a long period of time will eventually endanger human health[2] through enrichment and cause irreversible harm to human beings[3]. Therefore, a simple and efficient method for detecting heavy metals in foods is of crucial importance in ensuring the safety of consumers.

In order to study the pollution of heavy metals in foods, we chose tegillarca granosa as our research object. Tegillarca granosa is a common shellfish, widely distributed in shallow sea sediments in coastal areas of China and in Southeast Asia, with low activity of the physiological characteristics. Due to exposure to heavy metal-contaminated seawater, their bodies are prone to accumulate heavy metal ions and are not easily discharged. Their tissue and structural proteins are also susceptible to strong heavy metal contamination. Due to these unique characteristics, tegillarca granosa can be used as a suitable target for further study of heavy metal pollutions in food or the environment.

Traditional heavy metal detection methods, such as inductively coupled plasma mass spectrometry[4], flame atomic absorption spectroscopy[5] and graphite furnace atomic absorption spectrometry[6], requires a high degree of accuracy and sensitivity but along with complex

experimental procedures that have long detection cycles. Biosensor detection methods are portable, fast and efficient, but require complex sensor preparation processes[7]. Compared with traditional methods biosensor detection methods, spectrometry is also commonly used because of its accuracy and easy use. LIBS is one of the spectroscopic methods, which requires minimal sample pretreatment, so it is suitable for rapid detection of heavy metals in tegillarca granosa. In principle, this method can analyse any element regardless of whether it is a solid, liquid or gaseous[8].

However, the LIBS method has its limitations. Because the reality of the sample composition is complex, different chemicals will cause the overlapping of the characteristic spectral lines. Therefore, the LIBS cannot be used directly for detection alone. This paper presents a method combining LIBS with PCA-LDA algorithms in order to solve the problems of the spectral noise and spectral line drift[9] presented simply by LIBS. Our work is a proof of theoretical research aimed at identifying an appropriate pattern recognition algorithm to distinguish contaminated and safe tegillarca granosa. We hope that in the near future, using this combination of methods for direct detection of heavy metals can facilitate environmental monitoring and food detection.

2. Materials and methods

2.1. Sample Preparation

We prepared three groups of tegillarca granosa samples, respectively labeled Ⅰ, Ⅱ, Ⅲ, each group containing 30 samples. The same group of samples were cultured in the same container of seawater for breeding. In the Ⅰ, Ⅱ, Ⅲ were added to $PbCH_3COO \cdot 3H_2O$, $CdCl_2$ and $ZnSO_4 \cdot 7H_2O$ respectively, which were purchased for the Guangzhou chemical reagent factory, Guangzhou, China.

After more than 24 hours of precipitation, the seawater was filtered to remove the sand used to grow the tegillarca granosa in the tank. The seawater has a pH of 8.1±0.20, a temperature of 21.3±4.8°C, an oxygen content of >5 mg/L, and a salinity of 22%. The other experimental conditions are the same. Experimental conditions set is shown in Table 1.

Table 1. Experimental conditions set.

Group	**Heavy Metal**
Ⅰ	seawater with $PbCH_3COO \cdot 3H_2O$
Ⅱ	seawater with $CdCl_2$
Ⅲ	seawater with $ZnSO_4 \cdot 7H_2O$

After 10 days of culture, each sample was killed and dried. The dried samples were measured using a LIBS spectrometer to ensure that each sample was measured five times at different positions on the translation stage. Shown in Figure 1 is the tegillarca granosa LIBS original spectrum.

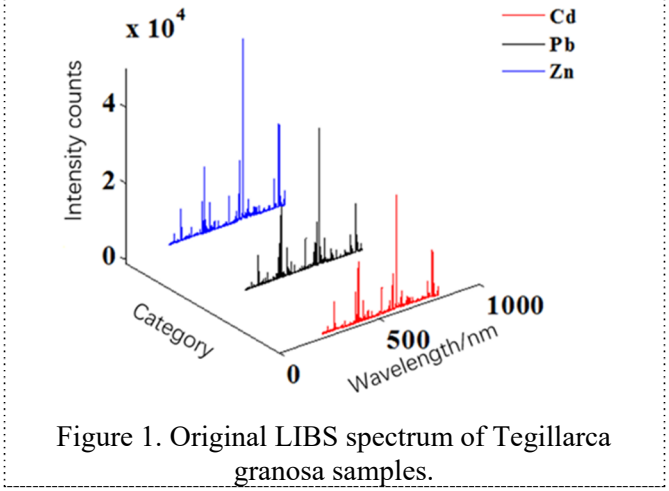

Figure 1. Original LIBS spectrum of Tegillarca granosa samples.

