Anonymous Referee #2

General comments

This manuscript studied the linear and nonlinear responses of last millennium climate models to volcanic and solar forcings. By testing i) the additivity and ii) the intermittency of the responses, the authors found i) additivity of the radiative forcings works up until roughly 50 year scales; and ii) the volcanic intermittency was much stronger than the solar intermittency, but the model responses were not very sensitive. Therefore, an important conclusion was reached, that is, linear stochastic models may be valid from over most of the macroweather range, from about 10 days to over 50 years. This study is new, and the conclusion is important. Therefore, I would like to recommend publishing this manuscript in Earth System Dynamics after a minor revision.

Authors: *We thank the referee for his positive evaluation and useful comments.*

Specific comments:

1. The paper is not well structured. In the current manuscript, there are “1 Introduction”, “2 Data and analysis”, “3 Method”, “4 Intermittency: multifractal trace moment analysis”, and “5 conclusion” five sections. The main results are shown in “3 Method”, and “4 Intermittency: multifractal trace moment analysis”. But you still can find some method description in “4 Intermittency: multifractal trace moment analysis”. When reading the manuscript, one may easy get lost. Therefore, I suggest the authors to improve the paper structure, such as i) add a new section as “Results”, and move the results shown in “3 Method” and “4 Intermittency: multifractal trace moment analysis” into the newly added “Results” section; ii) move the subsection “4.1 The Trace moment analysis technique” into the “Method” section, etc.

2. The scientific idea, as well as the results, are not well explained. The authors spent too much energy in reviewing other works, which seems to be too much in details, and not so relevant. Therefore, I would like to suggest the authors to shorten the paper and make it more compact. Some less relevant introductions can be put into supplementary Materials.

3. In the introduction, the authors summarized the scaling regimes of different time scales. They claim that the scaling behaviors is changeable. The “macroweather” regime (>10 days, $H<0$) can continue to time scales of 10-30 years (industrial) and 50-100 years (pre-industrial), after which a new $H>0$ regime is observed. They further introduce that the scaling picture has recently been extended to “macroclimate” ($H<0$, from about 80 to 500 kyr) and “megaclimate” regimes ($H>0$, from 500 kyr to at least 500 Myr). However, these results are based on the GCM controls runs and paleotemperature proxies, which may bring us with big uncertainties, or even biased scaling behaviors. I am not saying the changing scaling behaviors are incorrect, but one may need to be more careful when drawing a conclusion based on GCM control runs and paleotemperature proxies. Therefore, I would like to suggest the authors to at least mention the possible uncertainties (or even biases) in the GCM runs and paleotemperature proxies.

Technical corrections:
4. On page 1827, line 28, and on page 1828, line 1, the authors mentioned "Figure 2b (left)" and "Figure 2b (right)". Unfortunately, I cannot find in Figure 2b a left subfigure, nor a right subfigure. I guess it should be "Figure 2b (top)" and "Figure 2b (bottom)".
5. On page 1857, Figure 3a, the curve for "Multi-Proxies 1500-1900" is missing.
6. On page 1858, in the caption of Figure 3, it is confusing that there are surprisingly one sentence describing Figure 2. Line 3-4, ": : :Fig.2b left, "spliced" with a 10Be reconstruction with a 40 yr smoother, Fig. 2b right): : :" This sentence should be removed.

Authors: *We have removed the old section 3.2 and other review material that was not essential to out point. We have tightened up the introduction and given it more structure, and have made numerous other changes to improve the ms, based on the referee's comments. In addition, we thank the anonymous referee for his technical corrections, which helped to improve the manuscript in a revised version.*

Referee #1 (Comments by K. Rypdal)

GENERAL COMMENTS
Results, their relevance, and their validity
The results presented in this discussion paper are limited to assessing the linearity/non-linearity of the temperature response in two climate models, one model of intermediate complexity for the tropical Pacific (designed to describe ENSO), and one AOGCM, where the authors have confined themselves to studying results for mean northern-hemisphere land temperatures. The motivation for choice of models is not carefully explained, and their representativeness is not discussed.
The linearity issue is investigated by two methods:
(i) By considering solar, volcanic, and solar + volcanic forcing, and testing the additivity of the responses.
(ii) By testing the intermittency of the forcing and responses, assuming that in a linear system the intermittency in forcing and response should be the same.
By method (i) it is found that solar and volcanic responses in the models do not add up on time scales in the range 300-1000 yr. The result is based on neglecting the estimated correllation between responses to solar and volcanic forcing, respectively (section 3.4 and Fig. 3). This approximation is justified from the the statistical independence of solar and volcanic forcing. But in the model experiments these forcings are given as deterministic and do not vary over the statistical ensemble, so the estimated ensemble average over the product of these forcings is not zero. This approximation is unnecessary and may be the cause of the non-additivity result. If the authors believe it is not, they should estimate the Haar fluctuation of the sum of solar and volcanic responses directly, without using this approximation, and demonstrate that it does not change their result.

