Replies to Reviewer #2

November 2020

We would like to thank the Reviewer for the constructive set of comments provided. We provide a detailed reply (in red) to the individual comments (in *italics*) below.

Main points

1. *The assumption of stationarity is not sufficiently discussed. If a substantial regime shift occurs at some point during the simulation, I would expect that recurrences will only be observed within each regime but not between the two. Doesn't the estimated dimension then depend massively on the length of the time-series before and after the shift? Say we have simulated 1000 years before and 9000 years after a deglaciation phase. Won't the dimension in the first 1000 years be artificially increased just because there are fewer recurrence candidates? Wouldn't this result change completely if we had stopped the simulation 1000 years after the shift?*

The Reviewer raises a very good point about data stationarity. However, we believe that the Reviewer's comment combines two distinct issues. The first is how the length of the dataset affects the local dimension; the second concerns regime shifts in the data. The local dimensions depend on the number of "good" recurrences in the dataset chosen, where the definition of "good" will depend on the data and distance metric dist being used. Indeed a high dimension may correspond to few good recurrences. While very short data series do not allow to identify good recurrences, and hence valid estimates of the local dimension, once a "reasonably long" timeseries is obtained, our estimate of the local dimension is only weakly affected by the number of available years of observations. The meaning of "reasonably long" is hard to define formally, but we found empirically that several decades of daily data usually fit this definition when studying atmospheric data at synoptic or larger scales. Indeed, previous studies have outlined that the convergence of our estimators is relatively fast for daily atmospheric data in quasi-stationary conditions. For example, in Faranda et al. (2017b), we have analysed both the ERA-Interim (1979-2015) and NCEP/NCAR (1948-2015) reanalyses and found that they provide very similar values of local dimension and persistence, despite the different length of the time series. In Buschow and Friedrichs (2018), the authors analysed a 1000 year-long simulation from the simplified climate model PlaSim. Using daily data in stationary conditions they found that, as they increased the data used from a fraction of the total dataset up to the full 1000 years, the values of the dimension drifted very slowly.

Concerning the issue of non-stationarity, in Faranda et al. (2019a) we showed that our methodology is able to detect weak non-stationarities in the climate system, as for example is the case for the ongoing climate change. An abrupt regime shift poses a different challenge, and what is the limit of validity of our metrics for non-stationary systems is very much an open question. Our educated guess for the case put forth by the Reviewer is that, as long as each "regime" is sufficiently long to provide a reasonable number of "good" recurrences within it, then differences in the length of the different regimes may not be a deal-breaker for reliable estimates of d. Although we did not formally test this hypothesis on palaeoclimatic data showing bifurcations, we have indeed analysed the transitions between different jet regimes in both a conceptual coupled lattice map (Faranda et al., 2019) and a quasi-geostrophic model (Messori et al., in review). The results show that different values of d and θ successfully reflect these transitions, and thus may be interpreted as associated with different basins of attraction of the system. However, we currently have no formal arguments to support this observation. Clearly, if one is interested in mean dimension over the whole simulation, then the relative length of the pre- and post-shift simulated timeperiods will play a major role. However, upon inspection of a timeseries of d it would be natural to compute separate statistics before and after the regime shift (and hence shift in d). We now comment on the above points, which we recognise can be of relevance to a number of palaeoclimatic applications, in Sect. 4 of the revised text.

2. *The use of binary precipitation fields seems worrying: if I understand correctly, the distance measure is then effectively no longer a continuous but a discrete random variable. Do we know that the theoretical limit results apply in the discrete case?*

Precipitation data is a somewhat peculiar case since they have a fractal structure rather than that of a discrete variable (see e.g. Lovejoy and Schertzer, 1985). This holds even after binary discretisation is applied. The prp variable we analyse in our study is thus effectively a fractal spatial set (Langousis, 2009; Brunsell, 2010). Lucarini et al. (2012), have addressed the question of recurrences for fractal sets using iterated function systems which generate Cantor sets. Fractal data still have as limiting continuous model a generalized extreme value distribution, and therefore the estimates of the dimension and the persistence are possible for this kind of datasets. Indeed, in Faranda et al. (2017a), we have appplied the recurrence method to precipitation data extracted by the NCEP datasets and shown that the results have a physically meaningful interpretation. We thus have robust theoretical arguments supporting the use of our approach on discretised precipitation data. We have now added a short mention of this point in the text, when introducing the prp variable in Sect. 3.1.

