# Author Response to Referee 1 (RC1)

Manuscript No.: **egusphere-2026-2747**

**A Comparative Analysis of Clustering Algorithms for Characterizing Sea Surface Temperature and Kinetic Energy Variability in the Western Mediterranean**

Victor Rodríguez-Méndez[1], Enrico Ser-Giacomi[1], José J. Ramasco[1], Cristóbal López[1], and Emilio Hernández-García[1]

[1]Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB), Campus Universitat de les Illes Balears, 07122 Palma de Mallorca, Spain
Correspondence: Emilio Hernández-García (emilio@ifisc.uib-csic.es)

October 2026

We acknowledge Referee 1 for the careful reading and useful comments. We have tried to answer all of the referee's concerns. In the following we reproduce each of the referee's comments, followed by our response, and the changes to be introduced in the revised manuscript.

As a summary, the major changes introduced in this revision are:

- We have updated our title to *"A Comparative Analysis of Clustering Algorithms for Characterizing Sea Surface Temperature and Kinetic Energy Variability in the Western Mediterranean"* to better specify the scope of our work.
- We have performed a non-parametric bootstrap analysis to confirm that 4 clusters is optimal.
- We have included new Infomap calculations for SST. In this new version we thus provide calculations for both fields SST and KE with the three methods: K-means, SOM and Infomap.

We expect that with this revision the manuscript would be found suitable for publication in Ocean Science.

Emilio Hernández-García,
on behalf of all the Authors.

## Comment 1:

> *"Section 3 and the subsequent sections rely on the selection of an optimal number of clusters (four in this case). This decision is based on the Silhouette Score values shown in Figure 2. However, this figure, together with the supporting analysis, lacks a minimal assessment of statistical significance. As a result, it is not possible to guarantee that selecting four clusters instead of, for instance, nine makes any meaningful difference. Since the remainder of the analysis presented in the paper depends on this particular choice, this issue is highly relevant and undermines the validity of the conclusions. A straightforward solution would be to apply a bootstrap strategy to both the K-means and SOM methodologies. This could be done by iteratively subsampling the 12996 data samples (e.g., 70%), applying the clustering techniques, and computing the corresponding Silhouette Scores. Confidence intervals could*

*then be constructed and added to Figure 2 (and potentially Figure 3), allowing for a more robust selection of the optimal number of clusters, whether it is four or another value."*

**Author Response:**
Reviewer 1 is right in that statistical assessment of the silhouette scores is in order.
We evaluated cluster partition stability using a non-parametric bootstrap ($B = 100$ iterations, 70% random subsampling without replacement, $N_{\text{sub}} = 9{,}097$). K-means and SOM were retrained across all grid configurations ($2 \times 1$ to $6 \times 6$), and 95% confidence intervals were calculated (Figure 1). For SST, the mean silhouette score drops significantly from $0.570 \pm 0.003$ at $K = 4$ ($2 \times 2$) to $0.401 \pm 0.004$ at $K = 9$ ($3 \times 3$). For KE, the identical bootstrap resampling confirms the same structural knee point. Due to the turbulent nature of the surface velocity field, KE yields lower baseline silhouette scores overall ($0.142 \pm 0.016$ for K-means and $0.138 \pm 0.018$ for SOM at $K = 4$). Beyond $K = 4$, scores drop sharply to $0.064 \pm 0.028$ (K-means) and $0.042 \pm 0.025$ (SOM) at $K = 9$. The non-overlapping confidence intervals confirm that $K = 4$ marks an optimal knee point before artificial cluster fragmentation occurs.

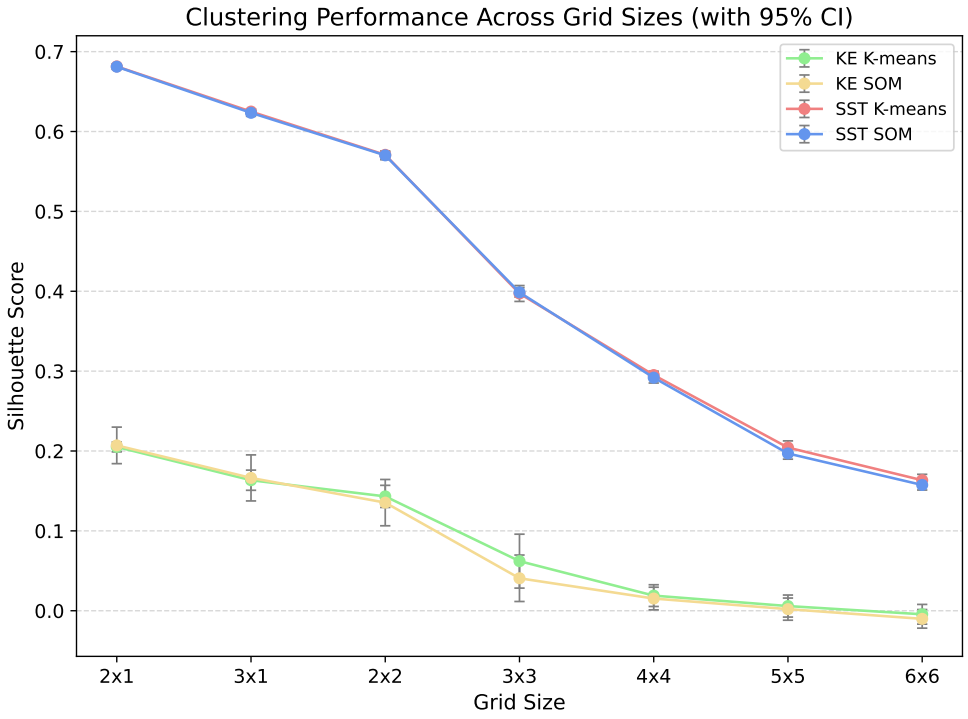



Figure 1: Mean silhouette scores for different clustering methods. Scores for KE and SST using K-means and SOM across varying grid sizes (number of clusters) and topologies. Error bars denote 95% confidence intervals from $B = 100$ bootstrap iterations (70% subsampling). The significant drop in mean silhouette score around four clusters for both K-means and SOM (with $2 \times 2$ topology in this case), particularly pronounced for SST, supports the selection of four clusters for subsequent analyses.

