Authors’ response to the Editor

We thank the Editor for giving us the opportunity to revise the manuscript.

Editor comment: *Based on the Reviewer comments and your replies, I encourage you to submit a revised version of your manuscript for further consideration in Earth System Dynamics. I would in particular urge you to reflect on how you can convincingly argue for the advantages and robustness of the approach that you propose, in line with some of the comments from Reviewer #2. This needs to be complemented by a clear exposition of what you can and cannot claim about a transitions and differences in datasets using your approach, to ensure it may be used with confidence by the research community.*

Authors’ response: In the revised version we have added detailed supporting information that is cited in the main text (Figs. C1-C12) that confirm the robustness of our approach.

In the conclusions, we stress the limitations of our technique: “Of course, not all the improvements in observation systems or data processing methodologies will modify the datasets in a way that can detected by our analysis technique. Other methods that return complementary information, such as cross-correlation and mutual information analysis (Figs. 7 and 8), need to be used for a more complete dataset comparison.”

At the end of the conclusions, we also clarify: “Since not every modification of the SST products has consequences that can be detected by our analysis technique, we propose using ordinal analysis in conjunction with other linear or nonlinear methodologies, to obtain a more complete characterization of the dynamics.”

Authors' response to Anonymous Referee #1

Referee comment: *I have reviewed the previous version of this paper, and I find the revised manuscript well written, and much clearer and easier to follow. The analysis is now better articulated and presented with greater detail. I believe the work is ready for publication after some small clarifications and minor edits. My comments below are minor and mostly high-level, aimed at clarifying a few parts of the text.*
Authors' response: We thank the reviewer for his/her positive opinion and the comments that can allow us to improve our work.

Referee comment: *Introduction.The authors should acknowledge the existence of other entropy quantifier for time series: e.g. the weighted permutation entropy by Fadlallah et al. (2013) (see https://journals.aps.org/pre/abstract/10.1103/PhysRevE.87.022911) as well as the work of Corso et al. (2020) (see https://pubs.aip.org/aip/cha/article/30/4/043123/211455/Maximum-entropy-principle-in-recurrence-plot).*
Authors' response: We agree with the reviewer that there are other entropy quantifiers and in the introduction of the revised manuscript we included a paragraph about this point, and cite several references (Fadlallah et al., 2013; Azami and Escudero, 2016; Politi, 2017; Prado et al., 2020; Falasca et al., 2020; Ikuyajolu et al., 2021; Novi et al., 2024; Paluš et al., 2024).

Referee comment: *I also think that the weighted permutation entropy introduced by Fadlallah et al. (2013) could be especially relevant to mention in the conclusion as a possible direction for future work. Specifically, given the spatial permutation entropy proposed in this manuscript, is it feasible to introduce weights to the spatial ordinal patterns in an analogous way to how Fadlallah et al. weighted temporal ordinal patterns? A "spatial weighted permutation entropy" could be an interesting generalization and may be worth briefly discussing as a potential extension.*
Authors' response: We agree with the reviewer that a spatial extension of Fadlallah et al. "weighted permutation entropy" can yield interesting results and is a natural continuation of the present work. In the revised manuscript we included a comment in the conclusions.

Referee comment: *Section 2: Data.The authors consider monthly SST anomalies. I suspect that these are anomalies to respect to the seasonal cycle, but this is not stated (apologies if I missed this). Please confirm that this is the case and add it in the paper.*
Authors' response: Yes, we analyze anomalies with respect to the seasonal cycle, and in the revised manuscript this is clarified in Section 2 Data.

Referee comment: *The authors chose two regions: the El Niño3.4 and the Gulf Stream. While both regions are of clear importance to the climate modeling community, I think it is valuable to add a few sentences clarifying why such regions were chosen for the general reader.*
Authors' response: We agree with the reviewer and in Section 2 Data of the revised manuscript we will included a sentence explaining why these regions were chosen:

These regions were chosen not only because of their importance for the global climate, but also because they display different spatio-temporal SST dynamics. SST in Niño region is governed by tropical dynamics, and in particular the SST dynamics results from ocean—atmosphere interactions leading to variability mainly on interannual time scales. On the other hand, the Gulf Stream dynamics, as one of the most intense western boundary currents, is governed by internal ocean dynamics and the extratropical winds across the basin, resulting in SST variability on several time scales, from fast changes due to atmospheric-driven heat fluxes to decadal shifts in spatial structure.

