Reviewer Comments and Responses

First, I would like to thank the reviewers and editor for the comments and supportive critique. I believe the submission is much stronger for it now.

The responses are organized per commenter and in blue, with concision in mind.

R1 Comments

1. The only mention to the state variable used to reconstruct the attractor properties is the sentence ''the MLD (=Mixed Layer Depth) will be taken as the representative system state of deep convection in the Gulf of Lion". I interpret that the time-dependent vector x(t) which is used in the dimension estimation method is the vector made with all values of MLD, one for each horizontal grid point in the reanalysis, at a given time. And that the Euclidean distance mentioned later acts on this vector. It seems to me that this is the reasonable interpretation of the methodology in the paper, but I feel that given the basic role this vector takes in the whole paper, its characteristics need to be more explicitly described: for example, to mention its dimensionality (number of grid points used). The need for this type of clarification is stressed by the fact that MLD configurations are never plotted, and only the 'maximum values' are mentioned plotted in Figs. 2 and 4-7. This exclusive reference to the 'max MLD' makes the reader to doubt if really the whole MLD configuration is being used as system state or just the time series of maxima.

Yes, you're right; I mistakenly forgot to supply that in the initial submission. An additional paragraph has been added to sec. 2.2 that provides this information and how the distance function is applied. The model MLD is a 238 by 137 by time (long. by lat. by time) array, with the spatial bound delimited as described in Fig. 1. I have also clarified how the maxima are calculated in sec. 2.4: daily for the Reanalysis and monthly for the Argo. I hope the clarifications made elucidate the methodology.

An additional figure, now Fig. 2, was added that compares the quantiles of the two datasets to better show that reanalysis data reasonably represents the observations, save for the extremes during deep convection.

2. The definition of 'max MLD' is also not very precise. The reanalysis provides daily values of MLD at each horizontal grid point, so that it seems natural to interpret the 'max MLD' as the maximal depth in the MLD spatial configuration each day. But this is never explicitly said and other interpretations could point to the maximum value in a space-time interval. This is further supported by the comment at the end of the caption of Fig. 2, which mentions 'monthly maximum and median' for the Argo data. Please explicitly state, for each data source, over which spatial, temporal, or spatio-temporal window are the maxima taken.

Yes, you're right. This is partially addressed in my response to your first comment, however I have additionally added clarification to the Fig. 2 caption (now Fig. 3).

3. Data on deep convection from the reanalysis and from Argo profiles are very different. The author 'corrects' reanalysis data by comparing them with Argo, Somot and Marginier data. But no details are given on how the correction is done, which set of time steps are 'removed' and under which criteria. It seems that 'too discrepant' data are 'removed' from

the reanalysis data set. But in the text and fig. 2, 19 years are said to be removed but in table 1 only 18 are moved from the 'deep-convection' to the 'not-deep-convection' category. Please explain in detail which objective criteria are used to 'correct' which time-steps (daily?) of the reanalysis, and the rationale for this.

You're right, the calculation of 18 years was a mistake and has been corrected to reflect the correct 19 years that were altered (see both the text and revised Tab. 1). Lines 111 and 112 of the original submission did provide the criteria for which time steps were removed, but it wasn't very clear. In the revised sec. 2.4, much more explanation has been given to describe which years were considered and kept as deep convections and why, as well was which time steps were removed and why.

4. Despite the large differences between reanalysis and Argo data, Fig. 3 seem to imply that they give very similar results for the dimension estimation. The reason for this needs some explicit explanation.

I believe this to be due to the statistical nature of the method. The number of points removed is around 10%, which is not nothing but is small. Since the bulk of the time remain, the dimensionality is largely unchanged. This has been expressed in the revised paper adjacent to now Fig. 4 (old Fig. 3).

5. There are essentially no details on the path between the MLD data and the final d estimation. A threshold value of 98% percentile for g1 is used without justification. Some sensibility analysis is in order to demonstrate that this is an appropriate threshold, for example by showing results for other values, or showing that the distribution of z for this threshold is indeed a Generalized Pareto (in fact, a simple exponential in this case), or other data that can convince the reader that the hypothesis needed for the validity of the methodology are indeed satisfied here to a good extent.

Yes, you're right. I've added an additional small section in the methodology (Sec. 2.5) exploring the sensitivity of this choice. It appears that the 90th, 94th, 98th, and 99th percentile all follow an exponential distribution (I've added clarification that the GPD is simplified to this for this case in Sec. 2.1). The larger the percentile used, the better it is for converging to the GPD through EVT, so it appears that the 98th percentile provides a good trade off to retain more data points for the z distribution than the 99th percentile threshold. (Now) Fig. 5 shows these results and it shows that there's only a small difference between the 98th and 99th percentile choice on the resulting T's as well (covered in the added section).

6. As a minor point, the color scales used in Fig. 7 b and c are quite confusing: First, at difference with panels a and d, here reddish/violet colors, easily confused, appear at both ends of the scale, which leads to difficult reading of the data in the plots. Second, the orientation of the colors is exactly opposite (red for low values in b, red for high values in c), which is really confusing. Please use in panels b and c a color scale of the type used in a or d.

You're right, I have changed the color maps used to not be as confusing at the limits for b) and c), as well as change them to colors that are more suitable for colorblind readers (now Fig. 9).

