Response to Editor Gabriele Messori

I have received two very discrepant Reviewer reports, one quite critical of your submission and the other recommending publication. Because of this, I have asked for a third Reviewer opinion, which suggests a number of updates before further consideration of your manuscript. Based on this, I am returning the manuscript to you for a further round of major revisions. Ensure that you provide detailed responses to both sets of Reviewer comments. Note that I would normally not consider further a submission requiring a third round of major revisions.

Authors' response:

We thank the Editor for this opportunity to revise our work. Regarding the comments of Reviewer 2, we have made our best effort to clarify that

- Not all changes in the methodologies used in ERA5 and NOAA translate in changes in SST anomalies that can be detected by our method.
- Our method only takes partial information (by considering relative but not absolute values), therefore, it can not extract all possible information nor it can detect all possible changes in ERA5 and NOAA. This is the usual situation that researchers in the field of "data analysis" are very familiar with: every dataset has its own difficulties and no analysis tool is complete, to obtain more complete information one should test diverse methods.
- Our method provides limited results (it is obviously unable to detect every change in ERA5 or NOAA) but it provides consistent, statistically significant and robust results, as demonstrated by extensive additional information that is included in the appendices.

Regarding the comments of Reviewer 3, we thank the reviewer for these comments that have allowed us to improve clarity of our work: we have completely revised the Section Methods, including a figure to schematically illustrate the steps of spatial ordinal analysis (Fig. 2). We thank the reviewer for the suggestion of extending the choice spatial orientations of the ordinal patterns. Indeed, this is fully in line with the flexibility of the ordinal approach: one can select the spatial scale (by choosing the pattern length and lag between grid points, L and delta respectively), and the orientation of the patterns (in this work, we use WE and NS because they are particularly fit for the equatorial dynamics of the Nino 3.4 region and they are also relevant for the Gulf Stream region). Of course, other orientations are possible and they can be even better for the Gulf Stream region; however, such a study is out of the scope of the present work.

We hope that the revised manuscript can be considered suitable for publication.

Sincerely yours,

Juan Gancio on behalf of all the authors

Response to Referee 2

The manuscript exposes the results of the computation of the Spatial Permutation Entropy (SPE) on two SST datasets (ERA5 and NOAA OI v2) over two regions (El Nino and Gulf Stream). The SPE is computed from "spatial patterns", which encode how neighbouring pixels values compare against each other, and the SPE is defined as an entropy over the distribution of these patterns. Spatial Mutual Information (SMI) on the distribution of patterns are also considered. I thank the authors for the precisions added with respect to the previous versions of the manuscript, which made some explanations clearer. I think in particular that the videos showing the evolution of the histogram of the patterns are useful to interpret the variations of the SPE.

Authors' response: We thank the reviewer for their positive feedback.
There are three ways in which the SPE can be used:
1. It can be used to detect transitions in a dataset, either through change points in H_{NS} and H_{WE} (Sec. 4.1), or through SMI_{NS} and SMI_{WE} (Sec. 4.2, 2 datasets required)
2. Variations of the SPE can be interpreted in terms of change of relative importance of spatial patterns, which the authors interpret as increase or decrease of the gradient patterns (Sec. 4.1)
3. Characterize the similarity of two datasets (how much the distribution of patterns is the same for the two datasets, Sec. 4.2)

Authors' response: While we agree in general with these three ideas, we point out that

1. SPE is a tool that can detect some transitions but not all, because, as it is done in many analysis tools, only takes into account partial information (relative data values).
2. Variations of SPE are due to changes in the probabilities of the patterns. Often but not always, captures the increase or decrease of the probabilities of the patterns that capture encode gradients.
3. SPE can characterize and quantify the similarity of two datasets in those aspects that can be captured by the patterns, that is, that do not take into account absolute data values.

This underlines the potential versatility of the tool, as emphasized by the authors in the manuscript. However, I think that concrete and pragmatic results (quantitative and qualitative) are lacking for the SPE to be used by other researchers in their work on different datasets.

Authors' response: We do not understand this comment about the lack of "*concrete and pragmatic results"*. We are convinced that other researchers can apply this tool on different datasets and will also obtain concrete information (they might find expected or unexpected differences when comparing datasets, or find expected or unexpected changes in the data).

