Mesoscale events classification in Sea Surface Temperature imagery

Marco Reggiannini[1 [0000−0002−4872−9541]], João Janeiro[3 [0000−0002−6241−8520]], Flávio Martins[2 [0000−0002−9863−6255]], Oscar Papini[1 [0000−0003−2069−5068]], and Gabriele Pieri[1 [0000−0001−5068−2861]]

[1] Institute of Information Science and Technologies, National Research Council of Italy, Pisa, Italy
{name.surname}@isti.cnr.it

[2] University of Algarve, Centre for Marine and Environmental Research – CIMA, Faro, Portugal
jmjaneiro@ualg.pt

[3] University of Algarve, Centre for Marine and Environmental Research – CIMA, Institute of Engineering ISE, Faro, Portugal
fmartins@ualg.pt

Abstract Sea observation through remote sensing technologies plays an essential role in understanding the health status of marine fauna species and their future behaviour. Accurate knowledge of the marine habitat and the factors affecting faunal variations allows to perform predictions and adopt proper decisions. This is even more relevant nowadays, with policymakers needing increased environmental awareness, aiming to implement sustainable policies. There is a connection between the biogeochemical and physical processes taking place within a biological system and the variations observed in its faunal populations. Mesoscale phenomena, such as upwelling, countercurrents and filaments, are essential processes to analyse because their arousal entails, among other things, variations in the density of nutrient substances, in turn affecting the biological parameters of the habitat. This paper concerns the proposal of a classification system devoted to recognising marine mesoscale events. These phenomena are studied and monitored by analysing Sea Surface Temperature images captured by satellite missions, such as Metop and MODIS Terra/Aqua. Classification of such images is pursued through dedicated algorithms that extract temporal and spatial features from the data and apply a set of rules to the extracted features, in order to discriminate between different observed scenarios. The results presented in this work have been obtained by applying the proposed approach to images captured over the south-western region of the Iberian Peninsula.

Keywords: Image Processing · Remote Sensing · Mesoscale Patterns · Sea Surface Temperature · Machine Learning · Climate change

1 Introduction

Evaluating the impact of climate change on coastal marine ecosystems may be a challenging task: near the coast, global drivers are modified by topography and by local atmospheric and oceanographic circulation patterns.

In particular, Ekman dynamics and large-scale thermocline processes control the coastal upwelling occurring at the Eastern Boundary Upwelling Ecosystems (EBUEs) [1, 2]; winds directed towards the Equator drive upwelling, which transports deeper, colder and nutrient-rich waters to the surface. As a result, these areas host the most productive ecosystems in the global ocean [3], playing a major role in the marine primary production and the worldwide fisheries. Moreover, it was recently shown that upwelled water's low long-term warming rates may provide thermal refugia, stabilize changes in species distributions and enhance local biodiversity [4].

According to related literature, more than 71% of coastal zones are experiencing a net heat gain due to global warming [5]. Yet, both positive and negative trends were observed in different upwelling ecosystems [6]. Therefore, it is surmised that every upwelling ecosystem reacts differently to the changing climate.

Among the world's EBUEs, the Iberia/Canary Current System (ICCS) is one of the least studied [7]. Despite a general circulation similar to other EBUEs, in ICCS the discontinuity imposed by the Mediterranean Sea, combined with the seasonality of the large-scale atmospheric circulation, has a profound impact on the regional oceanography. The region's continental shelf is characterized by a large number of topographical features, such as prominent capes, promontories and submarine canyons, whose spatial scales are tens to hundreds of kilometres [8]. All the above highlight the importance of sub-seasonal temporal scales and sub-basin spatial scales, which explain the observed oceanographic patterns.

The identification and cataloguing of upwelling regimes occurring in an EBUE are important achievements towards the characterization of the system. Traditionally this task has been performed subjectively by experts, analysing Sea Surface Temperature (SST) maps of the area of interest. This procedure is manageable if few tens or even hundreds of images are used, but it turns into an unfeasible task as the number of scenes approaches thousands of images, that is the typical order of magnitude when the purpose is to investigate climate-related changes.

