



**HAL**  
open science

## Automatic Detection of Moroccan Coastal Upwelling Zones using Sea Surface Temperature Images

Ayoub Tamim, Khalid Minaoui, Khalid Daoudi, Hussein Yahia, Abderrahman Atillah, Salma El Fellah, Driss Aboutajdine, Mohamed El Ansari

► **To cite this version:**

Ayoub Tamim, Khalid Minaoui, Khalid Daoudi, Hussein Yahia, Abderrahman Atillah, et al.. Automatic Detection of Moroccan Coastal Upwelling Zones using Sea Surface Temperature Images. International Journal of Remote Sensing, 2018. hal-01881880

**HAL Id: hal-01881880**

**<https://inria.hal.science/hal-01881880>**

Submitted on 26 Sep 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Automatic Detection of Moroccan Coastal Upwelling Zones using Sea Surface Temperature Images

Ayoub Tamim<sup>abc</sup>, Khalid Minaoui<sup>c</sup>, Khalid Daoudi<sup>d</sup>, Hussein Yahia<sup>d</sup>, Abderrahman Atillah<sup>e</sup>, Salma El Fellah<sup>c</sup>, Driss Aboutajdine<sup>c</sup> and Mohamed El Ansari<sup>b</sup>

<sup>a</sup>Higher Institute of Marine Fisheries (ISPM), Department of Marine Fisheries, Agadir, Morocco;

<sup>b</sup>Faculty of Science, LABSIV Computer Science, Ibn Zohr University, Agadir, Morocco; <sup>c</sup>Faculty of Science, LRIT Associated Unit to the CNRST, URAC 29, Mohammed V University, Rabat, Morocco;

<sup>d</sup>INRIA Bordeaux Sud-Ouest (GEOSTAT team), Talence, France; <sup>e</sup>Royal Centre for Remote Sensing (CRTS), Rabat, Morocco

### ARTICLE HISTORY

Compiled September 21, 2018

### ABSTRACT

An efficient unsupervised method is developed for automatic segmentation of the area covered by upwelling waters in the coastal ocean of Morocco using the Sea Surface Temperature (SST) satellite images. The proposed approach first uses the two popular unsupervised clustering techniques, *k*-means and fuzzy *c*-means (FCM), to provide different possible classifications to each SST image. Then several cluster validity indices are combined in order to determine the optimal number of clusters, followed by a cluster fusion scheme, which merges consecutive clusters to produce a first segmentation of upwelling area. The region-growing algorithm is then used to filter noisy residuals and to extract the final upwelling region. The performance of our algorithm is compared to a popular algorithm used to detect upwelling regions and is validated by an oceanographer over a database of 92 SST images covering each week of the years 2006 and 2007. The results show that our proposed method outperforms the latter algorithm, in terms of segmentation accuracy and computational efficiency.

### KEYWORDS

Sea surface temperature images; Upwelling; *k*-means clustering; Fuzzy *c*-means clustering; Cluster validity indices; Region-growing algorithm

## 1. Introduction

The canary current system is considered one of the four major eastern boundary currents, which are affected by the upwelling phenomenon. In particular, the Moroccan Atlantic coast, located in the central canary system, is exclusively dominated by persistent and variable upwelling throughout the year (Atillah et al. 2005). In fact, under the influence of northeasterly winds, the coastal upwelling takes place along the Moroccan continental shelf, which is evidenced at the surface by cold and usually nutrient-rich waters near the coast and by filaments extending hundreds of km offshore. The upwelling regions are shown to be loci of high biological activity and fishery production (Chassot et al. 2011).

The Sea Surface Temperature (SST) images obtained from the Advanced Very High Resolution Radiometer (AVHRR) sensor aboard the National Oceanic and Atmospheric Administration satellite (NOAA) are largely used to detect the thermal upwelling front located at the

boundary between two distinct temperature populations. Nevertheless, the large amount of SST images daily collected renders the visual interpretation made by oceanographers labour-intensive and time-consuming. So, the need for automatic tools of upwelling detection keeps growing.

A number of automated techniques have been proposed to detect the upwelling areas in oceanographic satellite images. Some of the most popular approaches include the use of the histogram-based separation (Nieto, Demarcq, and McClatchie 2005), which relies on the fact that upwelling fronts are usually regarded as boundaries between two water masses of constant temperature; neural networks (Chaudhari, Balasubramanian, and Gangopadhyay 2008), where the SST image is trained based on  $k$ -means segmentation results and a quantitative criterion is developed to test the existence of upwelling in each segmented image; hybrid method (Marcello, Marques, and Eugenio 2005), that has also been developed to identify the area covered upwelling waters, using region of interest histogram and region-growing process.

In the study carried out for the Portuguese coastal ocean (Nascimento et al. 2012), the investigators are specifically interested in locating the upwelling areas using the fuzzy  $c$ -means clustering. The algorithm used the Affinity Propagation-Fuzzy  $C$ -means (AP-FCM) clustering for the purpose of finding clusters of homogeneous and non-overlapping temperatures. Then a set of features is gathered from each cluster in order to detect the main upwelling front separating the cold waters near the coast and warmer offshore waters. The upwelling boundaries are delineated using measures of classification uncertainty. Compared to the preceding segmentation methodologies, the latter constitutes a natural advantage to classify the SST images, due to the presence of smooth thermal boundaries between the upwelling regions (Nascimento et al. 2012). Additionally, the algorithm does not require a relatively heavy pre-processing step in order to ensure an acceptable segmentation results. However, the AP-FCM algorithm suffers from the following shortcomings: 1) the quality of the segmentation results highly depends on the accuracy of the single IAP (Iterative Anomalous Pattern) algorithm (Mirkin, Boris 2005), used to determine the good number of clusters in SST images; 2) the use of control parameters, which should be adjusted to obtain an admissible segmentation. In this context, our proposed approach is designed to overcome these inconveniences and to achieve accurate segmentation of the main upwelling front in an efficient way and without tuning parameters.

Clustering techniques are among the unsupervised methods, which aim at partitioning objects into a number of clusters, showing a certain degree of closeness or similarity. One possible classification of clustering techniques can be according to whether an object belongs exclusively to a one cluster (hard clustering) (Likas et al. 2003), or to several clusters (fuzzy clustering) (Leski 2003). In the present study, we have respectively used the  $k$ -means and FCM algorithms as underlying hard and fuzzy clustering for the purpose of labeling the original SST images.

Whatever the clustering method is, one first needs to determine the right number of clusters, resulting of compact and well-separated groups. To achieve this end, we have used 19 validity functions that quantitatively evaluate individual groups of clusters partition, and extract the number of classes that better fits the oceanographic structures in SST images. The objective of this study is to provide a simple, fast and efficient algorithm for identification of the main upwelling front between cold and warm waters. The algorithm can be summarised by the following three steps:

- 1 Classification methodology, to obtain a labelled image, based on the two possible strategies of classification,  $k$ -means and FCM, with the number of clusters determined using 19 validity indices;

Year	January	February	March	April	May	June	July	August	September	October	November	December
2006	06, 14, 24, 28	03, 16, 25	05, 13, 15, 25	02, 07, 22, 25	04, 15, 22	01, 07, 17, 23, 30	07, 19, 27, 31	11, 19, 21, 31	10, 20, 25	01, 10, 19, 24	02, 14, 24, 26	06, 15, 26, 29
2007	02, 14, 20, 30	09, 14, 22	02, 12, 14, 22	04, 14, 22, 27	08, 11, 19, 30	08, 16, 19, 28	07, 15, 22	03, 05, 16, 25	04, 11, 14, 24	03, 12, 21, 29	01, 09, 17	02, 03, 17, 26

**Table 1.** Dates of the 92 SST images used in this study covering the years 2006 and 2007.

- 2 Segmentation methodology, to detect the main upwelling front, corresponding to strong variation between two consecutive clusters, using the simple arithmetic SST Mean collected from each labelled region;
- 3 Post-processing methodology, to extract and separate the upwelling area from the remaining noisy structures in offshore waters, using the region-growing process.

