General comments:

The authors propose combining two apparently contradictory analysis techniques to the outputs of a low (36) dimensional dynamical ocean - atmosphere model. The first, makes a nontrivial decomposition of the 36 dimensional signal into series with well-defined time scales, the second analyses the phase spaces assuming the existence of scale invariant properties. The justification and interpretation of this is opaque.
While the authors question the utility of conventional analysis techniques, at least the latter can be interpreted in straightforward manners. The interpretation of their results is nontrivial.

We thank the Referee for raising some points that can be helpful for improving the presentation and clarity of our findings. Most of all, we do not question the utility of conventional analysis techniques, but acknowledge their intrinsic limitations and attempt to explore the potentials of a combination of two "non-conventional" techniques to provide additional information.
We also want to stress that the two methods should not be seen as "apparently contradictory" as emphasized by the reviewer. The modes extracted in the first analysis step have no well-defined time scales but are instead characterized by scales that are time-dependent. This is one of the main novelties of the Empirical Mode Decomposition and its multivariate extension we used here (i.e., the MEMD) as compared to fixed-scale decomposition methods like wavelets. The extracted modes can be seen as representative of fluctuations at a typical scale that is the average of the instantaneous scales derived from a given mode via the Hilbert Transform. Moreover, the second analysis step, i.e., the generalized fractal dimensions, requires to have scale invariant properties in the phase-space of a given system, thus working (essentially) on measuring the geometrical properties of the system trajectory and information on how to reconstruct it by measuring the information dimension D_1 and q-tuplet correlations $D_{q>1}$. This means that there are no a priori constraints on understanding a system using D_q. Thus, the two methods are not contradictory but rather complementary.

In the following we provide replies (in italics, labelled by "A") to the Referee's detailed comments (in normal font, labelled by "C") that will be also thoroughly considered in a revised version of our manuscript.

Detailed comments:
C1. The notation is not easy to follow. Please explain the curly bracket notation used throughout:

$$\{\mathbf{s}(t)\}|_{t\in T} = \{s_1(t), s_2(t), \ldots, s_k(t)\}$$

On the left, a bold symbol "s" is used which is standard for indicating a vector. Why do the authors (apparently needlessly) add curly brackets and then an explicit restriction as a subscript?

Further, there is the bizarre looking symbol $(D_2^{\sum_j})$, that is also not adequately explained.

A1. *We thank the Referee for this suggestion. Indeed, we agree that the notation using curly brackets has been partly misleading, since the left-hand side of the equation was originally intended to represent a sequence of vectors, while the right-hand side was supposed to clarify the structure of each of those vectors composed of k scalar properties, the latter of which however was lacking clarity in our notation. This aspect will be clarified in our revised manuscript. Moreover, the "bizarre looking" symbol can be safely changed to $D_{q,\tau}$ following the notation used in Alberti et al.*

(Chaos, 2020). We will modify the corresponding parts of our manuscript also with a general attempt to be more precise when introducing notations in a revised version of our manuscript.

C2. When discussing the mathematical properties of the usual decompositions ("completeness, convergence, linearity, and stationarity") it is stated that "these conditions are not usually met when real-world geophysical data are analyzed". This is confusing since the mathematical properties of Fourier or other decompositions are valid irrespective of any application. I think the authors meant to question the appropriateness of such decompositions for their specific application? However, this is a mathematical question that cannot be answered without reference to a specific assumed mathematical framework. In the paper the authors do not analyze empirical data at all but rather model outputs. Contrary to real empirical series, their series are therefore taken from a well-defined mathematical framework given by dynamical systems theory. Please explain why standard decompositions are not adequate for studying such model outputs and why there is a need for them to be replaced by decompositions with quite nontrivial interpretations and properties.