Nowadays, with the growing amount of remote sensing observations, automated techniques have been gaining momentum, using tools such as two-dimensional wavelet transforms [9], neural networks [10], or edge detection algorithms [11]. A complete automation encompasses three main challenges: (1) the presence of noise, mainly due to clouds and other atmospheric phenomena; (2) the fact that gradients are weak and provide an excess of information, generating several unconnected borderlines and complicating the structure identification task due to the fine edge detail; (3) the substantial morphological variation that prevents from an accurate geometric representation and the absence of a valid analytical model for the structures [12].

The main objective of this work is to design and develop automatic methods capable of accepting massive datasets of oceanographic SST imagery as input and returning a classification of the images according to the different regimes of observable upwelling patterns. The identification of a specific temperature pattern is based on the extraction of quantitative features from the SST maps. Indeed the emergence of a certain pattern is usually highly correlated with peculiarities in the temperature spatial arrangement at time fixed (e.g. the presence of abrupt variations in the temperature values within a certain neighbourhood), as well as with the observation of specific temperature trends at fixed locations, providing insights about the flowing of water masses between points at different temperature values. In a previous work dedicated to this topic [13], a custom visualisation tool was developed to extract and visualise the time series of the SST signals related to a given number of fixed locations within an area of interest. Based on this result, a novel step is introduced in the pipeline, with the objective of processing the mentioned signal series to extract quantitative descriptors of the signals trend. The computed quantities are finally used to fulfil the classification task by implementing a set of rules that assign each set of time series to a specific class according to the numerical values of the computed features.

The proposed method will be applied to the South Iberian region, contributing to understanding the formation of upwelling filaments and the effects of climate change in this particular EBUE. In its current form, the metrics used (e.g. the signal variation rate and its deviation from the mean value) are able to identify different types of mesoscale features.

The paper is arranged as follows: Section 2 concerns a detailed description of the employed dataset and the related ground truth classification; Section 3 thoroughly reports on the developed processing pipeline and describes a relevant use case; Section 4 concludes the paper by discussing the outcomes of this work and providing a few considerations about future perspectives.

2 Materials

2.1 SST Satellite Data

The identification and classification of upwelling events in a marine ecosystem have been performed by processing SST maps of an area of interest. These maps are compiled using data coming from satellite missions. In particular satellite data from the years 2009 to 2017 has been retrieved from two sources: EUMETSAT's *Metop* programme [14] and NASA's *Aqua* satellite [15].

Satellites of the Metop programme gather data through the Advanced Very High Resolution Radiometer (AVHRR); the information is processed at level L2P and binned in a single netCDF-4 file every 3 minutes, for a total of 480 images per day covering the entire globe, with a spatial resolution of 1 km at nadir and an accuracy of 0.01 °C in the temperature measurement. We used data from the satellite Metop-A for the period 2009–2016 and from Metop-B for 2017.

The satellite Aqua uses the Moderate Resolution Imaging Spectroradiometer (MODIS) to gather data; the information is processed at level L2P and binned in

a single netCDF-4 file every 5 minutes, for a total of 288 images per day covering the entire globe, with a spatial resolution of 1 km at nadir and a temperature accuracy of 0.005 °C.

For both sources, only data covering the region of interest were downloaded. In particular, points with latitude between 35° and 40° N and longitude between 12° and 6° W were considered, resulting in 2–3 images per day at most. Moreover, since the data capture task is often prone to failures in registering SST values for areas that appear opaque (e.g. due to atmospheric events), a further selection has been performed by discarding images containing less than the 15% of the expected amount of data.

2.2 Types of Patterns

By looking at the SST images, four upwelling patterns have been identified:

1. a cold water filament going westwards, originating from the southward upwelling jet that runs along the western coast of Portugal;
2. a cold water filament going southwards, extending over Cape St. Vincent the upwelling jet mentioned above;
3. a clear stream of cool water running along the southern Iberian coast;
4. a warm countercurrent originating in the Gulf of Cádiz and running along the southern Iberian coast, eventually reaching Cape St. Vincent and turning northwards.

