https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

Open Access Earth System Science Data Discussions

Southern Europe and Western Asia Marine Heat Waves (SEWA-MHWs): a dataset based on macro events

Giulia Bonino[1*], Simona Masina[1], Giuliano Galimberti[2], and Matteo Moretti[1,2]

[1]Ocean Modeling and Data Assimilation Division, Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy.
[2]Department of Statistical Sciences, University of Bologna, Bologna, Italy

Correspondence: Giulia Bonino (giulia.bonino@cmcc.it)

Abstract. Marine Heat Waves (MHWs) induce significant impacts on marine ecosystems. There is a growing need for knowledge about extreme climate events to better inform decision-makers on future climate-related risks. Here we present a unique observational dataset of MHWs macro events and their characteristics over the Southern Europe and Western Asian (SEWA) basins, named SEWA-MHWs dataset (Bonino et al., 2022). The SEWA-MHWs dataset is derived from the European Space Agency (ESA) Climate Change Initiative (CCI) Sea Surface Temperature (SST) v2 dataset and it covers the 1981-2016 period. The methodological framework used to build the SEWA-MHWs dataset is the novelty of this work. Firstly, the MHWs detected in each grid point of the ESA CCI SST dataset are relative to a time-varying baseline climatology. Since intrinsic fluctuation and anthropogenic warming are redefining the mean climate, the baseline considers both the trend and time-varying seasonal cycle. Secondly, using a connected component analysis, MHWs connected in space and time are aggregated in order to obtain macro events. Basically, a macro event-based dataset is obtained from a pixel-based dataset without losing high resolution (i. e. pixel) information. SEWA-MHWs dataset can be used for many scientific applications. For example, we identified phases of the well-know MHW of summer 2003 and, taking advantage of statistical clustering methods, we identified the largest macro events in SEWA basins and we clustered them based on shared metrics and characteristics.

1 Introduction

Over the past decades, anomalous prolonged events of warm sea surface water, known as marine heatwaves (MHWs), have developed globally, from the open ocean to coastal regions, leading to serious consequences for marine ecosystems. The ecosystems are very sensitive to abrupt temperature changes and they can reach their “tipping point” (Lenton et al., 2008; Serrao-Neumann et al., 2016), which means entering into an unknown state which is completely distant from the previous one. Increased ocean temperatures imply a large number of ecological repercussions spanning from a coastwide onset of toxic algae (McCabe et al., 2016; Ryan et al., 2017) to dramatic range shifts of species at all trophic levels (Cavole et al., 2016; Sanford et al., 2019). These adverse conditions during MHWs can lead to substantial economic losses for important fisheries and aquaculture industries (McCabe et al., 2016; Cavole et al., 2016; Frölicher, 2019).

Some relevant examples of unprecedented ocean temperature anomalies are the 2011 Western Australia marine heatwave in the eastern Indian Ocean (Pearce et al., 2011), and the persistent 2014–2016 “Blob” in the North Pacific (Di Lorenzo

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

and Mantua, 2016). In the Mediterranean Sea, several works reported an anomalous warm sea surface temperature during the summer of 2003 (e.g., Olita et al., 2007; Marullo and Guarracino, 2003; Grazzini and Viterbo, 2003). Darmaraki et al. (2019) identified the MHWs of 2003, 2012, and 2015 as the basin-scale most severe surface events during the 1982–2017 period. Marbà et al. (2015), focusing on impacts of these extreme events on Mediterannean biota, reported a basin-scale MHWs during 1994 and 2009 and a regional MHW over the Adriatic, the Ionian and parts of the Levantine basin during the 1998. Other studies on ecological impacts identified MHWs over the western Mediterranean Sea during 2008 (Cebrian et al., 2011) and 2006 (Kersting et al., 2013), and over Adriatic Sea during 2009 (Di Camillo et al., 2013). Very limited information is available for MHWs in the Black Sea and in the Caspian Sea (e.g. Mohamed et al. (2022)). They are usually related to works where MHWs are detected and studied at global scale, such as Holbrook et al. (2020) and Sen Gupta et al. (2020). A basic definition of MHWs (i.e., IPCC SROCC) states they are extremely anomalous temperatures in the ocean (Pörtner et al., 2019). However, to compare events and study their impacts, they must also be identifiable, with clear start and end dates and measurable characteristics. Hobday et al. (2016) were the first to propose a definition for MHWs, according to which the temperature must be higher than a given percentile (e.g., 90th, relevant to a reference climatology) and must persist for at least five days. This definition has been widely adopted by the oceanographic community (e.g., Holbrook et al., 2019, Oliver et al., 2021, Smale et al., 2019). However, it is worth mentioning that this definition is characterised by flexibility in the choice of the set-up parameters (such as the climatology and the percentile threshold), thus limiting comparability among studies.

Most of the research conducted in this emerging field exploits this definition to study extreme events at individual locations (i.e. grid points, pixel). Nevertheless, the pixel-based MHWs events are likely connected in time and in space being part of the same extreme macro event. Very few studies put effort into defining macro events. Darmaraki et al. (2019) define the spatiotemporal extent of the MHW when a minimum of 20% of the Mediterranen basin is affected by pixel-based MHWs. More recently, Sen Gupta et al. (2020) used a semi-objective procedure to characterize and to detect globally the most extreme MHWs macro events and Woolway et al. (2021) used a connected component analysis to study extreme temperature macro events in the Laurentian Great Lakes. Macro event-based studies facilitate the investigation of MHWs drivers, which are not well understood (Holbrook et al., 2020). Driving mechanisms, which are usually seasonal and location-dependent, are related to oceanic and atmospheric forcing or a combination of both (e.g., Oliver et al., 2018, 2021; Holbrook et al., 2019; Frölicher and Laufkötter, 2018). Examples of key phenomena which cause these extreme temperatures are anomalous horizontal advection, anomalous heat fluxes, sea level pressure anomalies, reduced coastal upwelling, Ekman pumping or the re-emergence of warm anomalies from the subsurface (Schlegel et al., 2021; Holbrook et al., 2019, 2020). The time scale of these relevant physical drivers and processes involved in MHW emergence spans from days (e.g. anomalous heat fluxes), to weeks (e.g. blocking systems and atmospheric teleconnections), to months (e.g. re-emergence of warm anomalies from the subsurface) and years (e.g. climate modes and oceanic teleconnections) (e.g., Oliver et al., 2018, 2021).

Given the pronounced warming trend in recent years, Benthuysen et al. (2020) and Holbrook et al. (2020) suggest the need for a more comprehensive and consistent framework to report MHWs. To the best of our knowledge, there is no available MHWs macro events dataset in literature. Even thought MHWs are derived from Sea Surface Temperature (SST), whose observations are available, the processing to detect MHWs is computational and/or time demanding, especially for high resolution SST data.

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

In short, the current state of knowledge about MHWs requires addressing the need for more comprehensive efforts to document and to report these extreme events. The aim of this paper is to provide a unique dataset of MHWs macro events derived from the European Space Agency (ESA) Climate Change Initiative (CCI) Sea Surface Temperature v2 dataset. The dataset consists a daily dataset of MHWs macro events and their characteristics over the Southern Europe and West Asian (SEWA) basins, named as SEWA-MHWs dataset. We have focused on SEWA basins because they represent a well known "Hot Spot" region for climate change (Giorgi, 2006), which is currently affecting their high marine biodiversity, especially in the Mediterranean sea (Juza et al., 2022). In the future, we are consider to expand the dataset to other oceanic regions.

We generated the SEWA-MHWs dataset in a new consistent framework. In brief, we detected MHWs relative to a time-varying baseline climatology in each grid point of the ESA CCI SST dataset, then, using a connected component analysis, we aggregated the spatiotemporally connected MHWs in order to obtain macro events.

This paper is organized as follows: in Section 2, we present the data used to produce the SEWA-MHWs dataset and the studied area. In section 3 we describe the methodological framework applied to obtain the dataset and in section 4 we offer an example of its scientific application. Our conclusions and outlook of the work are summarized in section 5. Section 6 reports the availability of codes and data used to build SEWA-MHWs dataset.