## 2. Study area and data

In this study 92 AVHRR SST images obtained during the years 2006 and 2007 are used, covering the southern Moroccan Atlantic coast. The data is received and processed at the Royal Centre of Remote Sensing (CRTS) of Morocco, including radiometric, geometric and atmospheric corrections with cloud and land masks. The cloud overlay is generated using the multispectral radiance measurements algorithm tested over the AVHRR images (Stowe, Davis, and McClain 1999). Each SST image consists of  $770 \times 990$  pixels with a spatial resolution of  $1.1 \times 1.1$  km and each pixel is a temperature in  $^{\circ}\text{C}$ .

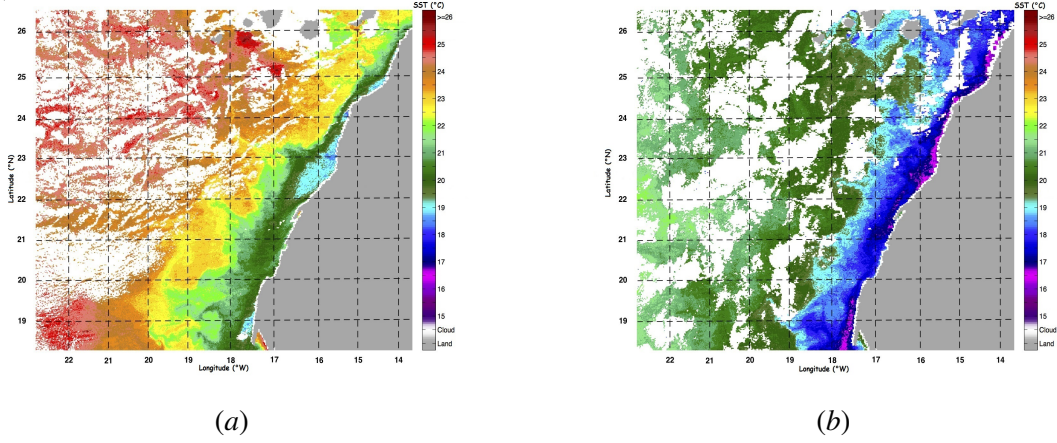
An automatic tool for selecting the SST images from the years 2006 and 2007 is developed due to the large amount of data and due to the fact that heavy cloud cover will exist in several image analysed. For this purpose, one image among the eight generated during the week is retained for the analysis if it has a maximum value of valid information (no clouds or missing pixels) between the coast and 200 km offshore in a direction perpendicular to the coastline. This is due to the fact that the Moroccan coastal upwelling is largely conditioned by the width and the direction of the continental shelf (Atillah et al. 2005). The dates of the 92 SST images selected from the years 2006 and 2007 are detailed in Table 1.

A colour scale of 26 levels is applied to each image in order to help the oceanographers for visual inspection of the upwelling area.

Fig. 1 shows two data, selected from our database of 92 SST images, illustrating two upwelling scenarios encountered throughout this study: 1) well-defined upwelling in terms of continuity along the coast, with a clear and visible thermal upwelling front separating cold waters near the coast and warmer offshore waters (Fig. 1(a)); 2) SST images where the presence of clouds and missing pixels over the upwelling area, masks the true underlying structure of upwelling (Fig. 1(b)).

## 3. Classification methodology

Simplicity and computational efficiency are major concerns in our task. We thus choose two widely used unsupervised clustering techniques,  $k$ -means and FCM, representing respectively hard and fuzzy clustering, to divide a given SST image into coherent and contrasted clusters. In order to make the paper self-contained, we briefly recall the preceding two algorithms.



**Figure 1.** SST images obtained on (a) 11 September 2007 and (b) 15 May 2006, showing two upwelling scenarios.

### 3.1. Clustering algorithms

Among the various hard clustering methods, the  $k$ -means algorithm (Jain Anil, and Dubes Richard 1988) is one of the simplest. Based on the temperature value in each pixel, the SST images is converted into a data set  $X = \{x_1, x_2, \dots, x_N\}$  of  $N$  pixels, and the algorithm allocates each data point to one of  $c$  clusters to minimize the within-cluster sum of squares:

$$\sum_{i=1}^c \sum_{k \in I_i} \|x_k - v_i\|^2, \quad (1)$$

where  $I_i$  is a set of objects in the  $i$ -th cluster and  $v_i$  is the mean for that points over the cluster  $i$ . In  $k$ -means clustering,  $v_i$  is called the cluster prototype:

$$v_i = \frac{\sum_{k=1}^{N_i} x_k}{N_i}, \quad x_k \in I_i, \quad (2)$$

and  $N_i$  is the number of objects in cluster  $I_i$ .

In the other hand, the fuzzy partition can be seen as a generalization of hard partition. In particular, the FCM algorithm (Bezdek James 1981) is based on the minimization of an objective function:

$$J = \sum_{i=1}^c \sum_{k=1}^N (\mu_{i,k})^m \|x_k - v_i\|^2, \quad (3)$$

where  $\mathbf{V} = [v_1, v_2, \dots, v_c]$  is a vector of cluster centers, and  $\|x_k - v_i\|^2$  is the Euclidean norm. The parameter ' $m$ ' controls the fuzziness of the clusters found by the algorithm. It must be given as a scalar greater or equal to one ( $m \in [1, \infty]$ ). For example, if  $m = 1$ , the clusters are separated by hard partitions and  $\mu_{i,k}^m = 0$  or 1. As ' $m$ ' increase, the partitions become more fuzzy. In the case of the Fuzzy clustering method, the default value of the weighting exponent  $m$  is 2.

The cluster centroids and the respective membership functions that solve the constrained

Measure	Notation	Definition	Optimal values
Partition coefficient	PC (Bezdek James 1981)	$\frac{1}{N} \sum_{i=1}^{N_c} \sum_{k=1}^N (\mu_{i,k})^2$	local maxima $_{2 \leq N_c \leq N-1}$ PC
Partition entropy	PE (Bezdek, Ehrlich, and Full 1984)	$-\frac{1}{N} \sum_{i=1}^{N_c} \sum_{k=1}^N \mu_{i,k} \log_2 \mu_{i,k}$	local minima $_{2 \leq N_c \leq N-1}$ PE
Dave index	MPC (Dave 1996)	$1 - \frac{N}{N_c-1} (1 - PC)$	local maxima $_{2 \leq N_c \leq N-1}$ MPC
Extended partition entropy	EPE (Zhang, Zhou, and Martin 2008)	$-\frac{1}{N \times \ln(N_c)} \sum_{i=1}^{N_c} \sum_{k=1}^N \mu_{i,k}$	local minima $_{2 \leq N_c \leq N-1}$ EPE
Chen and Linkens Index	P (Chen and Linkens 2004)	$\frac{1}{N} \sum_{k=1}^N \max_i(\mu_{i,k}) - \frac{1}{\sum_{i=1}^{N_c-1} \sum_{j=i+1}^{N_c} \sum_{k=1}^N \min(\mu_{i,k}, \mu_{j,k})}$	local maxima $_{2 \leq N_c \leq N-1}$ P

**Table 2.** Description of the 5 validity indices involving only the membership values.  $N$ : number of objects in the dataset  $X$ ;  $N_c$ : number of clusters in the dataset  $c$ ;  $\mu_{i,j}$  grade of membership of the  $j$ th element to the  $i$ th cluster.

optimization problem in equation (3) are given by the following equations:

$$v_i = \frac{\sum_{j=1}^N (\mu_{i,j})^m x_j}{\sum_{j=1}^N (\mu_{i,j})^m}, \quad 1 \leq i \leq N_c, \quad (4)$$

$$\mu_{i,k} = \frac{1}{\sum_{j=1}^{N_c} \left( \frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{\frac{2}{m-1}}}, \quad 1 \leq i \leq N_c, \quad 1 \leq k \leq N. \quad (5)$$

The goal of FCM algorithm is to iteratively improve a sequence of sets of fuzzy clusters in equation (3) by a simple iteration through the equation (4) and equation (5).

