CHLOROPHYLL A OF OPTICALLY COMPLEX COSTAL WATERS USING REGIONALLY SPECIFIC NEURAL NETWORK-BASED ALGORITHMS FOR MERIS FULL RESOLUTION DATA

Evangelos Spyrakos[1], Luis Gonzalez Vilas[1], Jesus Torres Palenzuela[1] and Nina Yarovenko[1]

1. University of Vigo, Department of Applied Physics, Vigo, Spain; jesu@uvigo.es

ABSTRACT

In typical case 2 waters an accurate remote sensing retrieval of chlorophyll a (chla) is still challenging. There is a widespread understanding that universally applicable water constituent retrieval algorithms are currently not feasible, shifting the research focus to regionally specific implementations of powerful inversion methods. In this study a series of algorithms for the retrieval of chlorophyll a in the Galician rias (NW Spain) from MERIS full resolution data were developed based on Multilayer perceptron (MLP) artificial neural networks and fuzzy c-mean clustering techniques (FCM) using different quality levels. The quality levels were given to the data set as a function of quality flags. All the NNs developed in this study showed high R2 and low root mean square error (RMSE) values in both training and validation sets. The algorithms were applied to six MERIS images delivered from the area at July 2008, in order to create chla maps. The best performance parameters were given for the NN trained with high-quality data using the most abundant cluster found in the rias. The transport of high phytoplankton biomass areas during the upwelling cycle was clearly captured in the images. Relatively high biomass "patches" were detected in detail inside the rias. There was a significant variation in the timing and the extent of the chla peak areas related to the winds and surface currents. A local-based algorithm for the chla retrieval from an ocean colour sensor with the characteristics of MERIS can be a great support in quantitative monitoring of chla and study of harmful algal events in Galician rias.

INTRODUCTION

Among ocean-colour derived data, chlorophyll a (chla) concentration is the most used product since it provides a good estimation of phytoplankton biomass and is common to almost all taxonomic groups (i). The phytoplankton community responds rapidly to environmental changes (2), which can cause visible changes in chlorophyll in the surface waters.

The estimation of chla concentration in the oceans from the first dedicated ocean colour scanner (CZCS), which was launched in 1978 and operated until 1986, provided useful information on the global distribution of chla but the quality of the data was limited (Robinson, 2004). Sea-viewing Wide-Field-view Sensor (SeaWiFS), Moderate Resolution Imaging Spectroradiometer (MODIS) and the most recent Medium Resolution Imaging Spectrometer (MERIS), which succeeded CZCS, are using more and narrower spectral bands and finer spatial resolution. MERIS provides data with a 300 m on-ground resolution in nadir (Full Resolution, FR) and has a spectral resolution of fifteen bands from visible to near infrared, supporting one of the mission objectives for delicate coastal zone monitoring (3).

Traditionally, chla is estimated using empirical algorithms based on the ratio between the radiance of blue and green light reflected by the sea. For the retrieval of chla from ocean colour sensors various empirical spectral-ratio algorithms (4, 5, 6, 7, 8) and semi-analytical models (9) were developed. In typical case II waters, where high concentrations of water constituents (coloured dissolved organic matter or CDOM, detritus) absorb strongly in the blue decoupling the phytoplankton absorbance, this ratio cannot be used for an accurate retrieval of chla (10).

Neural networks (NN) are a powerful tool for modelling multivariate, complex and non-linear data (11), and they can play an important role in order to achieve more accurate retrievals of water constituents in optically complex waters. Several authors have applied NN techniques to estimate water quality parameters from ocean-colour images (e.g.,12, 13). According to (14), ocean colour algorithms should be also specific for each region. The standard products for the estimation from MERIS data of chla, suspended particulate matter (SPM) and yellow substances developed by the European Space Agency (ESA) are currently based on NN algorithms (15, 16).

Due to the ecologic, economic and social importance of the Galician rias (NW Spain), study area of this work, the development of algorithms for an accurate estimation of chla is crucial. In this work, specific NN-based algorithms for chla retrieval from MERIS FR data were developed for this area. Firstly, a fuzzy c-mean cluster (FCM) algorithm was applied to remote sensing reflectance data to determine clusters that could be associated with different water types and define the scope of the NN. Secondly, Multilayer Perceptron (MLP) NN algorithms were trained and validated using in-situ data with different quality levels.

Finally, the best performance NN algorithm was applied to a short series of MERIS FR images acquired during an upwelling cycle to obtain chla maps. The temporal and spatial distribution of the chla patterns was analysed according to the environmental conditions in the area, showing the potential of the algorithms to map the possible algal blooms caused by the coastal upwelling.