**Manuscript Revision:**
The following text will be added to Section 3.1:

> *"To quantitatively evaluate the statistical significance of the cluster partitions, we implemented a non-parametric bootstrap framework across $B = 100$ iterations with 70% random subsampling without replacement ($N_{sub} = 9{,}097$). For each iteration, K-means and SOM models were retrained across all grid configurations (from $2 \times 1$ to $6 \times 6$), and the resulting 95% confidence intervals were computed (Figure 2). For SST, the mean silhouette score exhibits a sharp drop immediately following the $2 \times 2$*

> *($K = 4$) configuration, falling from* $0.570 \pm 0.003$ *to* $0.401 \pm 0.004$ *for the* $3 \times 3$ *grid. Similarly, for kinetic energy, while the higher spatial variability of the turbulent field results in lower overall silhouette scores (*$0.142 \pm 0.016$ *for K-means and* $0.138 \pm 0.018$ *for SOM at* $K = 4$*), the bootstrap analysis displays a comparable structural drop. Mean scores decline to* $0.064 \pm 0.028$ *and* $0.042 \pm 0.025$ *for the* $3 \times 3$ *grid, respectively. The non-overlapping confidence intervals confirm that this knee point marks a statistically significant shift in cluster cohesion.“*

and Figure 2 will be replaced by the version in this response, which displays the confidence intervals.

**Comment 2:**

> *“The results presented in Section 3.2 indicate that the four identified clusters essentially reproduce the seasonal cycle, characterized by four patterns, but provide limited additional insight. If this is the intended objective, it should be stated explicitly. However, I believe the paper would offer greater scientific value if the methods were applied to data with the seasonal cycle removed. As it stands, it gives the impression that relatively complex techniques are being used to recover “only” a seasonal signal. Assuming a more elaborated approach based on deseasonalized data, the existence of the four pattern cycle suggests that a simple Day-of-Year (DOY) climatology may not be sufficient to remove the seasonal signal. In this context, a PCA retaining two or three EOFs could be an appropriate strategy. In parallel, no consideration is given to potential decadal trends in SST or their influence on the analysis, which should also be addressed.”*

**Author Response:**
We agree with the reviewer's premise: if unsupervised clustering for SST only recovered the trivial domain-wide heating cycle, deploying these methods would offer limited value. However, the analyses in Section 3.2 identify, in accordance with the main objective of our work, persistent spatial configurations and horizontal gradients beyond simple global heating.

We addressed the reviewer's points as follows:

(a) **Spatial Detrending vs. Day-of-Year (DOY) Climatology (Section 3.2.2):**
To test whether the algorithms capture more than bulk warming, Section 3.2.2 applies clustering to SST snapshots after subtracting the daily spatial mean from each map, strictly setting each snapshot's spatial average to zero.
This operation is distinct from substracting a DOY climatology: Our daily spatial detrending ensures that all snapshots have zero mean temperature. Consequently, any emerging clusters arise strictly from the spatial distribution and internal horizontal gradients around that mean. The fact that these zero-mean fields still cluster into four seasonally coherent regimes is a non-trivial physical result: it proves that regional seasonality involves an internal reorganization of thermal fronts (e.g., NW–SE density fronts in winter vs. southern Atlantic water inflows in summer), rather than just uniform heating.

(b) **PCA/EOF Analysis vs. Clustering:** PCA is a dimensionality reduction method, not a clustering technique. It is intended to identify modes of dominant variability. The EOFs are spatial patterns of dominant variability, which are different objects from the persistent spatial configurations we are interested in here. PCA/EOF analysis can be used directly as a clustering tool if projection of the data on the leading EOFs leads directly to clean distinct clusters. Figure 2 shows, for the raw and particularly for the spatially detrended SST fields, that the projection form a single continuous cloud with no immediate groupings into discrete clusters. Linear PCA alone does not separate the data into discrete physical regimes. Of course, one can apply further clustering methods to the data projected onto the

EOFs, but this involves a loss of information, so that in our paper we apply the clustering methods directly on the unprojected spatial snapshots.

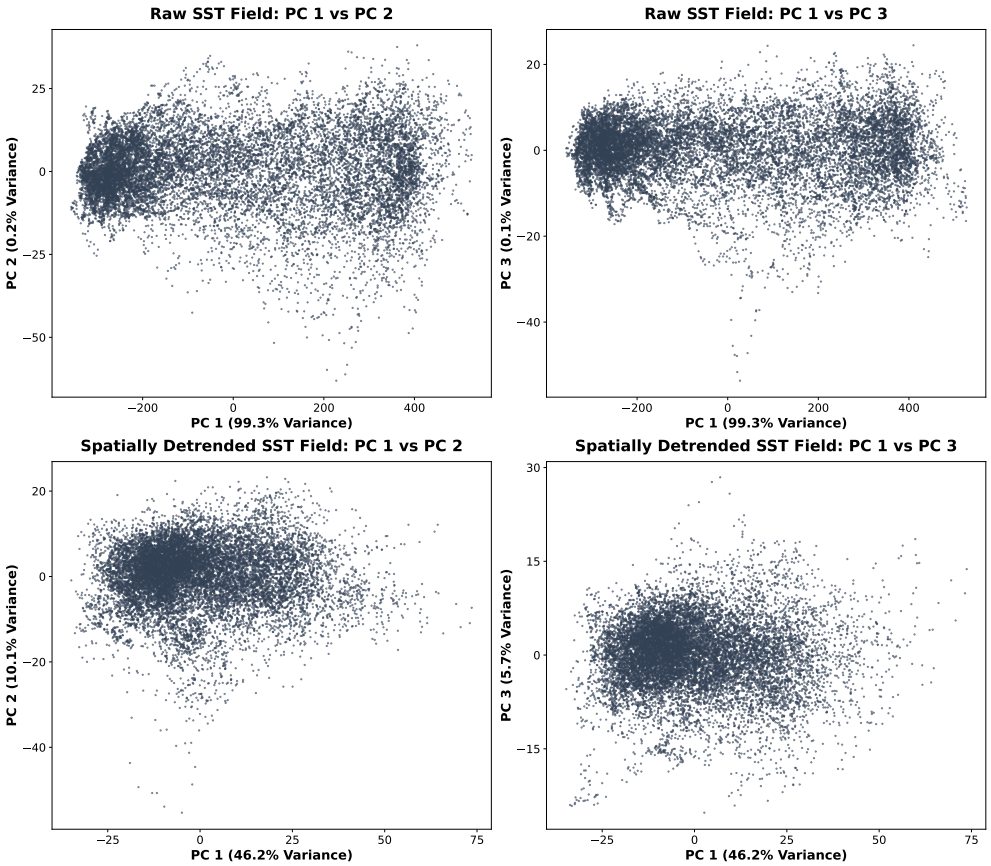



Figure 2: Projections of daily SST snapshots onto leading EOF modes (PC1 vs. PC2 and PC1 vs. PC3). Upper panels: Raw SST, dominated by PC1 (99.3% variance). Lower panels: Spatially detrended SST (PC1: 46.2% variance).