### 3.2. Cluster validity indices

A common problem in unsupervised clustering is the optimal choice of the number of classes, and cluster validity has been widely investigated for this purpose (Xu and Brereton 2005). In particular, Sousa et al. (2008) uses the cluster validity indices to extract the appropriate number of clusters in SST images.

A large number of validity functions exist in the literature, which can be grouped according to two different criteria: indices using only the membership values and have the advantage of being easy to compute; and indices which involve both the membership values and the dataset itself, which take into account the underlying connection in the data.

In our proposed strategy of determining the right number of clusters, 19 validity indices are selected and computed according to three criteria: well-established, implementation simplicity and computational considerations. In Table 2 and Table 3 we respectively present a detailed study of 5 widely used validity indices involving only the membership values, and a suite of 14 widely used validity indices involving the membership values and the dataset.

The indices are thus applied to our database of 92 SST images and the appropriate number of clusters is selected through a vote scheme.

### 3.3. Determining the optimal number of clusters

The 19 validity indices used in our study can be applied either in hard or fuzzy partition, and each of these index gives the optimal clustering scheme for the given image based on two criteria: compactness (measure of closeness of cluster elements) and separation (indicate how distinct two clusters are).

The foremost shortcoming encountered when attempting to detect the optimal number of cluster using the validity indices, is the nature of the data itself. In fact, we have noticed that

Measure	Notation	Definition	Optimal values
Partition index	SC (Bensaid et al. 1996)	$\sum_{i=1}^{N_c} \frac{\sum_{j=1}^N (\mu_{i,j})^m \ x_j - v_i\ ^2}{N_i \sum_{k=1}^{N_c} \ v_k - v_i\ ^2}$	local minima $_{2 \leq N_c \leq N-1}$ SC
Xie and Beni	XB (Xie and Beni 1991)	$\frac{\sum_{i=1}^{N_c} \sum_{j=1}^N (\mu_{i,j})^m \ x_j - v_i\ ^2}{N \min_{i,j} \ v_i - v_j\ ^2}$	local minima $_{2 \leq N_c \leq N-1}$ XB
Fukuyama and Sugeno	FS (Fukuyama and Sugeno 1989)	$\sum_{i=1}^{N_c} \sum_{j=1}^N \mu_{i,j}^m \ x_j - v_i\ ^2 - \sum_{i=1}^{N_c} \sum_{j=1}^N \mu_{i,j}^m \ v_i - \sum_{k=1}^{N_c} \frac{\mu_{k,j}}{N_c}\ ^2$	local minima $_{2 \leq N_c \leq N-1}$ FS
Kwon index	K (Kwon 1998)	$\frac{\sum_{i=1}^{N_c} \sum_{j=1}^N \mu_{i,j}^m \ x_j - v_i\ ^2 + \frac{1}{N} \sum_{i=1}^{N_c} \ v_i - \sum_{j=1}^N \frac{\mu_{i,j}}{N}\ ^2}{\min_{i \neq k} \ v_i - v_k\ ^2}$	local minima $_{2 \leq N_c \leq N-1}$ K
T index	T (Tang, Fuchun, and Zengqi 2005)	$\frac{\sum_{i=1}^{N_c} \sum_{j=1}^N \mu_{i,j}^m \ x_j - v_i\ ^2 + \frac{1}{N_c(N_c-1)} \sum_{i=1}^{N_c} \sum_{k=1, k \neq i}^{N_c} \ v_i - v_k\ ^2}{\min_{i \neq k} \ v_i - v_k\ ^2 + \frac{1}{N_c}}$	local minima $_{2 \leq N_c \leq N-1}$ T
Separation and compactness index	Z (Zahid and Essaid 1999)	$\frac{\sum_{i=1}^{N_c} \frac{\ v_i - \sum_{j=1}^N \frac{\mu_{i,j}}{N}\ ^2}{N_c} - \sum_{i=1}^{N_c} \sum_{i+1}^N \frac{(\sum_{j=1}^N \min(\mu_{i,j}, \mu_{i+1,j}))^2}{\sum_{j=1}^N \min(\mu_{i,j}, \mu_{i+1,j})}}{\sum_{i=1}^{N_c} \sum_{j=1}^N \frac{\mu_{i,j}^m \ x_j - v_i\ ^2}{\sum_{i=1}^{N_c} \mu_{i,j}^m} - \frac{\sum_{i=1}^{N_c} (\max_{k \leq i \leq N_c} \mu_{i,j})^2}{\sum_{i=1}^{N_c} \max_{k \leq i \leq N_c} \mu_{i,j}}}$	local maxima $_{2 \leq N_c \leq N-1}$ Z
Fuzzy hypervolume validity	FHV (Gath and Geva 1989)	$\sum_{i=1}^{N_c} \left[ \det \frac{\sum_{j=1}^N \mu_{i,j}^m (x_j - v_i)(x_j - v_i)^T}{\sum_{j=1}^N \mu_{i,j}^m} \right]^{\frac{1}{2}}$	local minima $_{2 \leq N_c \leq N-1}$ FHV
Average partition density	APD (Davies et Bouldin 1979)	$\frac{1}{N_c} \sum_{i=1}^{N_c} \frac{\sum_{j \in X_i} \mu_{i,j}}{\left[ \det \frac{\sum_{j=1}^N \mu_{i,j}^m (x_j - v_i)(x_j - v_i)^T}{\sum_{j=1}^N \mu_{i,j}^m} \right]^{\frac{1}{2}}}$	local maxima $_{2 \leq N_c \leq N-1}$ APD
Partition density	PD (Davies et Bouldin 1979)	$\frac{\sum_{i=1}^{N_c} \sum_{j \in X_i} \mu_{i,j}}{\text{FHV}}$	local maxima $_{2 \leq N_c \leq N-1}$ PD
PBMF index	PBMF (Pakhira, Bandyopadhyay, and Maulik 2004)	$\left( \frac{1}{N_c} \times \frac{\sum_{i=1}^N \mu_{i,j} \ x_j - v_i\ }{\sum_{i=1}^{N_c} \sum_{j=1}^N \mu_{i,j} \ x_j - v_i\ } \times \max_{i,j} \ v_i - v_j\  \right)^2$	local maxima $_{2 \leq N_c \leq N-1}$ PBMF
Bouguessa and Wang index	SCG (Bouguessa and Wang 2004)	$\frac{\text{trace} \left( \frac{\sum_{i=1}^{N_c} \sum_{j=1}^N \mu_{i,j}^m (v_i - \bar{v})(v_i - \bar{v})^T}{\sum_{i=1}^{N_c} \text{trace} \left( \frac{\sum_{j=1}^N \mu_{i,j}^m (x_j - v_i)(x_j - v_i)^T}{\sum_{j=1}^N \mu_{i,j}^m} \right)} \right)}{\sum_{i=1}^{N_c} \text{trace} \left( \frac{\sum_{j=1}^N \mu_{i,j}^m (x_j - v_i)(x_j - v_i)^T}{\sum_{j=1}^N \mu_{i,j}^m} \right)}$	local maxima $_{2 \leq N_c \leq N-1}$ SCG
Davies-Bouldin index	DB (Davies et Bouldin 1979)	$\frac{1}{N_c} \sum_{j,j \neq i} \max \left\{ \frac{\left[ \frac{1}{N_c} \sum_{i \in c_i} d(x, v_i) + \frac{1}{N_c} \sum_{i \in c_j} d(x, v_j) \right]}{d(v_i, v_j)} \right\}$	local minima $_{2 \leq N_c \leq N-1}$ DB
Dunn's index	DI (Dunn 1973)	$\min_{i \in c} \left\{ \min_{j \in c, j \neq i} \left\{ \frac{\min_{x \in c_i, y \in c_j} d(x, y)}{\max_{k \in c, x, y \in c} d(x, y)} \right\} \right\}$	local maxima $_{2 \leq N_c \leq N-1}$ DI
Calinski-Harabasz index	CH (Calinski and Harabasz 1974)	$\frac{\sum_{i=1}^{N_c} N_i d^2(v_i, \bar{v}) / (N_c - 1)}{\sum_{i=1}^{N_c} \sum_{x \in c_i} d^2(x, v_i) / (N - N_c)}$	local maxima $_{2 \leq N_c \leq N-1}$ CH

