EVALUATION OF A MARINE MESOSCALE EVENTS CLASSIFIER

Marco Reggiannini, Oscar Papini, Gabriele Pieri

Institute of Information Science and Technologies,
National Research Council of Italy,
Via G. Moruzzi 1, 56124 Pisa, Italy

ABSTRACT

Marine mesoscale phenomena are relevant oceanographic processes that impact on fishery, biodiversity and climate variation. In previous literature, their analysis has been tackled by processing instantaneous remote sensing observations and returning a classification of the observed event. Indeed, these phenomena occur within an extended time range, thus an analysis including time dependence is desirable. Mesoscale Events Classifier (MEC) is an algorithm devoted to the classification of marine mesoscale events in sea surface temperature imagery. By processing time series of satellite temperature observations MEC recognizes the considered area of interest as the domain of one out of a given number of possible events and returns the corresponding label. Objective of this work is to discuss the performance of the MEC pipeline in terms of its capability of correctly capturing the nature of the observed mesoscale process. The evaluation process exploited satellite remote sensing data collected in front of the Portuguese coast.

Index Terms— Mesoscale events, Sea surface temperature, Image processing, Environmental monitoring, Statistical classification, Upwelling classification

1. INTRODUCTION

Eddies, water filaments, upwelling and coastline countercurrents are examples of physical processes that affect marine habitats, their biodiversity and the related human activities, e.g. fishery. The analysis of these oceanographic patterns, called marine mesoscale events, is important since their occurrence and alternation also impact on the variations of the climate framework. Mesoscale phenomena can be observed and studied through Sea Surface Temperature (SST) data captured by low Earth orbit satellites. Indeed, the payload imagery captured by missions such as Metop and Aqua features the suitable level of detail in terms of both space and time scales of the observed phenomena.

The authors thank Prof. Flávio Martins and Dr João Janeiro from the University of Algarve, Centre for Marine and Environmental Research, for their support.

Previous attempts to implement the classification of marine mesoscale processes mainly concerned the identification and segmentation of upwelling events, usually based on the detection of the oceanic front pattern. This goal has been pursued following several approaches such as texture features based clustering [1, 2], segmentation based on mathematical morphology operators [3], gradient based boundary identification [4, 5] and more recent approaches such as deep learning based upwelling segmentation [6, 7].

MEC [8, 9] is an automatic procedure conceived to support oceanographers in the classification of a broad spectrum of mesoscale phenomena, not limited to the upwelling category alone. In a nutshell, MEC first subdivides an input SST map into a grid of squares, with a pace adapted to the scale of the sought phenomena, then it extracts for each square a number of statistical parameters describing the temperature trend over time. Finally, based on the identification of specific spatial and temporal patterns of the estimated statistics, it assigns a mesoscale event label to each square.

A major objective of this work is the evaluation of the MEC capability to classify mesoscale events. To the best of the authors' knowledge, the automatic classification of this type of phenomena in previous literature have been based on single instantaneous observations of the SST variable. The novelty proposed in MEC is to take into consideration an entire set of measurements captured within an extended time interval. This approach aims at increasing the classification robustness, thanks to the increased amount of data that enables to analyse broad portions of the event life cycle. The flip side of the coin is that the evaluation of MEC requires a novel format of ground truth that has to be generated ad hoc. This introduces possible issues concerning the definition of the usual descriptors employed for the assessment of the classifier performance (i.e. true/false positives and negatives)—in fact, a dedicated discussion on the topic is provided in the next sections.

The paper outline is the following: Section 2 concerns a concise description of the MEC essential features, Section 3 describes the conceptual framework adopted to evaluate the MEC performance and presents numerical results, Section 4 concludes the paper providing final considerations about the MEC performance and future developments.