Main comments:

About use 1. of the SPE: it is shown in Sec. 4.2 that the transitions detected by PELT on SPE time series are not all detected by PELT on SMI_{hist}, and none are detected in the time series of the Pearson's spatial cross-correlation coefficient (r) and of the Average Absolute Difference (AAD). I think that this is the major clue of the manuscript in favour of the SPE for detecting transitions. The contrast of SMI_{hist}, the AAD and r with the SPE is otherwise quite poor, so that it is not clear whether these tools could be complementary to the SPE (see below). The comparison of the SPE with the AAD and r is done when using the two datasets to detect transitions, and no equivalent comparison is done in the much more common case where only one dataset is available for the variables of interest.

Authors' response: The reviewer's comment "*The contrast of SMI_{hist}, the AAD and r with the SPE is otherwise quite poor, so that it is not clear whether these tools could be complementary to the SPE (see below)"* is answered below. Regarding the second point, "*no equivalent comparison*", in the new revised manuscript we modified Fig. 3 to include two new panels (for Niño and Gulf Stream regions) where we display H_{hist} of the two datasets.

I also wonder whether the AAD and r are the best choices to provide a point of comparison. It seems from lines 338-339 that techniques to detect transitions already exist in the geophysical literature, why did the authors not use the methods of these papers to provide points of comparison?

Authors' response: We do not know which are the "best choices" but *AAD and r* are well-known and widely used measures, and they are very natural choices to measure the difference of two spatial fields

Accordingly to the remarks added by the authors in this new version of the manuscript, there is no guarantee that the SPE detects all change points (there are actually some change points which were detected in some configurations of the SPE but not in others) so that the authors propose to use this tool in conjunction with other data analysis techniques.

Authors' response: This point was already explained in the first version of the manuscript: ordinal analysis takes into account partial information; therefore, it will capture some but not all change points.

Unfortunately, the authors do not suggest which other tools could be used in complement to the SPE. It is left to the user to further understand the weak points of the SPE to identify which tools should be used in addition to the SPE.

Authors' response: This is usually the case for researchers in the field of data analysis. Each dataset comes with its own challenges and difficulties and no analysis tools is able to extract complete information. Also, no analysis tool is appropriate for every dataset. However, in the original version of the main text (line 349 of the revised one), we did mentioned some complementary analysis. In this version we specified the reason: since the main information lost by the SPE concerns the absolute values of the data, complementary techniques should likely focus on that. The additional possible example of spatial Fourier analysis was also included.

A way of addressing this issue would be provide a quantification of the transitions detected.

Authors' response: This quantification is presented in Table A1.

Such tests can be done by generating synthetic datasets with known transitions.

Authors' response: SPE performance depends on the data, therefore, we could generate synthetic datasets with known transitions where SPE performs perfectly. Here we demonstrate SPE performance in two real datasets.

I think that, yet another possibility would be to count the fraction of transitions detected by the SPE in the ERA5 and in the NOAA datasets, after having listed all the changes in the methodology to produce these datasets (it seems reasonable to me that an almost exhaustive listing is doable since I expect datasets like ERA5 to be well-documented).

Authors' response: The fraction of transitions detected and missed depend on the significance threshold used, which can be more or less strict. We could of course try to "manipulate" this threshold to maximize the number of transitions that can be traced back to known changes in methodology, or to eliminate the apparently "spurious" transitions (those that cannot be traced back to known changes in methodology). Here we do not take that approach and use a threshold determined from surrogate data and from robustness analysis (as explained in lines 345-350 of the previous revised manuscript) to demonstrate that SPE is a useful "data-based tool" able to return relevant information, directly from the data (without using extra information). Capturing every change or modification that ERA5 has experience, bearing in mind that ERA5 integrates many sources from in situ and satellite observations, it is outside the scope of our work.

With synthetic datasets, one could also investigate which type of transitions is detected by the method and which are not.

Authors' response: We agree but it is outside the scope of the present work.

About use 2. of the SPE: the authors interpret variations of the SPE as increases or decreases of gradients in the SST. I think that this is an interesting potential use, but I think that the caveat for such a use should be refined. Indeed, as explained at lines 221-230, some low values of the SPE with delta = 8 in the NS direction of the El Nino region are not explained by gradients of the SST. It can be checked explicitly on the videos showing the histograms of patterns for ERA5 on this region that there are a lot of histograms which are not dominated by the patterns 0123 and 3210 (for example on 1998-06, 2002-11, 2010-03, 2010-05, 2015-06, 2015-10, 2025-01). I think that this is a nice example that a low SPE does not automatically mean that the 0123 and 3210 patterns dominate and that using the SPE in such a way would require more investigation to be able to confidently draw conclusions on datasets for which we do not whether there are gradients or not.

