Detection of Moroccan Coastal Upwelling in SST images using the Expectation-Maximization

Ayoub Tamim, Khalid Minaoui, Khalid Daoudi, Abderrahman Atillah, Driss Aboutajdine

Detection of Moroccan Coastal Upwelling in SST images using the Expectation-Maximization

Ayoub Tamim[1], Khalid Minaoui[1], Khalid Daoudi[2], Abderrahman Atillah[2], Driss Aboutajdine[1]
[1]LRIT Associated Unit to the CNRST-URAC n°29, Mohammed V-Agdal University, Rabat, Morocco
[2]INRIA Bordeaux Sud-Ouest (GEOSTAT team), Talence, France
[2]Royal Centre for Remote Sensing (CRTS), Rabat, Morocco

Abstract—**This paper proposes an unsupervised algorithm for automatic detection and segmentation of upwelling region in Moroccan Atlantic coast using the Sea Surface Temperature (SST) satellite images. This has been done by exploring the Expectation-Maximization algorithm. The good number of clusters that best reproduces the shape of upwelling areas is selected by using the two popular Davies-Bouldin and Dunn indices. Area opening technique is developed that is used to remove and discarded the residuals noise in offshore waters not belonging to the upwelling region. The complete system has been validated by an oceanographer using a database of 30 SST images of the year 2007, demonstrating its capability and robustness for precise detection of Moroccan coastal upwelling.**

Keywords*—*Upwelling, Sea Surface Temperature, Expectation-Maximisation, Davies-Bouldin index, Dunn index, Area opening

I. Introduction

The Moroccan Atlantic coast located in the central canary current system is characterized by persistent and variable upwelling throughout the year [3]. In fact, the upwelling is a wind-driven ocean process, which brings cooler, saltier, and usually nutrient-rich deep water upward to replace surface water displaced offshore due to Ekman transport. The upwelling waters stimulates the growth and reproduction of primary producers such as phytoplankton, and consequently play a major contribution in the fisheries management and also in the study of the oceanic circulation and dynamics [7], [4]. The upwelling phenomenon is evidenced at the surface by the presence of cool temperatures and high chlorophyll-a concentrations in onshore waters [2].

The increasing demand and rapid expansion of the satellite-derived products of the ocean have allowed the scientists to replace the labor intensive manual interpretation of the satellite data made by the oceanographers by an automatic detection [10], [6]. In particular, the infrared images of the ocean obtained from Advanced Very High Resolution Radiometer (AVHRR) sensor onboard the NOAA satellite are often rich in oceanographic structures and have been frequently used to study the sea surface temperature patterns related to upwelling activity [10], [6], [19].

A number of automated techniques and algorithms have been proposed in order to detect and localise the upwelling features in SST images. These algorithms vary in complexity from simple classification and thresholding methods [19], [13] to histogram analysis [15] and structure growing process [12]. These wide range of developed algorithm reflects the nature of the problem encountered when attempting to analyse the oceanographic data; namely cloud contamination, artifacts, strong morphological variation and the absence of valid analytical model for the upwelling structures.

Tamim et al. [19] were the first investigators to show that classification of the SST images using any of the two strategies of partitioning, fuzzy or hard, leads to similar and satisfactory results. The investigators were specifically interested in detecting the Moroccan coastal upwelling in infrared satellite imagery. The algorithm used respectively the Otsu's and Fuzzy *c*-means algorithms as underlying hard and fuzzy clustering strategies to generate a labelled images with homogenous and non-overlapping temperatures. Then region-growing process is used to automatically segment and extract the upwelling area from the remaining residuals offshore waters.

Nieto et al. [15] developed an edge detection algorithm for the automatic detection of oceanographic features in SST images. The algorithm is based on the analysis of the bimodality of the image, and the edge is considered to be present if the algorithm detect two different population. The result is series of thermal fronts which separating two waters masses of distinct temperatures.