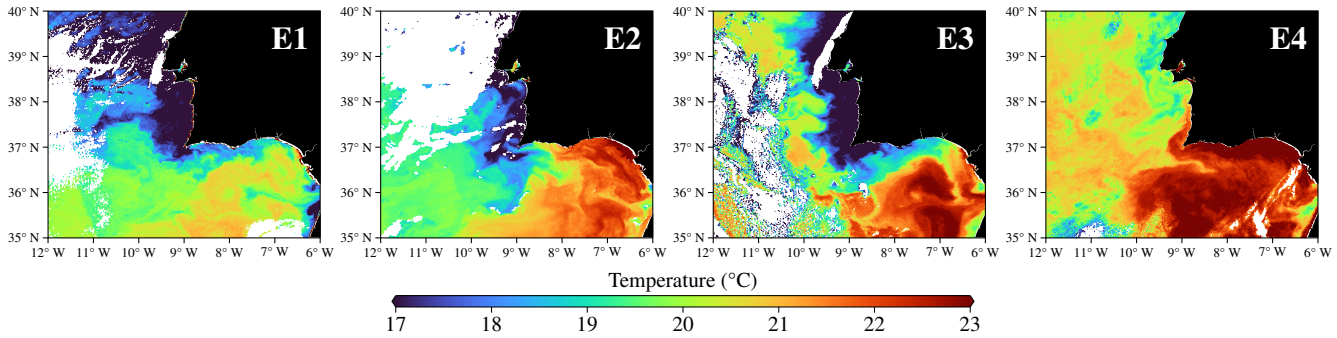

Fig. 1. Examples of patterns in the SST maps data.

2. MEC IN A NUTSHELL

This section provides a brief description of the processing pipeline of MEC applied to the coastal region in the south-west of the Iberian peninsula. The interested reader can find a more detailed explanation in [9].

2.1. Mesoscale Patterns

Due to the geomorphological peculiarities of this region [10], the upwelling phenomenon originating here gives rise to a variety of patterns that can be seen in SST satellite maps. In particular, experts identified four recurring temperature trends associated with four types of mesoscale events, depending on the direction taken by the upwelled water flow and the presence or absence of warm countercurrents. These events will be called E1, E2, E3 and E4 in the remainder of this paper. Figure 1 shows some examples of those patterns.

2.2. Satellite Data

For this work, two sources of SST data were used: the satellites of the *Metop* programme of EUMETSAT [11] and the satellite *Aqua* of NASA [12]. The data-collecting sensor is the Advanced Very High Resolution Radiometer (AVHRR) for Metop and the Moderate Resolution Imaging Spectroradiometer (MODIS) for Aqua. Both products are processed at level L2P in accordance with the GHRSST data processing specification and cover the entire surface of Earth. Only files from the years 2009 to 2017 that contain data within the region

$$A = [35^\circ\text{ N},\ 40^\circ\text{ N}] \times [12^\circ\text{ W},\ 6^\circ\text{ W}]$$

have been downloaded; moreover, since the sensor measurements is often affected by atmospheric noise or other failures, it is possible to have large portions of images without usable data, therefore the files containing less than 15% of the expected data (estimated using the declared spatial resolution of 1 km at nadir) have been pre-emptively discarded.

2.3. Time Series Computation

The area A has been divided in a 20×24 regular grid of squares $a_{i,j}$, $i = 0, \ldots, 19$, $j = 0, \ldots, 23$, with side length 0.25°. For the classification of an image with timestamp t_0, MEC collects all the images of the dataset with timestamps $t \in [t_0 - 15 \text{ days},\ t_0]$ and computes, for each t and for each square $a_{i,j}$, the spatial average of the SST $\overline{T}_{i,j}(t)$ at time t in the square. This way, a set of SST time series is obtained, one for each square $a_{i,j}$:

$$p_{i,j} = \{(t_k, \overline{T}_{i,j}(t_k)) \mid k = 1, \ldots, n_{i,j}\}.$$

Notice that the number $n_{i,j}$ of values in a series depends not only on the number of images with timestamps in $[t_0 - 15 \text{ days},\ t_0]$, but also on the quality of the SST data in those images in the square $a_{i,j}$.

The resolution of the grid, the amplitude of the time interval and in general the values of all the parameters involved in the MEC algorithm have been chosen after a careful analysis of the geophysical characteristics of the sought mesoscale events (see also [8] for further details).