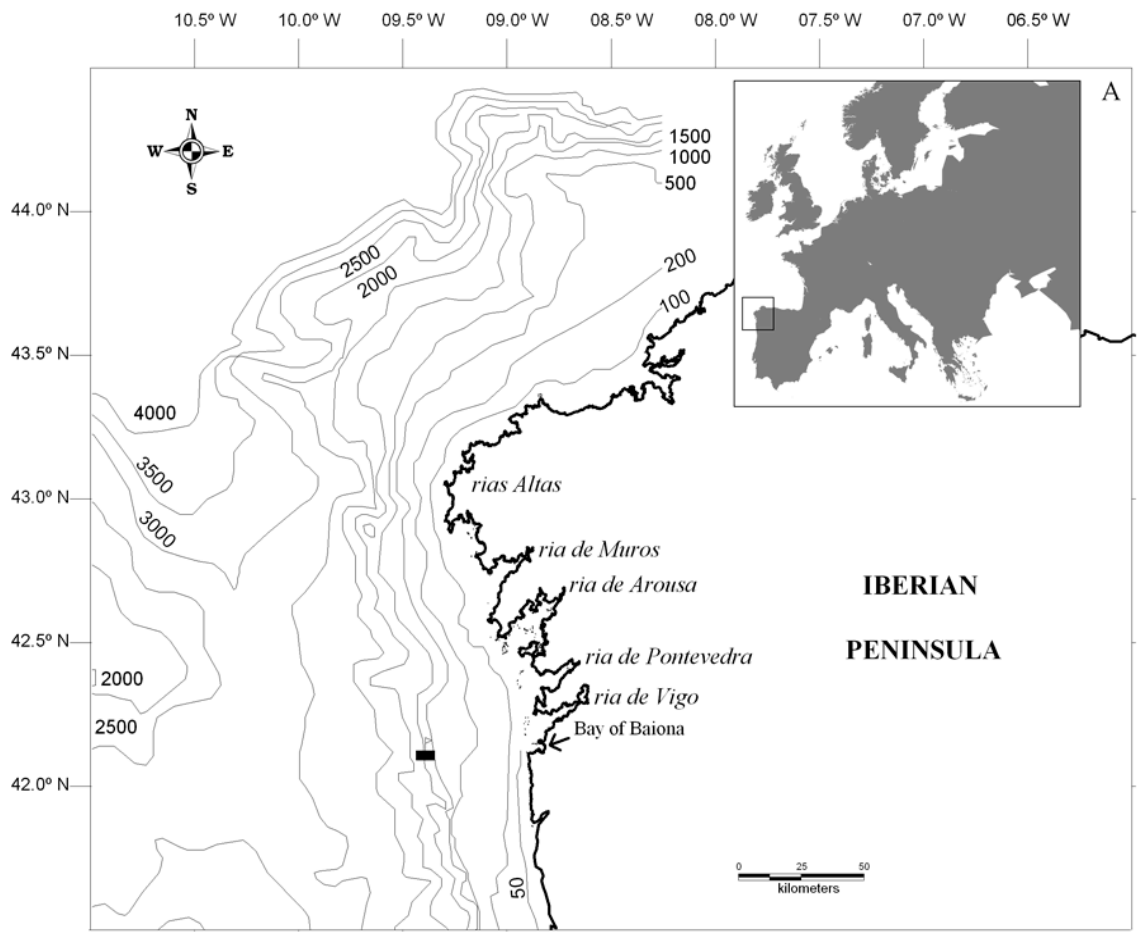

Figure 1: Galician coast and bathymetry of the area. From north to south the Rias Baixas: Muros y Noya, Arousa, Pontevedra and Vigo. The location of the Seawatch buoy station off Cabo Silleiro is shown by a black rectangle.

METHODOLOGY

Study area

The Galician rias are V-like coastal embayments located in the northwest part of the Iberian Peninsula (Figure 1). This study focused on the Rias Baixas, which include the southern and largest rias, from north to south: Muros y Noya, Arousa, Pontevedra and Vigo. They are characterized by a SW-NE orientation, strong tides and freshwater inputs from rivers situated in their innermost part. All these rias, except Muros, are connected to the open sea through two entrances, north and south, separated by inlands located at their external part. Rias vary in width from 1-3 km in their inner part to 8-12 km in their external part, covering an area of approximately 600 km^2 with depths ranging from 5 to 60 m (17). The Rias Baixas waters are optically complex and can be categorized as Case II (18). Ocean colour sensors with the characteristics of MERIS are considered a powerful tool for chla monitoring and detection of harmful algae events (HABs) on coastal areas such as the Rias Baixas. However, the number of studies using MERIS data on the Galician coast is limited (19,20,21,22).

