Potential Support Vector Machines for phytoplankton fluorescence spectra classification: Comparison with Self-Organizing Maps

Ismael F. Aymerich[1,2], Jaume Piera[1], Johannes Mohr[4], Aureli Soria-Frisch[3] and Klaus Obermayer[4]

[1] Marine Technology Unit (UTM-CSIC)
Pg. Maritim Barceloneta 37-49
Barcelona, 08003 (Spain)
ismaelf@utm.csic.es

[2] Universitat Politècnica de Catalunya (UPC)
Jordi Girona, 31
Barcelona, 08034 (Spain)

[3] Starlab
C. de l'Observatori Fabra s/n.
Barcelona, 08035 (Spain)

[4] Technische Universität Berlin (TU Berlin)
Franklinstr. 28/29, 10587
Berlin (Germany)

Abstract – *Evaluation of phytoplankton communities is an important task to characterize marine environments. Fluorescence spectroscopy is a powerful technique usually used for this goal. This study presents a comparison between two different techniques for fast phytoplankton discrimination: Self-Organizing Maps (SOM) and Potential Support Vector Machines (P-SVM), evaluating its capability to achieve phytoplankton classification from its fluorescence spectra. Several cultures representing different algae groups were grown under the same conditions and their emission fluorescence spectra were measured every day. Finally, the classification results obtained from both techniques, SOM and P-SVM, are presented. In the case of using emission fluorescence spectra, the results show that we are able to reduce the acquisition time required for some of the existing techniques, obtaining encouraging classification performance.*

Keywords – *Self_Organizing Maps (SOM), Potential Support Vector Machines (P-SVM), hyperspectral, phytoplankton fluorescence.*

I. INTRODUCTION

Fluorescence spectroscopy is a non-invasive technique able to measure water properties directly, and provide qualitative and quantitative information about phytoplankton. It has become in the last years a powerful tool to study phytoplankton communities' distribution. Several techniques use phytoplankton fluorescence spectroscopy to discriminate between different phytoplankton groups. Some of these techniques can achieve a high taxonomic discrimination but they are based on measurements that require excitation at different wavelengths. In [1] the possibility to use the information contained in emission fluorescence spectra to discriminate between several phytoplankton species was evaluated. In that case, Self-Organizing Maps was used and its performance was presented as a feasible technique to use in those studies, in which time acquisition is an important constraint, e.g. mobile platforms for high spatial resolution measurements.

In this work, a comparison between fluorescence spectra classification based on Self-Organizing Maps (SOM) and Potential Support Machines (P-SVM) is presented. SOM is a type of unsupervised artificial neural network commonly used for clustering, pattern recognition, classification and visualization of high dimensional data. A brief description of SOM and previous results using it for classification of phytoplankton from emission fluorescence spectra can be found in [1]. P-SVM is herein briefly described and finally, the results of both classification techniques are presented and discussed.

TABLE 1. Phytoplankton cultures under study

Species	***Division***	***Abbreviation***
Alexandrium minutum	Dinophyceae	Am
Thalassiosira weissflogii	Bacillariophyceae	Thwi
Dunaliella primolecta	Chlorophyceae	Duna
Isochrysis galbana	Prymnesiophyceae	Iso
Pleurochrysis elongata	Prymnesiophyceae	Pl

A. Potential - Support Vector Machines

P-SVM [2] is a supervised learning method used for classification and regression. As well as standard Support Vector Machines, it is based on kernels. Kernel Methods approach the problem by mapping the data into a high dimensional feature space, where each coordinate corresponds to one feature of the data items, transforming the data into a set of points in a Euclidean space. In that space, a variety of methods can be used to find relations between the data.

II. DISCUSSION AND RESULTS

Five different cultures representing the major algae divisions were selected (Table 1) and grown under the same conditions. Excitation-Emission Matrices were acquired every day.

SOM method was used as a first approach to phytoplankton discrimination from excitation and emission fluorescence spectra [1]. Herein we present the results using P-SVM, and we compare the results using both techniques.

Two data sets were selected: excitation spectra at a 680nm emission wavelength, and emission spectra from samples excited at 470nm (Fig. 1).

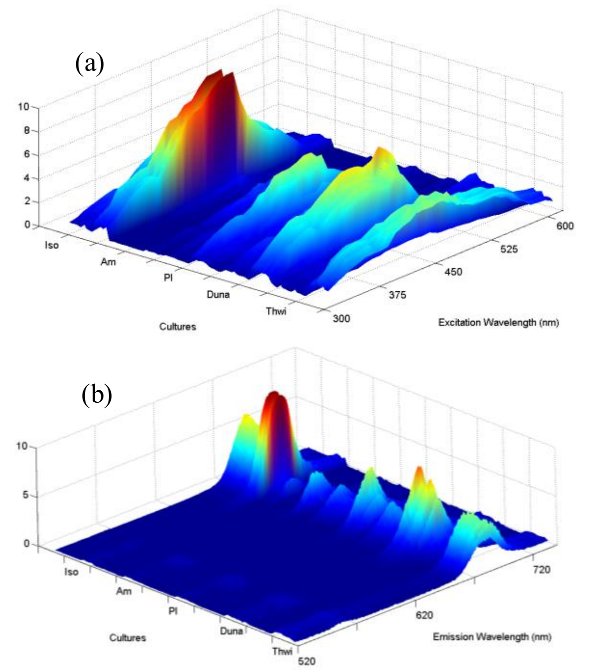

Fig. 1 (a) Example of training data set from excitation fluorescence spectra, (b) example of training data set from emission fluorescence spectra.