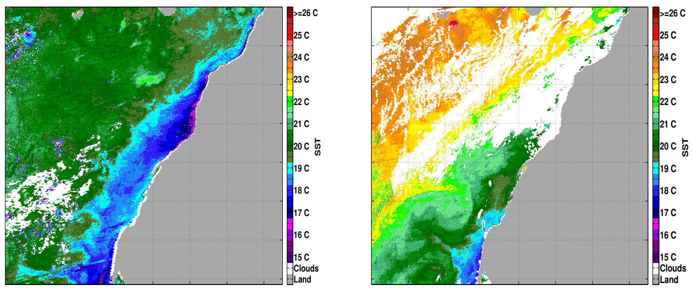
Marcello et al. [12] have combined two approaches in order to detect the upwelling area in SST images. The latter started with the initial structures detection to obtain preliminary coarse segmentation based on the structure analysis and thresholding stage, followed by structures growing step in order to attain the fine detail of the upwelling structures, based on the watershed transform. The algorithm has been tested and validated over database of multisensorial images.

The main aim of this paper is to provide a simple tool for automatic detection and segmentation of upwelling area in the coastal ocean of Morocco using the well-known classification algorithm.

The rest of this paper is organized as follows: Section 2 presents the database and the geographic area used in this study. The proposed methodology to detect and segment the upwelling area in the SST images are listed in Section 3. Section 4 highlights the results and the validation of the algorithm. Finally, Section 5 gives our conclusions and future work.

II. Study area and data

A total of 30 AVHRR SST images of the year 2007 covering the southern part of Moroccan Atlantic coast (spans from 20°50′ - 27°52′N and 13°5′ - 20°10′W) are used throughout this study in order to test and validate the applicability of our proposed approach for the automatic detection of upwelling region. The database were received and processed at the Royal



(a) (b)

Fig. 1: SST images obtained on (a) 2007-07-26 and (b) 2007-09-10, showing two upwelling scenarios.

Centre of Remote Sensing (CRTS) of Morocco including geometric, atmospheric and radiometric corrections.

The cloud and land masks are applied to each image in order to reduce the number of contaminated pixels in the input SST image, and they are generated by using a sequence of multispectral contrast, spectral, and spatial signature threshold tests to perform the classification of the pixel as cloud pixel [18]. The size of each SST images is 714×750 pixels with a spatial resolution of 1.1×1.1 and each pixel represent a temperature in Celsius degree. In order to help and facilitate the interpretation made by the users or oceanographers for the detection of upwelling areas in SST image, a color scale of 25 levels is applied for each image.

Fig. 1(a) shows two SST upwelling scenarios encountered throughout this study: 1) SST images with clear, visible and well-define upwelling structure in the Moroccan Atlantique coast (Fig. 1(a)); 2) SST images where the presence of cloud contamination and artifacts affect the continuity of upwelling structures along the coast (Fig. 1(b)). The gray color region on the right side of the SST images in Fig. 1(a) and (b) corresponds to Moroccan land, whereas the white pixels in the ocean correspond to cloud pixels.

III. The proposed algorithm

A. Clustering algorithm

The main aims of cluster analysis [11] is the classification of patterns and objects according to similarities among them, and then organizing data into clusters. Generally, the cluster is a group of patterns that are more similar to one another than to members of other clusters.

The clustering techniques [1] are among the unsupervised methods when the only data available are unlabelled, and no structural information about it is available. Different classifications approaches can be distinguished depending on the algorithm used: Partitioning, hierarchical, graph-theoretic, and methods based on objective function. In this work we are focused on the partitioning approach, especially for hard clustering methods. In fact, one possible classification of clustering methods can be according to whether the clusters are fuzzy or hard. Hard clustering methods are based on classical set theory, and require that an object either does or does not belong to a cluster. In contrast, Fuzzy clustering methods allow objects to belong to several clusters simultaneously, with different degrees of membership. In this study we have used the expectation-maximization algorithm as underlying hard clustering techniques.

The expectation-maximization (EM) algorithm [5] is considered to be an appropriate optimization algorithm for constructing proper statistical models of data. It provides a probabilistic clustering where each data element has a certain probability of being a member of any cluster. EM is widely used in applications such as computer vision, speech processing and pattern recognition [14]. In fact, EM starts with an initial estimate for the missing variables and iterates to find the maximum likelihood (ML) for these variables. Maximum likelihood methods estimate the parameters by values that maximize the sample's probability for an event. EM is typically used with mixture models.