Authors' response: We agree and to further stress this point, in the second revised manuscript we clarify (page 11): "However, low SPE values do not always imply that the 0123 and 3210 patterns dominate, and a careful inspection of the patterns' probabilities is needed in order to be able to confidently draw conclusions."

About use 3.: the authors finally use the SPE to compare the two datasets. Again, I find the conclusions somewhat vague. There are two conclusions: a) the datasets become more similar with time, b) the datasets are more similar at large scales.

Authors' response: We do not understand why these conclusions are somewhat vague to the reviewer because they are quantified in Fig. 7 (increase of the spatial mutual information with time) and in Fig. B13 (increase of SMI with the spatial lag, it was Fig. C10 in the previous revised manuscript).

I think that conclusion a) is to be expected, given the "significant advances in Earth observation systems due to the introduction of new satellite observations and new data processing methodologies" (lines 293-294). Stated in this way, conclusion b) should also be expected.

Authors' response: While the conclusions can be expected, differences remain as discussed in the text and clearly seen in Figs. 3b, 3d.

It would interesting to be able define a scale (possibly depending on the user's needs) where the datasets agree sufficiently (this scale would probably depend on time, since the datasets are more similar with time). I think that Fig. C10 provides an interesting point to start this analysis, and I think that developing this in the main text would support this way of using the SPE. Characterizing qualitatively the differences in the datasets could also be interesting, so that users could choose one or the other depending on the processes they study.

Authors' response: We agree that this is an interesting application and in the second revised manuscript we discuss Fig. C10 in the main text (now Fig. B13). While very interesting, a detailed study is out of the scope of the present work.

Minor comments:

I do not understand the explanation about the sudden drops of H_{WE} for delta = 8 in the lines 217-221: why uneven cooling/warming explain sudden drops? What does exactly mean "at the smaller scale the variations of SST are more uniform" (lines 220-221)? Is there a reference for that? Or is does it just mean that the SST values are very noisy at these scales?

Authors' response: Sudden drops are seen in years 1982, 1997 and 2008. In these years (as can be seen in Fig. B1), a WE gradient is formed due to the warming of the easter Pacific (this usually corresponds to a canonical El Niño, like in 1982 and 1997, but is also observed in 2008 at the end of a strong La Niña), which dominates at this long spatial scale, and in consequence, pattern "0123" prevails (thus, the entropy drops). In contrast, for $\delta=1$, the distribution of ordinal probabilities presents larger contributions of two patterns, "0123" and "3210", so the entropy is higher (see Fig. 5). We have re-written the explanation to clarify this point (page 12).

What can we conclude about the fact that there are correlations between the SST anomaly and the SPE for the El Nino region (lines 207-216), especially given that no correlation is found in the Gulf Stream region (lines 231-233) ?

Authors' response: In the revised manuscript we expanded the analysis of the correlations between SST anomalies and SPE in both regions, including a new figure (Fig. B8) that shows the temporal evolution of the entropies and SST anomaly in the Gulf Stream region, and a new table (Table 2) that displays the cross-correlation coefficients. From the correlation coefficients we can conclude that the agreement between the datasets (in terms of the magnitude and sign of the cross-correlation coefficient) is good in El Niño region, but not in The Gulf Stream region, and we speculate that such disagreement can be due to the string mesoscale variability in monthly SST induced by the Gulf Stream meandering (as seen in Fig. B2) in comparison to SST anomaly in El Niño region.

What can we conclude about the fact that a transition in 2016 is detected in SMI_{NS} with delta = 1 in the El Nino region (lines 251-252)? Is this related to some change in the datasets? Why was it not detected from Fig. 3?

Authors' response: The 2016 change point is also detected in the Gulf Stream region in SMI_{WE} with delta = 1 (see Fig. 7f). In both regions there is a rather small increase of SMI at 2016, which reveals a small increase in the consistency of the two datasets. Regarding Fig. 3, entropy variations are seen by eye inspection, and many of them are detected as "change points" by PELT algorithm (including variations in the entropies that occur at 2016); however these "change points" are considered not robust because a small decrease (or increase) of the penalty parameter results in a large increase (or decrease) of the number of change points detected.

Are the transitions reported in lines 259-264 already found in Sec. 4.1? If no, does that mean that it is better to have two datasets to compare to detect transitions in one of them, rather than computing the SPE on one dataset?