Authors: *Thanks for this suggestion. We implemented it (the revised fig. 3c) and it makes a little difference but doesn't change the conclusions.*

Method (ii) is based on the theoretical fact that if the response is linear, the response kernel is a perfect power-law function, and the forcing is perfectly multiscaling, then the intermittency is the same for response and forcing. If the intermittencies are different

the authors take it as a proof of nonlinearity of the response. However, there are at least two differents tests that need to be done before one can draw this conclusion:
(a) Theoretical and estimated scaling is not the same. In order to test that the estimated intermittency is the same for the actual forcing and the response from a linear power law response model, the authors should use such a model and apply the trace-moment analysis to the response computed using this model. If the trace moments are the same as for the forcing, they can proceed to the next step.

Authors: *Actually, the numerics are robust: enough tests have been done over the last thirty that we can have confidence in the trace moment technique (see e.g. [Lavallée et al., 1991] for extensive numeric tests). In actual fact, the effect here is so strong that one can detect by eye (fig. 1) the much lower "spikiness" or intermittency of the response when compared to the volcanic forcing. As indicated in the text, this was noticed over twenty years ago. Therefore I don't think that the basic result is in doubt.*

(b) In this step they should question their assumption of perfect power-law scaling of the linear response. It is well known that there must be a cut-off of this response at large time scales (Rypdal and Rypdal, 2014). A cut-off at scales from a few decades to a century can easily explain the difference in intermittency. The authors should test if introduction of such a cut-off (or use of other plausible response kernels) will change the trace moments in the linear model and make them more similar to the trace moments of the actual temperature signal.

Authors: *It is not at all well known that there is large scale truncation, indeed where is the evidence! All that is known is that there is a break in the scaling at some large scale between about one century and several millennia, probably depending on geographical location and epoch (see the reference in the text to the Holocene). This break is not synonymous with a truncation.*

Structure and style
The paper has the form of a broad review of work by Lovejoy and co-workers, spanning most of the 16 self-citations. Most of this material is irrelevant for the interpretation of the results developed in the present paper. There is hardly a need for another review of dr. Lovejoy's work in his field in addition to the monograph Lovejoy and Schertzer, 2013. In this review I restrict myself to those aspects that are relevant for the new results presented. It does not mean that I approve of everything that is not commented.

General judgement
The manuscript is not suitable as a research article in ESD in its present form. My reservations described in points (i) and (ii) above have to be addressed and proven wrong, and a drastic shortening of the manuscript is necessary. The authors should adhere to the principles for a regular research article.

Authors: *The review material was an attempt to explain the context of the problems in enough detail so that they could be understood in a fairly self-contained way. We have removed quite a lot of material in the new text and changed the structure, especially in the first part.*

Anonymous Referee #3

The authors analysis output from millennium experiments with the Zebiac-Cane model and the GISS model. They conclude that both models underestimate variability at centennial scaled compared to observations, and also observe a phenomenon of 'subadditivity' in the ZC model.
(1) One of the surprising findings featured in this article is the 'subadditivity' of the Zebiac-Cane model. When it is forced by both solar and volcanic forcings, the ZC model has a spectrum response close to the simulations with volcanic forcing only, as if the solar forcing had been ignored. The seasoned modeller would be tempted to attribute the result to a trivial mistake in the experiment design. Assuming that chances of mistakes have been checked and eliminated, we need to find an explanation to this result and discuss wisely its implications for our understanding of climate dynamics. We remember that the ZC model was developed specifically to study tropical Pacific interannual variability, and in particular the ENSO phenomenon. It does not have deep ocean dynamics, nor extratropical atmospheric dynamics, which are two processes which may significantly interplay with interdecadal variability. Lacking ocean modes of motions active at times scales over a few years, the use of the ZC model in a study focused on long-memory processes and non-linearity at time scales of several hundreds of years is highly contentious.

The inadequacy of the ZC model for spectral analysis at scales over decades is a case for rebuttal of the article.

Authors' response:

We agree that the ZC model is not the theoretically optimal model for this problem. However, as we indicated, there are no equivalent suites of models that are better: no Millenium simulations exist with the necessary suite of: solar, volcanic, solar plus volcanic simulations.

That being said, there are clearly sources of low frequency variability present in the ZC model. For example, using 360 year control runs, [Goswami and Shukla, 1991] showed that due to its internal variability, that the ZC model can generate very significant multidecadal and centennial low frequency variability due to the feedbacks between SST anomalies, low level convergence and atmospheric heating. In justifying his Millenium ZC simulations, Mann specifically cited model centennial scale variability as a motivating factor. Therefore, it isn't perhaps so surprising that we find sub-additivity at scales ≈ 50 years and longer, although we agree that the conclusions are not so strong on this point, and the source of the nonlinearity in the models needs to be pin-pointed.

References

Goswami, B. N., and Shukla J.: Aperiodic Variability in the Cane—Zebiak Model, J. of Climate, 6, 628-638, 1991.

Lavallée, D., Lovejoy, S., and Schertzer, D.:On the determination of the codimension function, in Non-linear variability in geophysics: Scaling and Fractals, edited by D. Schertzer and S. Lovejoy, pp. 99-110, Kluwer, 1991.