A2. *We thank the Referee for this important suggestion. As also highlighted in our reply to Referee #1 we need to clarify the sentence on properties met by real-world data (note that our manuscript does not exclusively utilize low-order model output, but also reanalysis data, which in our opinion would qualify as "empirical") for which linearity and stationarity assumptions are often not met. Indeed, we fully agree that mathematical properties of the decomposition methods themselves are surely valid irrespective of any application. As suggested by the Referee, we referred to the use of adaptive methods that can be justified to overcome some limitations of fixed-basis methods such as linearity and stationarity assumptions. Moreover, adaptive methods (as the MEMD) could be more suitable for reducing some mathematical assumptions and a priori constraints.*
Although we use the MEMD on a well-defined framework derived from dynamical systems theory, the reduced a priori constraints and the limited number of intrinsic components that can be visually inspected could be an advantage with respect to standard decompositions. Another advantage concerns the combination with generalized fractal dimensions: if we, for example, use Fourier decomposition we will have a large number of (harmonic) oscillating components at different fixed frequencies that should be summed up for exploiting our proposed procedure. Furthermore, if we, for example, use wavelets we will deal with some a priori assumptions on the decomposition basis onto which we are projecting our data that could produce misleading results in our procedure of evaluating fractal measures on a priori fixed scales. Thus, we do not question the appropriateness of conventional analysis techniques, but rather acknowledge that they (as well as any other) have intrinsic limitations in what we can learn from them.

C3. Also in the Methods section, it is stated that the authors "put forward a novel approach based on combining two different data analysis methods for Multivariate Empirical Mode Decomposition and generalized fractal dimensions". What is confusing is that while the MEMD analyzes time series in real space, in their application, the generalized fractal dimensions analysis is carried out in a quite different space - the phase space of each series. The result is that for each time series with characteristic time scale t, that the corresponding phase spaces are assumed to be scaling. In other words, while there are essentially no scaling properties in real space, it is assumed that there will be nontrivial scaling properties in the corresponding phase space. The approach is presumably justified if the characterizing these scaling properties via generalized fractal dimensions will help understand the system. At this point one wonders whether the conventional Fourier spectrum of each t scale series might have been easier to interpret, to understand. All this needs explanation, clarification.

A3. *We really appreciate this comment since it allows us to better underline our main aim. We are interested in investigating how phase-space properties (geometry, correlations) change when dynamical components at different mean scales with different dynamics are considered. In other words, we are interested in looking at the role of scale-dependent phenomena in defining the whole properties of a system. Global measures proposed in the past only allow us to investigate the statistical, topological, geometrical, scaling, properties of the whole system; conversely, our proposed approach allows us to investigate how the different scales contribute to the global properties of a system. Moreover, our framework also provides consistency with established measures for characterizing time series from an integral (not scale-resolved) perspective, since the scale-dependent measures we evaluate converge to the associated global measures as all scales are considered, i.e., when the full system dynamics, composed by all accessible scales, is reached, Within this framework, our approach could be promising for investigating scale-dependent properties, as measured by fractal dimensions, of the system. We are indeed interested in nonlinear variability characteristics at different time scales, thus employing for example Fourier decomposition would leave us with perfectly linear and stationary harmonic functions as components, which do not carry any information on nonlinear dynamics, unless when studying their mutual phase relationships, leaving out the high-order statistical properties and only focusing on the autocorrelation function (i.e., the second-order moment). Otherwise, by looking at the behavior of fractal dimensions we can explore how the different scales contribute to change the phase-space properties that cannot be highlighted by using the conventional Fourier spectrum.*

C4. In particular, when the generalized fractal dimensions are estimated, the authors need to show that there are indeed some phase space scaling properties. Using mathematical definitions such as eqs. 1-3 - where the small scale limits are taken - has only a formal validity when the definitions are applied to numerical model outputs, especially when the latter has been subjected to cubic spline interpolation which makes the small scales artificially smooth. In practice, one needs to display scaling behaviour over at least an order of magnitude or so in scale in order for any fractal dimension estimates to be convincing. The authors must therefore display some of their scaling plots - not just logarithmic slopes that have already been interpreted in terms of dimensions.