(c) **Decadal trends and Cluster Stability:** Because each daily SST snapshot is detrended individually by substracting its mean temperature, long-term warming does not distort the cluster patterns in our detrended analysis. Although our objective in the paper was to identify dominant regimes in the whole time-series while comparing the different clustering methods, we can additionally use our results to evaluate decadal changes, as the referee suggests. Figure 3 displays the results, that can be summarized as follows:

- Mean SST increased by $+0.201 \pm 0.054\,^{\circ}\mathrm{C}$/decade ($p = 6.84 \times 10^{-4}$).
- This warming shifts regime duration: the summer regime (C2) expands by $+5.6$ days/decade ($p < 0.05$), while the winter regime (C3) contracts by $-7.2$ days/decade ($p < 0.01$).
- Despite these decadal trends, the spatial-structure clusters remain stable. To show this we divided the 1987–2022 study period into two eras of equal duration (Era I: 1987–2004 and Era II: 2005–2022), and recomputed the K-means clusters for each of them. The high spatial Pearson correlation ($r \geq 0.81$) and spatial cosine similarity ($S_c \geq 0.999$) between the clusters in the two eras show that decadal warming alters regime duration, but their spatial patterns remain stable, probably anchored by the underlying topography.

**Manuscript Revision:**
In Section 3.2.2:

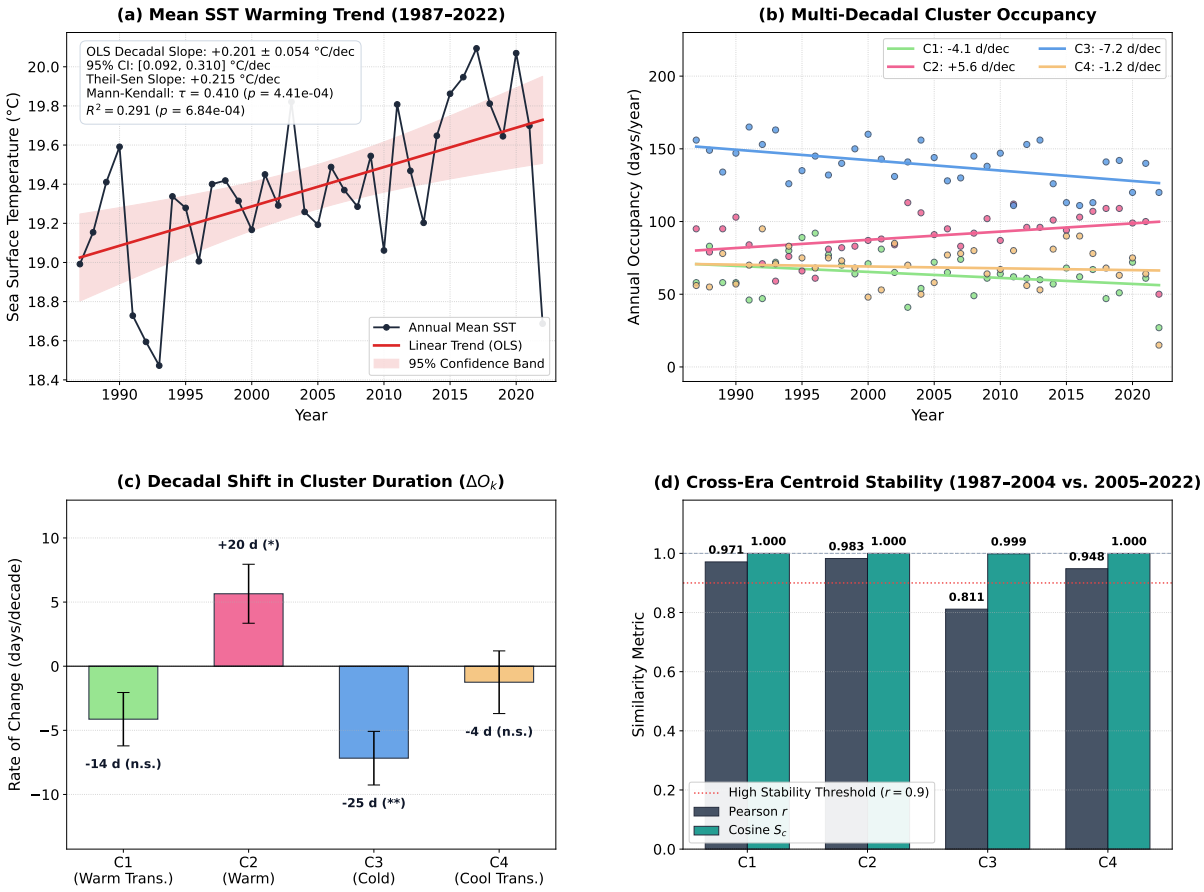



Figure 3: Multi-decadal warming, cluster occupancy dynamics, and centroid morphological stability in the study area (1987–2022). (a) Latitude-weighted annual mean SST with ordinary least-square fit ($+0.201 \pm 0.054\,^{\circ}$C/decade, $p = 6.84 \times 10^{-4}$) and 95% confidence band. (b) Annual occupancy time series for the SST K-means clusters C1–C4. (c) Decadal rate of change in cluster duration ($\Delta O_k$, days/decade) with standard error bars. (d) Cluster centroids similarity between Era 1 (1987–2004) and Era 2 (2005–2022) for the four clusters, via 2D spatial Pearson correlation ($r$) and cosine similarity ($S_c$) across 3,135 the ocean grid points.

> *"Although the raw SST cluster centroids in Figure 4 naturally reflect the dominant regional seasonal heating and cooling cycles, the clustering does not merely group snapshots by spatial mean temperature. As demonstrated by the detrended analysis in this section, subtracting the domain-averaged temperature from each daily snapshot preserves four distinct, seasonally locked spatial patterns. This indicates that seasonal regime shifts in the Ibiza Channel represent coherent reorganizations of internal horizontal gradients rather than just uniform large-scale thermal shifts."*

In section 3.2.1, end of section: :

> *"Over the 35-year study period (1987–2022), the domain-averaged SST exhibits a robust warming trend of* $+0.201 \pm 0.054\,^{\circ}C$ *per decade* ($p = 6.84 \times 10^{-4}$, *linear regression; Theil–Sen estimator:* $+0.215\,^{\circ}C$*/decade,* $p = 4.41 \times 10^{-4}$*). This warming signal is accompanied by temporal shifts in cluster occurrence. By tracking annual cluster occupancy, we observe a significant expansion of the warm regime (C2) at* $+5.6$ *days/decade* ($p < 0.05$, *net* $+20$ *days over 35 years) and a concomitant contraction of the cold regime (C3) at* $-7.2$ *days/decade* ($p < 0.01$, *net* $-25$ *days). Conversely, the transitional states (C1:* $-4.1$ *days/decade; C4:* $-1.2$ *days/decade) exhibit non-significant changes."*