2 ESA SST CCI data and Study Area

To generate the SEWA-MHWs dataset we used the European Space Agency (ESA) Climate Change Initiative (CCI) SST dataset v2.0 (hereinafter ESA CCI SST dataset). This dataset, available at https://catalogue.ceda.ac.uk/uuid/62c0f97b1eac4e0197a674870afe1ee6, provides global daily satellite-based SST data covering the period from September 1981 to December 2016. A detailed overview of processing updates, and of the history behind, for the ESA CCI SST dataset v2.0 is presented by Merchant et al. (2019). The ESA CCI SST v2.0 dataset is designed to provide a long-term, stable, low-bias climate data record derived from different infrared sensors, i.e., the AVHRR, (A)ATSR and SLSTR series of sensors (Merchant et al., 2019, 2014). These data are available at different processing levels: single-sensor data on the native swath grid (Level-2); uncollated single-sensor (Level-3U) and collated multi-sensor (Level-3C) gridded data; and optimally interpolated (Level-4) multi-sensor data. Here, only the Level-4 product is considered. The ESA CCI SST Level-4 product v2.0 consists of spatially complete (optimally interpolated) maps of global daily average SST at 20 cm nominal depth at 0.05° x 0.05° regular grid covering the period from September 1981 to December 2016. Level-4 SST data have been produced by running the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system (Donlon et al., 2012). In this work we focused on the Southern Europe and West Asian basins, namely we constructed our SEWA-MHWs dataset over the Mediterranean sea, the Black sea and the Caspian sea (see Figure 1a).

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

Open Access Earth System Science Data Discussions

3 Methodological Framework

In this Section we present the methodological framework developed to obtained the SEWA-MHWs dataset (Bonino et al., 2022). We first present the pixel-based definition, the baseline climatology estimation strategy and the metrics (Section 3.1). Then, we describe the macro event detection method and the detected 2003 MHW macro event (Section 3.2).

3.1 Pixel-based MHWs definition and their characteristics

We identified MHWs for each grid point of ESA CCI SST dataset following the definition of Hobday et al. (2016): "MHWs are identifiable events with start and end dates, a persistent duration of at least five days and anomalous warmer sea surface water relative to a threshold (90th or 99th percentile) in a 30-year baseline climatology". To build a globally consistent detection framework to study MHWs drivers and variability, it is therefore crucial the choice of a proper baseline climatology, along with the choice of the percentile threshold. To describe and to study MHWs it is also fundamental to define MHWs metrics and characteristics.

3.1.1 Baseline climatology estimation and threshold

Hobday et al. (2016) estimate the baseline climatology as a seasonally varying climatology calculated over 30-year period without removing long-term trend (hereinafter Hobday-method, see Hobday et al., 2016 for details). In contrast, we considered trend and time-varying seasonality in the baseline climatology estimation. As the definition states, MHWs detection depends on the underlying SST properties (Oliver et al., 2021). Recent studies show that increases in both the mean SST and the variability of SST due to global warming can lead to increase in warm temperature extremes (Pierce et al., 2012), so that by the late twenty-first century most of the global ocean will reach a permanent MHW state (Oliver et al., 2018; Holbrook et al., 2020; Frölicher et al., 2018). Persistent warming of SST and the increased frequency and intensity of extreme events indicate that the global ocean is experiencing unprecedented climate normals (Tanaka and Van Houtan, 2022). Thus, the rationale of considering trend and time-varying seasonality in the baseline climatology estimation is to take into account that the mean state of the ocean is changing over time due to natural variability and anthropogenic climate change.

We estimated the baseline climatology using a non-parametric time series decomposition algorithm, named Seasonal-Trend Decomposition Procedure Based on LOESS (hereinafter STL-method), designed and described by Cleveland et al. (1990). It is based on the non-parametric technique known as Locally Estimated Scatterplot Smoothing (LOESS), also commonly called Local Regression. The method estimates time-varying trend-cycle and seasonality for each time series. The STL algorithm that we applied to ESA SST CCI is implemented in a freely available Python function named STL (https://www.statsmodels.org/devel/examples/notebooks/generated/stl_decomposition.html) from the Statsmodel Python package (https://www.statsmodels.org/stable/index.html). Figure 2 shows the STL-method decomposition for an ESA SST CCI time series in the western Mediterranean. The main parameters to set for the STL algorithm are related to the periodicity of the seasonal signal to be extracted, and to some LOESS smoothing parameters for the trend and for the seasonality. We tuned our estimation of the trend with a smoothing window of 10 years. This choice produces a trend-cycle which captures both the long-term trend and low frequency

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

variability (Figure 2c). Regarding the seasonal cycle, we tuned its estimation with a yearly periodicity, smoothed over 5 years (Figure 2b). The time-varying amplitude of the seasonality captures increasing/decreasing trends in the seasonal variability of the time series. The sum of the trend-cycle and the seasonality is considered as our baseline climatology (Figure 1b, orange line).

The STL-method climatology (orange line, Figure 1b), in contrast to the Hobday-method climatology (black line, Figure 1b), varies in location, due to the trend-cycle, and in variability, due to the time-varying seasonality. The Hobday-method climatology results instead in a pure periodic seasonal climatology with a flat trend. In particular, the time series in Figure 1b shows an increased STL-method climatology during the 2012-2016 period, which is due to the fact that STL-method includes an increased trend in the estimation of climatology (Figure 1c, top panel). It is evident that SST time series could present different time evolution. They could show decreasing, oscillating or stationary trends in the mean and in the variance. One of the main advantages of the STL-method decomposition is that it allows to work on SST time series in a consistent framework, where the residuals, obtained by substracting the estimated trend-cycle and climatology from the corresponding observed SST values, are the relevant comparable series, and trend and seasonality form together the time-varying baseline climatology.

As suggested by Hobday et al. (2016), we computed the 90th percentile of the residuals and used it as threshold (Figure 2d) in order to detect MHW. In addition, two successive MHW events with a 2-day or less time break were considered as a continuous events.

It is worth noting that we did not consider pixels where sea ice can be present, so that MHWs are not available over the North Caspian Sea and the Azov Sea (Figure 1a).

3.1.2 MHWs characteristics

Following the approach of Hobday et al. (2016), for each detected MHW we calculated the following MHWs metrics and characteristics: duration (i.e. the time between the start and end dates), mean intensity and maximum intensity (i.e. the average and maximum temperature anomaly over the duration of the event), the index start, the index end and the index peak (i.e. indices at start, end and at the peak of the marine heatwave in times series of origin). Each detected MHWs was assigned to a category describing its severity (Hobday et al., 2018). These categories range from 1 to 4 and they are based on the maximum intensity in multiples of the 90th percentile exceedances, i.e., category 1 indicates the MHW peak intensity is >=1 times the value of the 90th percentile threshold, but less than 2 times. Category types are defined as 1=moderate, 2=strong, 3=severe, 4=extreme. Linked to the categories we also estimated the moderate, the strong, the severe and extreme MHWs duration (i.e. the number of days of a MHW in each category).

Figure 3 shows the spatial pattern of annual event count, mean duration, mean intensity, maximum intensity of the MHWs events over SEWA basins. The Black Sea, the north Caspian Sea, the Adriatic Sea, the Gulf of Lion and the Gibraltar Strait experienced the most frequent, shortest and most intense MHWs. The majority of the events over SEWA basins belong to the moderate category, reaching the 90% in the Black sea and in the Central Mediterranean Sea (Figure 4). All the basins experienced moderate MHWs, especially the Eastern Mediterranean, the North Adriatic and the Ligurian Seas. 5% of MHWs