A4. *We agree with this comment. To be clearer and more convincing, we display in Figs. 1 and 2 of this response letter (which will also be included in a revised version of our manuscript) the scaling behavior for the correlation integral for the two cases C=0.008 and C=0.015 at different timescales. We choose to show here only the correlation integral since it can be faster evaluated than other moments (cfr. Grassberger and Procaccia, 1983). We show here that there exists at least an order of magnitude in scale over which a scaling behavior is observed. A similar behavior is also observed when considering the reanalysis data as shown in Fig. 3 of this response letter for the different regions. Taking also into consideration a comment by Referee #1, we consider adding Supplementary Materials with more details on the computation of fractal dimensions and scaling plots to a revised version of our manuscript.*

We would further like to remark that the cubic spline interpolation does not produce artificially smoothed small scales since it does not act on the data themselves but only on local extreme values of the data to extract intrinsic oscillating components from the data. Thus, the shape of the raw data is not changed and generally the (M)EMD extracts scale-dependent components that are smoother as the largest scales are approached.

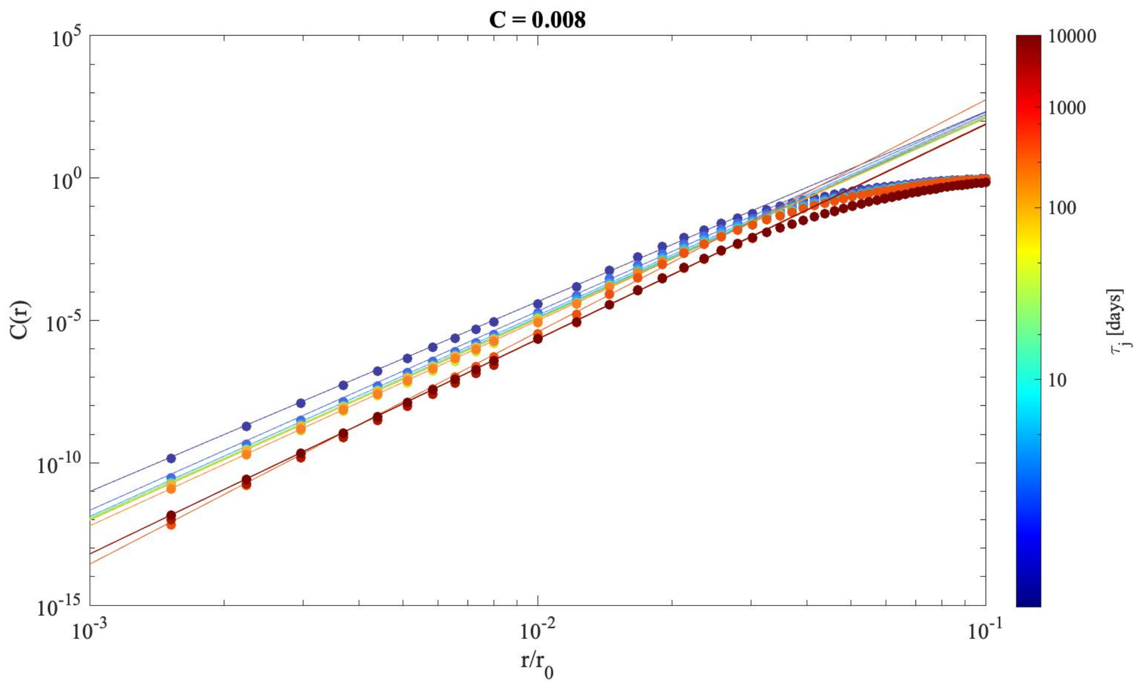

Fig. 1 The log-log scaling plots of the correlation integral C(r) as a function of r (normalized with the respect to the largest possible separation between points in the phase-space represented by r_0) at different scales represented by colors for the case C=0.008. The lines refer to the power law fit in the limit $r \to 0$.