2.2. Parameter Settings

Two-thirds of each sample was selected randomly as the training set, a total of 60 samples, and the remaining 30 samples were used as a prediction set. In PLS-DA and SIMCA, 10-fold venetian blinds were cross-validated within 20 main components and the data were normalized using auto scaling (centralization and normalization). PLS-DA uses bayes as the assignation criterion, and PCA-LDA uses Linear to identify it. The search range of SVM c, g is (2-5, 2-4, 2-3, 2-2, 2-1, 20, 21, 22, 23, 24, 25).

2.3. Principal Component Analysis (PCA) Algorithm

PCA is a commonly used dimensionality reduction technique[10]. The idea of PCA is to map m-dimensional features onto d-dimensional features orthogonally with the largest variance. This reconstructed d-dimensional feature is called the principal component. Below, we formulate this problem mathematically. Assume that the sample set is $X=\{x_1,x,...,x_m\}$. The projection of the sample point x_i on the hyperplane in the new space is $W^T x_i$. If the projections of all sample points can be separated as much as possible, the variance of the sample points after the projection should be maximized. The variance of the sample points after projection can be expressed as:

$$\sum_i W^T x_i x_i^T W \tag{1}$$

So the optimization goal can be written as:

$$\max_w tr(W^T XX^T W) s.t.\ W^T W = I \tag{2}$$

For the formula (2) using the Lagrange multiplier method can be obtained:

$$XX^T W = \lambda W \tag{3}$$

Therefore, only the eigenvalue decomposition of the covariance matrix XX^T is performed. Sort the obtained eigenvalues: $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$, and then take the feature vector corresponding to the first d eigenvalues to form $W=(w_1, w_2, ..., w_d)$, which is the solution to the principal component analysis.

2.4. Linear Discriminant Analysis (LDA) Algorithm

The basic idea of LDA is to project high-dimensional pattern samples to the best discriminant vector space[11], so as to achieve the effect of extracting the classification information and compressing the dimensions of the feature space. Below, we formulate this problem mathematically. Assume that the dimension of the low-dimensional space we project to is d, the corresponding base vector is $(w_1, w_2, ...w_d)$, and the matrix of the base vector is W, which is an $n \times d$ matrix. At this point our optimization objective function is:

$$\arg\max\ J(W) = \frac{\prod_{diag} W^T S_b W}{\prod_{diag} W^T S_w W} \tag{4}$$

Where $\prod_{diag} A$ is the product of the main diagonal elements of A, and W is a matrix of $n \times d$. The optimization process of $J(W)$ can be translated into:

$$J(W) = \frac{\prod_{i=1}^{d} w_i^T S_b w_i}{\prod_{i=1}^{d} w_i^T S_w w_i} = \prod_{i=1}^{d} \frac{w_i^T S_b w_i}{w_i^T S_w w_i} \tag{5}$$

Since W is a projection matrix obtained using a class of samples, its dimension d to the dimension d is a maximum of $k-1$.

3. Result and discussion

The LIBS data showed that there were much noises in the original spectrum, and the useful feature information was submerged in those noises, so the appearance of the characteristic peaks are greatly disturbed. It was difficult to distinguish the samples contaminated by different heavy metals directly, so we have determined that it is necessary to reduce spectral noises. The Group Ⅲ samples were spectral data, shown in Figure 2. However, this group of samples of tegillarca granosa samples contained Zn ions. We noticed that this noises data was removed with a threshold of 500, reducing the number of dimensions from the previous 30,267 dimensions to 2811 dimensions.

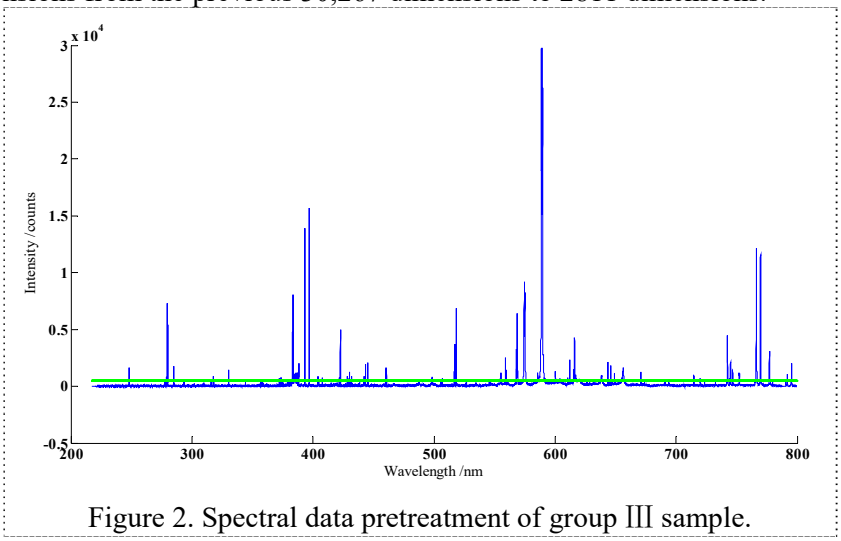

Figure 2. Spectral data pretreatment of group Ⅲ sample.