Referee comment: *Section 3: Analysis tools. By reading Bandt and Pompe (2002), it appears that the parameter L is connected to the embedding dimension of the underlying dynamical system. In the context of the Spatial Permutation Entropy introduced in this manuscript, is there any analogous physical interpretation for the parameters L and δ? It may be that these parameters primarily have a statisticarole. However, if a physical interpretation exists, it would be helpful to discuss this in this section.*
Authors' response: Indeed, in the original application of permutation entropy to time series analysis, L and δ allowed to embed the time series in a high dimensional space. In the context of the spatial approach, in our opinion these parameters play a similar role and we included a comment about this point in Sec. 3.2 of the revised manuscript.

Referee comment: *Related to above: is there a physical reason to use L = 4? I understand that exploring all possible values of L is beyond the scope of this work, and the results presented are already compelling with this choice, but it would be helpful to clarify if there is a general guideline to choose L.*
Authors' response: To the best of our knowledge, there is no general guideline to choose L, except for the limitation of having enough datapoints to define a sufficiently large number of ordinal patterns, in order to have good statistics to estimate the L! probabilities of the possible ordinal patterns. We used L=4 as a compromise to analyze long correlations with good statistics. However, we checked that L=3 and L=5 produced similar results (see Figs. C1-C5 in the revised manuscript).

Referee comment: *Section 4: Results. In Line 145: "To objectively quantify…" I appreciate that the authors used the PELT algorithm to identify shifts in entropy. However, many of the changes highlighted later in the manuscript are already visible by eye (which is positive and further reflect the utility of the spatial permutation entropy metric). It would be helpful to clarify (perhaps in this Section?) that while the PELT algorithm is a valuable and systematic tool, especially for future studies that may extend this analysis to many more regions, the main features in the present results are sufficiently clear to be identified through simple visual inspection. In other words, visual inspection provides a first-order confirmation of shifts in entropy, and the PELT algorithm serves as a helpful, complementary method that could be also further refined or expanded in future work.*
Authors' response: We agree with the referee that the main shifts in entropy are clearly identified by visual inspection, and PELT provides a complementary identification. We also agree that PELT performance should be, for future work, refined and better understood. In Section 4.1 we clarify that:

"Although many transitions can be identified by visual inspection, to objectively identify change points in the temporal evolution of the entropies (and in all the quantifiers used), we employed a well-known unsupervised change point detection (CPD) algorithm."

And in “Appendix A: Unsupervised CPD algorithm PELT” we explain the limitations of the algorithm.

Referee comment: *Line 167. The fact that changes can be sometimes identified in the H_{WE} direction but not in the H_{NS} direction in the Niño3.4 region ppears physically meaningful. I suggest the authors briefly highlight that ENSO is a dominant mode characterized by large zonal (rather than meridional) temperature gradient changes, and that the spatial permutation entropy is therefore most sensitive in the direction of the largest gradients. This would help the reader connect the directional differences in entropy changes to known physical mechanisms.*

Authors’ response: We really appreciate this comment of the reviewer that help us interpret the reason why, in the Niño3.4 region, some changes are identified in H_{WE} but not in H_{NS}. In Sec. 4.1 (page11) of the revised manuscript we included the following sentences:

“As El Niño develops in a waveguide between 5oS and 5oN, the SST anomalies develop particularly in that area, and the NS gradients are highly concentrated near the equator. Therefore, they can be seen with δ = 1, but if δ = 8 is used, the Ops can include the SST variations north and south of the equator (SST decreases north and south of the equator), mixing both gradients. This does not occur in the longitudinal WE direction, as the WE gradients have a larger spatial scale.”