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

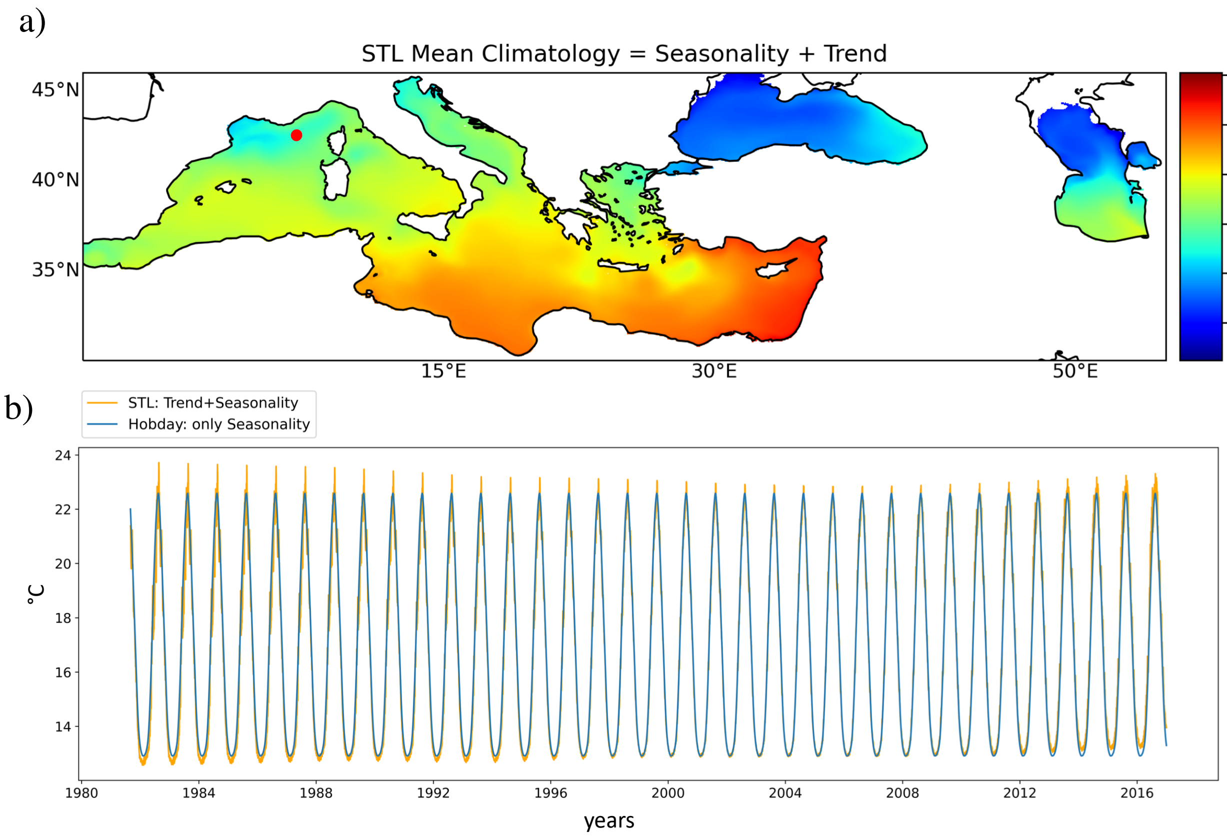

Figure 1. a) Mean SST climatology detected by STL: b) STL mean climatology (orange line) and Hobday mean climatology (black line) for a SST times-series in Western Mediterranean (red circle in Figure 1a)

are severe over the Eastern part of the Caspian Sea, South of the Gulf of Lion and the Gulf of Sidra. MHWs in category 4
155 (extreme) are very rare over the studied basins.

3.2 MHWs macro events detection and their characteristics

From the previous analysis we identified MHWs for each 4 km grid for the ESA SST CCI dataset and we obtained a pixel-based dataset. For each day, a binary map was generated, identifying which of the 4km grid points across the global ocean experienced a MHW. The 3D binary maps (i.e. time x longitude × latitude) were then used to identify spatiotemporally connected marine
160 heatwaves (i.e. grid points that are connected in 3D space in terms of MHWs occurrence). Specifically, following the approach of Woolway et al. (2021), a connected component analysis is used to identify a connected group of marine heatwave pixels which, in turn, are considered as part of the same MHW (i.e. a contiguous region simultaneously experiencing a MHW). For this work we used the distributed version of Label function (https://docs.scipy.org/doc/scipy/reference/generated/scipy.

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

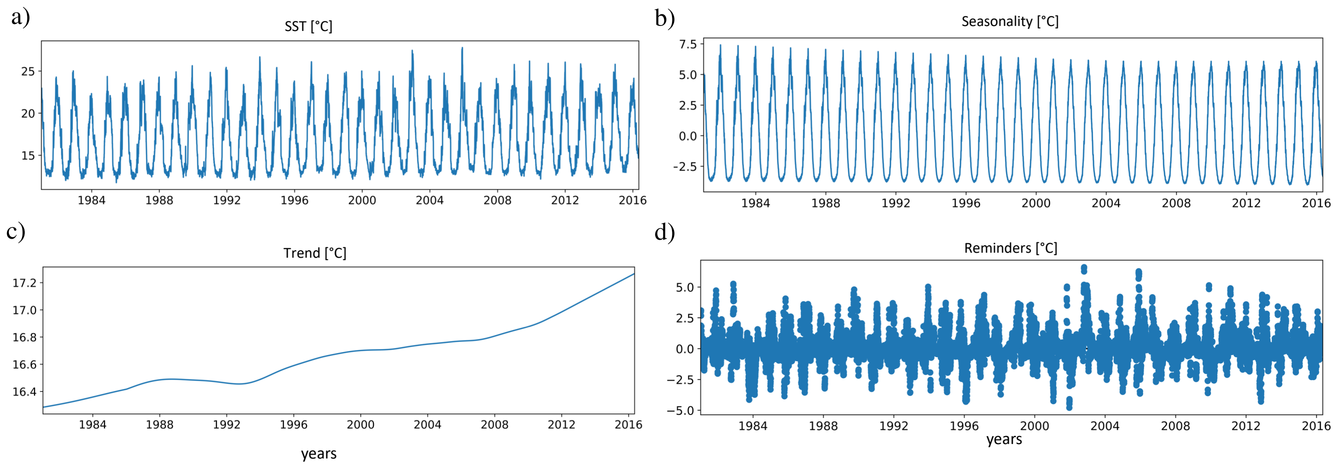

Figure 2. STL decomposition for a SST times-series in Western Mediterranean (red circle in Figure 1a): a) SST times-series, b) Time-varying seasonality, c) Trend-cycle, d) Residuals.

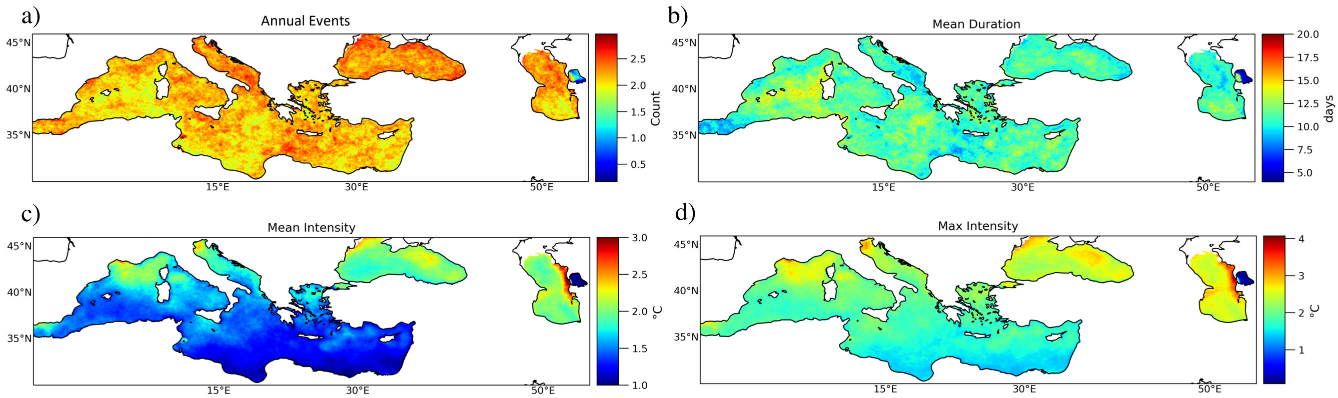

Figure 3. MHWs characteristics for each pixel: a) Number of events, b) Mean Duration, c) Mean Intensity, d) Maximum intensity.

ndimage.label.html#scipy.ndimage.label) from the python packaged named dask_image.ndmeasure (http://image.dask.org/en/latest/_modules/dask_image/ndmeasure.html#label). The label function calculates connectivity of features to their neighbors based on a structure element matrix, which defines the directions in which connectivity is inspected. In our case, the inputs of the function are the 3D binary maps chunked by time (the non-zero values in matrices, i.e. MHWs occurrence, in the matrices are counted by the algorithm as features and zero values are considered the background) and the orthogonal structuring element matrix that defines the features connection. Since the algorithm works in parallel, first, each chunk is independently labeled (i.e. connection in space). Then, the independent labels are made consecutive and merged along chunks' faces whenever connected (i.e. connection in time). The algorithm returns connected pixels with an unique label. Each of these unique labels represents

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

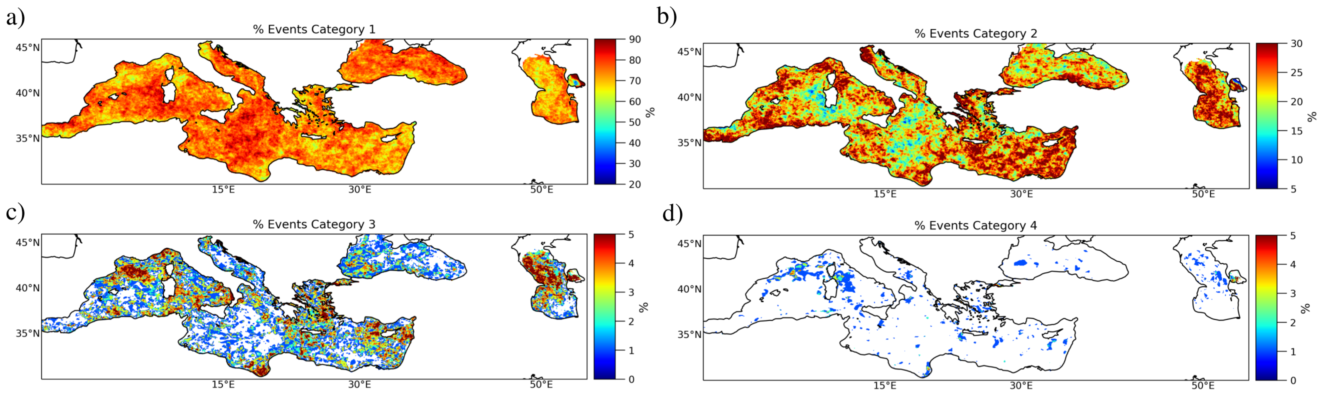

Figure 4. Percentage (%) of MHWs events in a) category 1, b) category 2, c) category 3, d) category 4

a macro event. After filtering out macro events with a maximum area lower than $64km^2$ (4 pixels), we found 68068 macro events over the SEWA region.