In section 3.2.2:

> *"To assess whether this climatic warming has distorted the spatial morphology of the patterns themselves, we divided the time series into two independent sub-periods (1987–2004 and 2005–2022) and recomputed the detrended SST anomaly centroids. The spatial cross-correlations between the sub-period centroids remain exceptionally high (*$r = 0.971$ *for C1,* $r = 0.983$ *for C2,* $r = 0.811$ *for C3, and* $r = 0.948$ *for C4; all* $p < 10^{-15}$*, with cosine similarity reaching or exceeding* $0.999$ *across all regimes. This decoupling demonstrates that the detected long-term trend in the Western Mediterranean operates primarily as a temporal duration shift—prolonging summer stratification while abbreviating winter cooling—while the spatial configuration of internal fronts remains essentially invariant, possibly because of topographic locking."*

## Comment 3:

> *"The inclusion of the InfoMap clustering technique should be reconsidered. The main concern is that the results are inconclusive (nearly 100% of samples are assigned to a single cluster). It is unclear whether including this method adds any value to the manuscript. Furthermore, it is applied only to one of the two variables (KE) without clear justification. Finally, the implementation details, particularly the choice of threshold (L156 and L323-325), appear arbitrary and lack proper justification."*

**Author Response:**
The objective of our paper is to identify persistent clusters of spatial configurations of SST and KE, and compare the peculiarities of different methodologies in doing so. We considered InfoMap because it is a method based on principles very different from other more standard approaches such as K-means and SOM, and in particular it is known to be free from a *resolution limit* (Fortunato and Barthelemy, PNAS 104, 36–41, 2007), meaning that it can detect clusters of rather small size, which exist in the dataset but remain invisible for most other methods. Our results for InfoMap applied to KE demonstrated this feature and its power as an anomaly detector, which we think is one of the important results in our paper. However, we agree with the referee in that our presentation would be more consistent if we apply the three methods to

the two datasets (SST and KE). In consequence we have performed additional computations of InfoMap for SST, which will be included in the revised version of the manuscript.

In order to keep consistency between the previous and the new InfoMap calculations across the whole paper, some methodological adaptations were needed. First we replaced the previously used cosine similarity by a normalized Euclidean affinity:

$$w_{ij} = 1 - \frac{d_{ij}}{d_{\max}}, \tag{1}$$

where $d_{ij}$ is the pixel-to-pixel Euclidean distance between daily snapshots $i$ and $j$, and $d_{\max}$ is the maximum distance across the entire time series. $w_{ij} \in (0, 1)$ is now the measure of similarity used between spatial configurations $i$ and $j$, both for SST and for KE.

Second, the similarity thresholds $\tau$ above which network nodes (i.e. spatial snapshots) $i$ and $j$ are considered to be connected in the InfoMap network (i.e. when $w_{ij} > \tau$) are now $\tau = 0.88$ for SST and $\tau = 0.59$ for KE. They were chosen specifically to obtain a four-cluster InfoMap partition. This establishes an objective, direct baseline for comparison with the $K = 4$ configurations selected for K-means and SOM from the bootstrap silhouette analysis.

With these adaptations, the results obtained via InfoMap are:

(a) **InfoMap for Sea Surface Temperature:**
InfoMap is now applied to the SST field using a threshold $\tau = 0.88$ to obtain four clusters. The algorithm identifies four balanced seasonal clusters (43.7%, 26.8%, 15.2%, and 14.3%) (Figure 4). These clusters show high concordance with K-means (diagonal Jaccard indices between 0.68 and 0.93) (Figure 6) and, by extension, with SOM. Thus, for SST InfoMap essentially reproduces the results of the other methods.

(b) **InfoMap for Kinetic Energy and Anomaly Detection:**
For KE, we use a threshold $\tau = 0.59$ to obtain four clusters. While the metric and threshold have changed with respec to the initially submitted manuscript, the resulting spatial patterns (Figure 5) remain practically identical to those originally obtained with cosine similarity: a dominant, low-energy background state (C2, 99.88%) and three rare, localized mesoscale regimes, including the southern jet (C4, 0.08%) and eddy structures (C1 and C3, 0.02% each).

**Manuscript Revision:**
We will include in the revised text the above results, the revised methodology, and the rationale for using InfoMap and the chosen thresholds:

In Section 2.4:

> *"In our application of InfoMap to SST and KE, daily snapshots represent network nodes. To evaluate pattern similarity consistently, link weights between nodes i and j are calculated using normalized Euclidean distance: $w_{ij} = 1 - d_{ij}/d_{\max}$, where $d_{ij}$ is the pairwise Euclidean distance and $d_{\max}$ is the maximum distance across the time series. To retain only the most significant connections, we construct the networks applying a threshold of* 0.88 *for SST and* 0.59 *for KE. These values are chosen in order to partition the graph into four community clusters, for direct comparison with the $K = 4$ configurations of K-means and SOM."*

In Section 3.2.4:

> *"To evaluate network-based community detection on SST, we applied the InfoMap algorithm to the SST data using a network link threshold of $\tau = 0.88$. This partitions the SST dataset into four communities or clusters (Figure 4): C1 (*14.3%*), C2 (*26.8%*), C3 (*43.7%*), and C4 (*15.2%*). The spatial centroids recovered by InfoMap closely match*

*those obtained via K-means, confirming that network flow analysis captures the same underlying regimes."*

In Section 3.3.3:

*"The centroids of the clusters of KE configurations obtained from InfoMap are shown in Fig. 5. The threshold used to retain significant links ($\tau = 0.59$) has been chosen to partition the dataset into exactly four clusters. It is known that InfoMap is not significantly constrained by a resolution limit requiring balanced module sizes (Fortunato and Barthelemy, 2007), meaning that it is able to detect relevant clusters even if composed of very few configurations. Indeed, nearly all configurations are assigned to cluster C2, containing 99,88% of them. The centroid of this cluster represents a calm background state with low energy levels and minor maxima along the Balearic Current. Note the similarity of this centroid with that of C3 from the K-means (Fig. 9) and SOM (Fig. 10) methods, and the low-energy levels of this cluster across all three approaches. In contrast, the centroids of InfoMap clusters C1, C3, and C4 isolate rare, highly structured mesoscale features that remain visible without being smoothed out into climatological means. In particular, the centroid in C4 displays an intense southern jet in the region of the Algerian Current and Algerian eddies (Sayol et al. 2013), demonstrating that InfoMap functions as an effective anomaly detector for fine-scale, high-energy circulation events.*
*To further understand the relevance of this 'intense south jet' energy configuration we applied K-means and SOM algorithms to subsets of the whole data set. We find that these two methods clearly identified a distinct cluster showing this feature when applied to the first 500 days of the dataset. This confirms that the configurations clustered by InfoMap as C4 represent real, physically consistent high-energy events rather than an artifact, despite comprising only 0.08% of the total data. K-means and SOM extract only the most dominant configurations, thus hiding this 'south jet' configuration behind the other, most prevalent, energy distributions found. The discrepancy between this clear isolation in InfoMap and its dilution in the full K-means and SOM results highlights the different sensitivities of the algorithms. While K-means and SOM tend to average out such rare, low-populated configurations into broader dominant regimes when applied to the full 17-year dataset, InfoMap acts as an anomaly detector, unveiling these fine-scale and localized dynamics. Similarly, InfoMap's C1 (0.02%) and C3 (0.02%) clusters identify high-energy mesoscale configurations ($> 0.30\ m^2/s^2$) characterized by strong jet-like flows, including intense transport through the Ibiza Channel and northern basin currents, which span over a hundred kilometers and remain obscured in the global averages of the other methods. This underlines the complementary nature of the different clustering methodologies used in this paper, where InfoMap proves essential for identifying the rare, high-energy tails of the oceanographic distribution. A further comparison is performed in the next section."*

In Section 3.4 (Cluster-analysis comparative for SST and KE: Jaccard index)

*"The Jaccard similarity matrices comparing the clustering results from K-means and SOM are presented in Figure 12 for both SST and KE data. In the upper panel (SST), we observe an exceptionally high degree of similarity between the clusters generated by SOM and K-means. The diagonal elements exhibit Jaccard indices close to unity, indicated by the deep purple shades. This near-perfect correspondence confirms that the two methods provide an identical and robust classification of SST configurations into the four seasonal regimes. To further validate network-based community detection,*

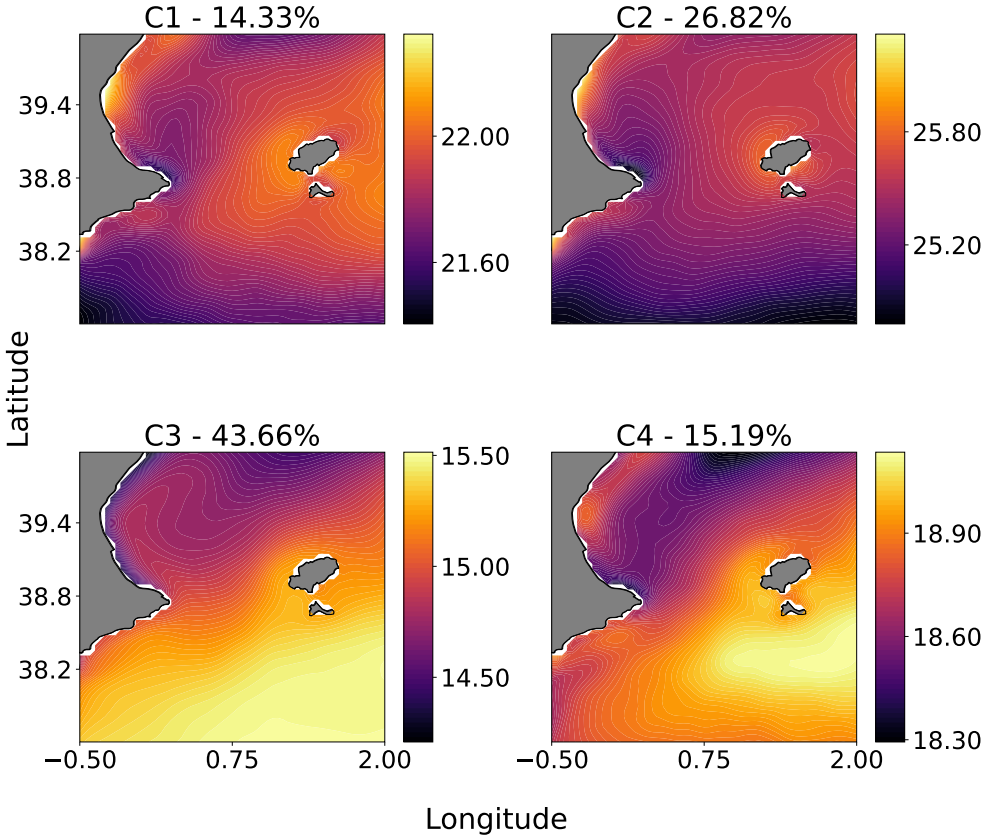



Figure 4: Spatial distribution of SST centroids for clusters identified using the InfoMap community detection method. Each panel shows the average spatial distribution of SST in each cluster (C1–C4), with the percentage indicating the proportion of total samples assigned to that cluster. The color bar indicates sea surface temperature in °C.