A more detailed description can be found in [13]. These four patterns will be called E1, E2, E3 and E4 respectively. Pattern E3 may be further divided into E3i, when the thermal gradient between the cool stream and the water in the Gulf of Cádiz is small, mostly occurring during winter; and E3u, with a more significant gradient. For our analysis we do not distinguish between these two subpatterns. Figure 1 shows some examples of the described patterns.

Based on this classification, a dataset with labels E1–E4 assigned to the corresponding SST maps has been provided by expert oceanographers. This represents the ground truth for the subsequent SST analysis.

3 SST Analysis

The main goal of the method proposed for the analysis of SST time series is to classify upwelling events exploiting the *dynamic* information contained in the temperature patterns, observed over a given time window. To this aim, the multiple SST signal sequences related to a given geographical area are extracted from the corresponding netCDF files and arranged in a single 2D plot, namely a *spaghetti plot* (see Figure 2d). The resulting visualization allows for a direct and clearer interpretation of the ensemble of the SST trends in the considered area of interest. The software dedicated to this analysis has been developed within a Python framework.

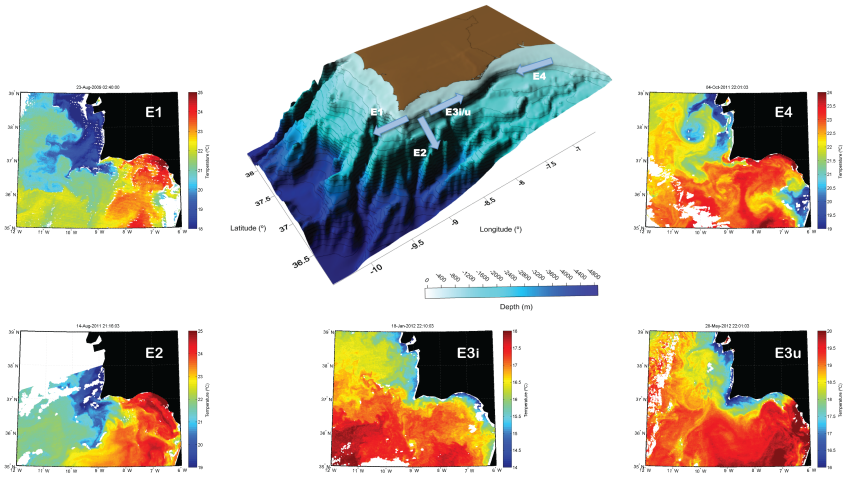

Figure 1. Mesoscale patterns in the south-western Iberian Peninsula (from [13]).

3.1 Spaghetti plot generation

The steps fulfilled to generate a spaghetti plot for a given geographical area A are described below:

1. the rectangular area A selected on the SST map is split into a grid of N_A disjoint squares a_j, each with fixed size (typically between 0.01 and 0.25 degrees in latitude/longitude):

$$A = \bigcup_{j=1}^{N_A} a_j;$$

2. at a given time t (recorded in the netCDF file) the mean value of the SST signal is estimated by averaging the available n_j SST values, located within the corresponding a_j:

$$\mu_j(t) = \frac{1}{n_j} \sum_{i=1}^{n_j} \mathrm{SST}_i(t);$$

3. the previous step is repeated for each a_j and for each t within the considered time window, eventually returning N_A time series of the averaged SST signal;
4. the spaghetti plot is finally generated by simultaneously plotting all the N_A signals within the same coordinate system.

Each square a_j, and the corresponding averaged signal $\mu_j(t)$, is colour-coded so that the differences in the signal trends observed in different squares can be easily recognised by visual inspection. The reader can refer to [13, 16] for further information about the generation of the spaghetti plots.