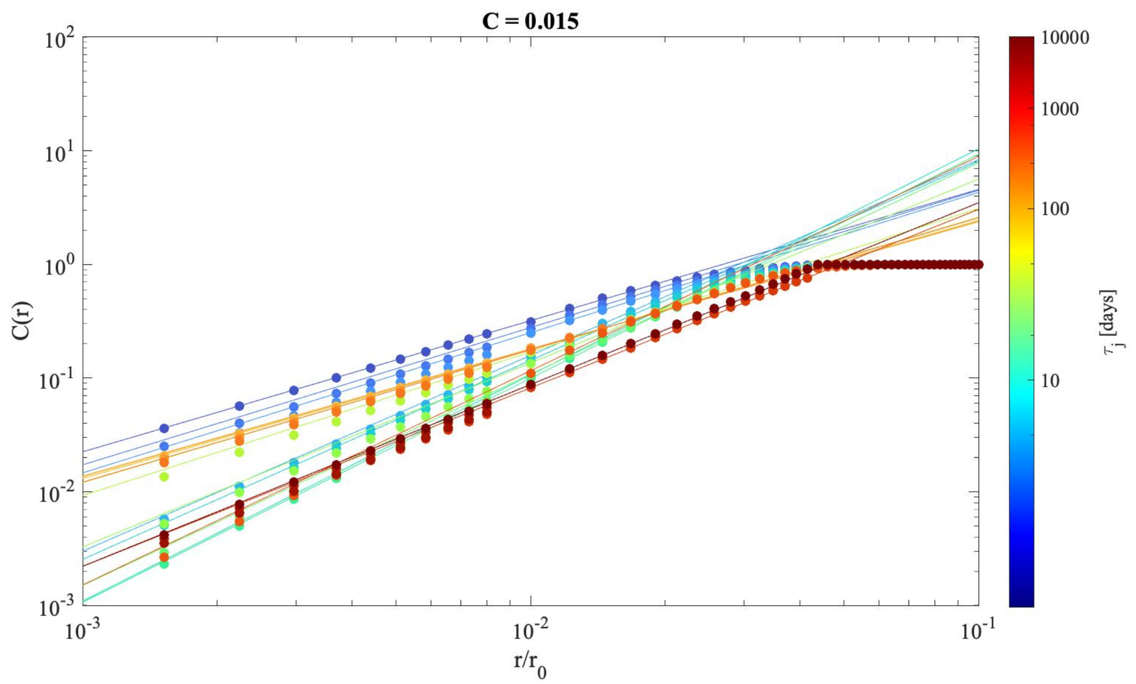

Fig. 2 The log-log scaling plots of the correlation integral C(r) as a function r (normalized with the respect to the largest possible separation between points in the phase-space represented by r_0) at different scales represented by colors for the case C=0.015. The lines refer to the power law fit in the limit $r \to 0$.

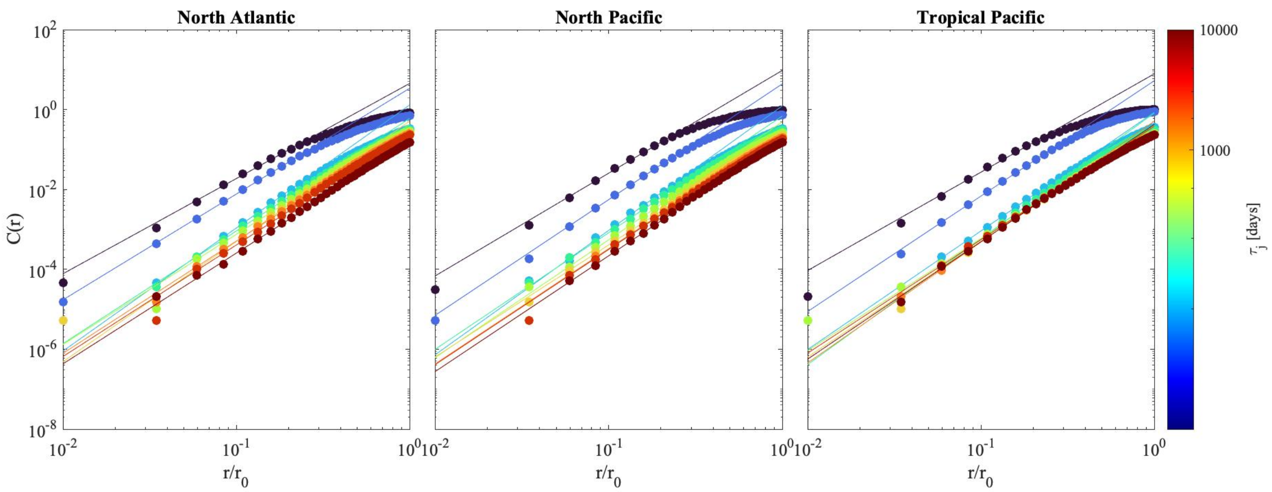

Fig. 3 The log-log scaling plots of the correlation integral C(r) as a function r (normalized with the respect to the largest possible separation between points in the phase-space represented by r_0) at different scales represented by colors for the reanalysis data.
The lines refer to the power law fit in the limit r→0.