After the original spectral data was de-noised, the optimal parameters of the training sets were learned by cross-validation. After this step, the optimal parameter determined by PCA-LDA was 16. However, the optimal parameter of PLS-DA is 9, and when using the SIMCA and SVM algorithms the optimal parameters are 6, 8, 3 and 0.0313, 0.0313 respectively. As shown in Figure 3, 4, 5.

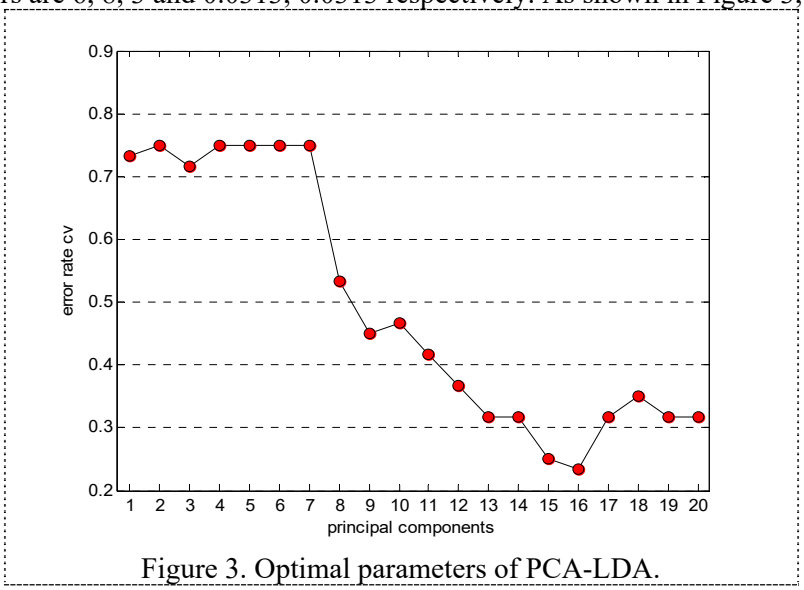

Figure 3. Optimal parameters of PCA-LDA.

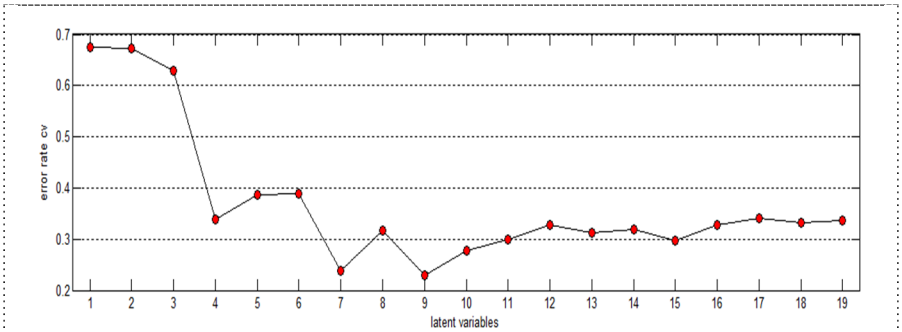

Figure 4. Optimal parameters of PLS-DA.