Authors' response: Not all changes in SPE (Sec. 4.1) are detected in SMI (Sec. 4.1) and vice-versa. Therefore, having two datasets to compare allows to calculate SMI, and SMI can yield information that complements that obtained from SPE analysis of a single dataset. This is because SPE and SMI encapsulate different information. SPE contains information about the regularity of the spatial structures of one field (spatial gradients) while SMI measures the similarity of the local structures of two fields.

I find Appendix A confusing:
1. If I understand correctly, the CPD algorithm used for the SMI is described in lines 352-361, while the one used for the SPE, the AAD and r is described at lines 349-350 and 362-376. Is that correct?

Authors' response: Yes – the algorithm is always the same, but the preprocessing of the signal, and the significance tests are different.

2. Are surrogates created for the SMI? I would think so from lines 341-347, but not from lines 352-361.

Authors' response: The change points detected in the SMI signals all correspond to changes in the linear trends, and thus, their significance can be tested with the Wald test. Therefore, there was no need to use surrogates of the SMI signals.

3. I do not really understand the point of the steps described in lines 368-376: from what I understand, the previous step allows to identify a value of P for which no false change points are detected, so why add another step?

Authors' response: In the "previous step" we identify a value of P by comparing changes in the signal with changes in the surrogates generated from the signal (changes in the surrogates provide a threshold of the penalty parameter, P, and in the original signal we consider "significant" only the changes that are detected with a penalty parameter higher than the threshold penalty). The additional step is to test the robustness of the changes with respect to variations of the penalty parameter.

4. Why make a difference between the points mentioned in the main text and those reported in Table A1? If change points are considered robust (and therefore reported in Table A1), why not consider them in the text?

Authors' response: The change points reported in Table A1 are all significant, but they are not all equally robust. In the text we discuss the most robust ones (R>=19). We remark that it is not the goal of this work to identify or characterize change points, but to demonstrate that spatial ordinal analysis and permutation entropies are analysis tools that provide useful information from spatio-temporal climatic data. Therefore, in the text, we limit the discussion to the most robust change points found. We also remark that 1) we do not claim that all the detected change points that are significant and robust are also understood (in the sense that they can be interpreted as due to some known change in the data) and 2) we do not claim that our method (i.e., combining spatial ordinal analysis with a change point detection algorithm such as PELT) can identify all change points.

5. The first two quartiles of P are the same than those of R (Eq. (A1)), so the lines 368-376 seem to simply describe that half of the change points are considered robust, is that correct? Since change points are supposed to correspond to something real, it is quite arbitrary to consider that half of them.

Authors' response: It is not correct because the first two quartiles of P and R are not the same because \bar{P^*} changes from signal to signal (determined from the analysis of surrogates of each signal) while the distribution of R is evaluated from all the detected change points that are significant, across all the signals. Regarding "*change points are supposed to correspond to something real*", in fact, it is arbitrary what is considered a "change point" and what is not, because this depends on a penalty threshold that can be selected arbitrary. In that sense, a small fluctuation can be a "change point" if the penalty threshold is low. This is why we use "surrogates" to determine the penalty threshold for each signal. To take this drawback into account, we perform a "robustness" analysis and keep only the change points that are robust with respect to variations of the penalty parameter.

Are there other methods to choose a suitable penalty parameter than the one described in Appendix A ? The original paper about PELT seems to expose some of them and Rocha and de Souza Filho (2020) (cited at line 339) seems to discuss methods to choose penalty functions. Why did the authors not use these functions? To have a better idea of the performance of the SPE to detect transitions, I think that it would be better not to use new methods to choose P.

Authors' response: Yes, there are several criteria to choose an appropriate penalty values, as well as several search algorithms and cost functions. The methodology used in our article is very similar to the one mentioned in Souza Filho (2020) as being the best performer of their study: CROPS. As in CROPS, we evaluate a large range of penalty parameters. However, CROPS is an algorithm for a fast and computationally efficient evaluation of this range of penalty values, but ultimately relies on manually choosing the optimal penalty value (usually graphically) for which perturbations of its value do not alter significantly the number of change points detected. Our approach is intended to avoid this kind of subjectivity in the implementation; thus, we provided a clear and quantifiable criterion for the selection of the penalty value. Additionally, our approach selects the most significant change points (those that are detected for a penalty values larger than a threshold), as an attempt to minimize false detections.