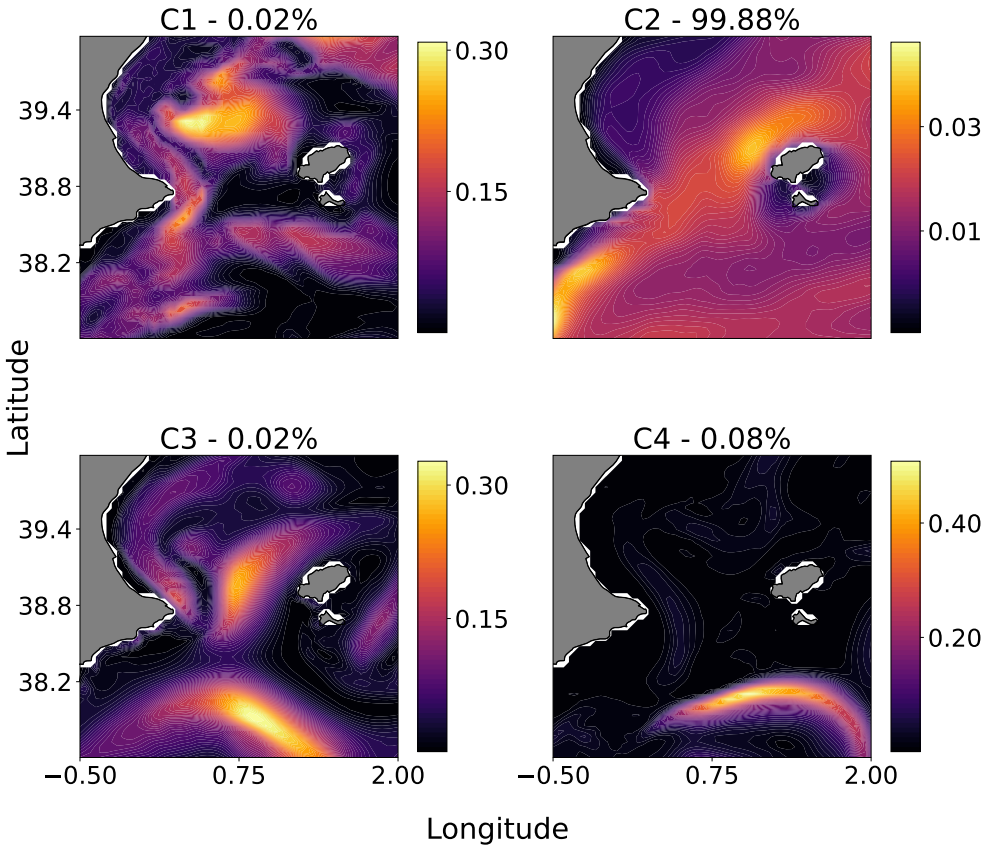



Figure 5: Spatial distribution of KE centroids for clusters identified using the InfoMap community detection method. Each panel shows the mean KE pattern of the populations classified in each cluster (C1–C4), with the percentage indicating the proportion of total samples assigned to that cluster. The color bar indicates the magnitude of KE in $m^2/s^2$.

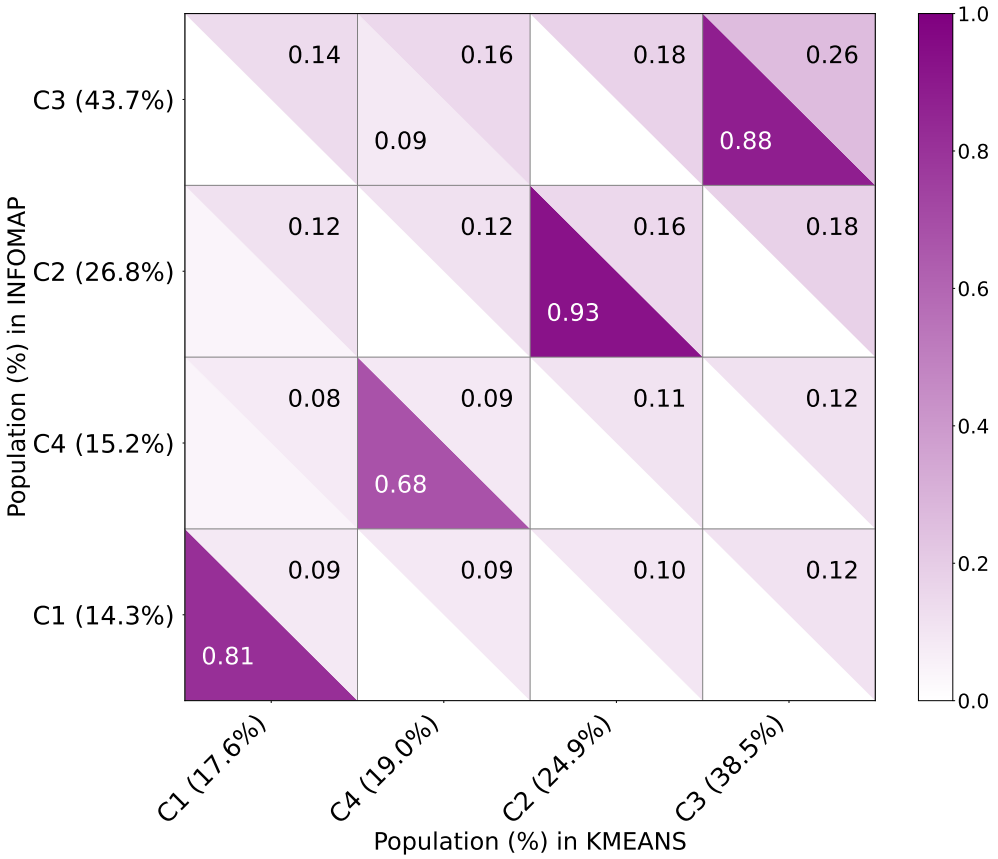



Figure 6: Jaccard similarity matrix comparing InfoMap and K-means clustering for SST. Matrix values range from 0.0 to 1.0, where darker purple indicates higher similarity (overlap). In each cell, the lower-left triangle represents the actual Jaccard index between the two methods, while the upper-right triangle represents a baseline Jaccard index calculated using randomized samples that preserve the original cluster sizes of both the K-means and InfoMap populations. The axes indicate the population percentage for each cluster.

*Figure 6 compares InfoMap against K-means for SST. The diagonal elements show strong Jaccard agreement ranging from* 0.68 *to* 0.93*, confirming consistency between the clustering results of the two methodologies. The lower panel of Figure 12 also reveals a strong consistency between K-means and SOM for KE, contrasting with the potential divergence often expected when clustering high-variability fields. While the Jaccard indices are slightly lower than those for SST, reflecting the higher complexity of the kinetic energy field, the diagonal dominance is clear, with values ranging from 0.66 to 0.87. This indicates that the dynamical regimes identified by SOM (the background state, the Balearic Current, the Algerian Current, and the anticyclonic eddy field) are robustly matched by the K-means centroids. A slightly lower agreement (0.66) is found in the cluster associated with the localized eddy activity (C4 in K-means, 16.8%; C4 in SOM, 13.4%), suggesting that while both methods identify this high-energy feature, the exact boundaries of the partition differ slightly due to the topological constraints of SOM versus the variance-minimization of K-means. Nevertheless, the overall picture is one of convergence, validating that the identified patterns are intrinsic to the physical system rather than artifacts of a specific algorithm.*
*Another observation from the figure is the consistently higher Jaccard indices in the lower-left triangles (representing the actual method comparison) compared to the upper-right triangles. The latter represent the indices obtained by comparing the clusters against randomized clusters that maintain the exact population sizes of the respective populations. The fact that the actual similarity scores exceed the randomized baseline confirms that the detected clusters, in both SST and KE, reflect significant patterns within the data and are not merely due to chance.*
*On the other hand, the clustering results for KE from InfoMap are very different from those of the other two methods. InfoMap assigns* 99.54% *of the configurations into a dominant background cluster, with the rest assigned to rare anomaly states. A Jaccard-matrix comparison of InfoMap KE clusters with the output of the other methods would yield near-zero diagonal values. SOM and K-means find clusters of most common configurations, whereas, if they exist, InfoMap can detect the presence of distinct clusters even when they represent a small minority of the states in the dataset. "*

In addition, Figures 4, 5 and 6 of the present reply will be incorporated into the revised manuscript. See also the response to comment 4 by referee 2, were we further elaborate in the comparison of InfoMap C4 cluster and clusters detected by the other methods in a small temporal portion of the dataset.