Figure 5 shows some examples of macro events detected by our procedure and their evolution in time. Each color identifies a different macro event. Focusing on the lilac event in Figure 5 we can appreciate the strength of the method. The macro event starts on October,5th in three different spots in the Aegean Sea, then it grows spatially reaching its maximum extension on October,19th almost occupying all the Eastern Mediterranean sea. Finally, it decays along the Libyan coast by November, 20th. Meanwhile, other macro events develop in the basins (e.g. blue label in the Caspian Sea, green label in the Black sea).

Our dataset is composed of daily fields of labels. Moreover, we extracted for all the grid points belonging to a given macro event, the MHWs characteristics (Section 3.1.2) and we obtained daily fields of MHWs characteristics. In particular we extracted daily fields of MHWs mean intensity, maximum intensity, index start, index end, index peak and categories. Basically, we moved from a pixel-based dataset to an event-based dataset without losing pixel information. The daily labels and the daily MHWs macro events characteristics form the SEWA-MHWs dataset.

3.2.1 Mediterranean MHW of 2003

In order to assess how the proposed method performs in a the case of a well-know event, we show in details how it detected the 2003 MHW over the Mediterranen Sea (hereinafter MED-MHW-2003). The detected macro event covered all the western Mediterranean and it lasted 302 days (Figure 6 a and b). Based on the number of active points (i.e. points which simultaneously experienced the labeled MED-MHW-2003, blue line in Figure 6a) and the mean intensity of the active points (orange line) we can distinguish and characterize the evolution of the MED-MHW-2003 in five phases:

1. Phase 1: it lasted from May to June 2003 and it expanded over the Central Mediterranean sea. The MHW mean intensity reached 2°C west of Sicily (Figure 6c).

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

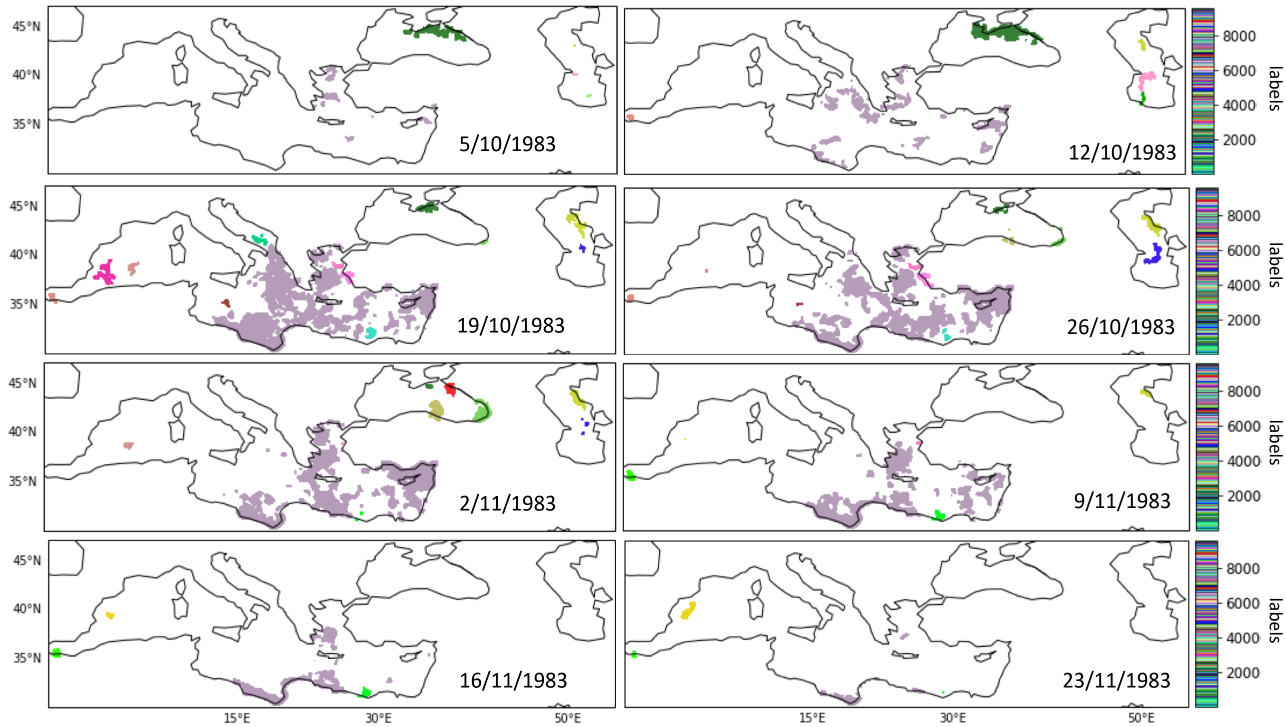

Figure 5. Labels of MHWs Macro-Events and their evolution in time, from 5th October 1983 to 23rd November 2021, over the Mediterranean Sea detected by the connected component analysis.

2. Phase 2: it lasted from June to middle July 2003 and it expanded over the Western, Central Mediterranean and the Adriatic seas. The MHW mean intensity reached 3-4°C in the Ligurian Sea, in the Tyrrhenian sea and on the Adriatic coast.
3. Phase 3: it lasted from middle July to August 2003 and it expanded over the Central Mediterranean sea. The MHW mean intensity reached 3°C to the west of Sardinia.
4. Phase 4: it lasted from August to middle September 2003 and it expanded over the Western Mediterranean sea. The MHW mean intensity reach 3-4°C in Gulf of Lion.
5. Phase 5: it lasted from middle September 2003 to February 2004 and it expanded over the Central Mediterranean sea. The MHW mean intensities are around 1°C.

Even though the evaluation of a MHW is strictly dependent on its definition, we compared our results for the MED-MHW-2003 with existing works in the literature. Analyzing SST data from 1985 to 2003 by NOAA-AVHRR sensors,Marullo and Guarracino (2003) reported a sea surface temperature warm anomaly of about 4°C in June of 2003 in the Gulf of Lion, Ligurian,

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

Tyrrhenian, Northern Ionian and Adriatic Seas similarly to MED-MHW-2003 phase2 (Figure 6b). Similarly to MED-MHW-2003 phase 3 and 4 shown in Figure 6b, they also observed the persistence of the event in July with weaker anomalies of about 3°C, and an intensification in August over the Gulf of Lion and the Ligurian Sea, with recorded anomalies of about 4°C. In September, as detected in MED-MHW-2003 phase 5, they observed as well a considerable weakening of the event, characterized by anomalies below 1°C. MED-MHW-2003 phase 1, 2 and 3 are also comparable with the analysis of Grazzini and Viterbo (2003) and Olita et al. (2007). Using the ECMWF forecast products and a regional 3-D ocean model, respectively, they described the event as a rapidly increase of SST anomalies over the Central Mediterranean Basin during May 2003, which, it intensified (with anomalies of about 3-4°C) and it expanded, covering the whole Mediterranean with the exception of the Aegean Sea, by the end of July. The data analysed by Sparnocchia et al. (2006) in the Ligurian Sea evidenced the development of a warm anomaly in the SST as reported by MED-MHW-2003 phase 2, 3 and 4. They detected a local SST anomaly up to 2-3°C, which built up at the end of May and persisted until August 2003.