Figure 2 shows the case of an event classified as E3 in the ground truth, together with the corresponding spaghetti plot. This will be exploited in the following as a case study to describe step by step the implemented processing stages and the corresponding output. The spaghetti plot has been computed dividing the region of interest in squares of size 0.25° and considering a 15 days window for the time range, with the last observation coinciding with the classified event. The specified resolution values represent input parameters to the spaghetti plot generation task. They have been set up through empirical considerations after testing alternative values, and assessing that the mentioned choice achieves a better agreement with the ground truth. Indeed, it can be noticed that the temperature curves diverge and decrease, starting at times t beyond 8th September, in agreement with the temperature trends of an E3 pattern.

3.2 Features extraction

Based upon the ground truth dataset, the corresponding set of spaghetti plots has been generated. The objective of the task described in this section is to investigate the discriminating properties of a number of features extracted from the signals. Generally speaking these features can be computed based on one or more SST sequences belonging to the same spaghetti plot. The goal is to define a set of rules to be applied to the extracted features, with the purpose of identifying the class of the observed mesoscale pattern.

As mentioned before, a single curve p_j in a spaghetti plot represents the spatially averaged trend of the signal in the j-th square of the grid. It is represented by the following array of n pairs, where n is the number of time samples:

$$p_j = \{(t_m, \mu_j(t_m)) \mid m = 1, \ldots, n\}.$$

Notice that n is not a constant value: it depends on the quantity and quality of the data actually exploitable in the netCDF file. Therefore, n can be considered as a reliability index, reflecting the confidence level associated with the estimators described hereafter.

The following statistics are computed:

1. the temporal *mean* of p_j,

$$\mu(p_j) = \frac{1}{n} \sum_{m=1}^{n} \mu_j(t_m);$$

2. the *standard deviation* of p_j,

$$\sigma(p_j) = \sqrt{\frac{1}{n} \sum_{m=1}^{n} (\mu_j(t_m) - \mu(p_j))^2};$$

3. the *linear regression coefficient* $\theta(p_j)$, defined as the slope of the straight line that better fits the curve p_j.