C5. In this regard, I could also add that figs. 9 and 11 are almost certainly largely spurious. This is because typically for moments of order q≈>3-4, the moments are completely dominated by a single hypercube (a "second order multifractal phase transition") so that for larger q, the values will depend sensitively on the exact details of the input series. Similarly for q<0 most if not all the values will likely be spurious essentially due to the statistics of the very sparsely populated regions of phase space (the very low probability regions, see e.g. the discussion in ch. 5 of [Lovejoy and Schertzer, 2013]). In other words over most of the range of moments given in the figure (-20<q<20), the dimensions are likely to be spurious.

A5. *We thank the Referee for raising this important point on the statistical significance of higher-order moments. We are aware that this is a crucial point, especially when working with scale invariant features measured via structure functions, detrended fluctuation analysis, and spectral methods (as for wavelets). To deal with this problem and to support the statistical significance of our results we have followed the approach also described in Ch. 5 of Lovejoy and Schertzer (2013) to evaluate the maximum moments as those derived from the tail of the cumulative distribution function of the data. Since we deal with the investigation of scale-dependent fractal dimensions, we evaluate the cumulative statistics at different scales and as shown in Figs. 4-6 in this response letter we observe that extreme fluctuations follow a power law decay leading to the divergence of the 6th-order moment and the 4th-order moment for C=0.008 and C=0.015, respectively. Thus we fix our range of moments -6<q<6 and -4<q<4 for C=0.008 and C=0.015, respectively, and we will modify accordingly Figs. 9-11 in a revised version of our manuscript. Similar results are also obtained for the reanalysis data (see Figs. 7-8 in this response letter), thus we fix here our range of moments to -3<q<3.*