- *Initialization step*: initialize the hypothesis $\theta^0 = (\mu_1^0, \mu_2^0, \cdots, \mu_K^0)$

$$\theta_k^0 = \mu_k^0 \tag{1}$$

Where K is the current number of Gaussians. σ is the standard deviation, θ^0 is the estimate at 0^{th} iteration and μ is the mean.

- *Expectation step*: estimate the expected values of the hidden variables z_{ij} (mean and standard deviation) using the current hypothesis $\theta^t = (\mu_1^t, \mu_2^t, \cdots, \mu_K^t)$

$$E(Z_{ik}) = \frac{\exp[-\frac{(x_i - \mu_k^t)}{2\sigma^2}]}{\sum_{j=1}^{K} \exp[-\frac{(x_i - \mu_j^t)^2}{2\sigma^2}]} \tag{2}$$

Where t is the number of iteration, $E(z_{ik})$ is the expected value for the hidden variables (namely mean and standard deviation), k is the dimension, σ is the standard deviation.

- *Maximization step*: provides a new estimate of the parameters.

$$\mu_k^{t+1} = \frac{\sum_{i=1}^{n} E(z_{ik}) x_i}{\sum_{i=1}^{n} E(z_{ik})} \tag{3}$$

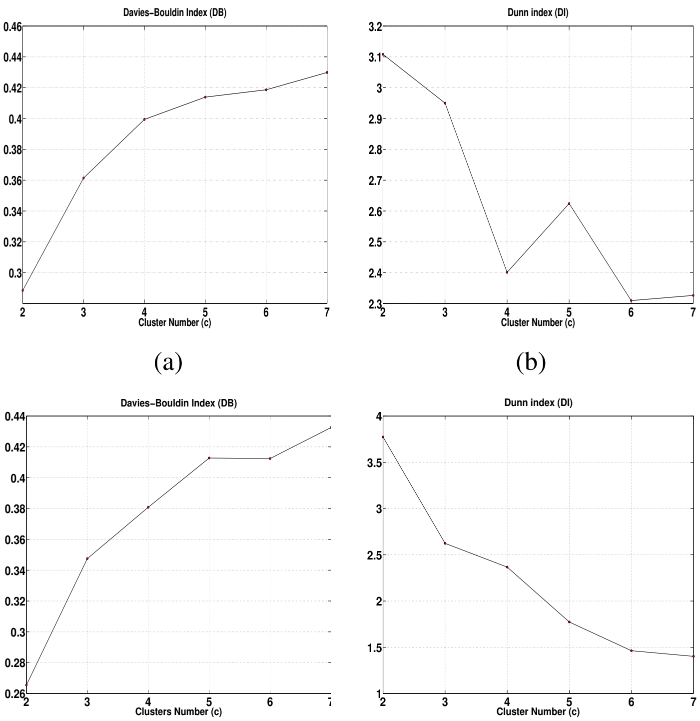
- *Convergence step*: if $\|\theta^{t+1} - \theta^{t+1}\|$ stop (finish iteration); otherwise, go to step 2.

The hidden variables are the parameters of the model. In our case we use mixtures of Gaussians; hence our hidden variables are the mean and standard deviation for each Gaussian distribution. We start with an initial estimate of those parameters and iteratively run the algorithm to find the maximum likelihood for our estimates. For convergence we run it number of times so that the values stop increasing.

B. Optimal Number of Clusters

Once the collection of hard c-partitions labeled images are generated, one needs to determine the optimal number of clusters, c^*-partition, which better reproduces the shape of the SST patterns. Indeed, the clusters validity indices are widely used to evaluate the c-partitions and then extract the optimal number of clusters.