**Table 3.** Description of the 14 validity indices involving the membership values and the dataset.  $N$ : number of objects in the dataset  $X$ ;  $N_c$ : number of clusters in the dataset  $c$ ;  $\mu_{i,j}$ : grade of membership of the  $j$ th element to the  $i$ th cluster;  $v$ : center of  $X$ ;  $c_i$ : the  $i$ -th cluster;  $N_i$ : number of objects in the cluster  $c_i$ ;  $v_i$ : center of  $c_i$ ;  $d(x, y)$ : distance between  $x$  and  $y$ ;  $X_j$ : set of data points that are within a pre-specified region around cluster centroid  $v_j$ .

Image	PC	PE	EPE	MPC	P	SC	XB	FS	K	T	Z	FHV	APD	PD	PBMF	SCG	DB	DI	CH	$C^*$
11 September 2007	-	-	2, 6	4, 6	4, 6	-	2, 5	3, 5	2, 5	-	-	3, 6	-	3, 6	-	-	2, 6	2, 5	-	6
15 May 2006	2, 4	-	2, 4	2, 4	2, 4	4, 6	2, 4	-	2, 4	2, 4	4, 6	2, 4	-	2, 4	3, 5	2, 4	-	2, 4	-	4

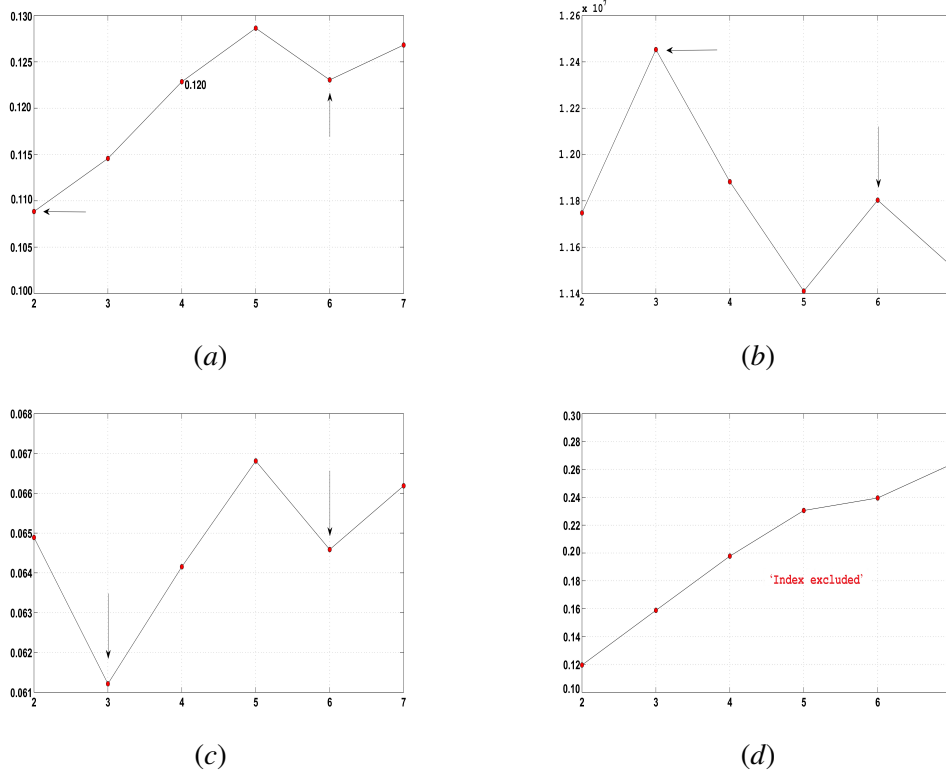
**Table 4.** Optimal number of clusters,  $C^*$ , using 19 validity indices to the images in Fig. 1.

none of the abovementioned indices correctly recognizes the optimal cluster numbers ' $c^*$ ' for all our 92 SST images. As Pal and Bezdek (1997) stated, no matter how good your index is, there is a data set out there waiting to trick it. Indeed, the imprecise nature of remote sensing SST's in general and the upwelling patterns in particular, which are typically characterized by the interaction of a complex set of upper ocean and atmospheric boundary layer processes having spatial and temporal scales ranging from meters to hundreds of km and from seconds to several days, can make the weighted voting scheme a very difficult task.

In this sense, we took the decision to use an equally weighted vote and the same amount of influence over the outcome for all the 19 validity indices used in this study, and therefore select the appropriate number of cluster based on the decision of these indices.

The proposed algorithm for determining the number of clusters in each image using the 19 validity indices can be summarised by the following steps:

- Application of the  $k$ -means and FCM algorithms to each SST image with the number of clusters varying from  $C_{\min} = 2$  to  $C_{\max} = 7$ . Our experiments show indeed that a number of clusters larger than  $C_{\max} = 7$ , leads to multiple broken of one or more compact clusters.
- Calculation of the 19 validity indices for each  $C$ -partition and extraction of the local extrema for each index.
- Selection of the best  $C^*$ -partition for each image, being the one that have the maximum number of occurrences of the collected local extrema.

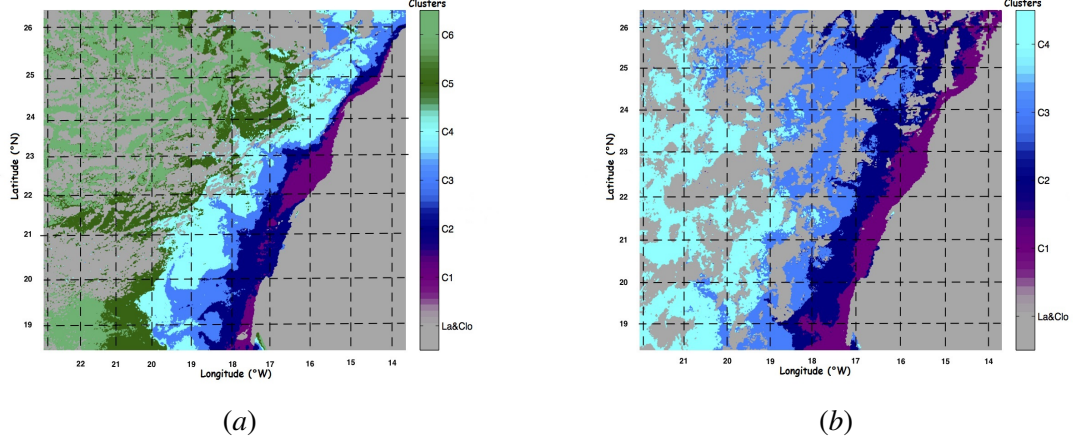


**Figure 2.** Results of four validity indices for the image in Fig. 1(a), using the FCM algorithm with the number of clusters varying between  $C_{\min} = 2$  and  $C_{\max} = 7$  (the black arrows indicate the position of the local extrema). Fig. 2(d) and Fig. 2(a) show respectively the result of the Partition entropy (PE) and the Extended partition entropy (EPE). Fig. 2(b) and Fig. 2(c) show the result of the Partition density (PD) and Fuzzy hypervolume validity (FHV).