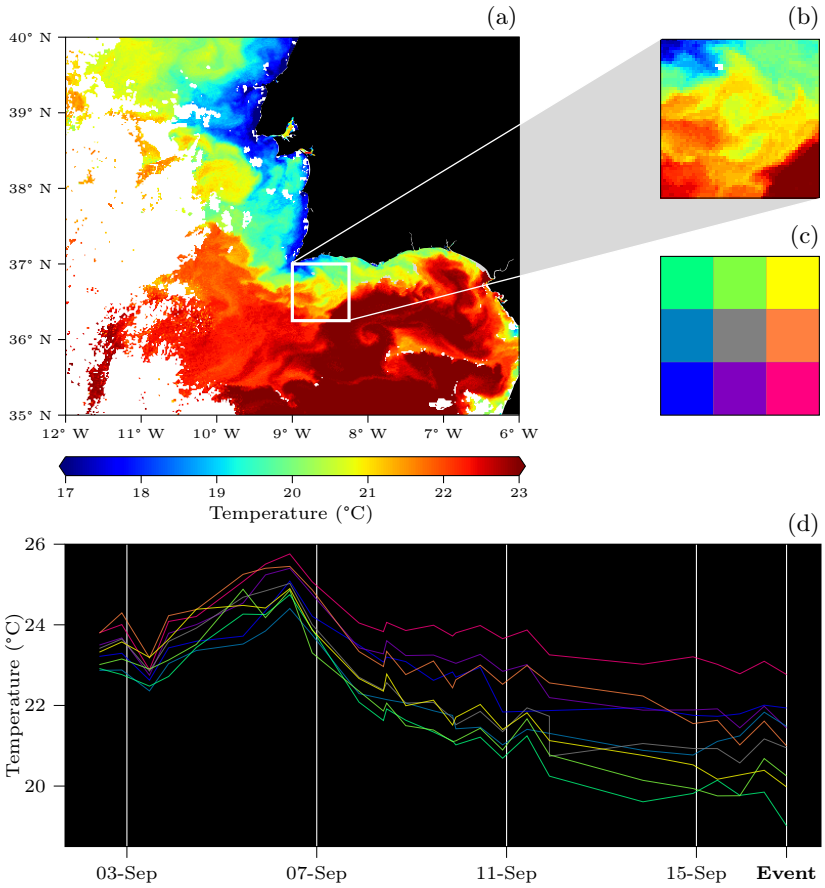

Figure 2. Event of 16 September 2016 at around 21:35 UTC. (a) SST map at the date of the event; (b) detail of the SST in the reference area for the spaghetti plot (latitude between 36.25° and 37° N, longitude between 9° and 8.25° W, resolution 0.25°); (c) reference grid; (d) generated spaghetti plot.

The three statistics are meant to describe in first approximation the local behaviour of the SST in the square a_j. Additional parameters, such as the coefficients of the quadratic regression for the curve p_j, have been taken into consideration, but finally discarded since they don't seem to capture the SST trend, as their integration in the pipeline didn't improve the classification performance. Notice that the choice of the linear regression coefficient $\theta(p_j)$ *does not imply* that we assume a linear correlation between SST and time during an upwelling event: that value is interpreted as a descriptor of the SST trend in the square a_j.

Figure 3 shows the statistics computed for the case study under investigation.

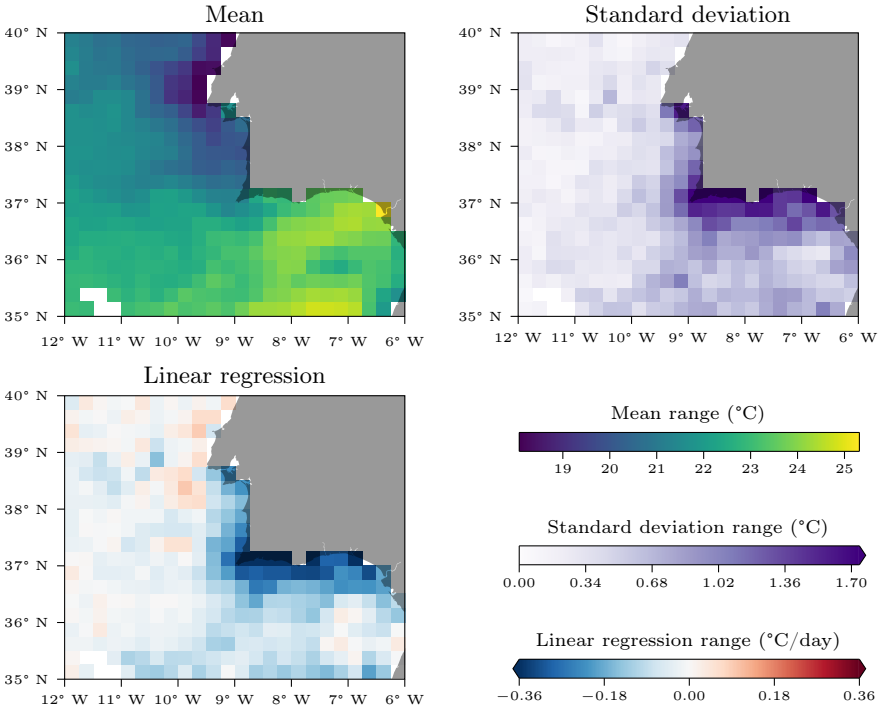

Figure 3. Computed statistics for the period between 2 and 16 September 2016.

3.3 Classification rules

Considering a square a_j in the area of interest, a score array (e_1, e_2, e_3, e_4), with values e_i normalized in $[0, 1]$, is defined. The value e_i represents a belief index for the corresponding event Ei to have occurred inside a_j at the end of the considered time range. Each e_i is obtained by applying a set of conditional rules to the statistics described in Section 3.2, computed inside both a_j and the neighbouring squares. The rules are modelled on the *a priori* knowledge of the oceanographic patterns, so that the score e_i is increased (by a fixed amount) only if the behaviour of the features μ, σ and θ, inside and in the neighbourhood of the square a, matches the one observed in the case of an Ei pattern. A qualitative description of these rules is reported below.