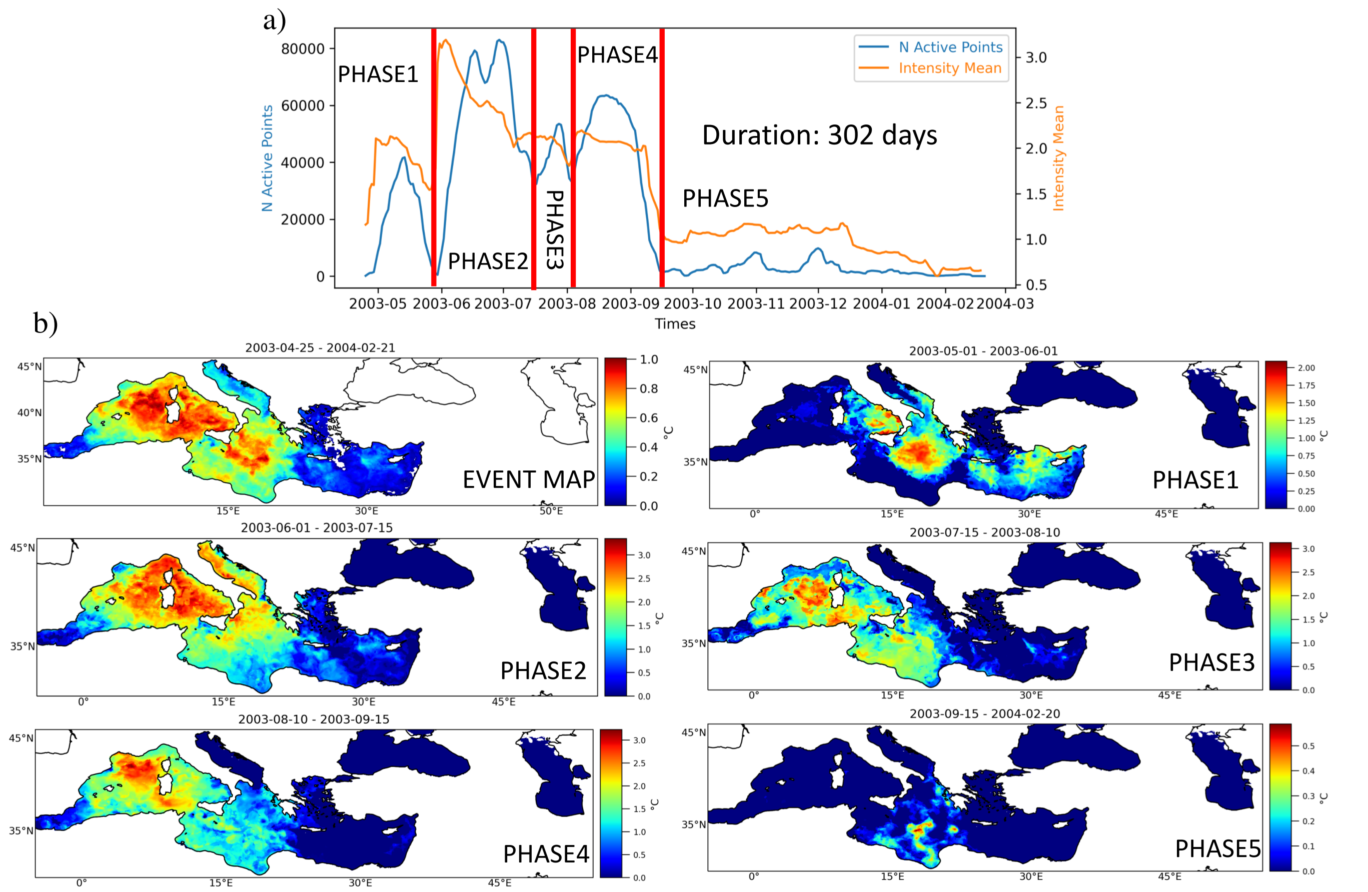

Figure 6. a) Active points (blue line) and mean intensity of active points (orange) during the 2003 MHW, b) Average of the mean intensity during the 2003 MHW phases.

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

4 Scientific Application

4.1 Spatial/Agglomerative clustering of MHWs Macro-Events

In order to highlight the added value of the SEWA-MHWs dataset we studied the largest macro events identified by our methodology. Specifically, we classified and we aggregated MHWs macro events that share characteristics taking advantage of statistical clustering methods. Following the work by Stefanon et al. (2012), who identified and classified continental heat waves over Europe during the 1950-2009 period, we devised a clustering procedure to explore similarities and dissimilarities among macro events included in the dataset. Our clustering technique consists of 4 steps:

1. For each identified MHW macro event, we extracted the maximum area extension reached. We retained MHWs macro events with area greater than 100 000 km2, which is around 25% of the SEWA basins. We identified 187 macro events.

2. For each day belonging to one event, we extracted the MHW mean intensity for all the grid points which belong to the macro event and we set to nan the others grid points. So that, for each Macro-Event we obtained daily maps of mean intensity.

3. All daily maps belonging to one macro event are averaged producing 'event maps'. So that, we obtained one event map of mean intensity for each macro event.

4. An agglomerative hierarchical clustering algorithm (Gordon, 1999) is applied to the event maps. At the initial step, each event map forms a cluster. The two 'nearest' clusters are then merged by pair into a new cluster. We used the cosine distances to measure the distances between clusters. In general, the definition of the cosine distance is particularly suited to distinguish between different spatial patterns of the intensities as it tends to increase as the number of pixels shared by the event maps of two macro events decreases (see Stefanon et al. (2012) for additional details). We used a Python algorithm to perfom the clustering (https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html).

Figure 7 shows the average silhouette score of our clustering approach. The silhouette score is a measure of the average distance between a member in one cluster to the members in the neighboring clusters and thus provides a way to assess cluster separation (Kaufman and Rousseeuw, 2009). Therefore, a large average silhouette score generally indicates large separating distances between the resulting clusters, and hence better clustering results. We varied the number of clusters between 2 and 14. We concluded that the optimal number of clusters is 6. Moreover, using the silhouette score we also assessed the quality of this optimal solution (Figure 7b). We computed the silhouette scores of the data points in each cluster when the number of cluster is 6. Different colors correspond to different clusters. Red dashed line shows the average silhouette score for this solution. Most data points in the 6 clusters have a silhouette score larger than the average score, which indicates a favorable clustering result. Negative values represent possibly misclassified events.

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

Open Access Earth System Science Data Discussions

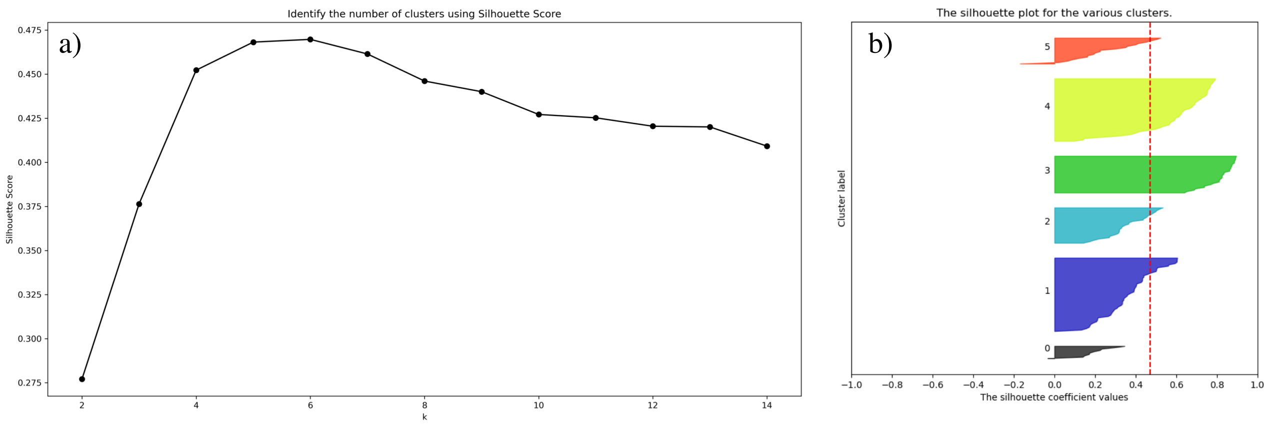

Figure 7. a) Average Silhouette score for Agglomerative clustering performed using cosine distances, b) silhouette analysis for agglomerative clustering performed using cosine distances on data with n clusters=6.

4.2 MHWs Macro-Events clusters and their characteristics

Figure 8 shows the typical marine heatwave patterns for each identified cluster. The patterns of each cluster are represented by the average of the mean intensity of all the event maps which belong to that cluster. Boxplots for area maxima, duration,intensity maxima and intensity means for MHWs macro events for each cluster are available in Figure 9, the box shows the quartiles of the dataset distribution while the whiskers extend to show the rest of the distribution, except for diamonds that are determined to be “outliers”. Moreover, Figure 10 shows the number of MHWs macro events by season and by decades for each cluster.