MERIS chla algorithms for Rias Baixas

The development of the regional algorithms for chla retrieval from MERIS data was based on approximating a set of water-leaving radiance reflectance data derived from the images to the corresponding chla concentrations obtained using in-situ data. Multilayer perceptron (MLP) artificial neural networks (NN) were implemented because of its ability to approximate a set of input data to the corresponding output data. Moreover, MLP does not make assumptions about data distribution and it is useful for modelling multivariate, complex and nonlinear data (23).

Chla concentrations were derived from two different data sets. The first one consisted of 181 analytical measurements from the monitoring programme conducted by INTECMAR between 2002 and 2006 in the four rias. Chla concentrations were determined by spectrofluorometry using water samples collected from the surface to a depth of 4 metres. The second data set included 46 data points derived from water samples collected from the surface to a depth of 4 metres during 5 boat campaigns carried out in 2007 and 2008 on cloud-free days in the ria of Vigo and ria of Arousa. The pigments were extracted and separated using a High Perfomance Liquid Chromatography (HPLC) method with a reverse phase C8 (more details in 24).

Fifteen MERIS FR level-1b images over the study area (for 2002-2004 and 2006-2008) were acquired for the development of the algorithms. Dates were selected considering the availability of ground data and cloud coverage conditions. The Beam-4.6 (Brockmann Consult and contributors, Germany) software's smile correction and the atmospheric correction algorithm developed by (16) were applied to the images to obtain water-leaving radiance reflectance data. Land pixels and pixels suspected of being affected low clouds or fog were masked from the flags derived from Beam software.

Data from available MERIS images were extracted and linked to the in situ database so that the time difference between the image pass and the sampling was lower than two hours. Only sampling points derived from cloud-free scenes that were not affected by sun-glint were considered to be valid match-up data (107 points from INTECMAR dataset and 43 points from campaigns dataset). Reflectance and geometry values (including sun zenith, view zenith and difference between view and sun azimuths) were obtained for each sampling point, computing the median of 9 pixels (approximately 0.8 km^2) around the pixel containing the exact geographical location to reduce MERIS instrument noise. Moreover, the effect of possible mixed pixels with extremely high or low values is reduced using the median instead of the mean. Previously masked pixels were excluded from the median computation and the number of included pixels was extracted as a quality flag ranging from 9 (highest quality) to 1 (lowest quality). Low quality values show that the sampling point is located near coastal, cloudy or foggy areas, and therefore reflectance values could be af-

fected. The effect of the quality flag was evaluated by developing the neural models for two different quality levels.

FCM technique is a fuzzy clustering algorithm that divides a dataset into a specific number of clusters so that each data point can belong to more than one cluster with a membership degree (between 0 and 1). Different FCM algorithms were applied to MERIS reflectance values linked to the in-situ measurements with the aim of determining the number of spectral clusters in the dataset. More details about methodology and results are provided in (21). In this study, the best FCM algorithm was used for establishing the scope of the NN for chla retrieval. Although in theory a different NN might be developed for each cluster, in practice only one of the clusters (Cluster#1) provided a representative and sufficiently large dataset.

The basic architecture of the MLP networks used in this work is shown in Figure 2. It includes an input layer with 14 input nodes (11 reflectance values and 3 geometry values), two hidden layers (each one with four nodes) and an output layer with a node associated to the desired output, i.e. the chla concentration. Three different NN, with this structure but with different input datasets and number of data points, were developed: a preliminary network including the entire dataset (NN#1), another network using all data points belonging to the main cluster (NN#2), and finally other network including data points belonging to the main cluster with a quality flag equal to 9 (NN#3), so that points near the coastline were excluded.