Table 4 shows the local extrema extracted from the 19 validity functions, by using FCM algorithm to the images in Fig. 1. Fig. 2 illustrates the graphic chart of 4 validity indices, among the 19 used in this study, of the image in Fig. 1(a).

According to the local minima of the curve in Fig. 2(a), the EPE index yields  $C^* = 2$  and  $C^* = 6$  as good cluster number estimate. In Fig. 2(b) and Fig. 2(c), the PD and FHV indices show that  $C^* = 3$  and  $C^* = 6$  may be another good estimates according respectively to the local maximums and local minimums of the two curves. The PE index in Fig. 2(d) yields a monotonic tendency of the cluster number  $C$ . The validity indices that present a monotonic tendency have been excluded from the process of determining the appropriate number of clusters, because the extreme value in that case is generally not a good estimate (Wu and Yang 2005). After a voting scheme the indices calculated over the two original images indicate respectively, that  $C^* = 6$  and  $C^* = 4$  are the optimal number of clusters. Fig. 3 illustrates the clustering results thus obtained over the images in Fig. 1 using the FCM algorithm.





**Figure 3.** Visualization of the clustering results using the FCM method to the images in Fig. 1, with respectively  $C^* = 6$  and  $C^* = 4$  clusters.

#### 4. Segmentation of upwelling area

Once the classification algorithm is performed, the image is partitioned into  $C^*$  clusters:

$$\mathbf{I} = \bigcup_{c=1}^{C^*} \mathbf{I}_c \quad (6)$$

The goal of the segmentation is then to find the frontier between cold and warm waters, that is to find the clutter  $\tilde{c}$  where there is a significant difference between the temperatures of  $\mathbf{I}_{\tilde{c}}$  and  $\mathbf{I}_{(\tilde{c}+1)}$ . The upwelling area is then defined as:

$$\mathbf{I}_{\text{upw}} = \bigcup_{c=1}^{\tilde{c}} \mathbf{I}_c \quad (7)$$

To do so, we compute the difference between the mean temperatures of consecutive clusters:

$$\Delta\mu_{\mathbf{I}_c} = |\mu_{\mathbf{I}_c} - \mu_{\mathbf{I}_{c+1}}|, \quad 1 \leq c \leq C^* - 1, \quad (8)$$

where  $\mu_{\mathbf{I}_c}$  is the main temperature value of the cluster  $\mathbf{I}_c$ . We define  $\tilde{c}$  as the cluster where  $\Delta\mu_{\mathbf{I}_c}$  is maximum:

$$\tilde{c} = \operatorname{argmax}_c \Delta\mu_{\mathbf{I}_c}, \quad c = 1, \dots, C^* - 1 \quad (9)$$

by convention, we consider that  $\tilde{c} = 1$  if  $C^* = 2$ . The fourth column in Table 5 shows the mean difference between each two consecutive clusters gathered from the labelled images in Fig. 3. The main front in the first image is detected between the clusters  $\mathbf{I}_3$  and  $\mathbf{I}_4$ , while in the second image, the transition cluster is taken between the second and the third cluster.

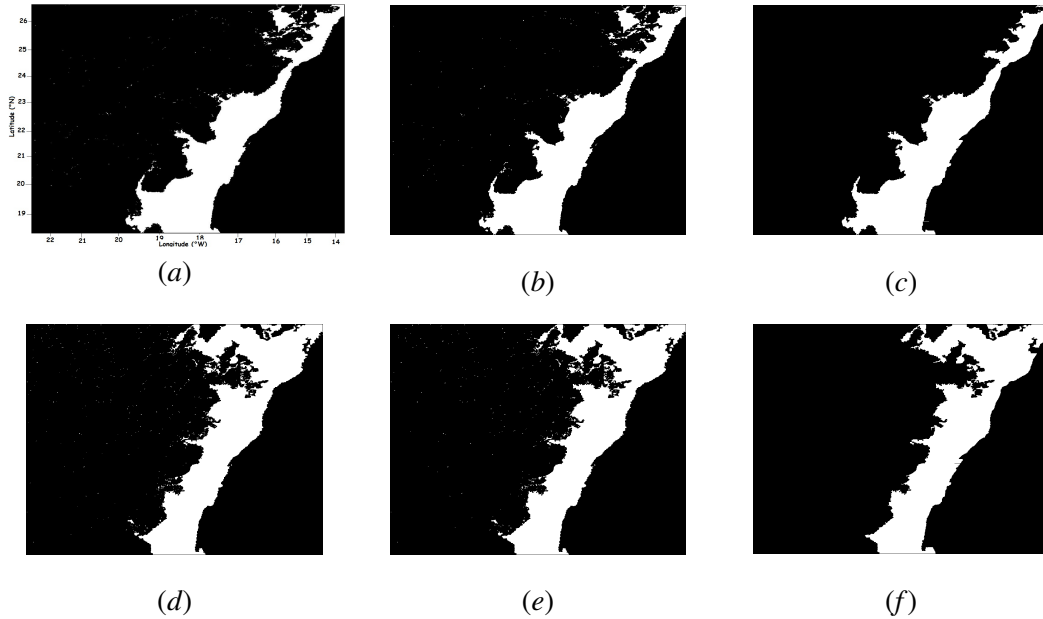
Images	$I_c$	$\mu_{I_c}$ (°C)	$\Delta\mu_{I_c}$ (°C)
11 September 2007	$I_1$	19.658	1.264
	$I_2$	20.922	0.479
	$I_3$	21.401	<b>1.757</b>
	$I_4$	23.158	0.863
	$I_5$	24.021	0.693
	$I_6$	24.714	–
15 May 2006	$I_1$	17.606	1.277
	$I_2$	18.882	<b>1.305</b>
	$I_3$	20.188	1.043
	$I_4$	21.230	–

**Table 5.** Statistical Data for the labelled images in Fig. 3. Bold number represents the cluster transition of the main front.

## 5. Post-processing

The region covered by upwelling waters corresponds to the first  $c$  clusters  $c \leq \tilde{c}$ , with the last,  $\tilde{c}$ , being the cluster transition between two water masses. The problem encountered in the extraction of upwelling zone, is the presence of cold and noisy structures in offshore waters.

The most popular methods used to remove the noisy structures in SST images, are the morphological operators (Soille 2003; Holyer and Peckinpaugh 1989; Marcello, Marques, and Eugenio 2005). We have used the cleaning operator to our problem, but the results are not satisfactory. Indeed, despite the good performance in terms of reducing the isolated pixels, not all the remaining pixels are removed, especially the big isolated groups of pixels in the images (Fig. 4(b)(e)).



**Figure 4.** Upwelling area automatically retrieved. (a) and (d) show the interference between the upwelling area and the noisy structures in left side of the figure. (b) and (e) show respectively the binary images results of the cleaning operator applied to (a) and (b). (c) and (f) show the results of the region-growing algorithm to the images (a) and (b).