The longest and the largest macro events, in terms of area maxima, belong to cluster 2 which spans over the Western-Mediterranean and the Adriatic sea, with the exception of the Aegean Sea. The maximum intensities are located in the Gulf of Lion and the Ligurian Sea and they decrease magnitude towards the Adriatic Sea. The Aegean sea, instead, emerges as separate cluster, it counts 27 macro events (cluster 3 in Figure 8). The maximum intensities are located west of Cyprus Island and they decrease magnitude around the Greek Archipelago, showing mean intensity of about 1.3°C and maximum intensity of about 2.5°C (Figure 9). The Aegean sea, to a lower extent, is also part of other two identified clusters, cluster 5 and cluster 6. The former expands in the Black sea, where it shows its maximum intensity, while the latter includes the Adriatic Sea and the the Ionic sea. Both of these two clusters experienced strong macro events with maximum intensity of about 3.5°C (Figure 9). Apart from cluster 6, the Central Mediterranean sea is also home of the smallest cluster, consisting of 10 macro events confined from the Strait of Sicily to the eastern Libyan coast (cluster 1 in Figure 8). The maximum intensities are located south of Sicily, with mean intensity of about 1.3°C and maximum intensity of about 2.5°C (Figure 9). A completely isolated cluster groups the 28 macro events over the Caspian Sea (cluster 4 in Figure 8). The maximum intensities are located along the western coast of the basin. The strongest, in term of intensity, macro events belong to this cluster, reaching 2.3 and 4.5°C of mean and maximum intensity, respectively (Figure 9).

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

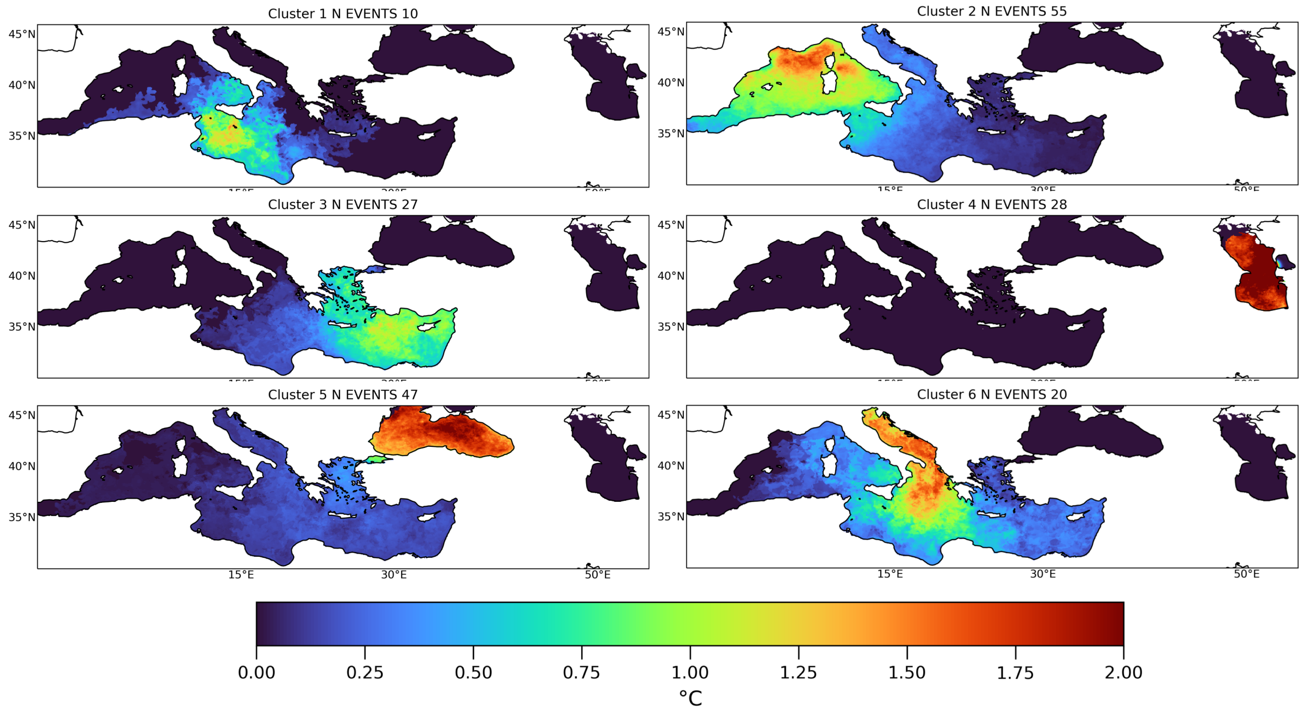

Figure 8. Patterns of the obtained clusters represented by the average of the mean intensity of all the event maps which belong to each cluster.

Except for cluster 6 and cluster 1, it is interesting to highlight that the majority of the events are during the summer season, while the winter season counts only few macro events for each cluster (Figure 10). Moreover, we do not report an increasing number of MHWs during the last decades (2011-2016) with respect to the past, as is instead reported in Dayan et al. (2022). Actually, the majority of the macro events in almost all the clusters occurred during the first two decades of the studied period (Figure 10). This is likely linked to the fact that we considered the trend in the baseline climatology estimation (see section 3.1.1).

Our clustering methodology seems to be efficient in distinguishing different spatial patterns of MHWs macro events over the SEWA basins. The macro events result geographically confined in the closed basin (i.e Caspian sea) and in the sub-basins of the Mediterranean Sea, however highlighting relations between them.

5 Summary and Outlook

In this work we presented a dataset of Marine Heatwaves macro events and their characteristics over the Southern Europe and Western Asia basins during the 1981-2016 period, named SEWA-MHWs dataset. We obtained the dataset analysing observed sea surface temperature provided by the European Space Agency, the ESA CCI SST v2.0 dataset (Merchant et al., 2019). Briefly, we defined MHWs in each 4x4km grid point of the ESA CCI SST dataset, then, using a connected component analysis,

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

Open Access Earth System Science Data Discussions

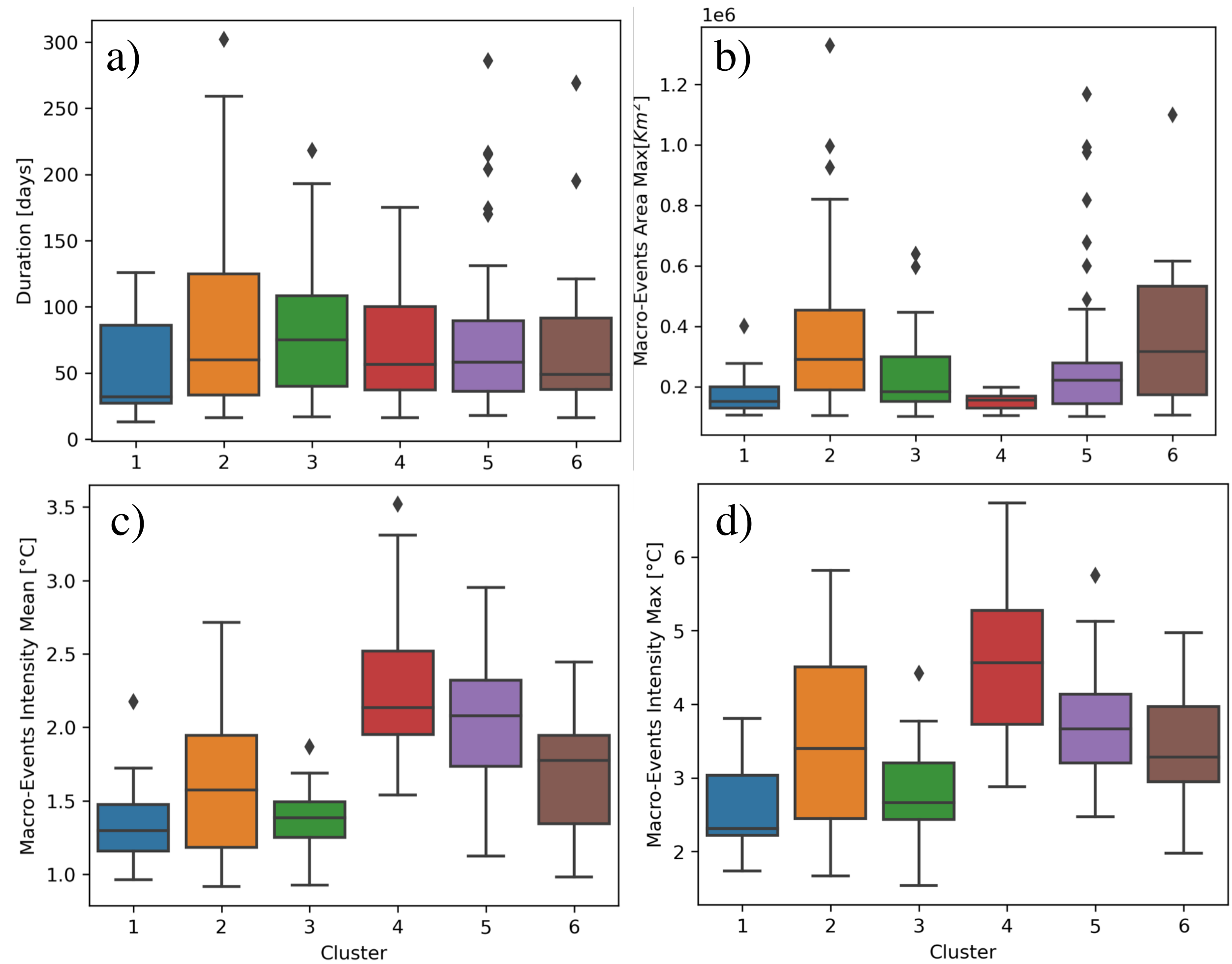

Figure 9. Boxplots for a) Duration, b) Area, c) Intensity Mean and d) Intensity Maxima for each cluster.

we aggregated the spatiotemporally connected MHWs (i.e. grids that are connected in 3D space in terms of marine heatwave occurrence) in order to obtain MHWs macro events. As a result, the SEWA-MHWs dataset, consists in daily field of MHWs macro events and their characteristics.