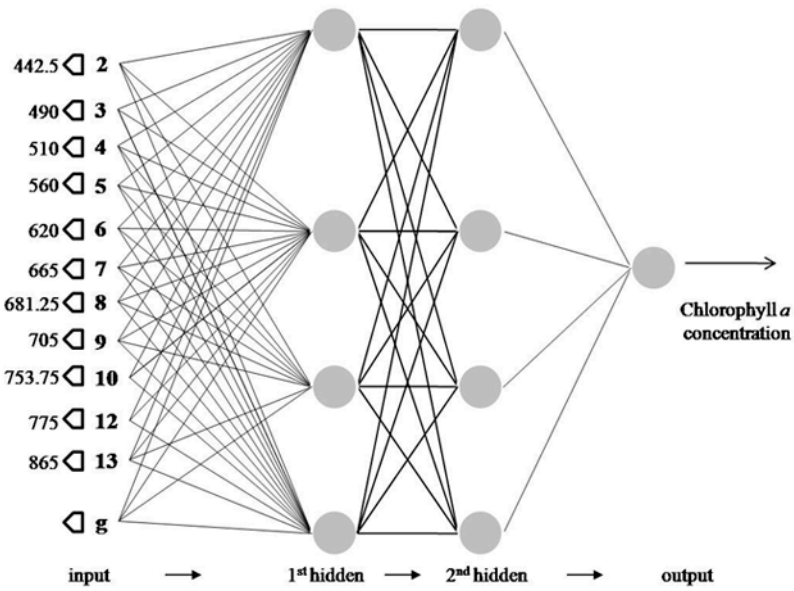

Figure 2: Basic architecture of the MLP networks used in this work. g is the geometry of the images. The inputs are the MERIS remote sensing reflectance values for 11 bands centred on different wavelengths (nm) and the geometry of the images (g).

The complete dataset for each network was divided into two subsets: the training set, with approximately 80% of the data points and used for training the NN, and the validating set, with the remaining records necessary for validating the algorithm. Both subsets were created including records from all the images and covering the entire range of variation of the chla concentration in order to make these subsets random and significant for the whole dataset. The performance of the neural models was improved by applying leave-one-out cross-validation in the training phase (25).

A set of parameters comparing the observed chla concentration (ChlO) and the retrieved one using the algorithm (ChlM) were computed, from both the training and validation sets, to evaluate the

model fitting. In addition to the coefficient of determination (R2) between ChlO and ChlM, the following parameters were obtained:

- Mean prediction error (MPE) between ChlO and ChlM. MPE is defined as:

$$MPE = \frac{1}{N}\sum_{i=1}^{N} PE_i$$

where $\mathrm{PE_i} = \mathrm{ChlO_i} - \mathrm{ChlM_i}$ is the prediction error and N is the number of data points.

- Variance of the prediction errors (VAR):

$$VAR = \frac{1}{N-1}\sum_{i=1}^{N}(PE_i - MPE)^2$$

- Root mean square error (RMSE) between ChlO and ChlM:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} PE_i^2}{N}}$$

- Relative RMSE

$$\mathrm{Rel.\,RMSE} = \sqrt{\frac{RMSE}{\frac{1}{N}\sum_{i=1}^{N} ChlO_i}}\,100$$

MPE is useful for finding out if a model tends to underestimate (high positive values) or overestimate (high negative values) the observed chla concentrations. RMSE and relative RMSE quantify the absolute error (in mg m^{-3}) and relative error respectively, while VAR was used as measurement of error variability.

Chlorophyll mapping

The NN algorithm with the best performance measurements was applied to a series of 6 MERIS Full Resolution images acquired in July 2008 over the study area to obtain chla maps. Beam 4.2 (Brockmann Consult and contributors, Germany) software was used for pre-processing the images, including smile correction, atmospheric correction (16) and masking of coastline, land, clouds, invalid reflectance and areas significantly affected by sun glint(beyond a solar zenith angle limit of 60°).

Before applying the chla algorithm to leaving radiance reflectance values, FCM algorithm was used for obtaining classification images, in which each pixel was assigned to the cluster with the highest value in its corresponding membership function. The NN was only applied to those pixels belonging to the cluster that had been previously used for its development. Chla maps were then remapped using the standard Mercator projection with a fixed grid of 890 by 890 pixels, ranging from 42o 04' N to 42o 40' N latitude and from 8o 32'W to 9o 32'W longitude, which covers approximately 3.1 x 103 km2.

Finally, wind speed and direction data were observed at a Seawatch buoy station located off Cape Silleiro (42° 7.8′N, 9° 23.4′W) and daily upwelling indexes (Iw) were estimated by Bakun's method (26) using these data. The meteorological station was selected because it is considered to be fairly representative of the study area (27).

RESULTS AND DISCUSSION

Chlorophyll algorithms

The FCM algorithm with the maximum degree of separability between clusters was achieved with three clusters. This algorithm allowed us to assign one cluster to each one of the 150 valid data points considering their highest membership grade. Cluster#1 is the largest cluster, with approximately 80% of the values, and it was the only suitable for the development of the neural model since, in addition to its size, it covered the entire range of chla (from 0.03 to 7.94 mg m^{-3}) observed in the whole data set. Moreover, Cluster#1 was the most frequent cluster in the classification images and on average, 70% of the pixels over the study area belonged to it. Cluster#2 and Cluster#3 seem to represent non-typical situations in the Galician rias, and pixels belonging to theses clusters appeared close to the coastline and at the innermost and outermost parts of the rias.