1. Increase e_1 if:
   (a) the SST trend $\theta(p_j)$ inside a_j is negative;
   (b) inside the eastern neighbouring squares, the SST trend is lower than in a_j;
   (c) the SST average value $\mu(p_j)$ inside a_j is lower than SST in both the northern and southern neighbouring squares.
2. Increase e_2 if:
   (a) the SST trend inside a_j is negative;

   (b) inside the northern neighbouring squares, the SST trend is lower than in a_j;
   (c) the SST average value inside a_j is lower than SST in both the eastern and western neighbouring squares.
3. Increase e_3 if:
   (a) the SST trend in the neighbourhood of a_j is negative;
   (b) inside a_j the SST trend and its average value are larger than in the north-western neighbours, and smaller than in the south-eastern neighbours.
4. Increase e_4 if:
   (a) the SST trend in the neighbourhood of a_j is positive;
   (b) inside a_j the SST trend and its average value are smaller than in the south-eastern neighbours, and larger than in the north-western neighbours.

Additional considerations affect the final scores:

- if the SST variation $\sigma(p_j)$ is large (namely $\sigma(p_j) \geq 1\,^{\circ}\mathrm{C}$), increase either e_1, e_2 and e_3 (if the SST decreases) or e_4 (if the SST increases);
- if a_j is globally either "cold" (in case of events E1, E2 and E3) or "warm" (event E4) with respect to all the other squares in the area of interest, boost the corresponding scores;
- if a_j is too near the coast (i.e. less than 3 squares, that is *circa* 75 km away from the coast), penalize (halve) the scores e_1 and e_2; if a_j is too far from the coast (i.e. more than 3 squares), penalize e_3 and e_4.

The application of the described classification rules to the case study is displayed in Figure 4, where the four scores have been computed for each square in the grid.

In order to classify a_j, the maximum score $e_m = \max\{e_1, e_2, e_3, e_4\}$ is considered and, in case it is larger than a certain threshold (empirically decided), the square is marked with the corresponding "Em" colour label. If none of the scores exceeds the threshold, no label is assigned. Referring to our case study, the final outcome of the classifier is shown in Figure 5, representing a heatmap where each square is labeled according to this rule, together with the percentage of the available SST data.

4 Discussion and Conclusion

In this work, a methodology for the analysis of SST time series has been proposed, with the objective of automating the classification of upwelling events by exploiting the dynamic information observed in the SST patterns. This ongoing study involves the analysis of large imagery datasets, using expert knowledge to contour the positions of the mesoscale feature, and eventually aiming at its characterisation.

The current results are promising and show patterns of differentiation among different mesoscale events occurring in the analysed area. Throughout Section 3,