Referee comment: *Related to what asked above in “Section 3: Analysis tools”. It appears that \delta clearly carries some physical meaning rather than just statistical. Smaller \delta allow to characterize changes at small spatial scales while larger \delta are more linked to larger, homogeneous changes driven by climate change. The two reanalyses then differ in terms of small spatial variability while capturing the same large scale warming signal.*

Authors’ response: Yes, indeed, δ allows to tune the spatial scale covered by the ordinal pattern and our results show that the two reanalysis products differ in short spatial scales while are consistent in the long scales. In the revised manuscript we modified the conclusions to more clearly explain this idea.

Referee comment: *It would be then useful to briefly describe this when introducing the spatial entropy tool. This could also be important to clarify that different \delta allows us to quantify different aspects of the dynamics: some researchers may be more interested in quantifying large scale differences rather than small scale variability.*

Authors’ response: We agree with the reviewer and in the introduction, we will stress that a main advantage of the ordinal methodology for the analysis of a climatological dataset (or for the comparison of two datasets) is the possibility of selecting the spatial lag δ according to the interest of the research: to capture changes (or to identify differences) in short or in long spatial scales.

When introducing the spatial entropy we now clarify this point (“Moreover, by selecting the pattern orientation and by tuning L and δ, we are able to tune the spatial scale and orientation at which the datasets were compared.)

We have also modified the conclusions to remark this point (“By tuning the spatial lag δ, we were able to tune the spatial scale at which the datasets were compared”).

Referee comment: *Section 4.2: Line 207. The two datasets should indeed capture the same large scale global warming signal, therefore leading to small differences in H_{WE} and H_{NS}. If this is indeed the case I would add a small comment. Something of this kind perhaps: "…which indicate that the differences found between ERA5 and NOAA occur mainly at short time scales and that the large scale, low-frequency warming signals are correctly identified in both."*
Authors' response: We thank the reviewer for this comment. We modified the text (Section 4.2) to clarify that differences between ERA5 and NOAA occur at short time and spatial scales, while at long scales, the warming signals are consistent in both datasets.

Referee comment: *Given the comments above, I suggest the following analysis: it would be interesting to examine how the spatial entropy of temperature anomalies changes after removing a linear trend from the raw data. I am curious whether, in this case, the results obtained with small and large δ become more similar. If the authors prefer not to perform this analysis, please provide a justification.*
Authors' response: The analysis was performed after removing the seasonal cycle but not the linear trend. In the revised manuscript we include supporting information to show that removing the linear trend has almost no effect in the agreement between the datasets, but reduces the trends in SPE for δ=8 (Figs. C6-C9).
We also performed the analysis on the "raw data" (without removing the seasonal cycle) and as shown in Figs. C6-C9, depending on the region and method used, it can return additional information. Specifically, the variation of the entropies in the top row of Figs. C6 and C7 is consistent with the fact that in the equatorial Pacific, the seasonal cycle is significant in the WE direction and therefore manifests itself on both short and long scales. In the Gulf Stream, Figs. C8 and C9, the seasonal cycle is more significant in the NS direction because the current is nearly zonal. Therefore, it can be seen in NS when using δ=1 (panel a in Fig. C8), but not in the WE direction (panel b in Fig. C8). However, when δ=8, the distances are large enough for the seasonal cycle to also manifest itself also in the WE direction (panel b in Fig. C9). In the revised manuscript we include this discussion in the caption of Fig. C9.

Referee comment: *Figures 6 and 7 are somewhat hard to follow. I suggest adding, on each panel, an indication of whether it corresponds to the Niño3.4 or Gulf Stream region. This would make the figures much clearer. I understand that this information is described in the caption, but given the large number of panels, providing this clarification directly on each panel would greatly improve readability.*
Authors' response: We agree with the reviewer that this modification will improve readability and in the revised manuscript we modified these figures accordingly.