A number of validity indices for hard clustering exist in the literature [22], [21], all of which have a common goal to find the clustering which results in compact clusters that are well separated. In particular, the Davies-Bouldin (V_{DB}) [8] and



(a) (b) (c) (d)

Fig. 2: Results of the validity indices with EM algorithm using respectively the V_{DB} and V_{DI} indices to the SST images in Fig. 1(a) ((a) and (b)) and Fig. 1(b) ((c) and (d)).

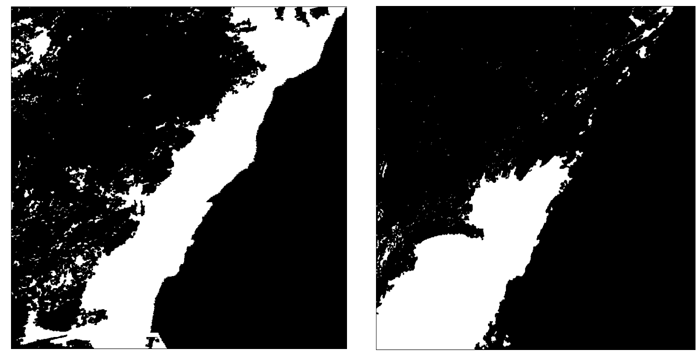
Dunn's (V_{DI}) [9] indices have been frequently used due to their computational efficiency and effectiveness for estimating the good number of clusters [16].

Davies-Bouldin index is a function of ratio of the sum of within cluster scatter to between cluster scatter [16]. The optimal cluster number c^* is found by solving $\min_{c_{min} \leqslant c \leqslant c_{max}} V_{DB}(c)$ to produce the best clustering performance for data. In other hand, the Dunn index is a ratio of within cluster and between cluster separations. The number of clusters that maximizes the V_{DI} index is taken as the optimal number of clusters, $\max_{C_{min} \leqslant C \leqslant C_{max}} V_{DI}(c)$. Consequently, these two indices are used in the current work to assess the quality of the obtained hard c-partitions. Hence, the best c-partition is obtained by taking respectively the global minimum and the global maximum of V_{DB} and V_{DI} indices with respect to $c = 2, 3, ..., c_{max}$.

The results of the two validity indices (V_{DB} and V_{DI}), applied to the original SST images in Fig. 1 using the EM algorithm with number of cluster varying from $c_{min} = 2$ to $c_{max} = 7$, are shown in Fig. 2.

As shown, the absolute extremum of the four curves is reached by the cluster number $c* = 2$, which is consistent with the suitable number of cluster to reproduce the areas covered by upwelling waters for SST images in Fig. 1, and also for all the images analyzed in this study.

Once the labeled image is generated using the expectation-maximization algorithm with the optimal number of clusters, $c^* = 2$, the upwelling area is defined as the clusters with the lowest mean value, based on the fact that the upwelling region is characterized by a cold temperature waters compared to the offshore warmer waters [19].



(a) (b)

Fig. 3: Binary image result after the application of the expectation-maximization algorithm the two SST images in Fig. 1, with the automatic selection of the number of clusters using the Davies-Bouldin and Dunn's indices.

Fig. 3 shows the binary image results after the application of the expectation-maximization algorithm with the automatic selection of the number of clusters using the Davies-Bouldin and Dunn's indices.

C. Segmentation algorithm

The binary images results in Fig. 3 include the upwelling area plus additional noise structures in offshore waters not belonging to the upwelling region, mainly due to the cloud contamination pixels which are not properly masked by the cloud detection algorithm used in this study [18]. Therefore, a segmentation procedure is required consisting on the application of the area opening algorithm to remove pixels and contiguous regions faraway from the coast, not belonging to the upwelling waters. The area opening is a transformations using a structuring element which locally adapts its shape to the image structures, and therefore filtering out small regions (lower than a given threshold) without damaging the remaining structures in the images [20]. An evaluation has been carried out for this study using the 30 SST AVHRR images, in order to select a single threshold value that remove the noisy structures from the images. After an extensive evaluation, good results are obtained by the threshold value of 200 pixels. So, each regions less than 200 pixels, will be discarded from the given image. Fig. 4 shows the results of the area opening algorithm applied to the binary images of Fig. 3 with the thresholding value fixed at 200 pixels. As we can see, all the remaining pixels in offshore waters have been removed and separated from the upwelling area.