The methodological framework used to build SEWA-MHWs dataset is the novelty of this study with respect to the existing literature (e.g. Darmaraki et al., 2019; Sen Gupta et al., 2020; Oliver et al., 2021). Firstly, the detected MHWs in the ESA CCI SST dataset are relative to a time-varying baseline climatology, which considers both the trend and seasonal variability to mimic the changing of the climate mean state. Secondly, the connected component analysis allow us to aggregate the MHWs connected in time and in space and to pass from a the pixel-based dataset to an event-based dataset without loosing high resolution (i.e. pixel) information. This approach, differently from the previous studies, provides the time evolution of the

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

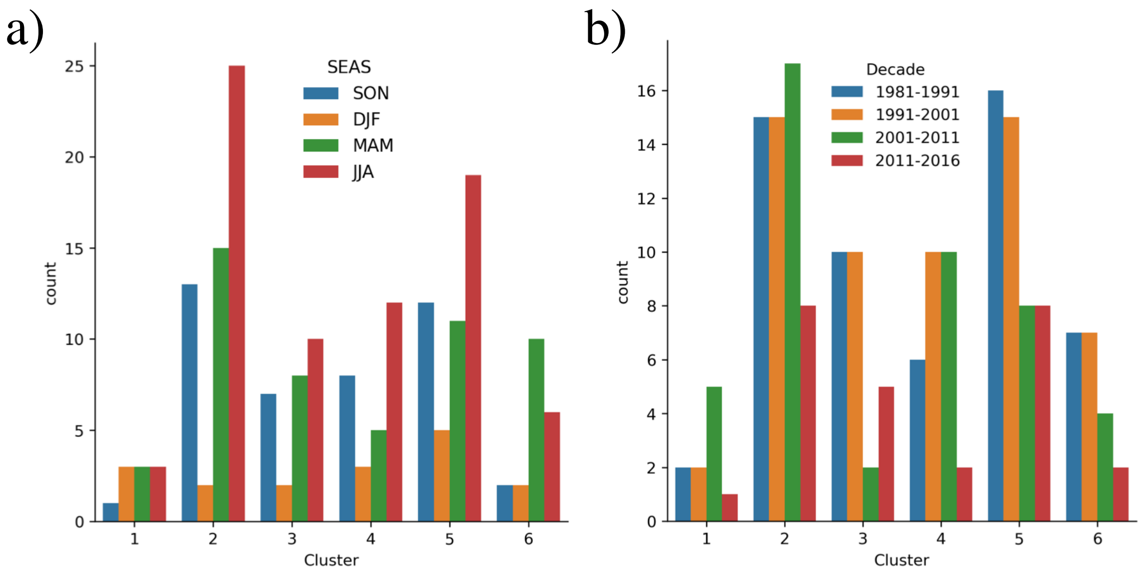

Figure 10. a) MHWs macro events by season for each cluster; b) MHWs macro events by decades for each cluster.

event at the basin scale. Even though the evaluation of a MHW is strictly dependent on its definition, we demonstrated that our method is efficient in detecting MHWs Macro-Events. The well-known MHW of 2003 in the Mediterranean sea is comparable with the records in the existing literature (e.g. Olita et al., 2007; Sparnocchia et al., 2006; Marullo and Guarracino, 2003; Grazzini and Viterbo, 2003).

To the best of our knowledge, the SEWA-MHWs dataset is the first effort in literature in archiving extreme hot sea surface temperature macro events. The advantages of the availability of a MHWs macro events dataset are to avoid waste of computational and/or time resources to process SST data to detect MHWs, and to build a consistent framework which would increase comparability among MHWs studies. The SEWA-MHWs dataset can be used for many scientific applications. For instance, we efficiently clustered the biggest SEWA-MHWs macro events that share common features characteristics in order to report and to characterize typical spatial patterns of MHWs over SEWA basins. Indeed, the employed clustering method was able to distinguish different spatial patterns of MHWs macro event.

The SEWA-MHWs dataset is also suitable for regional and coastal MHWs studies, due to its high resolution, and, it is expandable to all the ocean basins up to reach a global coverage. Moreover, the synergistic use of SEWA-MHWs dataset with other model outputs and observation data could help to fill the knowledge gaps about the drivers and the marine ecosystems impacts of these extreme events. Recently, compound events have become of particular interest, i.e., when conditions are extreme for multiple potential ocean ecosystem stressors such as temperature and chlorophyll (Gruber et al., 2021, Le Grix et al., 2021). On top of that, this synergistic use of SEWA-MHWs dataset with other datasets could facilitate the building of

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

prediction scheme using, for example, deep machine learning approaches. These techniques need large and high resolution datasets to be trained and tested.

In a broader perspective, the novel science resulting from the aforementioned exploitation of the SEWA-MHWs dataset could be transferred into solutions and advanced decision support systems for society.

6 Data availability

The SEWA-MHWs dataset, that consists in daily fields of MHWs macro events and their characteristics, is stored in the Zenodo archive (https://doi.org/10.5281/zenodo.7153256, Bonino et al. 2022)

Author contributions. GB and SM conceived the study. GB, SM, GG and MM discussed and defined the methodological framework. GB and MM set up the code for STL decomposition and for the MHWs Macro-Events detection. GB performed the clustering, all the analysis and wrote the manuscript. GB, SM, GG, MM interpreted the results. SM and GG revised and contributed to improving the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This research has been funded by the European Space Agency (ESA) as part of the FEVERSEA Climate Change Initiative (CCI) fellowship (ESA ESRIN/Contract No. 4000133282/20/I/NB). We acknowledge the CMCC Foundation for providing computational resources.

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

References

Benthuysen, J. A., Oliver, E. C., Chen, K., and Wernberg, T.: Advances in understanding marine heatwaves and their impacts, Frontiers in Marine Science, 7, 147, 2020.

Bonino, G., Masina, S., Galimberti, G., and Moretti, M.: Southern Europe and Western Asia Marine Heat Waves (SEWA-MHWs): a dataset based on macro events, Zenodo[dataset],https://doi.org/10.5281/zenodo.7153256, 2022.

Cavole, L. M., Demko, A. M., Diner, R. E., Giddings, A., Koester, I., Pagniello, C. M., Paulsen, M.-L., Ramirez-Valdez, A., Schwenck, S. M., Yen, N. K., et al.: Biological impacts of the 2013–2015 warm-water anomaly in the Northeast Pacific: winners, losers, and the future, Oceanography, 29, 273–285, 2016.

Cebrian, E., Uriz, M. J., Garrabou, J., and Ballesteros, E.: Sponge mass mortalities in a warming Mediterranean Sea: are cyanobacteria-harboring species worse off?, PLoS One, 6, e20 211, 2011.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I.: STL: A seasonal-trend decomposition, J. Off. Stat, 6, 3–73, 1990.

Darmaraki, S., Somot, S., Sevault, F., and Nabat, P.: Past variability of Mediterranean Sea marine heatwaves, Geophysical Research Letters, 46, 9813–9823, 2019.

Dayan, H., McAdam, R., Masina, S., and Speich, S.: Diversity of marine heatwave trends across the Mediterranean Sea over the last decades, in: Copernicus marine service ocean state report, issue 6, Journal of Operational Oceanography, vol. 15, pp. 49–56, 2022.

Di Camillo, C. G., Bartolucci, I., Cerrano, C., and Bavestrello, G.: Sponge disease in the Adriatic Sea, Marine Ecology, 34, 62–71, 2013.

Di Lorenzo, E. and Mantua, N.: Multi-year persistence of the 2014/15 North Pacific marine heatwave, Nature Climate Change, 6, 1042–1047, 2016.

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., and Wimmer, W.: The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system, Remote Sensing of Environment, 116, 140–158, https://doi.org/10.1016/j.rse.2010.10.017, 2012.