3. *Depending on the domain, we might have many time-steps with identical zero precipitation. What happens to theta and d in such cases? Shouldn't the persistence be effectively infinite and the dimension zero if two subsequent time steps are exactly identical?*

The Reviewer raises a very good point. We had initially decided not to discuss this in the paper as we deemed it a technical subtlety, but we now realise that this omission may cause difficulties to those readers who may wish to apply the methodology we propose. If there are two timesteps in the dataset which are identical, this leads to the "distance" between the state of interest and the recurrence being 0. In practice, this would not preclude identifying a finite threshold to identify recurrences and then applying our algorithm. However, from a theoretical standpoint, in the limit of an infinitely long timeseries the existence of identical states in the chosen variable underscores a $d = 0$ for those states. In our calculations, we therefore assign $d = 0$. Since these states have the dimension of a point and do not reflect any dynamical information, we have decided to exclude them from our calculations in the revised study. A second, related, issue is that all the "good" recurrences may be identical. This is for example the case for a day with very little precipitation, whose closest recurrences are all the days without precipitation. In this case it is not possible to compute a meaningful recurrence threshold and we again discard the datapoint. Our algorithm for computing θ is also affected by these issues, and in order to obtain valid estimates of persistence one should revert to a naïve calculation of the average number of consecutive identical timesteps. In our analysis, we have decided to discard these states, for consistency with the calculation of d. This choice has led to minor changes in some of the figures, but does not alter any of our qualitative conclusions. Concerning the final part of the Reviewer's comment, θ can only be zero at a fixed point of the system, i.e. if all successive timesteps bring no change to the state of the system. A trivial example is a pendulum in its equilibrium position (or the equilibrium climate of a hypothetical planet at 0 K without any external energy input). As such, having two (or more) successive timesteps which are identical does not imply infinite persistence. We have now included a brief discussion of these technical yet very important points in Sect. 2.2.

4. *Please include significance tests for your composites (Fig. 3, 5, 6). As it is, we don't know which of these patterns might just be random chance.*

We fully agree with the Reviewer on this point. We have now computed the one-sided 5% significance bounds for the positive and negative precipitation anomalies respectively, in the three figures in the main text (and the corresponding figures in the Appendix) by bootstrap resampling with 1000 iterations. The results of this test are now shown in the figures. We have also updated the text to include a discussion of the significance of the anomalies when referring to the figures.

Minor points

5. *Neither the abstract not the Motivation chapter gives the reader any idea what the dynamical indicators actually do. "Different dynamical properties" (l.9) is too vague. Please add at least an intuitive explanation of what kinds of properties you mean.*

We have added a short qualitative interpretation of the three metrics to the introductory section. In the abstract, we have added some dynamical systems keywords, such as "persistence" and "instantaneous", which were previously missing and may help the readers form a better picture of the contents of the paper. We then expand upon these terms in the main text.

6. *L.20 "new challenges" I'm no expert on this but you cite Paleo-simualtions going back at least to 1996, this is hardly a "new" issue.*

We understand the apparent contradiction. We have now clarified in the introduction that the challenge arises from the exponential increase in the amount of data generated by numerical simulations of (palaeo)climates, which has gone hand in hand with the development of ever more complex and highy-resolved models. The study we cite from 1996 is indeed one of the earliest examples of numerical simulations for the MH Green Sahara episode, and likely produced an amount of data which is orders of magnitudes smaller than the data produced by a modern numerical climate model, such as the one used to perform the simulations analysed here.

7. *L.30-33 this paragraph is copied nearly verbatim from the abstract, maybe instead you could give some more explanation of what the dynamical indicators actually do. You mention that theta and alpha are bounded, what about d?*

As part of the new qualitative interpretation of the three metrics we added to the introductory section (see our reply to comment #5 above), we have rephrased this paragraph. Concerning the bounds of the indicators, d is a positive real number. Technically $0 \leq d \leq +\infty$ with $d = 0$ being the dimension of a point and $d = +\infty$ being the dimension of an unbounded infinite dimensional system with no attractor (e.g. a brownian motion in infinite dimensions). Moving to the persistence, $0 \leq \theta \leq 1$. $\theta = 0$ is the limiting case of a fixed point and $\theta = 1$ the case of points immediately leaving the neighborhood of ζ. Finally $0 \leq \alpha \leq 1$ with $\alpha = 0$ being the case of non co-recurring variables and $\alpha = 1$ the case of perfect synchronization. We now state these bounds and their interpretation in Sect. 2.2.