Referee comment: *4.3. Summary and robustness of detected points. Line 255, "Wald test": please add a citation.*
Authors' response: In the revised manuscript we added the following citation "Fahrmeir, Ludwig; Kneib, Thomas; Lang, Stefan; Marx, Brian (2013). Regression: Models, Methods and Applications" section 5.1.3)."

Referee comment: *I feel this section interrupts the flow of the paper. I would suggest moving it to an appendix, though the authors may choose to retain it in the main text if they prefer.*
Authors' response: In the revised manuscript we moved the section to Appendix B.

Referee comment: *Section 5: Conclusions. There is currently strong interest in the climate community in developing neural climate emulators. Recent work has highlighted* *the limitations of such tools, aiming to motivate improved neural network architectures or data-driven strategies (see, e.g., https://arxiv.org/abs/2510.04466). I think the proposed method could be further adopted as a novel metric to explore discrepancies in emerging AI models. The authors could add one sentence on this as a potential direction for future work.*

Authors' response: We again thank the reviewer for this comment. Indeed, a very useful application of the ordinal method will be to identify discrepancies in AI models. In the revised manuscript we mention this in the conclusions as a possible direction of research.

Authors' response to Anonymous Referee #2

Referee comment: *This work presents the results of the computation of the Spatial Permutation Entropy (SPE) on two different sea-surface temperature datasets (ERA5 and NOAA OI v2), for two different regions (Nino3.4 and Gulf Stream regions). This tool has not yet been applied on climate data in this setting and it allows to detect some temporal transitions in the datasets. There are mainly two parts to the results: the first part exposes how to detect temporal transitions in the datasets from temporal transitions in the SPEs (Sec. 4.1) and the second part compares the two datasets on the basis of their respective SPEs (Sec. 4.2).*

Overall, it seems also that this tool can provide some interesting insights about the spatiotemporal structure of the datasets, but precise conclusions and concrete results about the efficiency of the method are lacking or, at least, not clearly exposed. I would not recommend publication without major revisions because of this.

Authors' response: Our main conclusion is indeed that the Spatial Permutation Entropy (SPE) analysis provides new information about the spatiotemporal structure of climate datasets, and we demonstrate this by analyzing two SST products in two relevant geographical regions. Specifically, we show that two SST products consistently reflect climate warming effects in long spatial and temporal scales, while we found some discrepancies in short scales.

A concrete statement about the "*efficiency of the method*", it would require the definition of efficiency in a concrete task. However, in our study, we are demonstrating that SPE is able to compare and detect differences in the two datasets not limited to a point-by-point comparison (such as the cross-correlation or the absolute difference), but analyzing the behavior by selecting the spatial scale of the analysis (by tuning two parameters: the spatial lag, δ, and the length of the pattern, L), and by selecting the spatial orientation of the analysis (by using ordinal parameters orientation NS or WE orientation).

Even thinking about the combination of SPE with PELT to build a sort of change detector, to measure any efficiency we would need to know all of the "changes" in the dataset that are supposed to give discontinuities that should impact the SST, which is not realistic (see the answer to the next comment).

Moreover, as any method for analyzing complex spatiotemporal data, SPE captures partial information (in the same way as the analysis of the distribution of values of a climatological field gives no information about the scales of temporal or spatial oscillations). In this sense, we argue that SPE is a complementary analysis tool (we have revised the conclusions to clarify this point) that does not replace standard methods such as Fourier analysis. We also argue that SPE has been applied to analyze data in different scientific areas (images, art works, textures, EEG and cardiac data) and we show here that it also returns relevant information when applied to SST spatio-temporal data.

Referee comment: *Major comments: The authors claim that the proposed method (SPE + PELT) allows to detect temporal transitions in datasets. There is an effort to estimate the robustness of the detection of change points but a general assessment of the success rate is missing: if the goal is to provide a method to detect transitions in a given dataset, there should be an estimation of the number of transitions that the method will indeed detect. There is no estimation of the number of changes that the method did not detect (one transition is actually detected from the time series of SMI_{NS}, but not from*

H_{NS}, lines 216-217), even though the detected changes are all linked to some changes in the methodology to produce the datasets.