2.4. Features Extraction

In the next step of the MEC classification pipeline, the information of the SST trend in a square $a_{i,j}$, contained in the corresponding series $p_{i,j}$, is condensed in a number of statistical features. In particular, for each series $p_{i,j}$ the temporal mean $\mu_{i,j}$, the standard deviation $\sigma_{i,j}$ and the linear regression coefficient $\theta_{i,j}$ are computed, where

$$\mu_{i,j} = \frac{1}{n_{i,j}} \sum_{k=1}^{n_{i,j}} \overline{T}_{i,j}(t_k),$$

$$\sigma_{i,j} = \sqrt{\frac{1}{n_{i,j}} \sum_{k=1}^{n_{i,j}} (\overline{T}_{i,j}(t_k) - \mu_{i,j})^2},$$

and $\theta_{i,j}$ is the slope of the straight line that better interpolates the points in $p_{i,j}$.

2.5. Image Classification

For each square $a_{i,j}$, an array $e_{i,j} = (e^1_{i,j}, e^2_{i,j}, e^3_{i,j}, e^4_{i,j})$ is computed. Each $e^k_{i,j} \in [0,1]$ is an index representing how much an event of type Ek is believed to have occurred inside $a_{i,j}$ at time t_0. It is computed by applying a series of conditional rules to the features described in Section 2.4; in particular, the scores for $a_{i,j}$ depend not only on the values $\mu_{i,j}$, $\sigma_{i,j}$ and $\theta_{i,j}$ but also on the values of μ, σ and θ of the squares in a neighbourhood of $a_{i,j}$.

The final step of the MEC classification process is the production of a heatmap highlighting the squares that are assigned to each type of event by the following procedure: given the score array $e_{i,j}$, the maximum score $e^m_{i,j} = \max\{e^1_{i,j}, e^2_{i,j}, e^3_{i,j}, e^4_{i,j}\}$ is considered and, if $e^m_{i,j} \geq 0.6$ and $a_{i,j}$ belongs to a zone where it is possible to have an Em-type event, then $a_{i,j}$ is marked with the "Em" label; otherwise, no label is assigned to $a_{i,j}$. Figure 2 represents an example of an SST map with the corresponding classification map as returned by MEC.

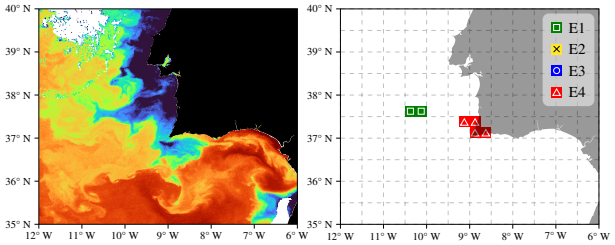

Fig. 2. Example of classification result. Left: SST map of the area A on 5 October 2017 at around 21:40 UTC; the temperature range is the same as Figure 1. Right: corresponding heatmap.

3. MEC PERFORMANCE EVALUATION

In order to evaluate the MEC performance, 1175 Metop images with timestamp t within the years 2016 and 2017 have been considered. For each image $I(t)$, the same grid as the one in Section 2.3 has been defined and, after a visual evaluation of the SST map, each square has been tagged with a label "E1", "E2", "E3", "E4", or "no event" depending on which type of event is recognizable in the geographical location of the square, thus obtaining a ground truth map $I_{\mathrm{gt}}(t)$ of the same kind as the heatmap returned as classifier output (hereafter denoted by $\hat{I}(t)$).

A first naive evaluation could be performed by comparing, for each time t and for each event type Ei, the MEC output with the ground truth in a square-by-square manner: each square in the Ek event detection zone is considered a true positive if and only if it is marked with the k label both in $\hat{I}(t)$ and in $I_{\mathrm{gt}}(t)$. With this convention, if TP_k, FP_k, FN_k and TN_k are the number of true positives, false positives, false negatives and true negatives respectively, we can define some typical evaluation parameters (precision, recall, F-score, accuracy) for each event class k in the usual way

$$p_k = \frac{\mathrm{TP}_k}{\mathrm{TP}_k + \mathrm{FP}_k}, \quad r_k = \frac{\mathrm{TP}_k}{\mathrm{TP}_k + \mathrm{FN}_k},$$
$$F_k = \frac{2p_k r_k}{p_k + r_k}, \quad a_k = \frac{\mathrm{TP}_k + \mathrm{TN}_k}{\mathrm{TP}_k + \mathrm{FN}_k + \mathrm{TN}_k + \mathrm{FP}_k},$$