Technical comments:

Appendix B seems to have redundant explanations with Appendix A, please merge the two.

Authors' response: Thanks, both appendices have been merged.

I find the notation to report coefficients and p-values in the caption of Fig. 4 not very clear.

Authors' response: We have modified the caption to clarify the notation.

Fig. C10: is there a reason why the results displayed were computed with L = 3 instead of L = 4 as in the rest of the paper?

Authors' response: Fig. C10 is now Fig. B13. We use L=3 to remove any possible effect due to lower statistics when the spatial lag is large. However, qualitatively similar results we found with L=4, as we show in the new Figs. B14.

Line 100: k is not an integer, so k \in [1, ..., L!] is not really correct

Authors' response: k refers to the label of the ordinal pattern and we have re-written this sentence in the following way to clarify this point:

Labelling the ordinal patterns (symbols) of length L from k=1 to k=L! (for L=3, k=1 corresponds to pattern 012, k=2 to pattern 021, etc., as illustrated in Fig. 2c), then, n(k) is the number of times the kth symbol appears in the symbolic sequence.

Line 154: "we have also performed the analysis…" ("the" is missing)

Authors' response: Thanks, corrected.

Line 324: "Both the size…" (no comma after "both")

Authors' response: Thanks, corrected.

Line 416-417: the sentence "The first one corresponding to a change point with linear trend before and after, which survives detrending, and the second one to a trend/no-trend transition" misses verbs.

Authors' response: Thanks, corrected.

Lines 189-198: The p-value of the coefficient from the fit in panel (a) of Fig. 4 implies that the coefficient can be considered to be 0. But this paragraph is a little misleading about that (it seems to say that all entropies follow the same kind of trend).

Authors' response: Thanks, we have corrected the sentence.

Line 385: "panel a" → "panel (a)", same for "panel d"

Authors' response: Thanks, corrected.

Lines 221-230: it would maybe be easier to understand if it is said explicitly that the patterns with delta = 8 span more than half of the length in the NS region for the selected region.

Authors' response: Thanks, we have modified the sentence accordingly.

Lines 254-255: "at long scales, warming signals are consistently identified in both, ERA5 and NOAA": how exactly are identified the warming signals in Fig. 6? Is this related to what is discussed about Fig. 4?

Authors' response: We are referring to the linear trends in the temporal evolution of the entropies and we have modified the sentence to avoid confusion.

Lines 264-265: it should be said the transition here is in addition to the 2007 one reported just above.

Authors' response: We have modified the text to clarify this point.

Line 305: what is meant by "that are consistent with the two datasets"?

Authors' response: We have modified the text to specify that the same trends are found in both datasets.

Line 350: "ADD" → "AAD"

Authors' response: Thanks, corrected.

Line 355: "non" → "none"

Authors' response: Thanks, corrected.

Line 362: "ADD" → "AAD"

Authors' response: Thanks, corrected.

Response to Referee 3

General comment:

The authors present a multi-scale approach founded on permutation entropy metrics and change point analysis to identify and characterise changes in spatiotemporal datasets with focus on the Nin˜o 3.4 and Gulf Stream regions for two SST datasets, the ERA5 SST and NOAA OISST v2 products (anomalies SSTA and/or detrended anomalies dSSTA?). They recover known discontinuities or transitions in both datasets indicating that the methodology is potentially capable of discovering potential issues due to 1) changes in the Earth's observing system with the introduction of new observing platforms (e.g. MeteOp-A to MeteOp-C) and 2) changes driven by modifications of the data analysis workflow (e.g. OSTIA error covariance update). The authors also apply spatial permutation entropy (SPE) to tease out quantitative differences between datasets and compare SPE feature discovery to standard distance metrics (e.g. the average absolute and correlation distances). Adding SPE to the standard suite of diagnostics to characterise changes in widely used data products is important as it 1) will increase confidence in the data products for specialised use cases such as extreme event detection and regime shift identification, 2) could alert practitioners to potential issues with their workflow if the SPE methodology is part of regular data quality protocols, and 3) may allow potentially new insight in understanding geophysical processes and their impact.

Authors' response: We agree with the reviewer that adding SPE to the suite of diagnostics tools available will have the (potentially important) applications suggested by the referee.