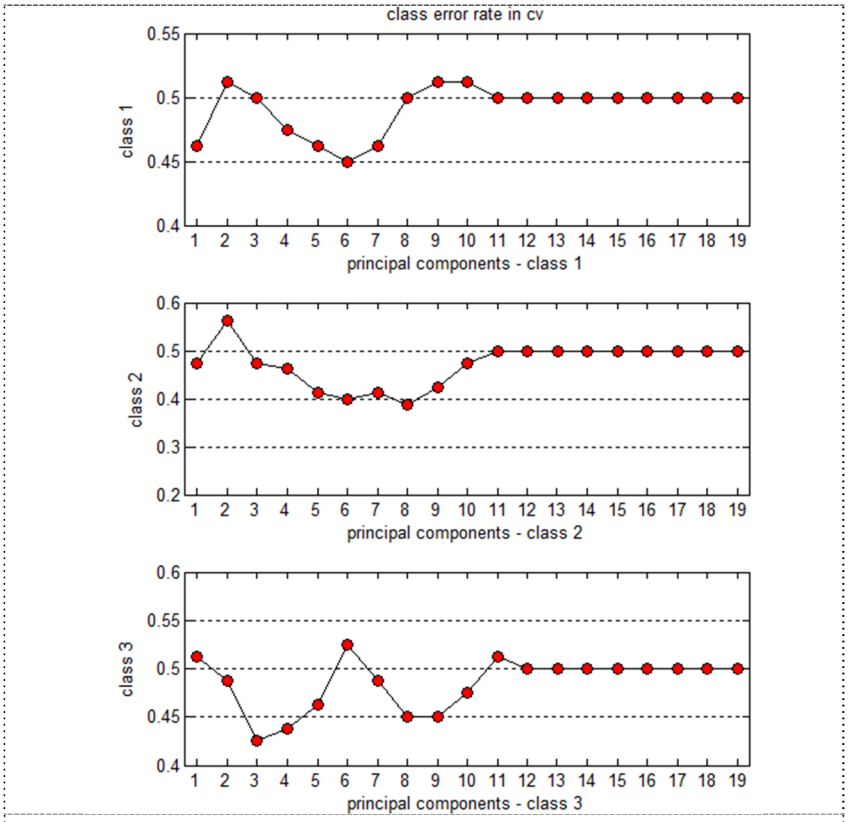

Figure 5. Optimal parameters of SIMCA.

The results of the remaining 30 samples from the prediction set showed that the discriminant accuracies of PLS-DA, SIMCA and SVM were 57%, 20% and 33% respectively. Whereas the PCA-LDA algorithms achieved the best discrimination performance at an 87%. Because the LIBS data dimension is very high and causes useful feature lines to be submerged in the noise, using PCA for dimensionality reduction and using these data points as inputs can improve the discriminant performance of the LDA classifier. The analysis shows the search range of SVM needs better prior knowledge as the poor search range causes the data to fall into local convergence, so the accuracy of SVM is only at a 33%. For the SIMCA algorithm, the predictive samples may or may not satisfy a plurality of different types of feature spaces. After testing, results showed the predictive set may not belong to any class, so its recognition rate was lower than 33% of the theoretical random average probability.

4. Conclusion

In summary, this study provides a rapid identification method for identifying the types of heavy metal pollutions in seafood. Compared with PCA-LDA, PLS-DA, SIMCA and SVM, the results show that the classification accuracy of the PCA-LDA algorithm is the best, with a discrimination rate of 87%. Therefore, combining PCA-LDA algorithm with a portable LIBS spectrometer offers the possibility of rapid on-site identification of food safety, marine environment monitoring and protection.

References

[1] Storelli, M. M. (2001) Heavy metals in the aquatic environment of the Southern Adiartic Sea,Italy Macroalgea, sediments and benthic species. Environment International, vol. 26, no. 7, pp. 505-509.

[2] Hamdoun, H., Van-Veen, E., Basset, B., Lemoine, M., Coggan, J., Leleyter, L., Baraud, F. (2015) Characterization of harbor sediments from the English Channel: assessment of heavy metal enrichment, biological effect and mobility. Mar Pollut Bull, vol. 90, no. 1-2, pp. 273-80.

[3] Takeshi, H., Kazuyoshi, T. (2010) Heavy Metal-Induced Oxidative DNA Damage in Earthworms: A Review. Applied and Environmental Soil Science, vol. 2010, pp. 1-7.

[4] Beauchemin, D. (2002) Inductively coupled plasma mass spectrometry. Analytical Chemistry, vol. 74, no. 12, pp. 2873-2893.

[5] Bazzi, A., Kreuz, B., Fischer, J. (2004) Determination of Calcium in Cereal with Flame Atomic Absorption Spectroscopy. An Experiment for a Quantitative Methods of Analysis Course. Journal of Chemical Education, vol. 81, no. 7, pp. 1042-1044.

[6] Wang, G., Yuan, J., Gong, B., Matsumoto, K., Hu, Z. (2001) Immunoassay by graphite furnace atomic absorption spectrometry using a metal chelate as a label. Analytica Chimica Acta, vol. 448, pp. 165-172.

[7] Ghica, M. E., Carvalho, R. C., Amine, A., Brett, C. M. A. (2013) Glucose oxidase enzyme inhibition sensors for heavy metals at carbon film electrodes modified with cobalt or copper hexacyanoferrate. Sensors and Actuators B-chemical, vol. 178, pp. 270-278.

[8] Wang, Z., Feng, J., Li, L., Ni, W., Li, Z. (2011) A multivariate model based on dominant factor for laser-induced breakdown spectroscopy measurements. Journal of Analytical Atomic Spectrometry, vol. 26, no. 11, pp. 2289-2299.

[9] DeLucia, F. C. (2005) Laser-induced breakdown spectroscopy (LIBS): a promising versatile chemical sensor technology for hazardous material detection. IEEE Sensors Journal, vol. 5, no. 4, p. 9.

[10] Abdi, H., Williams, L. J. (2010) Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, vol. 2, no. 4, pp. 433-459.

[11] James, G. M., Hastie, T. (2001) Functional linear discriminant analysis for irregularly sampled curves. Journal of The Royal Statistical Society Series B-statistical Methodology, vol. 63, no. 3, pp. 533-550.