8. *Section 2.2: Maybe mention that x(t) corresponds to the sea level pressure or precipitation field from the example before.*

We have added this, as suggested by the Reviewer.

9. *Eq.3 looks like alpha was asymmetric with respect to x and y because the denominator contains only x, but nu(g(x) $\geq$ sx) = nu(g(y) $\geq$ sy) = 1 - q, correct ? Maybe make that more clear.*

Since excedances in both x and y are defined relative to the same high quantile, the Reviewer is absolutely correct. We have clarified this point in the text in Sect. 2.2.

10. *L. 155 how do you arrive at these definitions of Monsoon and Pre-Monsoon? Are those the present day conditions?*

These definitions are indeed taken from the present-day climatology of the West African Monsoon. We use them as reference periods to highlight the changes in the timing of the monsoonal onset and decay in the Green Sahara or Green Sahara and Reduced Dust simulations. We now specify this in the text.

11. *Fig.1 there is almost no visible difference between b and c, maybe plot the difference between GS-PD / GS-RD and CNTL instead?*

We have now added a new panel (d) to the figure which, as suggested by the Reviewer, shows the difference in precipitation between the two Green Sahara simulations. We briefly comment on this in the revised text. We have also implemented a corresponding change in Fig. A3 (Fig. A4 in the original submission).

12. *Fig. 2 Please explain more specifically what a) and b) are telling us besides the shift in monsoon onset and ending, both of which we already see in c). In particular, how do you interpret the fact that the maxima in d shift from blue to red, but the decrease in theta (b) happens nearly at the same time in all three curves.*

We argue that the three panels of Figure 3 (Figure 2 in the original submission) provide largely complementary information. For example, Fig. 3c does not give any indication as to the mechanisms driving monsoonal precipitation in the three simulations. The increased persistence in the MH_{GS+PD} and MH_{GS+RD} simulations, shown in Fig. 3b, immediately points to the fact that the role of transient atmospheric features is likely weakened compared to the MH_{CNTL} run. This hypothesis would then need to be verified with detailed analyses of atmospheric dynamics, but the computationally inexpensive θ metric is nonetheless valuable in pointing to it as an interesting aspect to investigate further. Similarly, having ascertained that d is sensitive to the monsoon's onset, its interannual variability can be used to quantify the variability of the monsoon's onset within each model simulations. If one were using more conventional analysis techniques, this would likely require defining and computing a monsoon onset index. It is further an aspect which could be easily overlooked, seeing as an inspection of a simple seasonal cycle of precipitation does not evidence the pre-monsoon season as being particularly variable, while for d it is the season displaying the largest variability (see blue shading in Fig. 3a, c). Also in response to comment #1 by Reviewer #1, we have updated the initial part of Sect. 3.2 to reflect more explicitly the information provided by the first two panels of Fig. 3. Concerning the last part of the Reviewer's comment, we would argue that there is a large shift also in the decrease in θ across the three simulations. Indeed, looking at Fig. 3c, the blue curve shows a drop between days ~100 to 135, and a rise between days ~275 to 300. The orange and red curves display a drop between days ~60 to 100, and a rise between days ~290 to 315. The shift in timing between the blue curve on the one hand and the orange and red curves on the other hand, is comparable to the shift seen in panel (a) for the increase/decrease in d at the onset/end of the monsoon season. The maximum d for the blue curve is achieved in the early monsoon season, at a time when the other two curves show a local (albeit not absolute) maximum. We therefore do not see any inconsistency between the relative changes in d and θ curves across the three simulations.

13. *Also in Fig. 2 there is no appreciable difference between the red and yellow curve. Do these systems have different dynamics or not?*

Previous analysis of these same simulations (e.g. Gateani et al., 2017) and studies from other authors (e.g. Thompson et al., 2019) suggest that, compared to the effect of Saharan Greening, the dust reduction under a Green Sahara scenario only has limited impacts on the atmospheric circulation. Moreover, the same studies have shown that the changes in atmospheric (thermo)dynamics leading to increased monsoonal precipitation in the MH_{GS+PD} and MH_{GS+RD} simulations are similar in nature. Indeed, the type of changes seen between the MH_{CNTL} and MH_{GS+PD} and MH_{CNTL} and MH_{GS+RD} simulations are very similar. We now discuss this briefly in section 4, highlighting that this shows how our approach may not be indicated for cases where the analysis focuses on similar climates displaying comparable (thermo)dynamical properties.