Results of neural networks models are summarized in Table 1, including the number of data points and the validation criteria computed for validation and training datasets. Validation data were not included in the learning procedure, so that parameters computed from validation subset are expected to be worse. The algorithms developed using Cluster#1 (NN#2 and NN#3) outperformed the network using the entire data set (NN#1). Results from NN#3, using only high-quality data points (quality flag equal to 9), were also better than those ones derived from the entire cluster dataset.

Table 1: Summary of the validation parameters computed using the training and validation data sets for each one of the three different NNs developed in this study.

NN	Dataset	N	R^2	MPE	VAR	RMSE	RMSE%
NN#1 Complete dataset (N =150)	Training Validation	120 30	0.77 0.63	-0.03 0.08	0.66 1.04	0.81 1.00	67 74
NN#2 Cluster#1 dataset (N = 119)	Training Validation	92 24	0.86 0.78	0.02 0.02	0.68 0.99	0.68 0.99	60 72
NN#3 Cluster#1, Quality =9 (N = 83)	Training Validation	66 17	0.97 0.86	0.01 -0.14	0.10 0.57	0.32 0.75	41 66

Finally, results from NN#3 were compared with the obtained ones from the Case II Regional (C2R) algorithm that has routinely been used for MERIS data (15). C2R performed worse than NN#3, since it showed a nonlinear relationship with in situ data (R2 = 0.00) and higher RMSE value (2.62 mg m^{-3}). Moreover, C2R tended to overestimate the chla concentration (MPE = -0.46), in particular with low chla values obtained from images acquired during winter months. The poor results of C2R may be explained by the relatively low chla concentrations (mainly <3 mg m^{-3}) observed in the study area, considering that the scope of C2R covers a much wider range of chla.

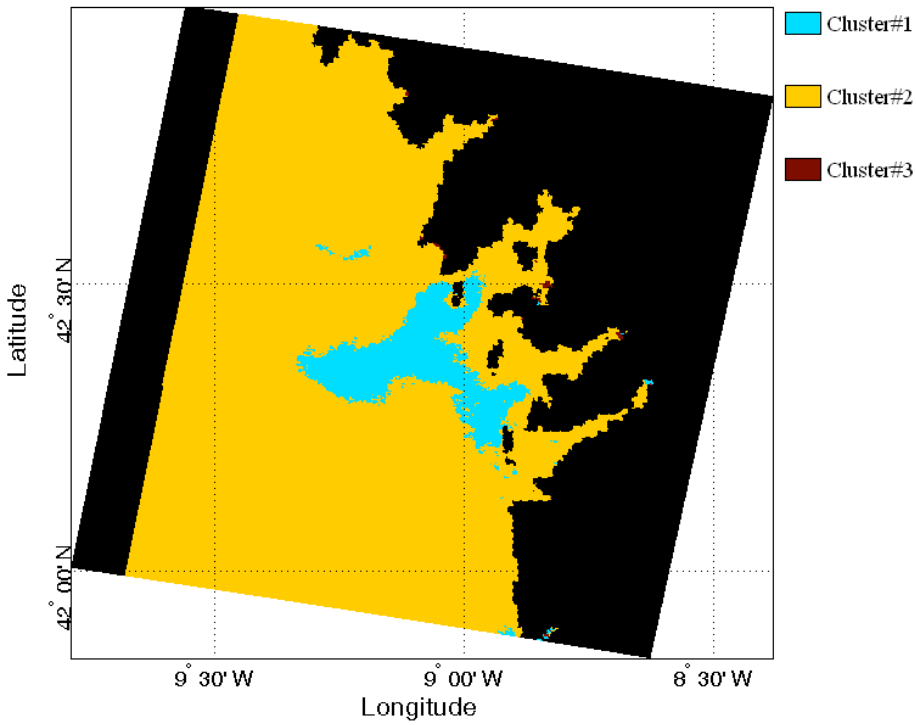

Figure 3: Classification of a MERIS image derived from the study area on 17/11/2003. The 3 classes identified using the FCM are shown

Chlorophyll mapping

Figure 3 shows an example of classification image, indicating the cluster value for each pixel, derived from the 3-cluster FCM algorithm. In theory, chlorophyll maps with soft transitions might be created by blending chla concentrations derived from cluster-specific algorithms considering memberships grades to each cluster (28). However, the best performance algorithm (NN#3) was only developed for Cluster#1, so that classification images were only useful for identifying areas where this cluster is dominant and the model may therefore be applied to obtain reliable results. Classification images showed that Cluster#1 is dominant in the rias Baixas and adjacent area for almost all the MERIS images.