IV. Experimental Results and Discussion

The performance of the proposed methodology of segmentation of Moroccan coastal upwelling in SST images has been tested by using the visual inspection made by the oceanographer based on the scientific knowledge of the Moroccan atlantic coast. The in-situ measurement, e.g., buoys data, can give us a sounder support to the results obtained

(a) (b)

Fig. 4: Retrieved upwelling areas using the area opening algorithm of the two binary images results of Fig. 3.

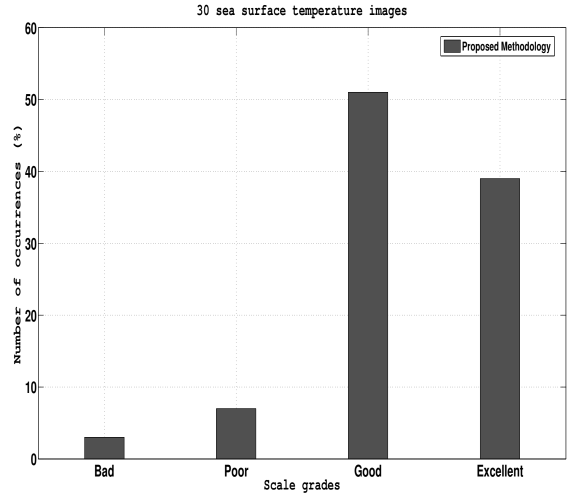


Fig. 5: Qualitative interpretation made by the oceanographer of the 30 SST images for the proposed approach using the Expectation-Maximization.

by the proposed algorithm. However, these measures are very complex to establish and would be extremely challenging for a use in coastal zone context, because of lack of these measures for all the satellite data and for every specific region in the study area. In this sense, it is a well-known fact that validation in the case of ocean data is often and truly performed by assessing the results by professional oceanographers [19], [13], [17], which has been done in the case of this work.

The oceanographic interpretation is applied in order to check if the proposed approach correctly identify the upwelling area over the 30 SST images. For this purpose, four grades are used: "Bad", "Poor", "Good" and "Excellent". The grade "Bad" is attributed when the upwelling area is not well-delimitated, and "Excellent" is reached when the upwelling region is correctly segmented. The results of the oceanographer interpretation are presented in Fig. 5.

As we can see, the Fig. 5 shows that the grades "Bad" and "Poor" reached respectively the values 3% and 7%, resulting in a total of 10% of 30 SST images. In other hand, the values of 51% and 39% are reached respectively by the grades "Good" and "Excellent", making a total of 90%. After this evaluation of the segmentation methodology, over this representative database, we can conclude that the proposed algorithm has provided satisfactory results.

V. Conclusions and Future Work

In the present research we have presented a classification methodology for automatic detection and segmentation of Moroccan coastal upwelling using the SST images. The algorithm is started with the application of Expectation-Maximization algorithm followed by cluster validity indices, for the purpose of generating labeled image with homogeneous groups of pixels and no-overlapping classes. Then a segmentation procedure is developed to eliminate isolated components, in offshore waters, not belonging to the upwelling structure, based on the area opening algorithm. The algorithm is applied and tested over 30 AVHRR SST images on board the NOAA satellite serie and evaluated using the scientific knowledge of the Moroccan costal upwelling provided by the oceanographer. The proposed algorithm has demonstrated a great performance for automatic detection of upwelling region in coastal region of Morocco for majority of images used in this study. This work can be continued in order to investigate the interannual variability of the upwelling in terms of upwelling activity, and also the objective of using this segmentation methodology to other satellite data products, such as the remotely sensed sea-surface chlorophyll field, to give us a sounder support to the results already obtained.

References

[1] M.R. Anderberg. *Cluster Analysis for Application*. Academic Press, New York, 1973.

[2] David M. Anderson and Warren L. Prell. A 300 kyr record of upwelling off oman during the late quaternary: Evidence of the asian southwest monsoon. *Paleoceanography*, 8(2):193–208, 1993.