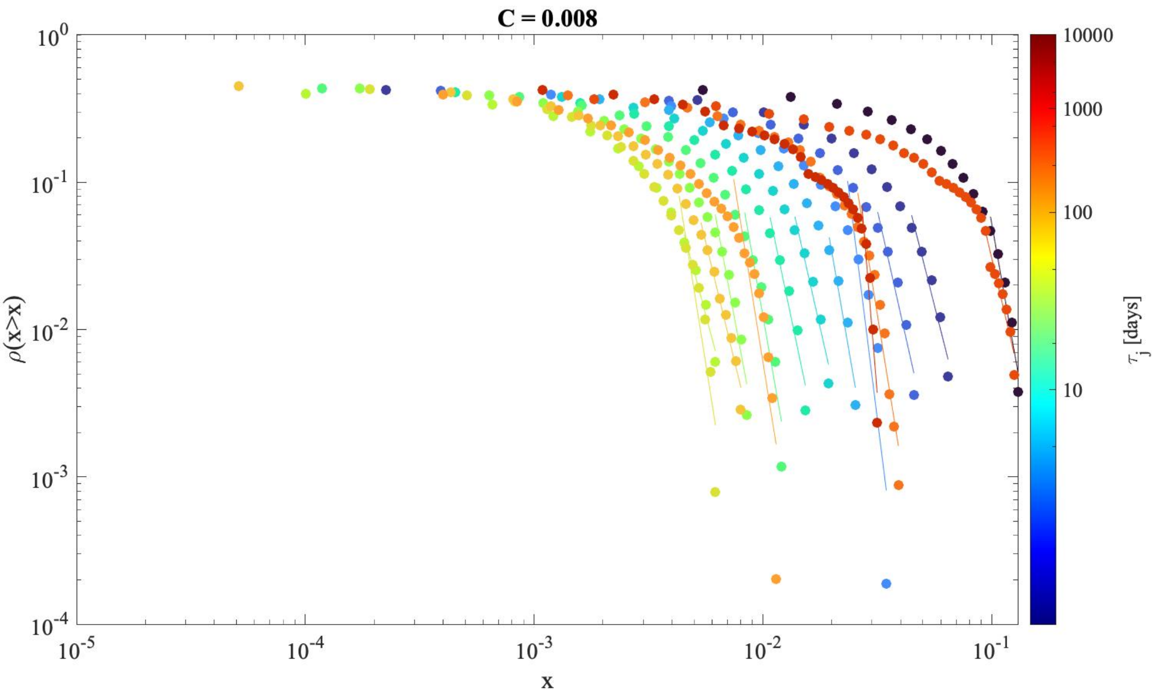

Fig. 4 The cumulative distribution function at different scales as reported by different colors for the case C=0.008. The lines refer to the power law fit of the tail.

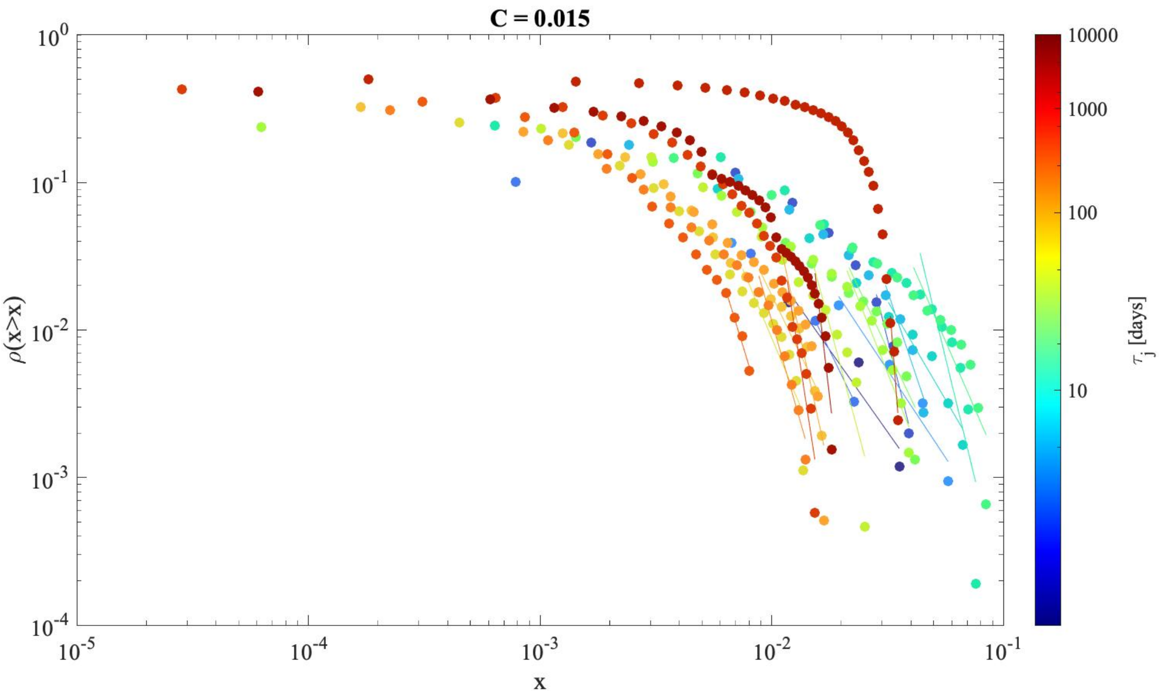

Fig. 5 The cumulative distribution function at different scales as reported by different colors for the case C=0.015. The lines refer to the power law fit of the tail.