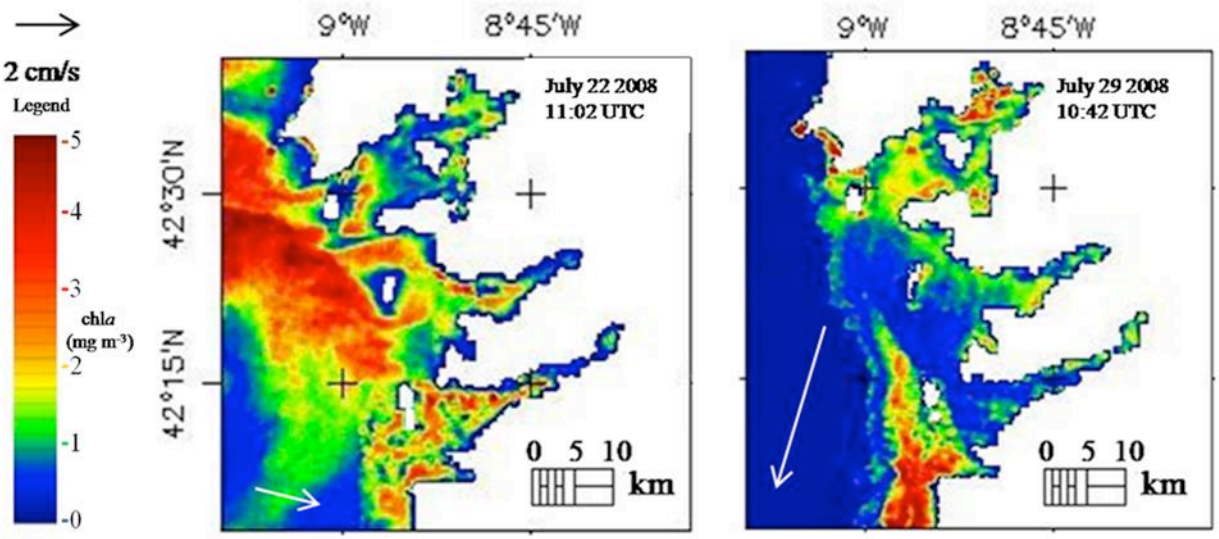

Figure 4: Two chla maps for MERIS FR data derived during the upwelling cycle of July 2008 in the study area. Land and clouds were masked and they appear with white colour.

Chlorophyll maps derived from the MERIS images using NN#3 are shown in Figure 4. In this study we show only 2 of the 6 chla images. The full description of the upwelling events in comparison with the changes appeared in the MERIS chla images are presented in detailed in (29).

The images allowed us to identify three different meteorological and oceanographic states, related to an upwelling cycle, as it is shown in Figure 5. Each state lasted from nine to eleven days.

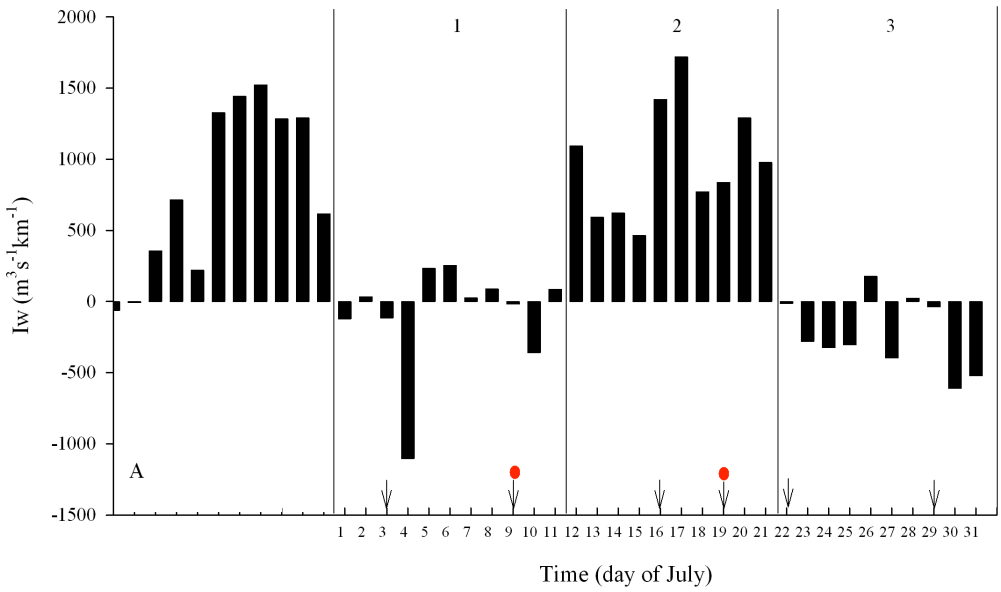

Figure 5: Daily upwelling index off the Rias Baixas

State 3 (between July 22 and July 31) started with an abrupt decrease of upwelling-favourable to zero wind on July 22 followed by a strong relaxation (Figure 5). On July 22, high clha concentrations up to 5 mg m-3 were recorded as a consequence of the presence of cold and chlorophyll-rich