R2

1. The author does not provide a clear enough explanation as to why he chose the fractal attractor as the main subject of study. Other methods and approaches exist for studying the predictability of geophysical processes. From my view point, the author needs to explain why the fractal attractor was chosen as the subject of this study.

In truth, the fractal attractor is just the base on which the methodology is built. d, the local dimensionality, describes the attractor, but also describes the predictability of the system at different time steps. Studying the attractor is inherent to the methodology. However to stress the predictability side of the analysis, I've added "Predictability and" to the beginning of the title (I believe doing this is possible still at this stage).

To address the comment regarding other methods, to my knowledge I'm only aware of one other method used in oceanography to ask this question: ensemble modelling. And this is the only other example looking at chaos in the Gulf of Lion (Waldmann et al. 2018; in the article's reference list). I have added an additional paragraph discussing the trade offs of ensemble modelling vs this method in Sec. 2.1 (at the end). The main trade off is: this method can look at the chaotic variability of each time step, whereas ensemble is less suited for this. And it is much more computationally intensive, as it requires multiple simulations; this one does not. However, this method loses the spatial information, as its eaten up by the distance calculation. This doesn't happen with ensemble modelling as it is used in other studies (see again Waldmann et al. 2018).

2. A reader unfamiliar with the authors' papers, the developers of method used in this study, can only assume that other methods will not allow a detailed investigation of the predictability of all phases of oceanic deep convection. Therefore, some explanations should be provided.

This comment is very similar to the first comment, I believe, and is answered by my response there: with the additional paragraph in Sec. 2.1 (see the above response).

3. The third comment concerns the discussion of the results. The author needs to clarify which parameter values are considered "good" and which are ''bad''. It may be difficult for the reader to interpret Figures 4 and 5. It is clear that ocean convection is chaotic in nature due to several factors. Readers should be confident that if the local correlation dimension exceeds a certain (critical) value, the system's behaviour becomes more chaotic than before. However, this does not allow one to draw conclusions about the predictability of the system in either case. One can only perform a comparative analysis, nothing more.

There isn't really a defined threshold for the local dimensionality, in terms of what is good or bad. At the core, it describes the local geometric dimension of the fractal attractor, e.g. 2 for locally planar. When the dimensionality explodes, it suggests a purely stochastic process. In this sense, the limits provide a ground truth. And this is discussed in the results for (now) Fig. 9 around the very extreme values (subplot c)).

However, in a relative sense, it describes the relative predictability of a given state (which is found at a given location on the attractor's phase space). For this study, we only need the relative case because the interesting result is that the restratification period is the more chaotic/less predictable period, relatively, than the deep convection period. I'd argue for many in the oceanic deep convection field, the deep convecting phase would be traditionally the

considered the least predictable phase (of the 3).

To help clarify this in the text, I've added language to stress these aspects primarily in the methodology and results.

4. Next comment: The author uses mixed layer depth as the primary parameter characterizing the convention. However, there are some other parameters used for this purpose. Therefore, the author could provide a ranking of the relative importance of input characteristics. This would robustly demonstrate that mixed layer depth alone is the most significant predictor, justifying its sole use in this study.

For tracking deep convection, the mixed layer depth in the ocean is the traditional metric used. It describes the depth from the surface to the limit where the upper water mass meets a strong gradient. This captures what happens during deep convection: the homogenization of the vertical water mass from the surface to the deeper parts of the sea. You could use this methodology directly on the 3D temperature and salinity gradients, but you're not really gaining anything as the mixed layer depth is derived from these fields. But, it would add computation cost. I have added similar language to the introduction and methodology sections to clarify this to the reader (introduction: 2nd paragraph, methodology: Sec. 2.1).

5. It would be interesting to know whether the obtained results can be scaled to other regions of the world's oceans.

Yes, it can be. I haven't added anything to article regarding this point, as it feels tangential, yet an interesting comment. If you can work around the data limitations to compute the statistics for the global scale, then it might bring some very interesting results, depending on your desired question. But, that is main limit I experience, the size of the data and the memory available in the accessible compute servers.

EC 1

1. The results depend strongly on the reanalysis data and the Argo profiles, both of which carry uncertainties (model biases, sampling gaps, etc.). Since these data directly feed into the dimensionality estimates, the results must also be uncertain. At the moment, this is not really addressed beyond the dataset revision. I suggest the authors discuss these uncertainties more explicitly and, if possible, show how they affect the results (e.g. with confidence ranges or sensitivity/ robustness tests).

This is also questioned by reviewer 1 and in response to their questions (mainly comment 5) I've added a sensitivity analysis question that looks at the choice of percentile for threshold, which is the main factor that can be controlled with this methodology that can change the results. The choice of the 98th percentile for the threshold appears to be appropriate. This is more pointed, however to really investigate the uncertainties of this method, you need additional models, which I believe to be outside the scope of this work. This is touched upon in the discussion section of the article in paragraph 2. Performing this additional study I believe would be valuable for the community, however.

2. In addition, reproducibility is not fully ensured, as also mentioned by referee 1: while the data sources are public, the code and workflow used for the local correlation dimension analysis are not available. I suggest that the authors provide code or at least detailed

instructions, ideally mentioned in the Data availability section, so that others can reproduce the results.

Yes; I have added a link to the git repository in the data availability section where the code will be added when the article is ready (in case additional changes are made during revisions).