**Comment 4:**

*"Based on the current title of the manuscript, one would expect a broader range of clustering methodologies. However, only two to three methods are applied, and no discussion is provided regarding alternative approaches. This issue could be partially addressed by modifying the title. Nevertheless, at a minimum, the manuscript should include a discussion of other available clustering methods. Additionally, it may be worth reconsidering whether the term "Surface Ocean Variability" in the title is appropriate, or if a more specific formulation such as "SST and KE Variability" would be more accurate."*

**Author Response:**
We thank the referee for this comment. The title has been updated to "*A Comparative Analysis of Clustering Algorithms for Characterizing Sea Surface Temperature and Kinetic Energy Variability in the Western Mediterranean*". Also, Section 1 now mentions alternative clustering methods (DBSCAN, hierarchical agglomerative clustering, and Fuzzy C-means) and notes their computational limitations ($\mathcal{O}(N^2)$ to $\mathcal{O}(N^3)$ scaling) and sensitivity to high-dimensional continuous

fields ($M = 3{,}136$, $N = 12{,}996$).
**Manuscript Revision:**
Section 1:

> *"Beyond partitioning and topological neural networks, various unsupervised classification frameworks exist, including density-based spatial clustering (e.g. DBSCAN, Ester et al 1996), agglomerative hierarchical clustering (Ward, 1963) and soft clustering such as Fuzzy C-means (Bezdek, 1981). While density-based techniques can identify arbitrarily shaped clusters without prescribing cluster counts, they struggle in high-dimensional feature spaces where distance metrics become equidistant and require uniform density thresholds ill-suited for continuous geophysical fields. Similarly, hierarchical clustering scales quadratically or cubically in computational complexity ($\mathcal{O}(N^2)$ to $\mathcal{O}(N^3)$), rendering pairwise distance matrices computationally intractable for multidecadal daily synoptic records ($N = 12{,}996$). Consequently, this study focuses on three structurally distinct paradigms: geometric partitioning (K-means), topologically constrained mapping (SOM), and network-flow information compression (InfoMap)."*

### Comment 5:

> *"P3L73: To what extent do the presented results depend on the selected data product? What are the limitations?"*

**Author Response:**
The CMEMS Mediterranean Sea Reanalysis (`MEDSEA_MULTIYEAR_PHY_006_004`) uses NEMO v3.6 and OceanVAR 3D-Var assimilation (Escudier et al, 2021). Its 1/24° resolution ($\sim$ 4–5 km) resolves the regional Rossby radius of deformation ($R_D \approx$ 10–15 km), capturing mesoscale dynamics (Northern Current, Balearic Current, Algerian eddies). Submesoscale processes ($<$ 10 km), internal waves, and diurnal mixed-layer changes are filtered out by the grid resolution and daily averaging. Thus, the data used are an optimal product for comparing how the different methodologies obtain information from surface large- and meso-scale (spatial scales larger than 10 km and time scales longer than several weeks) dynamics in the ocean. Smaller scales would not be well resolved in the dataset, and our clustering methods can not use them to refine the partition of ocean configurations.

### Comment 6:

> *"P6L145: It is not discussed why InfoMap is applied only to KE and not to SST"*

**Author Response:**
In the revised version we apply InfoMap to both KE as to SST. See response to comment 3.

### Comment 7:

> *"P7L156: See general comments regarding the threshold"*

**Author Response:** As explained in the response to comment 3, we choose the affinity threshold to construct the InfoMap networks for SST and KE in order to obtain 4 clusters in each, ensuring a direct comparison with the $K = 4$ partitions of K-means and SOM. This leads to $\tau = 0.88$ for SST and $\tau = 0.59$ for KE.

### Comment 8:

> *"P8L196-199: This is very speculative given the actual results"*

**Author Response:**
The reviewer refers to our argumentation on the choice of number of clusters and their topology. We revised the text to remove speculative reasoning. The discussion is now grounded directly in bootstrap metrics: increasing the SOM lattice beyond $2 \times 2$ ($K = 4$) significantly reduces silhouette scores ($0.570 \pm 0.003$ at $K = 4$ vs. $0.401 \pm 0.004$ at $K = 9$; Figure 1), confirming that larger grids artificially fragment continuous regimes. More details in the response to comment 1, including the text to be included in the revised manuscript.

### Comment 9:

> *"P10Figure4: Is this result expected, given that the seasonal cycle typically captures most of the SST variance? Also see general comments on the seasonal cycle"*

**Author Response:**
While raw SST centroids reflect the annual cycle, clustering does not simply group days by mean temperature. In Section 3.2.2 (Figures 6 and 7), subtracting daily spatial means recovers four seasonal spatial anomaly patterns, showing that the different clusters do not just respond just to the different seasonal global temperatures, but also to reorganizations of internal horizontal gradients.
**Manuscript Revision:**
In section 3.2.1:

> *"Although the raw SST cluster centroids in Figure 4 naturally reflect the dominant regional seasonal heating and cooling cycles, the clustering does not merely group snapshots by spatial mean temperature. As demonstrated by the detrended analysis in Section 3.2.2, subtracting the domain-averaged temperature from each daily snapshot preserves four distinct, seasonally synchronized spatial anomaly patterns. This indicates that seasonal regime shifts in the Ibiza Channel represent coherent reorganizations of internal horizontal thermal gradients rather than uniform large-scale heating or cooling."*

### Comment 10

> *"P10L221: Is this statement correct? Please explain how and why."*

**Author Response:**
K-means cluster labels C1–C4 were identified with the corresponding SOM neurons by maximizing pairwise Jaccard similarity ($J_{ii} > 0.95$; Section 3.4, Figure 12a). The ordinal ranking of the centroids from lowest to highest domain-mean temperature (C3, C4, C1, C2) directly reflects the regional annual cycle, where winter convective mixing and heat loss produce the lowest surface temperatures in C3, summer solar insolation establishes the maximum temperatures in C2, and C4 and C1 capture intermediate transitional states.

