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.

Comments.

- 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).
    - 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.
- 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.
    - 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.

- 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 statistical role. However, if a physical interpretation exists, it would be helpful to discuss this in this section.
    - 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.

- 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.
    - 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.
    - 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.
        - 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.
    - 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.”
    - 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.
    - 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.

- 4.3. Summary and robustness of detected points.
  - Line 255, “Wald test”: please add a citation.
  - 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.

- 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.