Frölicher, T. L.: Extreme climatic events in the ocean, in: Predicting Future Oceans, pp. 53–60, Elsevier, 2019.

Frölicher, T. L. and Laufkötter, C.: Emerging risks from marine heat waves, Nature communications, 9, 1–4, 2018.

Frölicher, T. L., Fischer, E. M., and Gruber, N.: Marine heatwaves under global warming, Nature, 560, 360–364, 2018.

Giorgi, F.: Climate change hot-spots, Geophysical research letters, 33, 2006.

Gordon, A. D.: Classification, CRC Press, 1999.

Grazzini, F. and Viterbo, P.: Record-breaking warm sea surface temperature of the Mediterranean Sea, ECMWF Newsletter, 98, 30–31, 2003.

Gruber, N., Boyd, P. W., Frölicher, T. L., and Vogt, M.: Biogeochemical extremes and compound events in the ocean, Nature, 600, 395–407, 2021.

Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Feng, M., et al.: A hierarchical approach to defining marine heatwaves, Progress in Oceanography, 141, 227–238, 2016.

Hobday, A. J., Oliver, E. C., Gupta, A. S., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Holbrook, N. J., Moore, P. J., Thomsen, M. S., Wernberg, T., et al.: Categorizing and naming marine heatwaves, Oceanography, 31, 162–173, 2018.

Holbrook, N. J., Scannell, H. A., Sen Gupta, A., Benthuysen, J. A., Feng, M., Oliver, E. C., Alexander, L. V., Burrows, M. T., Donat, M. G., Hobday, A. J., et al.: A global assessment of marine heatwaves and their drivers, Nature Communications, 10, 1–13, 2019.

Holbrook, N. J., Sen Gupta, A., Oliver, E. C., Hobday, A. J., Benthuysen, J. A., Scannell, H. A., Smale, D. A., and Wernberg, T.: Keeping pace with marine heatwaves, Nature Reviews Earth & Environment, 1, 482–493, 2020.

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

Juza, M., Fernández-Mora, À., and Tintoré, J.: Sub-Regional Marine Heat Waves in the Mediterranean Sea From Observations: Long-Term Surface Changes, Sub-Surface and Coastal Responses, Frontiers in Marine Science, 2022.

Kaufman, L. and Rousseeuw, P. J.: Finding groups in data: an introduction to cluster analysis, John Wiley & Sons, 2009.

Kersting, D. K., Bensoussan, N., and Linares, C.: Long-term responses of the endemic reef-builder Cladocora caespitosa to Mediterranean warming, PLoS One, 8, e70 820, 2013.

Le Grix, N., Zscheischler, J., Laufkötter, C., Rousseaux, C. S., and Frölicher, T. L.: Compound high-temperature and low-chlorophyll extremes in the ocean over the satellite period, Biogeosciences, 18, 2119–2137, 2021.

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., and Schellnhuber, H. J.: Tipping elements in the Earth's climate system, Proceedings of the national Academy of Sciences, 105, 1786–1793, 2008.

Marbà, N., Jordà, G., Agusti, S., Girard, C., and Duarte, C. M.: Footprints of climate change on Mediterranean Sea biota, Frontiers in Marine Science, 2, 56, 2015.

Marullo, S. and Guarracino, M.: L'anomalia termica del 2003 nel mar Mediterraneo osservata da satellite, Energia, ambiente e innovazione, 6, 48–53, 2003.

McCabe, R. M., Hickey, B. M., Kudela, R. M., Lefebvre, K. A., Adams, N. G., Bill, B. D., Gulland, F. M., Thomson, R. E., Cochlan, W. P., and Trainer, V. L.: An unprecedented coastwide toxic algal bloom linked to anomalous ocean conditions, Geophysical Research Letters, 43, 10–366, 2016.

Merchant, C. J., Embury, O., Roberts-Jones, J., Fiedler, E., Bulgin, C. E., Corlett, G. K., Good, S., McLaren, A., Rayner, N., Morak-Bozzo, S., et al.: Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI), Geoscience Data Journal, 1, 179–191, 2014.

Merchant, C. J., Embury, O., Bulgin, C. E., Block, T., Corlett, G. K., Fiedler, E., Good, S. A., Mittaz, J., Rayner, N. A., Berry, D., et al.: Satellite-based time-series of sea-surface temperature since 1981 for climate applications, Scientific data, 6, 1–18, 2019.

Mohamed, B., Ibrahim, O., and Nagy, H.: Sea Surface Temperature Variability and Marine Heatwaves in the Black Sea, Remote Sensing, 14, 2383, 2022.

Olita, A., Sorgente, R., Natale, S., Gaberšek, S., Ribotti, A., Bonanno, A., and Patti, B.: Effects of the 2003 European heatwave on the Central Mediterranean Sea: surface fluxes and the dynamical response, Ocean Science, 3, 273–289, 2007.

Oliver, E. C., Donat, M. G., Burrows, M. T., Moore, P. J., Smale, D. A., Alexander, L. V., Benthuysen, J. A., Feng, M., Gupta, A. S., Hobday, A. J., et al.: Longer and more frequent marine heatwaves over the past century, Nature communications, 9, 1–12, 2018.

Oliver, E. C., Benthuysen, J. A., Darmaraki, S., Donat, M. G., Hobday, A. J., Holbrook, N. J., Schlegel, R. W., and Sen Gupta, A.: Marine heatwaves, Annual review of marine science, 13, 313–342, 2021.

Pearce, A., Jackson, G., Moore, J., Feng, M., and Gaughan, D. J.: The "marine heat wave" off Western Australia during the summer of 2010/11, 2011.

Pierce, D. W., Gleckler, P. J., Barnett, T. P., Santer, B. D., and Durack, P. J.: The fingerprint of human-induced changes in the ocean's salinity and temperature fields, Geophysical Research Letters, 39, 2012.

Pörtner, H.-O., Roberts, D. C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., and Weyer, N.: The ocean and cryosphere in a changing climate, IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, 2019.

Ryan, J., Kudela, R., Birch, J., Blum, M., Bowers, H., Chavez, F., Doucette, G., Hayashi, K., Marin, R., Mikulski, C., et al.: Causality of an extreme harmful algal bloom in Monterey Bay, California, during the 2014–2016 northeast Pacific warm anomaly, Geophysical Research Letters, 44, 5571–5579, 2017.

https://doi.org/10.5194/essd-2022-343
Preprint. Discussion started: 10 October 2022

Open Access Earth System Science Data Discussions

Sanford, E., Sones, J. L., García-Reyes, M., Goddard, J. H., and Largier, J. L.: Widespread shifts in the coastal biota of northern California during the 2014–2016 marine heatwaves, Scientific Reports, 9, 1–14, 2019.

Schlegel, R. W., Oliver, E. C., and Chen, K.: Drivers of marine heatwaves in the Northwest Atlantic: The role of air–sea interaction during onset and decline, 2021.

Sen Gupta, A., Thomsen, M., Benthuysen, J. A., Hobday, A. J., Oliver, E., Alexander, L. V., Burrows, M. T., Donat, M. G., Feng, M., Holbrook, N. J., et al.: Drivers and impacts of the most extreme marine heatwave events, Scientific reports, 10, 1–15, 2020.

Serrao-Neumann, S., Davidson, J. L., Baldwin, C. L., Dedekorkut-Howes, A., Ellison, J. C., Holbrook, N. J., Howes, M., Jacobson, C., and Morgan, E. A.: Marine governance to avoid tipping points: Can we adapt the adaptability envelope?, Marine Policy, 65, 56–67, 2016.

Smale, D. A., Wernberg, T., Oliver, E. C., Thomsen, M., Harvey, B. P., Straub, S. C., Burrows, M. T., Alexander, L. V., Benthuysen, J. A., Donat, M. G., et al.: Marine heatwaves threaten global biodiversity and the provision of ecosystem services, Nature Climate Change, 9, 306–312, 2019.

Sparnocchia, S., Schiano, M., Picco, P., Bozzano, R., and Cappelletti, A.: The anomalous warming of summer 2003 in the surface layer of the Central Ligurian Sea (Western Mediterranean), in: Annales Geophysicae, vol. 24, pp. 443–452, Copernicus GmbH, 2006.

Stefanon, M., D'Andrea, F., and Drobinski, P.: Heatwave classification over Europe and the Mediterranean region, Environmental Research Letters, 7, 014 023, 2012.

Tanaka, K. R. and Van Houtan, K. S.: The recent normalization of historical marine heat extremes, PLOS Climate, 1, e0000 007, 2022.

Woolway, R. I., Anderson, E. J., and Albergel, C.: Rapidly expanding lake heatwaves under climate change, Environmental Research Letters, 16, 094 013, 2021.