as well as their aggregated versions

$$p = \frac{\sum_k \mathrm{TP}_k}{\sum_k (\mathrm{TP}_k + \mathrm{FP}_k)}, \quad r = \frac{\sum_k \mathrm{TP}_k}{\sum_k (\mathrm{TP}_k + \mathrm{FN}_k)},$$
$$F = \frac{2pr}{p+r}, \quad a = \frac{\sum_k (\mathrm{TP}_k + \mathrm{TN}_k)}{\sum_k (\mathrm{TP}_k + \mathrm{FN}_k + \mathrm{TN}_k + \mathrm{FP}_k)},$$

obtaining the results shown in Table 1.

Table 1. Results of the direct comparison between the classifier output and the ground truth.

	Precision	Recall	F-score	Accuracy
E1	0.059	0.113	0.078	0.947
E2	0.033	0.068	0.045	0.937
E3	0.151	0.163	0.157	0.867
E4	0.083	0.211	0.119	0.872
Aggregate	0.098	0.150	0.118	0.914

This method compares the classifier output, which—as described before—is based on the analysis of the available images in $[t_0 - 15 \text{ days}, t_0]$, with the ground truth image relative to time t_0. However, this evaluation does not consider all the ground truth information existing in the temporal window around t_0. For example, suppose that in some images in $[t_0 - 15 \text{ days}, t_0]$ an event of type Ek is clearly visible in a square $a_{i,j}$, but the image $I(t_0)$ has no SST data, thus $a_{i,j}$ has no label in $I_{\mathrm{gt}}(t_0)$. Assuming that MEC correctly identifies the dynamic pattern of the event, it will assign the label Ek to $a_{i,j}$, so that $a_{i,j}$ would be a false positive with respect to the naive evaluation.

Analogously, the naive approach does not consider the ground truth information existing in the spatial neighbourhood of a given square $a_{i,j}$. Within the 15 days temporal window, the dynamic evolution of a mesoscale event causes fluctuations in its shape and position. This variability, intrinsic to the oceanographic phenomenon, impacts on the final result of the classification. In particular it may occur that a square $a_{i,j}$ is classified with a label that conflicts with the corresponding ground truth, while in the neighbourhood of $a_{i,j}$ ground truth and classifier are in agreement. In this case, the evaluation through the naive approach discussed before would entail an increase in the amount of misclassifications.

The above presented arguments suggest that the classifier output and the ground truth are not directly comparable. To fix this, a different approach to the evaluation is presented: let

$\ell_k(t)$ be the number of squares in $\hat{I}(t)$ that have been classified with the label corresponding to the k-th class. Analogously $\ell_k^{\mathrm{gt}}(t)$ is the number of squares in $I_{\mathrm{gt}}(t)$ that have been manually tagged as belonging to class k. Let us define

$$O_k(t) = \begin{cases} 1 & \text{if } \ell_k(t) > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $O_k(t) = 1$ means that the classifier output is interpreted as an Ek event in $I(t)$.

Similarly to (1), the ground truth boolean sequence is defined as

$$O_k^{\mathrm{gt}}(t) = \begin{cases} 1 & \text{if } \exists\, s \in [t - 15 \text{ days}, t] \text{ s.t. } \ell_k^{\mathrm{gt}}(s) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

These newly introduced variables allow performing a comparison between the results of the classification and the reference knowledge so to resolve the spatial and temporal disparities that arise when the evaluation is performed through the naive approach.

The performance of the classifier can then be assessed based on the traditional true/false positives and true/false negatives counting, where in this case for an event of type Ek an image $I(t)$ is considered positive (respectively negative) if $O_k(t) = 1$ (respectively 0), and the classification is considered true (respectively false) if $O_k(t) = O_k^{\mathrm{gt}}(t)$ (respectively $O_k(t) \neq O_k^{\mathrm{gt}}(t)$). The classifier performance according to this new approach is presented in Table 2.

Table 2. MEC performance.