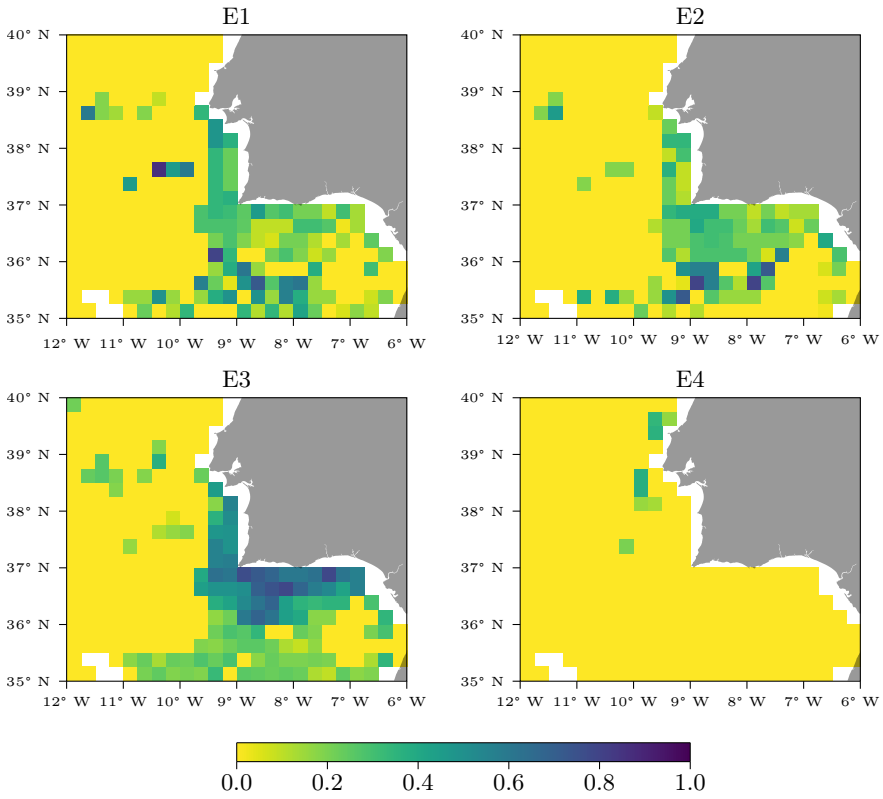

Figure 4. Scores given to each square of the grid for each type of event.

we presented a case study of an upwelling event that occurred on 16 September 2016 and was classified as type E3 by expert oceanographers. Looking at the map of Figure 5, we observe that the results of our analysis are aligned with the ground truth, since the majority of the squares have been classified as "E3" (the blue ones), and they are located within the geographical zone where events of type E3 are usually detected. Notably, the squares classified as E3 have a high reliability value n, i.e. they feature a percentage of valid data higher than the other labelled squares. Concerning the latter squares, the output of the classifier in those cases apparently conflicts with the ground truth, but this can be explained by the fact that around 16 September 2016, different types of SST events are observed (e.g. an E1-type event occurs on 17 September). This indeed confirms the correctness of the classification procedure that, employing a "time series based" approach to the analysis, accordingly identifies multiple typologies of events that develop within the considered time window in the neighbourhood of the main event.

The test and validation of the proposed algorithm are carried out and will continue as part of the activities of the EU H2020 project NAUTILOS [17].

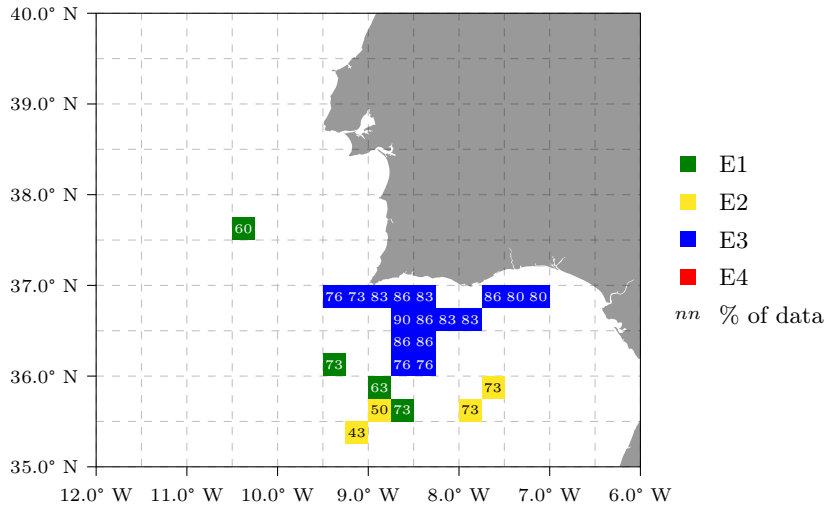

Figure 5. Labels given to each square of the grid, depending on their scores.

Acknowledgements This paper 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[4]).