Authors' response: The main changes in SPE and in SMI over time are clearly visible, and we use the PELT algorithm to provide a complementary identification. We argue that our work is not about a method to detect transitions, but a method to characterize complex spatio-temporal data. Some main transitions seen in the evolution of SPE and SMI are confirmed by PELT, and can be attributed to changes in the methodologies and data used to construct the SST products.
Regarding "*a general assessment of the success rate is missing*". This is because the ordinal method detects some changes, but surely not all, because, as explained before, the method does not use all the available information (for instance, the ordinal patterns are defined in terms of relative values), therefore, we speculate that changes not detected by ordinal analysis can be found by using other analysis tools. We have revised the conclusions to clarify this point.
Regarding "*there should be an estimation of the number of transitions that the method will indeed detect*". Unfortunately, we don't know how we can estimate such number. Over the years, several changes have been done in the methodologies and data used to construct the SST products. But not all of these changes can be expected to have significant consequences in the SST.

Referee comment: *The two datasets considered in this work are also compared, but I do not really understand what is the conclusion of this comparison. The detected change points (except one) are those detected from H's time series, so is there any additional conclusion with respect to Sec. 4.1? It is mentioned that there are small-scales differences between the two datasets (which can be expected, to some extent), can we conclude that the datasets are not reliable at these scales? If yes, how to identify a scale above which the datasets agree sufficiently ? I also wonder how does this technique compare with a Fourier or spectral analysis of the datasets.*

Authors' response: A main conclusion of the comparison of the two datasets is that as time passes, the two datasets become more similar because their spatial mutual information (SMI) increases (Fig. 7), the difference (AAD) decreases and the cross-correlation (r) increases (Fig. 8). The added value of the SMI analysis is the detection of the change in the evolution of the SMI in 2007, which is not seen in Fig. 8 that reveals a gradual increase of the agreement of the two products over time. Another important conclusion is that differences between ERA5 and NOAA are larger at short scales. This is clear in the new Fig. 7 that has the same vertical scale in all panels. We have also confirmed that with L=3 the same differences are found, shown in Fig. C5 of the supporting information.

"*is there any additional conclusion with respect to Sec. 4.1?*"
Because in Sec. 4.1 we analyzed each dataset separately, the additional conclusion that we obtain in Sec. 4.2 is that the two products tend to improve their agreement over time. The variation of SMI calculated with ordinal analysis (panels a-h in Fig. 7) is consistent with the variation of SMI calculated from the data values (without transforming the data points to ordinal patterns, panels i and j in Fig. 7). The increase of SMI over time is expected because the two SST products have improved over time, especially due to better remote observations.

In addition in Sec. 4.1 we found that the evolution of H_WE with $\delta=1$ in the Gulf Stream region (Fig. 3d) differs in the last 10 years: H_WE calculated from ERA5 increases, while

H_WE calculated from NOAA decreases. However, the corresponding SMI (Fig. 7b) increases in the last 10 years. Since SMI depends on both entropies as well as on the joint entropy (Eq. 4), and the entropy variations compensate (one increases and the other decreases), we can conclude that the joint entropy decreases, which in turn indicates that the joint distribution of ordinal patterns becomes narrower. In other words, this reveals the increase of the probability that the same pattern occurs, at a given time and geographical location, in the two datasets. In the revised manuscript we clarify this point in Sec 4.2 and also include, as supplementary material, figures that display the temporal evolution of the probability that, in a grid point in El Niño or in the Gulf Stream region, the spatial ordinal pattern (in the NS or WE direction) is the same in the two datasets.

Regarding "*can we conclude that the datasets are not reliable at these scales?"* No, our results do not allow us to conclude this. Our results indicate that the datasets were partially inconsistent in the past, but are increasingly consistent. For instance, in Figs. 8c and 8d we see that the cross-correlation r displayed large fluctuations, which have become smaller in the last few years (as r approaches 1). Complementing this information, in Fig. 7 we see a clear increase of their mutual information (SMI).