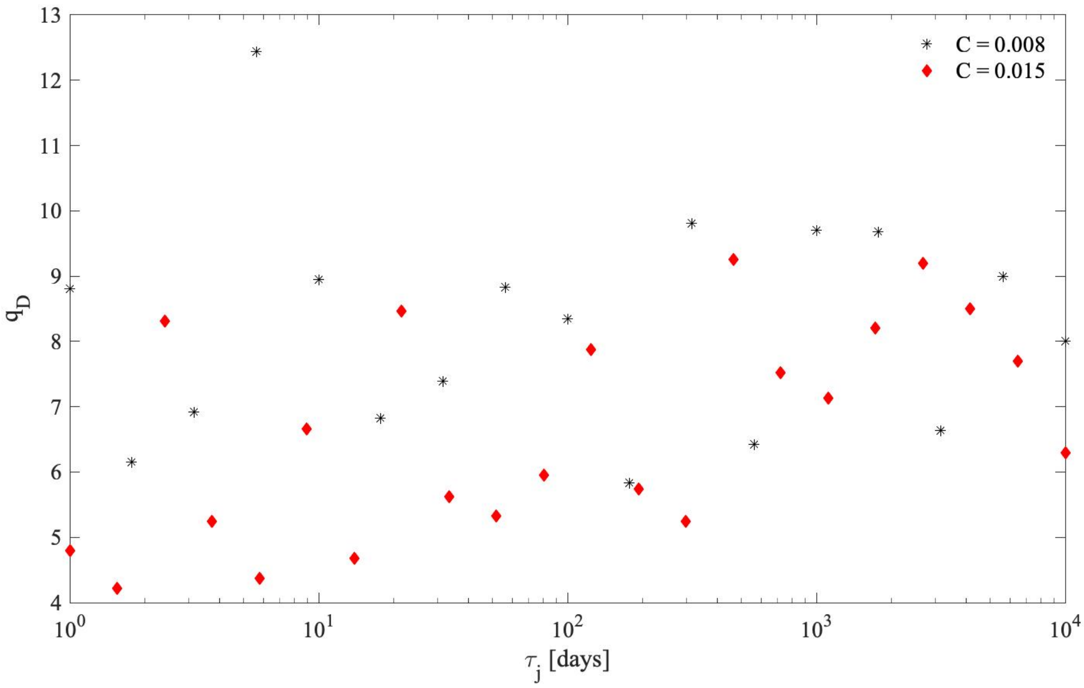

Fig. 6 The power-law scaling exponent q_D as a function of the different scales for the case C=0.008 (black asterisks) and C=0.015 (red diamonds). The minimum q_D has been chosen to set the range of statistically significant moments.

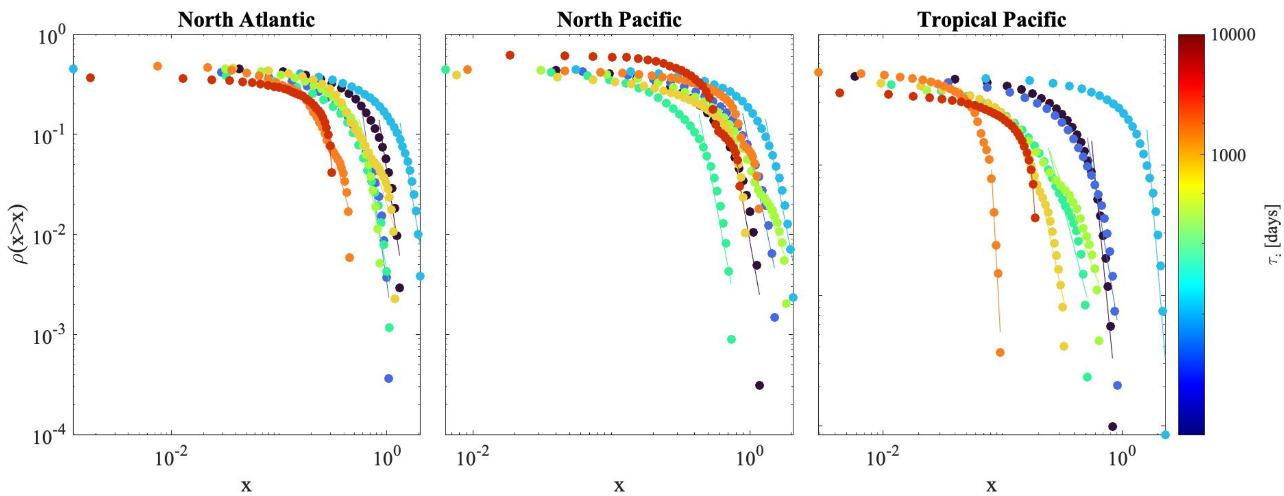

Fig. 7 The cumulative distribution function at different scales as reported by different colors for the reanalysis data. The lines refer to the power law fit of the tail.