waters resulting from the previous 10 days of upwelling. The high chla concentration was extended from the northern offshore area to the interior of the rias, with an alongshore gradient that probably reflect both the persistence of upwelling in the north and an earlier relaxation in the south. On July 29, chlorophyll map is indicative of the strong relaxation, with near-zero chlorophyll in the offshore region and moderately high values bound to the coast.

CONCLUSIONS

Results of this work show the capability of neural network algorithms to forecast chlorophyll concentrations on the Galician coast from MERIS images, considering the necessity of regionally specific models. According to the in situ data, our model improve significantly other previously techniques and allow us to obtain reliable chlorophyll maps using almost every image. These maps may be useful for studying the evolution of local oceanographic processes, which also may be related to algal blooms in the area.

NN#3 cover the entire range of the temporal variations of chlorophyll concentrations (0.03-7.94 mg m^{-3}) that have been recorded in the study area. The main limitations are that it requires good MERIS images (cloud free and without sunglint) and the masking of Cluster#2 and Cluster#3 areas. The neural models might be improved using TSM and CDOM data. Moreover, if sufficient points were available, it may be possible to develop a neural model for Cluster#2 and #3, or to improve the Cluster#1 model by adding new data. Chlorophyll maps would become more complete by blending different algorithms for different clusters.

This work also shows the capability of MERIS images to obtain a more detailed description of chlorophyll distribution in the Galician rias and the adjacent coastal areas during a summer upwelling cycle, due to their fine spatial resolution and the precise atmospheric correction. The application of a specifically developed algorithm provides accurate chlorophyll mapping of the interior of the Rias Baixas and the detection capability of relative high biomass “patches”. Maps show that spatial patterns of phytoplankton distribution in the study area can be complex, so that areas of high chlorophyll concentrations observed using the satellite might even be missed by in situ monitoring programmes.

ACKNOWLEDGEMENTS

MERIS data were obtained through ESA/ENVISAT project AO-623. We thank the "Technological Institute for the Control of the Marine Environment of Galicia (INTECMAR)", for providing us with chlorophyll data. This work was partially funded by the European Commission’s Marie Curie Actions (project 20501 ECOsystem approach to Sustainable Management of the Marine Environment and its living Resources [ECOSUMMER]) through a grant supported ES.

REFERENCES

1 Jeffrey S, R Mantoura R & S Wright, S, 1997. Phytoplankton pigments in oceanography: guidelines to modern methods. (Paris, France: UNESCO Publishing).

2 EC.2000. Directive of the European Parliament and of the Counsil 2000/60/EC establishing a framework for community action in the field of water policy. Official Journal of the European Communities, L327/321.

3 Robinson, I, 2004. Measuring the Oceans from Space: The principles and methods of satellite oceanography. (Chichester, UK: Spinger).

4 Muller-Karger F E, C R McClain, R N Sambrotto & G C Ray, 1990. A comparison of ship and coastal zone color scanner mapped distribution of phytoplankton in the southeastern Bering Sea. Journal of Geophysical Research, 95: 483-499.

5 Aiken J, G F Moore, C C Trees, S B Hooker & D K Clark, 1995. The SeaWiFS CZCS-type pigment algorithm. In: SeaWiFS Technical Report Series. Greenbelt MD, 34.

6 O'Reily J E, S Maritorena, D A Siegel, M C O'Brien, D Toole & B G Mitcell, 2000. Ocean color chlorophyll a algorithms foe SeaWiFS, OC2 and OC4. Version 4. In: NASA technical memorandum 2000-206892 edited by S B Hooker & E R Firestone, 11 49 pp.

7 McClain C R, G C Feldman, & S B Hooker, 2004. An overview of the SeaWiFS project and strategies for producing a climate research quality global ocean bio-optical time series. Deep-Sea Research II, 51: 6-42

8 Brown C A, Y Huot, P J Werdell, B Gentili & H Claustre, 2008. The origin and global distribution of second order variability in satellite ocean color and its applications to algorithm development. Remote Sensing of Environment, 112: 4186-4203.