14. *In Fig.5 a) and b) (also Fig. 6) it is impossible to tell what the actual values of the contours are because there are only negative anomalies. Maybe add labels? In any case this figure in particular needs a significance test in order to decide which patterns are actually worth interpreting.*

Concerning the contours, the Reviewer is entirely right. We now specify the range of SLP and Z500 anomalies shown in each of the panels in the figure captions. We have opted not to add labels directly in the figures as these either covered the precipitation anomalies (if on a white background) or were not easily legible (if on a transparent background). Regarding significance, following Major Comment #4, we have added significance bounds on the precipitation anomalies. We focus on these since we are interested in detecting significant changes in monsoonal precipitation and then being able to relate these to specific SLP and Z500 patterns. The interest in the latter is not to have a locally significant anomaly in terms of magnitude (which is what standard significance tests verify), but rather a spatially coherent pattern which can explain the significant precipitation anomalies.

15. *Section 4: You say that your method can complement "other", "conventional" approaches but never name any of these other techniques. Can you give an example of a standard method with similar goals as yours? Perhaps PCA? That might help readers grasp what (approximately) your method does. You could also discuss some similarities or differences, highlighting what sets your approach apart.*

We have added a paragraph on this issue in Sect. 4, as suggested by the Reviewer. In previous work, we have compared: (i) the information extracted with d and θ to that obtained via the analysis of weather regimes (e.g. Faranda et al., 2017b; Hochman et al., 2019), often obtained by PCA or clustering methods; and (ii) the results obtained from α with those extracted from Canonical Correlation Analysis (De Luca et al., 2020b). Concerning (i), we found that the dynamical indicators provide an (almost) continuous counterpart to the heavily discretised PCA/weather regime description of atmospheric variability. We specifically found that, for a North Atlantic domain, compositing days falling in different quadrants of the d–θ space allows to recover the four canonical weather regimes (Faranda et al., 2017b). At the same time, the dynamical systems metrics provide a wealth of additional information on how the atmosphere moves within and between these weather regimes. Concerning (ii), we found that the information derived from α largely overlaps that of the CCA, yet that α was able to better capture the footprint of co-variability for extreme events (De Luca et al., 2020b). Although not applied in this specific study, the definition of α can be easily extended to a multivariate case beyond two variables, while the CCA framework requires more complex adaptations (such as partial CCA). Finally, we remark that while statistical techniques can provide valuable information on the correlation structures of recurrences, the dynamical indicators are rooted in the causal structure of the underlying dynamics of the system (e.g. a low dimension does not just point to a specific metastable state of the dynamics – or a principal component – but also informs that this state is in a more predictable region of the attractor).

16. *L. 285 can you please be a little more specific than "several good recurrences"? Very roughly how much data do I need for this method? How can I check if I have sufficiently "good" recurrences?*

While we do not have a definitive theoretical answer to this question, we have devised some procedures to test the recurrence statistics. In Faranda et al. (2011), when first applying the technique to numerical data issued from dynamical systems, we verified via a Lilliefors test whether the distribution achieved by recurrences after application of the logarithmic weight is a Gumbel law. Since deviations from this law are observed at finite time for all statistical datasets (see, e.g. Gomes and De Haan, 1999), in more recent studies (e.g. Faranda et al., 2017b) we tend to prefer sensitivity tests for the obtained values, for example by reducing the length of the datasets and repeating the estimates to check their stability. This is also the strategy chosen by Buschow and Friedrichs (2018). We now included a brief description of these issues in Sect. 4

As a final note, we would like to highlight a further change that we have implemented in the paper beyond those outlined in the replies to the Reviewer. Specifically, we have decided to remove Fig. A3. This was the only figure showing year-round geographical anomalies. It was barely mentioned in the text, since our analysis of the geographical anomalies focuses on the rainy season, and upon reviewing the manuscript we did not think it contributed with meaningful information to the overall discussion.

Additional References not Cited in the Study

- Gomes, M. Ivette, and Laurens De Haan. "Approximation by penultimate extreme value distributions." Extremes 2.1 (1999): 71-85.