[3] A. Atillah, A. Orbi, K. Hilmi, and A. Mangin. Produits opérationnels d'océanographie spatiale pour le suivi et l'analyse du phénomène d'upwelling marocain. *Geo Observateur*, 14:49–62, 2005.

[4] A. Bakun. Fronts and eddies as key structures in the habitat of marine fish larvae : opportunity, adaptive response and competitive advantage. *Scientia Marina*, 70(S2):105 – 122, 2006.

[5] P. S. Bradley, U. Fayyad, and C. Reina. Scaling em (expectation maximization) clustering to large databases. *Technical Report, Microsoft Research Centre, Redmond, USA*, 1998.

[6] J. F. Cayula and P. Cornillon. Edge detection algorithm for SST images. *Journal of Atmospheric and Oceanic Techology*, 9:67 – 80, 1992.

[7] E. Chassot, S. Bonhommeau, G. Reygondeau, K. Nieto, J. J. Polovina, M. Huret, N. K. Dulvy, and H. Demarcq. Satellite remote sensing for an ecosystem approach to fisheries management. *Ices Journal of Marine Science*, 68(4):651–666, 2011.

[8] D. L. Davies and D. W. Bouldin. A cluster separation measure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI*, 1(2):224–227, April 1979.

[9] J. C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3):32–57, 1973.

[10] R. J. Holyer and S. H. Peckinpaugh. Edge detection applied to satellite imagery of the oceans. *IEEE Transactions on Geoscience and Remote Sensing*, 27(1):46 – 56, 1989.

[11] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988.

[12] J. Marcello, F. Marques, and F. Eugenio. Automatic tool for the precise detection of upwelling and filaments in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 43:1605–1616, 2005.

[13] S. Nascimento, P. Franco, F. Sousa, J. Dias, and F. Neves. Automated computational delimitation of SST upwelling areas using fuzzy clustering. *Computers and Geosciences*, 43:207–216, 2012.

[14] A.V. Nefian, Luhong Liang, Xiaobo Pi, Liu Xiaoxiang, Crusoe Mao, and Kevin Murphy. A coupled hmm for audio-visual speech recognition. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, volume 2, pages II–2013–II–2016, May 2002.

[15] K. Nieto, H. Demarcq, and S. McClatchie. Mesoscale frontal structures in the canary upwelling system: New front and filament detection algorithms applied to spatial and temporal patterns. *Rem. Sens. Environ.*, 123:339–346, 2005.

[16] M. K. Pakhira, S. Bandyopadhyay, and U. Maulik. Validity index for crisp and fuzzy clusters. *Pattern Recognition*, 37(3):487–501, 2004.

[17] Fatima M. Sousa, Susana Nascimento, Hugo Casimiro, and Dmitri Boutov. Identification of upwelling areas on sea surface temperature images using fuzzy clustering. *Remote Sensing of Environment*, 112(6):2817 – 2823, 2008.

[18] Larry L. Stowe, Paul A. Davis, and E. Paul McClain. Scientific basis and initial evaluation of the clavr-1 global clear/cloud classification algorithm for the advanced very high resolution radiometer. *Journal of Atmospheric and Oceanic Technology*, 16(6):656, 1999.

[19] A. Tamim, K. Minaoui, K. Daoudi, H. Yahia, A. Atillah, M. F. Smiej, and D. Aboutajdine. A simple and efficient approach for coarse segmentation of Moroccan coastal upwelling. In *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European*, pages 1–5, Sept 2013.

[20] Luc Vincent. Morphological area openings and closings for grey-scale images. In *Shape in Picture*, volume 126, pages 197–208. Springer Berlin Heidelberg, 1994.

[21] Kuo-Lung Wu and Miin-Shen Yang. A cluster validity index for fuzzy clustering. *Pattern Recognition Letters*, 26(9):1275 – 1291, 2005.

[22] Yun Xu and Richard G. Brereton. A comparative study of cluster validation indices applied to genotyping data. *Chemometrics and Intelligent Laboratory Systems*, 78(12):30 – 40, 2005.