References

1. Messié, M., Ledesma, J., Kolber, D.D., Michisaki, R.P., Foley, D.G., and Chavez, F.P.: Potential new production estimates in four eastern boundary upwelling ecosystems. Progress in Oceanography 83(1), 151–158 (2009). DOI: 10.1016/j.pocean.2009.07.018
2. Ramajo, L., *et al.*: Upwelling intensity modulates the fitness and physiological performance of coastal species: Implications for the aquaculture of the scallop *Argopecten purpuratus* in the Humboldt Current System. Science of The Total Environment 745, 140949 (2020). DOI: 10.1016/j.scitotenv.2020.140949
3. FAO: *The State of World Fisheries and Aquaculture 2018. Meeting Sustainable Development Goals*. 2018.
4. Varela, R., Lima, F.P., Seabra, R., Meneghesso, C., and Gómez-Gesteira, M.: Coastal warming and wind-driven upwelling: A global analysis. Science of The Total Environment 639, 1501–1511 (2018). DOI: 10.1016/j.scitotenv.2018.05.273
5. IPCC: Summary for Policymakers. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate. Ed. by H.-O. Pörtner et al. Cambridge University Press (2019). DOI: 10.1017/9781009157964.001
6. Varela, R., Álvarez, I., Santos, F., de Castro, M., and Gómez-Gesteira, M.: Has upwelling strengthened along worldwide coasts over 1982-2010? Scientific Reports 5, 10016 (2015). DOI: 10.1038/srep10016

[4] https://www.nautilos-h2020.eu/

7. Chavez, F.P., and Messié, M.: A comparison of Eastern Boundary Upwelling Ecosystems. Progress in Oceanography 83(1), 80–96 (2009). DOI: 10.1016/j.pocean.2009.07.032
8. Relvas, P., *et al.*: Physical oceanography of the western Iberia ecosystem: Latest views and challenges. Progress in Oceanography 74(2), 149–173 (2007). DOI: 10.1016/j.pocean.2007.04.021
9. Liu, A.K., Peng, C.Y., and Chang, S.Y.-S.: Wavelet analysis of satellite images for coastal watch. IEEE Journal of Oceanic Engineering 22(1), 9–17 (1997). DOI: 10.1109/48.557535
10. Kriebel, S., Brauer, W., and Eifler, W.: Coastal upwelling prediction with a mixture of neural networks. IEEE Transactions on Geoscience and Remote Sensing 36(5), 1508–1518 (1998). DOI: 10.1109/36.718854
11. Simpson, J.J.: On the accurate detection and enhancement of oceanic features observed in satellite data. Remote Sensing of Environment 33(1), 17–33 (1990). DOI: 10.1016/0034-4257(90)90052-N
12. Lea, S.M., and Lybanon, M.: Automated boundary delineation in infrared ocean images. IEEE Transactions on Geoscience and Remote Sensing 31(6), 1256–1260 (1993). DOI: 10.1109/36.317437
13. Reggiannini, M., Janeiro, J., Martins, F., Papini, O., and Pieri, G.: Mesoscale Patterns Identification Through SST Image Processing. In: Proceedings of the 2nd International Conference on Robotics, Computer Vision and Intelligent Systems — ROBOVIS, pp. 165–172. SciTePress (2021). DOI: 10.5220/0010714600003061
14. OSI SAF: Full resolution L2P AVHRR Sea Surface Temperature MetaGRanules (GHRSST) - Metop, (2011). DOI: 10.15770/EUM_SAF_OSI_NRT_2013
15. NASA/JPL: GHRSST Level 2P Global Sea Surface Skin Temperature from the Moderate Resolution Imaging Spectroradiometer (MODIS) on the NASA Aqua satellite (GDS2), (2020). DOI: 10.5067/GHMDA-2PJ19
16. Papini, O., Reggiannini, M., and Pieri, G.: SST Image Processing for Mesoscale Patterns Identification. Engineering Proceedings 8, 5 (2021). Presented at the 16th International Workshop on Advanced Infrared Technology and Applications — AITA. DOI: 10.3390/engproc2021008005
17. Pieri, G., *et al.*: New technology improves our understanding of changes in the marine environment. In: Proceedings of the 9th EuroGOOS International Conference. EuroGOOS (2021)