Major issues

1. To make the paper more accessible to practitioners in the weather and climate research, the different methodologies for PE, SPE and SMI should be clearly explained, especially in section 3, where the definition of the Shannon entropy H or PE, Equation 1, should carry all the necessary indices and time dependency and all expressions derived from it. Contrasting the temporal, spatial and spatio-temporal PE approaches and illustrating them with simple examples would also be instructive and would add clarity to the expose even if these have been shown elsewhere. Clear distinctions between symbol length, total sequence length and the probability of expression of different but a finite number of patterns for each L and δ. Ditto for a precise definition of the joint entropy term in Equation 4.

Authors' response: In the second revised manuscript we clarify these definitions and we have re-done Fig. 2 to include simple examples to illustrate the difference between the spatial and temporal PE approaches. We have also explained in more detail Eq. (1), used a more explicit notation for Eqs. (2) and (3), as well clarified how the number of spatial OP depends in the dimension of the region and the L and δ parameters. A new Eq.(7) explains the calculation of the joint entropy.

2. Following on point 1, targeted examples for both El Niño and La Niña snapshots in SSTA or detrended SSTA (or large meander excursions of the Gulf Stream) for example could be valuable.

Authors' response: We also include in the new Figs. B1 and B2 (in Appendix B) targeted examples for El Niño and La Niña snapshots as well as snapshots in the Gulf region.

3. Clarifications should also be given regarding how $AAD(t)$ and the Pearson's spatial crosscorrelation $r(t)$ are computed.

Authors' response: In the second revised manuscript we clarify that no subsampling is done of Xij and Yij. We remark that $\delta>1$ does not imply any subsampling as all the datapoints are used to define ordinal patterns, but with $\delta>1$ patterns are defined in term of data points that are not neighbors. For example, for $L=3$ and $\delta=2$, a pattern is determined by (x1, x3, x5), another pattern is determined by (x2, x4, x6), another by (x3, x5, x7), (x4, x6, x8) etc.

4. The choice of spatial ordinal pattern orientations, North-South and East-West, for the Gulf Stream box is probably harder to justify than for the Nin˜o 3.4 region, where asymmetries along these directions are relatively well understood given what we know of the equatorial dynamics of ENSO and associated phenomena – Kelvin and Rossby waves propagation, meridional mode and tropical instability waves. Have you tried to check if directions along and across the mean position (approximately SW to NE and NW to SE) of the Gulf Stream current system changes your results? Either by sharpening the points already discovered (i.e. 2007 and 2013) and possibly making the 2021 shift significant!

Authors' response: We thank the reviewer for this suggestion, which points at exploiting the versatility of spatial ordinal analysis. The study the reviewer suggest is a natural extension of the present work, as explained in our previous responses. However, the goal of the present work is not to identify or characterize change points, but to demonstrate that spatial ordinal analysis and permutation entropies are analysis tools that provide useful information from spatio-temporal climatic data. Considering patterns with other orientations (and also, geographical regions with other orientations) will be a natural continuation of the present work and we have now included a comment about this point in the conclusions.

Minor issues

Are all the results presented derived from sea surface temperature anomalies only? If so, a clear statement should be made upfront in the data section.

Authors' response: We have clarified this point. Now the Data section starts as: "We consider only monthly SST anomalies with respect to the seasonal cycle…".

In Figure 1, the spatial standard deviation could be slightly darker for both Niño 3.4 and the Gulf Stream. Its time evolution and variability as function of ENSO phases may explain some of the behaviour seen in the SPE metrics.

Authors' response: We have darkened the spatial standard deviation in Fig.1. Spatial ordinal analysis detects nonlinear spatial correlations (specifically, the spatial ordering of the relative values of SST anomaly), which are not detected when analyzing the evolution of the spatial standard deviation nor in its variability, because both remain unchanged when the datapoints are spatially shuffled.

The parameters L and δ and labels indication the regions should be appear directly on all subplots as this would make the plots more readable, or indicated in left for rows or top margins for columns.

Authors' response: We have labeled every subplot with all the information required to read the plots.

The symbol length, here L=4, should be mentioned in the caption of Table 1 for clarity.

Authors' response: Done

Line 333, change 'non' to 'none'.

Authors' response: Thanks, typo corrected.

Overall consideration

A more precise mathematical description would make the paper more digestible. Targeted illustrations for ENSO or Gulf Stream phases could be used to illustrate the methodology and make the paper more appealing, the impact of the lag parameter and potential sampling issues clearer.

Authors' response: We have revised the manuscript according to these comments, and we hope that the modifications made make the paper more appealing and clearer.