### Comment 11:

> *"P12L250-251: How does this compare with the suggested PCA approach?"*

**Author Response:**
See the response to comment 2. In EOF phase space (Figure 2), raw SST is dominated by PC1 (99.3% variance). The spatial EOF configurations are patterns of maximum variability, not the persistent patterns we are interested in here. Subtracting the daily spatial mean removes

spatially uniform heating. In this detrended SST, PC1 variance drops to 46.2%, and Cluster C3 resolves along an orthogonal branch on negative PC2 (10.1% variance). Projection of the data on this reduced space does not produce separate clusters, and data are specially intertwined in the detrended case. Thus, further clustering techniques would be needed to split the data proyected on EOF/PCA space. It is more adequate to apply these techniques to the original, unprojected, dataset, and this is what we do in this paper.

**Comment 12:**

> *"P14Section 3.3: Temporal distributions of clusters are missing in this analysis. Is there a reason?"*

**Author Response:**
Yes. The reason is that, unlike SST, which is phase-locked to the annual solar forcing, kinetic energy is rather driven by turbulence, shear instabilities along the Balearic Front, and episodic shedding of Algerian eddies. These processes operate on synoptic-to-mesoscale timescales without a well-defined seasonal phase, so Day-of-Year timelines do not show regular annual banding. Partition consistency for KE is assessed using multi-decadal Jaccard similarity (Section 3.4).
**Manuscript Revision:**
Section 3.3:

> *"In contrast to sea surface temperature, which exhibits strong seasonal phase-locking driven by annual atmospheric heat fluxes (Figure 5), surface kinetic energy is governed by mesoscale turbulence, baroclinic shear instabilities along the Balearic Front, and the intermittent shedding of Algerian eddies. Because these processes operate on synoptic-to-mesoscale timescales (days to months) that occur irregularly throughout the year, a multi-year Day-of-Year timeline for KE does not yield coherent seasonal bands. Consequently, temporal distribution plots are omitted for KE, and cross-method stability is evaluated using the multi-decadal Jaccard similarity framework (Section 3.4; Figure 12b)."*

**Comment 13:**

> *"P16L308: This statement is speculative ("significant. . . ")'*

**Author Response:**
Qualitative wording has been replaced with the calculated diagonal Jaccard similarity values ($0.66 \leq J_{ii} \leq 0.87$; Figure 12b).
**Manuscript Revision:**
Section 3.3.2:

> *"A comparison with the K-means results reveals strong cross-method agreement, as quantified by diagonal Jaccard similarity indices ranging between 0.66 and 0.87 (Figure 12b)."*

**Comment 14:**

> *"P17Section 3.3.3: See general comments on the removal of this section"*

See response to comment 3. Given that one of the objectives of our work is to compare the strengths of different clustering methods, we find very appropriate to discuss here InfoMap, which does not suffer from the *resolution limit* of other methods, meaning that it can detect

clusters with very different occupations. We have included in this revision the analysis of SST by InfoMap, which does not give results substantially different from the other two methods tested. But for KE the power of InfoMap in detecting real but not very frequent energetic events is revealed. See also response to comment 4 of referee 2, where we identify the energetic InfoMap C4 cluster as representing a situation occurring in the 500 first days of the dataset.

**Manuscript Revision:**

In section 3.3.3:

> *"In contrast to K-means and SOM, which distribute snapshots of KE across four balanced regimes, InfoMap assigns 99.54% of the observations to a single dominant background cluster (C2). The remaining three clusters isolate rare, localized, high-energy mesoscale configurations: C4 (0.31%), C1 (0.11%), and C3 (0.05%). This distribution reflects the distinct optimization objectives of the algorithms. K-means and SOM minimize global within-cluster variance, tending to absorb low-frequency states into broader clusters. InfoMap minimizes the description length of information flow, allowing it to detect communities without being constrained by cluster-size balance. InfoMap thus serves as a targeted anomaly detector for rare but structurally coherent circulation events."*

## Comment 15:

> *"P18L347-349: "So "rare" events are "similar" enough to be classified into specific cluster(s)? Wouldn't their nature place them outside any cluster?"*

**Author Response:**

Random noise snapshots would form isolated singletons. In contrast, extreme high-energy events represent recurrent physical states (e.g., the pulse of the southern Algerian Current jet, which can be identified in the dataset during the first 500 days) with consistent spatial structures. They are rare, and invisible to some clustering methods, when compared against the full 12996-days dataset. But in the network construction used by InfoMap, these snapshots form internally connected subgraphs that trap random-walk probability flow, enabling InfoMap to identify them as forming a distinct community.

**Manuscript Revision:**

In section 3.3.3:

> *"The grouping of infrequent states – rare events – by InfoMap is governed by the network constructed from the similarity between states. Community assignment – clustering – is determined by minimizing the 'Map-Equation description length' (Rosvall and Bergstrom, 2008). Random configurations arising from stochastic noise would yield disconnected network nodes. In contrast, the extreme high-energy states observed here represent recurrent hydrodynamic modes. Because these events recur with consistent spatial geometries, they form an interconnected topological module with high internal link density. This topology traps the random-walk probability flow generated by InfoMap, enabling this method to isolate these coherent dynamical regimes rather than scattering them as disconnected noisy configurations."*

## Comment 16:

> *"P21L401: This is difficult to sustain given the limited number of techniques applied and the specific results obtained with InfoMap"*

**Author Response:**

The conclusion has been revised to avoid ranking algorithms, describing instead their complementary uses: geometric partitioning (K-means/SOM) for dominant regimes, and network community detection (InfoMap) for rare anomaly detection.

**Manuscript Revision:**

Section 4:

> *"Rather than establishing a universal ranking across algorithms, our findings demonstrate the complementary roles of distinct clustering frameworks for oceanographic feature identification. Geometric partitioning via K-means and topologically constrained mapping via SOM provide consistent classifications of dominant, high-occupancy climatological regimes and their seasonal transitions. Conversely, network-flow community detection via InfoMap provides a targeted framework for anomaly detection, capturing rare, structurally coherent circulation events that reside in the heavy-tailed extremes of the kinetic energy distribution."*

### Comment 17:

> **Referee Comment:**
> *"P21L404-406: Specify what these implications are"*

**Author Response:**

Section 4 now mentions applications in operational monitoring, ecological management, and climate tracking.

**Manuscript Revision:**

Section 4:

> *"This framework supports practical applications such as real-time regime classification, monitoring high-energy oceanographic events, characterizing physical conditions relevant to fish spawning and larval retention, and tracking regional climate change through annual shifts in seasonal regime durations."*