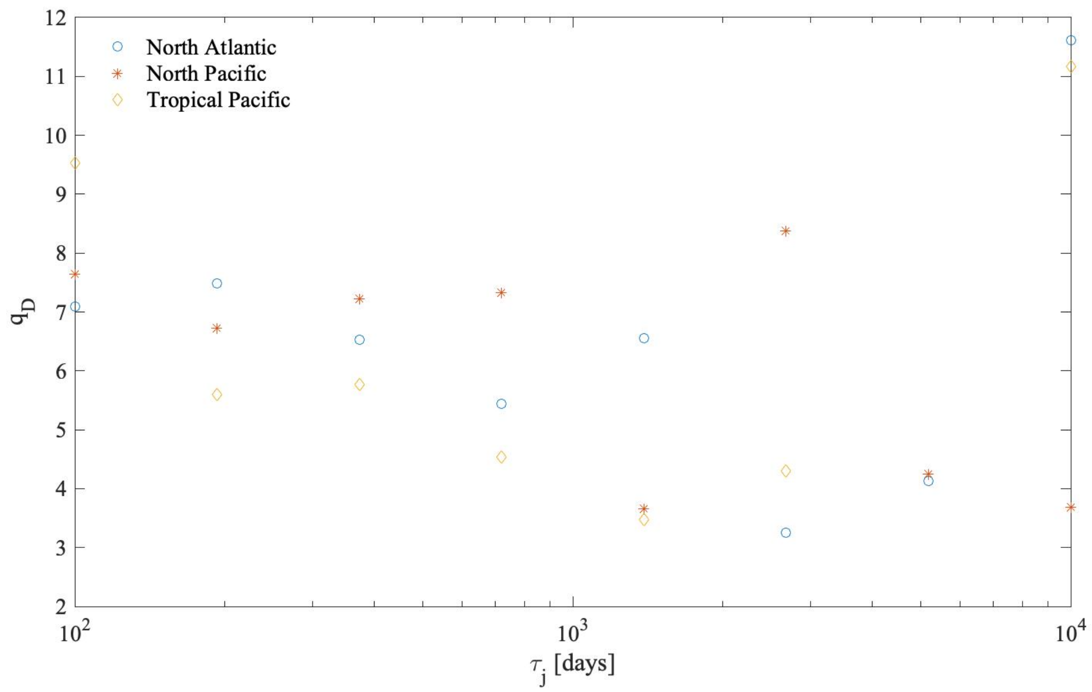

Fig. 8 The power-law scaling exponent q_D as a function of the different scales for the reanalysis data. The minimum q_D has been chosen to set the range of statistically significant moments.

C6. Finally, the interpretation of the key figures 5-8 is not at all obvious. Calling these characterizations "topological, geometric" is unhelpful and/or misleading since they are actually statistical exponents without any straightforward relationship to the phenomenon under study. The authors could note that whereas a white noise signal would give a correlation dimension equal to the dimension of the phase space itself (it is space filling), that a Brownian motion in a space $d \geq 2$ has a constant dimension = 2.

A6. *We thank the Referee for this comment. We will work on further improving the clarity of our manuscript, especially when introducing some key concepts and/or describing key features. We are referring to topological and geometrical since some measures are able to give us information on phase-space properties. For example, D_0 is a measure of the filling of the phase-space, thus providing a measure on the coverage of the phase-space by the studied system's dynamics, D_1 provides a measure of the information gained on the phase-space with a given accuracy ε, and the $D_{q>1}$ provide measures of q-tuplet correlations, i.e., mutual dependence, between phase-space points. This explains why we used the terms topological and geometrical in our manuscript.*

Reference:

Lovejoy, S., and Schertzer, D., The Weather and Climate: Emergent Laws and Multifractal Cascades, 496 pp., Cambridge University Press, 2013.

Thanks a lot for this reference that we will consider in a revised version of the manuscript.