9 Garder K L, F R Chen, Z P Lee, S K Hawes & D Kamykowski, 1999. Semianalytic moderate-resolution imaging spectrometer algorithm for chlorophyll a concentration and absorption with bio-optical domains based on nitrate-depletion temperatures. Journal of Geophysical Research, 104: 5403-5421.

10 Gons H, 1999. Optical Teledetection of Chlorophyll a in turbid inland waters. Environmental Science & Technology, 33: 1127-1132.

11 Thiria S, C Mejia, F Badran & M Crepon, 1993. A neural network approach for modelling non-linear transfer functions: Application for wind retrieval from spaceborne scatterometer data. Journal of Geophysical Research, 98(12): 22827−22841.

12 Zhang T, F Fell, Z -S Liu, R Preusker, J Fischer & M –X He, 2003. Evaluating the performance of artificial neural network techniques for pigment retrieval from ocean color in Case I waters. Journal of Geophysical Research, 108: 3286-3298.

13 Shahraiyni H T, S B Shouraki, F Fell, M Schaale, J Fischer, A Tavakoli, R Preusker, M Tajrishy, M Vatandoust & H Khodaparast, 2009. Application of an active learning method to the retrieval of pigment from spectral remote sensing reflectance data. International Journal of Remote Sensing, 30: 1045-1065.

14 Dransfeld S, A R Tatnall, I S Robinson & C D Mobley, 2004. A comparison of Multi-layer multi-linear regression algorithms for the inversion of synthetic ocean colour spectra. International Journal of Remote Sensing, 25(21): 4829-4834.

15 Doerffer R & H Schiller, 2007. The MERIS Case 2 water algorithm. International Journal of Remote Sensing, 28: 517-535.

16 Doerffer R & H Schiller, 2008. MERIS Regional Coastal aln Lake Case 2 Water Project - Atmoshperic Correction ATBD. (GKSS Research Center 21502 Geestacht).

17 Vilas F, A Bernabeu & G Méndez, 2005. Sediment distribution pattern in the Rías Baixas (NW Spain): main facies and hydrodynamic dependence. Journal of Marine Systems, 54: 261–276.

18 Morel A & L Prieur, 1977. Analysis of variations in ocean color. Limnology and Oceanography, 22: 709-722.

19 Torres-Palenzuela J M, L Vilas-González & A Mosqura-Giménez, 2005. Correlation between MERIS and in-situ data for study of Pseudo-nitzschia spp. toxic blooms in Galician coastal area. In: Proceedings of the 24th Symposium of the European Association of Remote Sensing Laboratories, edited by Marinko Olui (Rotterdam: Millpress.University Press) 489-497.

20 Torres-Palenzuela J M, L Vilas-González & A Mosqura-Giménez, 2005. Detection of Pseudo-nitzschia spp. toxic blooms using MERIS images on the Galician coast, 1915-1919 (European Space Agency (Special Publication)

21 González Vilas L, E Spyrakos & J M Torres Palenzuela, 2011. Neural network estimation of chlorophyll *a* from MERIS full resolution data for the coastal waters of Galician *rias* (NW Spain). Remote Sensing of Environment, 115: 524-535.

23 Cheng B & D M Titterington, 1994. Neural networks: a review from a statistical perspective. Statistical Science, 9: 2–54.

24 Zapata M, F Rodríguez & J L Garrido, 2000. Separation of chlorophylls and carotenoids from marine phytoplankton: A new HPLC method using a reversed phase C8 column and pyridine-containing mobile phases. Marine Ecology Progress Series, 195: 29-45.

25 Fukunaga K, 1990. Introduction to Statistical Pattern Recognition (Academic Press).

26 Bakun A, 1973. Coastal upwelling indices, west coast of North America 1946-71. In: NOAA Technical Report (Seattle: US Dept. Commerce), 103.

27 Herrera J L, S Piedracoba, R A Varela & G Roson, 2005. Statial analysis of the wind field on the western coast of Galicia (NW Spain) from *in situ* measurements. Continental Shelf Research, 25: 1728-1748.

28 Moore T S, J W Campbell & M D Dowell, 2009. A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product. Remote Sensing of Environment, 113, 2424−2430.

29 Spyrakos E, L Vilas Gonzalez, J Torres Palenzuela & E D Barton, 2011. Remote sensing chlorophyll a of optically complex waters (rias Baixas, NW Spain): Application of a regionally specific chlorophyll a algorithm for MERIS full resolution data during an upwelling cycle. Remote Sensing of Environment, doi:10.1016/j.rse.2011.05.008.