	Precision	Recall	F-score	Accuracy
E1	0.849	0.534	0.655	0.584
E2	0.474	0.291	0.360	0.505
E3	0.856	0.474	0.610	0.511
E4	0.730	0.606	0.663	0.718
Aggregate	0.753	0.481	0.587	0.579

When measured by the new evaluation approach, the classifier performance improves considerably at the expense of a reduction in the accuracy parameter. Indeed, the large accuracy values reported in Table 1 are caused by the unlabelled squares, which represent the vast majority both within the ground truth and the classified images, leading to an excessive amount of true negatives found by the naive approach.

4. CONCLUSIONS

Mesoscale phenomena play an essential role in defining and shaping coastal ecosystems, and their systematic analysis has proven to be of utter importance in the fields of maritime monitoring and climate change studies.

This work concerns MEC, a classifier of mesoscale events detected as patterns in SST data. Unlike previous similar tools, the core of MEC is designed to extract and elaborate dynamic information from these data. Therefore, starting from usually employed performance descriptors (precision, recall, F-score, accuracy), we defined a novel evaluation procedure that takes into account the peculiarities of this system to assess the effectiveness of the classifier. The results achieved by MEC according to this method are presented and briefly discussed.

This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101000825 (NAUTILOS; https://www.nautilos-h2020.eu).

5. REFERENCES

[1] Kiran K. Simhadri, S. Sitharama Iyengar, Ronald J. Holyer, Matthew Lybanon, and John M. Zachary Jr., "Wavelet-based feature extraction from oceanographic images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, pp. 767–778, 1998.

[2] Anass El Aouni, Khalid Daoudi, Khalid Minaoui, and Hussein Yahia, "Robust Detection of the North-West African Upwelling From SST Images," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 4, pp. 573–576, 2021.

[3] Sankar Krishnamurthy, S. Sitharama Iyengar, Ronald J. Holyer, and Matthew Lybanon, "Histogram-based morphological edge detector," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 4, pp. 759–767, 1994.

[4] Andrew G. P. Shaw and Ross Vennell, "A Front-Following Algorithm for AVHRR SST Imagery," *Remote Sensing of Environment*, vol. 72, no. 3, pp. 317–327, 2000.

[5] Xueming Zhu, Ziqing Zu, Shihe Ren, Miaoyin Zhang, Yunfei Zhang, Hui Wang, and Ang Li, "Improvements in the regional South China Sea Operational Oceanography Forecasting System (SCSOFSv2)," *Geoscientific Model Development*, vol. 15, pp. 995–1015, 02 2022.

[6] Oliverio J. Santana, Daniel Hernández-Sosa, and Ryan N. Smith, "Oceanic mesoscale eddy detection and convolutional neural network complexity," *International Journal of Applied Earth Observation and Geoinformation*, vol. 113, pp. 102973, 2022.

[7] Renlong Hang, Gang Li, Mei Xue, Changming Dong, and Jianfen Wei, "Identifying Oceanic Eddy With an Edge-Enhanced Multiscale Convolutional Network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 9198–9207, 2022.

[8] Marco Reggiannini, João Janeiro, Flávio Martins, Oscar Papini, and Gabriele Pieri, "Mesoscale Events Classification in Sea Surface Temperature Imagery," in *Machine Learning, Optimization, and Data Science. Proceedings of the 8th International Workshop LOD 2022, Revised Selected Papers, Part I*. 2023, vol. 13810 of *Lecture Notes in Computer Science*, pp. 516–527, Springer.

[9] Gabriele Pieri, João Janeiro, Flávio Martins, Oscar Papini, and Marco Reggiannini, "MEC: A Mesoscale Events Classifier for oceanographic imagery," *Applied Sciences*, vol. 13, no. 3, 2023, article no. 1565.

[10] Paulo Relvas, Eric D. Barton, Jesús Dubert, Paulo B. Oliveira, Álvaro J. Peliz, José C. B. da Silva, and A. Miguel P. Santos, "Physical oceanography of the western Iberia ecosystem: Latest views and challenges," *Progress in Oceanography*, vol. 74, no. 2, pp. 149–173, 2007.

[11] OSI SAF, "Full resolution L2P AVHRR Sea Surface Temperature MetaGRanules (GHRSST) - Metop," 2011.

[12] NASA/JPL, "GHRSST Level 2P Global Sea Surface Skin Temperature from the Moderate Resolution Imaging Spectroradiometer (MODIS) on the NASA Aqua satellite (GDS2)," 2020.