Thus, based on the fact that all the segmented pixels pertaining to the upwelling must have connectivity with the coastline, we used the region-growing process (Chang and Li 1994), using three seed points near the coastline that correspond to the upwelling sources. Since we apply the region-growing to the binary image, the algorithm does not require any adjustment

of some threshold value, generally used in this process. This constitutes another important advantage of our methodology. Fig. 4(c) and (f) display the results of the region-growing method, and it can be appreciated the precise elimination of the spatial isolated pixels, not belonging to coastal upwelling.

Fig. 5(a),(b),(d) and (e) shows the results of upwelling areas, automatically contoured with our proposed method, using respectively the  $k$ -means and FCM algorithms.

We shall call IAP-FCM-TCT, the algorithm used in Nascimento et al. (2012). The latter provides labelled images with the corresponding number of clusters in each image by using the IAP-FCM process. The main upwelling front separating the two water masses is detected using the transition-cluster threshold procedure (TCT).

The results of the IAP-FCM-TCT method applied to the original images in Fig. 1 are depicted in Fig. 5(c) and (f). The upwelling area is automatically contoured with the black bold colour.

The Moroccan Atlantic coast, belonging to the Canary upwelling system, is characterized by strong and persistent thermal structures (Atillah et al. 2005; Nykjaer and Van Camp 1994) including upwelling fronts which separate the cold coastal waters and warmer off-shore oceanic waters, and also include the upwelling filaments which, represent the extensions of upwelling fronts in open ocean. The upwelling filaments are very energetic, and their knowledge is very important to study the oceanic circulation and fisheries management and exploitation.

The Fig. 5(a) and Fig. 5(b) (also the Fig. 5(d) and Fig. 5(e)) shows the same segmentation results and provide satisfactory results with a precise detection of the upwelling structure and upwelling filaments over the original image in Fig. 1(a) (see also the Fig. 6 where the segmented image in the Fig. 5(a) is zoomed in to concentrate on the region of filaments structure). On the other hand, in the Fig. 5(c) and Fig. 5(f) (algorithm of Nascimento et al. (2012)), the main thermal upwelling front is well identified, but the detail corresponding to the upwelling filaments is not preserved, as shown in the Fig. 7 where the segmented image in the Fig. 5(c) is zoomed in to concentrate on the region of filaments structure.

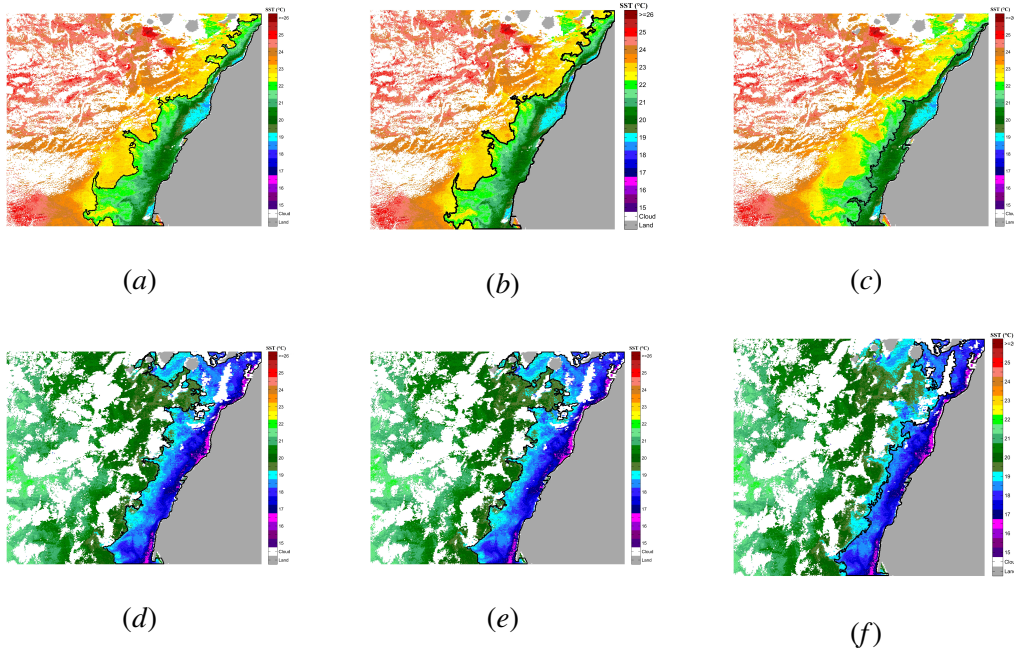
## 6. Experimental study and results

After rejecting images with heavy cloud occlusions and artefacts over the region of interest, a complete database composed of 92 SST images is generated, keeping one image in each week during the years 2006 and 2007. The goal of this section is two-fold. First, the comparison between the two clustering methods used in this work,  $k$ -means and FCM, for deciding which clustering algorithm will perform better on our data. Second, the comparison of our approach and the “IAP-FCM-TCT” algorithm (Nascimento et al. 2012) developed to detect the main upwelling front in SST images.

It is necessary to mention that, the in-situ measurements, e.g., the buoy data, of the upwelling fronts can give us a sounder support to the results obtained by the proposed approach. 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 SST satellite images used in this study and for every specific region of 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 (Nascimento et al. 2012; Tamim et al. 2013; Sousa et al. 2008), which has been done in the case of this work.

The oceanographers evaluation is performed in order to check if the delimitation of the detected area with the proposed method matches visually with the true upwelling area.

The fuzzifier and the test of convergence parameters in the FCM algorithm are performed

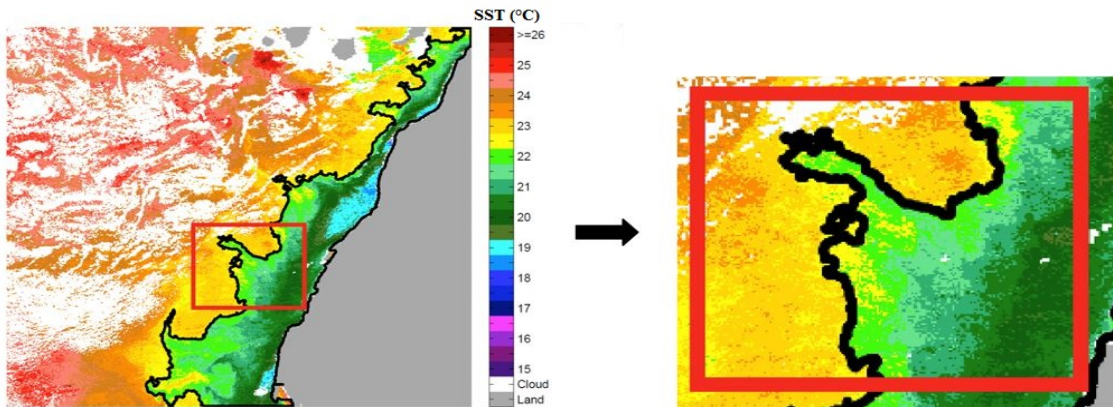


**Figure 5.** Upwelling zone automatically extracted from the original images using  $k$ -means (a)(d), FCM (b)(e) and IAP-FCM-TCT algorithm (c)(f).