In the first case, using excitation spectra, training and test data sets were chosen doing repeated random sub-sampling validation, and the results of 10 different classifications were averaged. In order to evaluate the performance of both techniques, the confusion matrices from the classification step were obtained, and the index Kappa was computed following Eq. 1 [1].

$$K = \frac{n\sum_{k=1}^{r} X_{kk} - \sum_{k=1}^{r} X_{k+}X_{+k}}{n^2 - \sum_{k=1}^{r} X_{k+}X_{+k}} \quad (1)$$

Table 2 shows an example of confusion matrix using P-SVM. The averaged Kappa index resulted in that case is 0.3636. SOM performed better in this case obtaining K=0.6629.

The next step was to try the same procedure but using emission fluorescence spectra instead of excitation spectra. Again, training and test data sets were chosen doing repeated random sub-sampling validation, and the results of 10 different classifications were averaged. The results in that case increase slightly, obtaining K=0.4839, while using SOM we obtained 0.6568.

As it happened utilizing SOM, these poor results could be due to the similarity of the emission fluorescence spectra among the different classes studied. In order to enhance the fluorescence spectra differences between the algae, a derivative analysis has been applied to the emission spectra. Derivative analysis has demonstrated to be a powerful tool to enhance differences [3], although it is high sensitive to noise. For this reason, the noise of the spectra has been also reduced using a wavelet denoising technique [4].

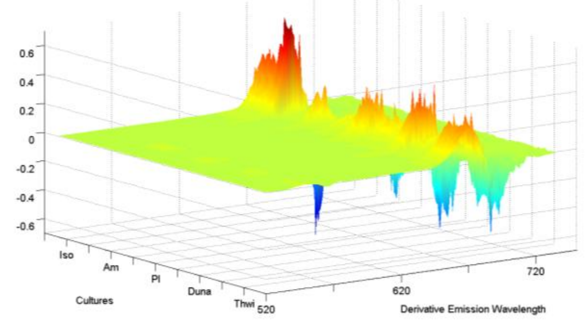

Figure 2. Example of training data set from derivative emission fluorescence spectra.

Once the data were pre-processed (Fig. 2), P-SVM was used again to classify the different spectra into the correct class. The index Kappa obtained this time was 0.7141. In contrast to the results obtained with SOM (K=0.6992), the performance of P-SVM using pre-processed data are even better than those obtained with SOM.

TABLE 2. Confusion matrix

	Predicted					
	Thwii	*Duna*	*Pl*	*Am*	*Iso*	Σ (sum)
Thwi	**7**	0	11	0	1	**19**
Duna	0	**12**	8	0	0	**20**
Pl	0	3	**9**	7	2	**21**
Am	0	0	2	**18**	0	**20**
Iso	4	5	0	0	**6**	**15**

III. CONCLUSIONS

The performance of P-SVM for fluorescence spectra classification was evaluated. For this purpose Excitation-Emission matrices from 5 different cultures were measured every day. Two different data sets were prepared for this study: one containing excitation spectra, while the second one contains emission spectra. The results were compared with those obtained with SOM.

The best performance was obtained with the P-SVM and using derivative analysis in order to enhance subtle differences between spectra (K=0.7141). If no pre-processing is used, the P-SVM results (0.4839) were worse than the results obtained with the SOM method (0.6568).

Although the pre-processing step helps to achieve higher classification accuracy with both techniques, it also shows the kernel method as a feasible technique to achieve phytoplankton discrimination from its emission fluorescence spectra, reducing the acquisition time needed by other techniques.

ACKNOWLEDGMENTS

This research has been supported by the project ANERIS (PIF-015-1), funded by the Spanish National Research Council (CSIC).

REFERENCES

[1] Aymerich, I. F., Piera, J. and Soria-Frisch, A. “A rapid technique for classifying phytoplankton fluorescence spectra based on self-organizing maps,” *Appl. Spectrosc.* 63, 716-726, 2009.

[2] S. Hochreiter and K. Obermayer, "Support vector machines for dyadic data," *Neural Computation*, 18 (6), 1472-1510, 2006.

[3] W. L. Butler, and D. W. Hopkins, “Higher Derivative Analysis of Complex Absorption Spectra,” *Photochem. Photobiol.* 12, 439-450, 1970.

[4] Piera, J., Quesada, R., Manuel-lazaro, A., Del Rio, J., Shariat Panahi, S. and Olivar, G., “Wavelet denoising technique for high-resolution CTD data. Characterization of turbulent oceanic flow,” *Sensors. Proceedings of the IEEE.* 3: 1468-1471, 2004.