Regarding "how to identify a scale above which the datasets agree sufficiently?" We thank the reviewer for this question. We studied the variation of SMI, considering the first 10 years and last 10 years of the period analyzed, as a function of the spatial lag δ. The results are included in the revised manuscript, in the supporting information, Fig. C10, were we see that:

- The similarity between the two datasets for the same δ in the two periods is quite different, being more similar in the second period (note the different vertical scale).
- The agreement between the datasets improves with δ.
- In general, the two datasets are more similar in the Gulf than in the Pacific, which can be due to the fact that the North Atlantic is the area with the highest concentration of in-situ measurements globally (https://ioos.noaa.gov/community/global/).
- In the Gulf Stream region, the similarity of the two datasets for the NS orientation saturates for δ=5, that is, when the NS ordinal patterns cover 4 degrees, meaning the submesoscale (0.1-10 degrees), an intermediate scale between typical ocean turbulence and eddies.
- In contrast, in the case of Niño 3.4, there does not appear to be a scale at which the similarity of the datasets stops increasing.

Regarding "I also wonder how does this technique compare with a Fourier or spectral analysis of the datasets". Ordinal analysis provides complementary information because it uses a symbolic technique (the relative values of L data points are encoded in ordinal patterns). In contrast, Fourier analysis is a linear method. For the SST datasets analyzed, SPE analysis provides information that cannot be obtained by Fourier analysis because the definition of the ordinal pattern allows to select a specific scale. In the revised manuscript we clarify this point.

Minor comments:

Referee comment:

1. A smaller SPE is systematically interpreted as a more pronounced gradient because ordered patterns like '0123' would be more frequent. This is probably correct in most *cases, but is it always true ? The fact that non-ordered patterns like '2031' become more frequent is also consistent with a smaller SPE, in contradiction with the more pronounced gradient interpretation. Did previous studies on SPE establish that we can confidently interpret a decrease in the SPE as an increase of the gradient ?*

This is an interesting point, and as the referee said, the fact that a decrease in SPE is essentially determined by an increase in gradients is correct "in most cases". To have non-ordered patterns to dominate the distribution we would need a spatial structure (or temporal structure in the case of PE computed using temporal patterns) with a high regularity at the spatial scale sampled by the ordinal patterns. This is usually not the case in SST data, which usually present high variability at many different scales, and not a few well defined spatial modes. Therefore, we can safely state that the evolution of the entropy is mostly ruled by the formation of large-scale gradients.
To further assess this point, in the supporting information, Fig. C11 displays the histograms of ordinal patterns, in El Niño region and in the Gulf Stream region, at a given time. Low entropy values occur when the most expressed patterns are those that encode gradients (0123 and 3210), and the variation in their frequencies of occurrence ultimately determines the temporal variation of the entropy. We also include 8 videos as supplementary information (https://doi.org/10.5281/zenodo.18223157), that show how the probabilities change over time, for δ=1 and δ=8, in the two regions and for the two datasets analyzed.

Referee comment:

2. I feel like some context about the physics of the SST in the studied regions is missing to understand some of the interpretations. For example, why asymmetries in the increase of the SST lead on one hand to a decrease of H_{WE} in the El Niño region, and to an increase of H_{WE} in the Gulf stream region on the other hand ? Why the behavior of H_{WE} is expected to be different for delta = 1 and delta = 8 for El Niño (lines 194-200) ? Why H_{NS} with delta = 8 cannot capture the NS gradients appropriately (lines 198-200) ?

Regarding "*Why asymmetries in the increase of the SST lead on one hand to a decrease of H_{WE} in the El Niño region, and to an increase of H_{WE} in the Gulf stream region on the other hand?*: the different effects of asymmetrical heating, in the El Niño, the differential heating enhances the temperature gradient across the region, favoring monotonous symbols in the distribution and therefore reducing the entropy. In the Gulf Stream region the opposite happens because differential heating decreases the gradient in the area, producing a distribution of ordinal patterns that is more and more uniform, therefore increasing the entropy. This was discussed in the original manuscript (lines 177-178 and 181-184), and we revised the text to clarify these points (now lines 193-202).