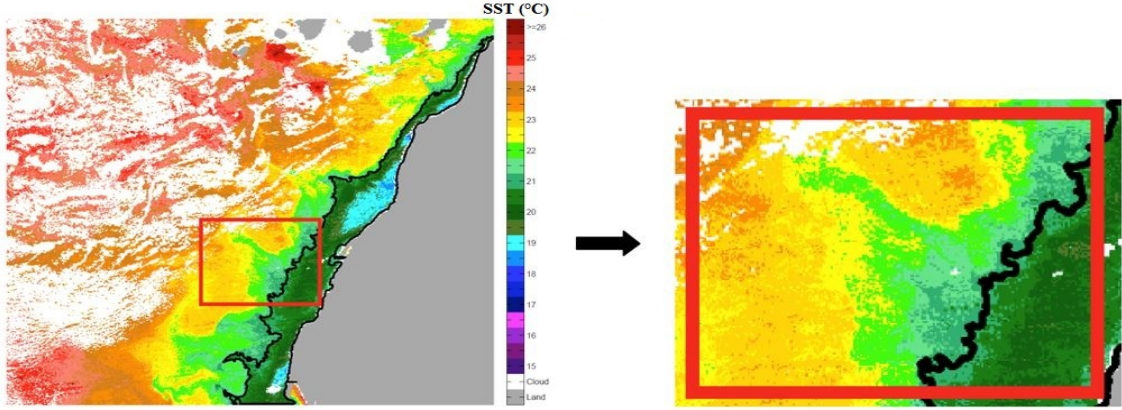
using respectively  $k = 2$  and  $\epsilon = 10^{-5}$ .

### 6.1. Experiment 1

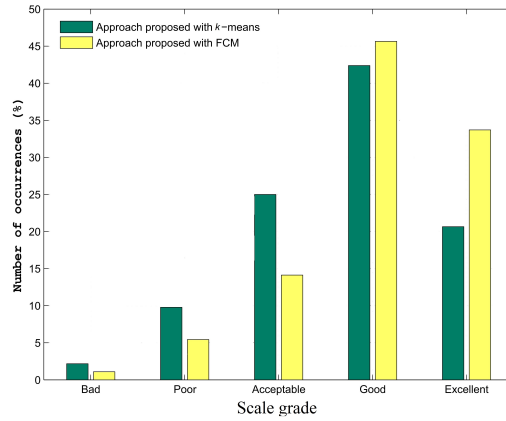
In this experiment, we evaluate the performance of each of the two unsupervised clustering techniques used in this work. The two graphs in Fig. 8 and Fig. 9 show the oceanographer’s evaluation of the 92 SST images using 5 grades: “Bad”, “Poor”, “Acceptable”, “Good” and “Excellent”. “Bad” is attributed when the location of the main upwelling front separating the upwelling and no-upwelling water is not well-identified, and “Excellent” is assigned when it is correctly identified.



**Figure 6.** Detection of the upwelling structure and filaments using our proposed method with FCM algorithm, zoomed in to concentrate on the region of filaments structure.



**Figure 7.** Detection of the upwelling structure and filaments using the Nascimento et al. (2012) algorithm, zoomed in to concentrate on the region of filaments structure.



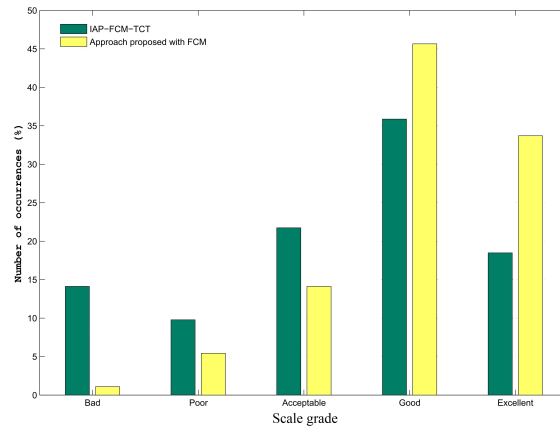
**Figure 8.** Comparison between the two proposed approach with *k*-means and FCM, for the segmentation of the main upwelling front using the oceanographer’s evaluation over the 92 SST images.

The segmentation results in Fig. 8 reveal that 42% and 21% marks where respectively Good and Excellent for *k*-means, which thus make a total of 63% for these two grades. For FCM, a value of 80% is achieved by the addition of the two grades Excellent (34%) and Good (46%). For the grades Bad and Poor, the total is 6% for FCM and 12% for *k*-means. So we can conclude based on the oceanographer evaluation, that the FCM algorithm provides globally a better accuracy than *k*-means, so it can be selected as our reference clustering technique.

## 6.2. Experiment 2

The next experiment focuses on the comparison of the proposed approach with the IAP-FCM-TCT algorithm. The latter is used with the scatter threshold parameter  $\tau = 10^{-3}$  for the fuzzy segmentation process, and respectively the threshold values  $\tau_d = 4.5 \times 10^5$  and  $\tau_c = 0.52 \times 10^{05}$  for the features TDiff and CCard used to determine the main upwelling front (TCT algorithm). These parameters are chosen in accordance with the values used in the article Nascimento et al. (2012).

The graphic result in Fig. 9 shows that our proposed segmentation is more accurate and reliable compared to “IAP-FCM-TCT” algorithm. Indeed, the oceanographer attributed 34%



**Figure 9.** Comparison between the proposed approach and IAP-FCM-TCT algorithm for the segmentation of the main upwelling front using the oceanographer’s evaluation on the 92 SST images.

Excellent to our segmentation approach against 18% to IAP-FCM-TCT algorithm. For the grade Good, the values of 46% and 36% are reached respectively for the proposed method and IAP-FCM-TCT algorithm. For the grade Bad, a value of 1% is achieved by our approach against 14% by the IAP-FCM-TCT algorithm.

These results show that our approach significantly outperforms the IAP-FCM-TCT algorithm, in terms of segmentation accuracy, and it is worth to highlight that our approach does not require the adjustment of some tedious thresholds. Thus, our method has additionally the major advantage of being efficient and easy to implement. More importantly, the upwelling zone detection quality obtained by our proposed method suggests that they can be used to obtain a first approximation of the upwelling area and serve as a basis for a further and subsequent step in the analysis of the SST images (spatial and temporal variability of the upwelling intensity and upwelling extension), and also discuss and compare the results with that of the article of Benazzouz et al. (2014).

The scientific explanation behind the limitation of the method proposed by Nascimento et al. (2012) for detecting the upwelling filaments in our region of interest can be the following:

- The method proposed by Nascimento et al. (2012) is not well suited for our database covering the Moroccan Atlantic coast. In fact, each region has its own characteristics and our region compared with the Portugal coastal ocean is characterized by a strong variation in the upwelling intensity and extension that persist all around the years (Atillah et al. 2005; Nykjaer and Van Camp 1994; Marcello, Marques, and Eugenio 2005).
- The method proposed by Nascimento et al. contains several tuning parameters, which affects the segmentations results, and its hard to know the best value of these parameters for our region.

## 7. Conclusion

This paper deals with the problem of developing of an automatic tool for segmentation of upwelling region in SST images. The proposed method starts with automatic identification of the appropriate number of clusters in each image using several validity indices over  $k$ -means and FCM classifications. Afterwards, the difference of arithmetic mean between consecutive clusters is applied to provide an indicator of the position of the main upwelling front, sep-

arating cold waters near the coast and warmer offshore waters. The proper segmentation of the upwelling and filaments, however, is interfered by the presence of clouds and remaining waters in the image. For this purpose, the region-growing process is applied to filter out noisy structures and thus provides the final segmentation.

We compared our method to a popular algorithm using a database of 92 SST images from years 2006 and 2007, covering the southern Moroccan Atlantic coast. The evaluation was carried out visually by an oceanographer, and show that the method yields a more accurate and reliable upwelling segmentation. Moreover, our method is threshold-free and easy to implement.