Regarding "*Why the behavior of H_{WE} is expected to be different for delta = 1 and delta = 8 for El Niño (lines 194-200) ?"* Our argument is that, for δ=1, L=4, the ordinal patterns span 1 degree, and the SST variations at these distances are relatively small. On the other hand, for δ=8, L=4, the patterns span 6–7 degrees, allowing us to observe the horizontal gradients associated with ENSO SST anomalies. For this reason, entropy decreases in the latter case. The gradients in the WE direction are larger than those in the NS direction, which explains the different behavior the entropies for δ=1 and δ=8.

Regarding "*Why H_{NS} with delta = 8 cannot capture the NS gradients appropriately?:*

We thank the reviewer for the questions as it allows us to better explain the idea. NS gradients are detected with δ=8, but not as clear as with δ=1.

El Niño develops in a waveguide around the equator, between 5°S and 5°N. This means that SST anomalies develop particularly in that area, and the NS gradients are highly concentrated near the equator. Therefore, they can be clearly seen with δ=1. If δ=8 is used, 6-7 degrees are considered, which can include the SST variations north and south of the equator (SST decreases north and south of the equator, mixing both gradients).

Referee comment:

3. There is no uncertainty quantification on the SPE, so that we do not know if some observed trends are really relevant. For example, are the trends in H_{WE} in Fig. 4a and 4d really significant ? In addition, these trends are qualitatively consistent with the mentioned asymmetries in the increase of the SST, but are these asymmetries really the cause of these trends ? (lines 177-184). It would be interesting to analyze which patterns become prevalent to give some sound basis for these explanations.

The estimation of entropy from the frequencies of the symbols has a variance that scales as 1/n, where n is the number of samples. Then, in the most under-sampled case (delta=8) the variance is of the order of 1/1440 ~ 7E-4, which is much smaller than the variance of the time series in Fig. 4 in the manuscript. Therefore, SPE uncertainties should not affect significantly the trends estimation of those time series and their significance levels.

Regarding "*are the trends in H_{WE} in Fig. 4a and 4d really significant ?"* We have performed a statistical test to confirm that the trends are significant, and we included the results in the figure captions of the revised manuscript.

Regarding "*but are these asymmetries really the cause of these trends? It would be interesting to analyze which patterns become prevalent to give some sound basis for these explanations"*: we have inspected the histograms to determine which patterns are prevalent (see Fig. 3 in this document) and we found that the trend patterns become more prevalent over time, but their probabilities fluctuate over the seasons. In the resubmission we include 8 videos as supplementary information (https://doi.org/10.5281/zenodo.18223157), that show how the probabilities change over time, for δ=1 and δ=8, in the two regions and for the two datasets analyzed, covering the time interval when the two datasets are available (1981-2025).

Referee comment:

4. The PELT algorithm contains a penalty parameter P which must be chosen and a method is proposed to test with respect to P the robustness of the detection of change points. I find it difficult to understand precisely the different steps of the method, and therefore to assess its effectiveness. For example, how is used the 99.5th percentile of P (line 344) ? How precisely are the surrogates used in this method ?*

The surrogates time series are used to compute the 99.5th percentile of P* that fixes the threshold for the change point significance. We have clarified this point in the revised manuscript, page 17.

Change points are considered robust if their R is larger than the median value of R (lines 350-351), does that mean that half of the points change points are robust ?

Yes, this was an attempt to make a global comparison of the various significant change points detected by PELT. Some change points have a relatively low robustness score, and they are significant in the sense that the surrogate timeseries did not display almost any change points. However, we acknowledge that using the median is not the best strategy. Therefore, in the revised manuscript we present the R score as it is, without doing a binary categorization into robust/non-robust labels (in the revised manuscript we modified Table A1).

Referee comment:

5. It would increase the readability if all details about the results presented (which value of delta is used, which region is considered) were indicated on the Figures themselves (in particular for Fig. 3, Fig. 4, Fig. 5, Fig. 6, Fig.7 and Fig. 8), next to each subplot. Ideally, this would appear even if there is only value of delta or one region considered in the Figure. This information is available in the captions, but the comparison of Figures would be easier if they were more visually explicit.

We have modified the figures to include all details in the figures.

Referee comment: *Technical comments*

Explicit formulas for H_{WE} and H_{NS} would be appreciable, since these are the main quantities in the paper.

The entropies are calculated with Eq. (1) (line 102) and in the revised manuscript we include, in Sec. 3.2 explicit formulas: H=H_{WE} when the ordinal patterns have WE orientation, and H=H_{NS} when they have NS orientation

Line 101: is the j index of p_j(t) the same index as the j index of X_{ij} ?

No, and in the revised manuscript we changed p_j to p_k to avoid confusion.

Lines 145-157: it would increase the readability if this paragraph on the details of PELT and its penalty parameter is moved to the appendix (so that everything about PELT is in the appendix)

In the revised manuscript we moved the paragraph to the appendix.

Lines 185-186: which dataset is used here to compute the SST anomaly ? Could this influence the results ?

In Fig. 5 we have used NOAA OI v2 to compute the temperature anomaly. The two datasets are highly correlated at the basin scale, as shown in Fig. 1, so using one or the other to compute the NINO34 index (black curve) gives very similar results. In the revised manuscript we clarify this point in the caption of Fig 5.

Line 191: I do not understand what 'reflecting the north-south gradients that occur as the equatorial zone is warmer than the north-south edges of the region' means. How can there be a uniform pattern across the region if the central part is warmer than the edges?

We are talking here about two north-south gradients, one in the northern hemisphere, and one in the southern. One contributes to an increase of 0123 pattern, the other to 3210 pattern. Overall, the prevalence of these two patterns decreases the entropy. In the revised manuscript we rephrased the sentence clarify this point in page 11.

Line 199: why \delta = 8 is equivalent to each word spanning 6° ? I would say that each word spans 2°, in agreement with what is written in line 205.

For delta = 8 the 4 datapoints that define a pattern with L=4 are separated by 2º, and therefore, the geographical distance covered by the pattern is 3 x 2= 6º. We revised the sentence to clarify this point, also in page 11.

Lines 185-200: has this analysis been done for the Gulf stream region ? Are the conclusions the same ?

The analysis was indeed carried out, but SST anomalies in the Gulf region, even if aggregated at the basin scale, do not present a smooth behavior as in the ENSO region. In the revised manuscript included a comment about this point at the end of Sec. 4.1.

Line 205: write → right

Thanks, typo corrected.

Lines 216-222: El Niño is written differently 3 times

Thanks, in the revised manuscript we write El Niño consistently.

Line 221: should Fig. 7e be Fig. 7i ?

Thanks, typo corrected.

Line 223: should 'relative constant' be 'relatively constant' ?

Thanks, typo corrected.

Line 238: there seems to be a verb missing in the sentence 'Increased cloud coverage in this region during winters could difficult infrared measurement'

In the revised manuscript we corrected the sentence.

Line 243: the acronym CPD should be explained before the appendix

Thanks, we introduce the acronym in sec. 4.1

Lines 245-285 (Sec. 4.3): the discussion in this Section seems to be partly redundant with Appendix A, and difficult to understand without reading first this Appendix. Could it be moved to the Appendix ?

We thank the reviewer for the suggestion and in the revised manuscript we move the discussion to Appendix B.

Line 291: what models are referred to ?

We are refereeing to the models used in the reanalysis and in the revised manuscript we modified the sentence to clarify this point as: “We used the temporal variations of SPE calculated with NS or WE OPs, H_{NS} and H_{WE} respectively, to compare ERA5and NOAA OI v2 SST anomalies in two key region”.