## References

- Atillah A, Orbi A, Hilmi K, and Mangin A. 2005. "Produits opérationnels d'océanographie spatiale pour le suivi et l'analyse du phénomène d'upwelling marocain." *Geo Observateur*. 14: 49–62.
- Askari F, and Zerr B. 1998. "A neural network architecture for automatic extraction of oceanographic features in satellite remote sensing imagery." *OCEANS, IEEE*. 2: 1017–1021.
- Benazzouz, A., Mordane, S., Orbi, A., Chagdali, M., Hilmi, K., Atillah, A., Pelegr, J.L. and Herv, D., 2014. "An improved coastal upwelling index from sea surface temperature using satellite-based approach The case of the Canary Current upwelling system". *Continental Shelf Research*, 81, pp.38-54.
- Bensaid A M, Hall L O, Bezdek J C, Clarke Laurence P, Silbiger M L, Arrington J A, and Murtagh R F. 1996. "Validity-guided (re)clustering with applications to image segmentation." *Fuzzy Systems, IEEE Transactions on*. 4(2): 112–123.
- Bezdek James C. 1981. "Pattern Recognition with Fuzzy Objective Function Algorithms." *Plenum Press*. New York, USA.
- Bezdek James C, Ehrlich Robert, and Full William. 1984. "FCM: The fuzzy *c*-means clustering algorithm." *Computers & Geosciences*. 10(2): 191–203.
- Bouguessa, M. and Sheng-Rui Wang. 2004. "A new efficient validity index for fuzzy clustering." *Machine Learning and Cybernetics, 2004. Proceedings of International Conference on*. Shanghai. 3: 1914–1919.
- Calinski T, and Harabasz J. 1974. "A dendrite method for cluster analysis." *Communications in Statistics*. 3(1): 1–27.
- Cayula J F, and Cornillon P. 1992. "Edge detection algorithm for SST images." *Journal of Atmospheric and Oceanic Technology*. 9(1): 67–80.
- Chang Yian-Leng, and Li Xiaobo. 1994. "Adaptive image region-growing." *Image Processing, IEEE Transactions on*. 3(6): 868–872.
- Chassot E, Bonhommeau S, Reygondeau G, Nieto K, Polovina J J, Huret M, Dulvy N K, and Demarcq H. 2011. "Satellite remote sensing for an ecosystem approach to fisheries management." *ICES J. Mar. Sci.* 68(4): 651–666.
- Chaudhari S, Balasubramanian R, and Gangopadhyay A. 2008. "Upwelling detection in AVHRR sea surface temperature (SST) images using neural- network framework." *IEEE International Geoscience Remote Sensing Symposium*. 2: 926–929.
- Chen Min-You, and Linkens D A. 2004. "Rule-base self-generation and simplification for data-driven fuzzy models." *Fuzzy Sets and Systems*. 142(2): 243–265.
- Dave Rajesh N. 1996. "Validating fuzzy partitions obtained through *c*-shells clustering." *Pattern Recognition Letters*. 17(6): 613–623.
- Davies David L, and Bouldin Donald W. 1979. "A Cluster Separation Measure." *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 1(2): 224–227.
- Dunn J C. 1973. "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters." *Journal of Cybernetics. Taylor & Francis*. 3(3): 32–57.
- Frisch Uriel. 1995. "Turbulence : The Legacy of A. N. Kolmogorov." *Cambridge University Press*
- Fukuyama Y, and Sugeno M. 1989. "A new method of choosing the number of clusters for the fuzzy

- c-means method." *Proc. Fifth Fuzzy Systems Symp.* 13(8): 247–250.
- Gath I, and Geva A B. 1989. "Unsupervised optimal fuzzy clustering." *Pattern Analysis and Machine Intelligence, IEEE Transactions on.* 11(7): 773–780.
- Holyer R J, and Peckinpaugh S H. 1989. "Gradient-based edge detection and feature classification of sea-surface images of the southern california bight." *Rem. Sens. Environ.* 27(1): 46–56.
- Jain Anil K, and Dubes Richard C. 1988. "Algorithms for Clustering Data." *Prentice-Hall, Inc., Upper Saddle River, NJ, USA.*
- Kwon S H. 1998. "Cluster validity index for fuzzy clustering." *Electronics Letters. Institution of Engineering and Technology.* 34(22): 2176–2177.
- Jacek Leski. 2003. "Towards a robust fuzzy clustering." *Fuzzy Sets and Systems.* 137(2): 215–233.
- Aristidis Likas, Nikos Vlassis, and Jakob J. Verbeek. 2003. "The global  $k$ -means clustering algorithm." *Pattern Recognition.* 36(2): 451–461.
- Marcello J, Marques F, and Eugenio F. 2005. "Automatic tool for the precise detection of upwelling and filaments in remote sensing imagery." *IEEE Transactions on Geoscience and Remote Sensing.* 43: 1605–1616.
- Mirkin Boris. 2005. "Clustering For Data Mining: A Data Recovery Approach." *Chapman & Hall/CRC Computer Science.*
- Nascimento S, Franco P, Sousa F Dias J, and Neves F. 2012. "Automated computational delimitation of SST upwelling areas using fuzzy clustering." *Computers and Geosciences.* 43: 207–216.
- Nieto K, Demarcq H, and McClatchie S. 2005. "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.
- Nykjaer L and Van Camp L. 1994. "Seasonal and interannual variability of coastal upwelling along northwest Africa and Portugal from 1981 to 1991." *Journal of Geophysical Research: Oceans.* 99 C7: 14197-14207.
- N.R. Pal, J.C. Bezdek, Correction to "On Cluster Validity for the Fuzzy  $c$ -Means Model", *IEEE Trans. Fuzzy Systems* 5 (1997) 152-153.
- Pakhira Malay K, Bandyopadhyay Sanghamitra, and Maulik Ujjwal. 2004. "Validity index for crisp and fuzzy clusters." *Pattern Recognition.* 37(3): 487–501.
- Soille Pierre. 2003. "Morphological Image Analysis: Principles and Applications." *Springer-Verlag New York, Inc.*
- Sousa Fatima M, Nascimento Susana, Casimiro Hugo, and Boutov Dmitri. 2008. "Identification of upwelling areas on sea surface temperature images using fuzzy clustering." *Remote Sensing of Environment.* 112(6), 2817–2823.
- Stowe Larry L, Davis Paul A, and McClain E Paul. 1999. "Scientific Basis and Initial Evaluation of the CLAVR-1 Global Clear/Cloud Classification Algorithm for the Advanced Very High Resolution Radiometer." *J. Atmos. Ocean Technol.* 16(6): 656.
- Tamim A, Minaoui K, Daoudi K and Yahia H, Atillah A, Smiej M F, and Aboutajdine D. 2013. "A simple and efficient approach for coarse segmentation of Moroccan coastal upwelling." *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European.* 3(6): 1–5.
- Tang Yuangang, Sun Fuchun, and Sun Zengqi. 2005. "Improved validation index for fuzzy clustering." *American Control Conference, 2005. Proceedings of the 2005.* Portland, OR, USA. 2(22): 1120–1125.
- Wu Kuo-Lung, and Yang Miin-Shen. 2005. "A cluster validity index for fuzzy clustering." *Pattern Recognition Letters.* 26(9): 1275–1291.
- Xie Xuanli Lisa, and Beni G. 1991. "Validity measure for fuzzy clustering." *Pattern Analysis and Machine Intelligence, IEEE Transactions on.* 13(8), 841-847.
- Xu Yun, and Brereton Richard G. 2005. "A comparative study of cluster validation indices applied to genotyping data." *Chemometrics and Intelligent Laboratory Systems.* 78(1): 30–40.
- Zahid N, Limouri M and and Essaid A. 1999. "A new cluster-validity for fuzzy clustering." *Pattern Recognition.* 32(7): 1089–1097.
- Zhang C, Zhou Y, and Martin T. 2008. "A Validity Index for Fuzzy and Possibilistic  $c$ -means Algorithm." *Proceedings of IPMU.* 10(2): 877–882.