Manuscript prepared for Earth Syst. Dynam. Discuss.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 21 January 2016

Review of revised paper by Lovejoy and Vorotsos

K. Rypdal[1] **and M. Rypdal**[1]

[1]Department of Mathematics and Statistics, UiT The Arctic University of Norway, Norway

Correspondence to: Kristoffer Rypdal (kristoffer.rypdal@uit.no)

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

This referee report will comment on the revised manuscript and the authors' reply to the first referee report. Martin Rypdal has joined in to speed up and secure the quality of the review process.

We have been asked to help the editor make a final decision about publication of this paper in ESD. Our general position is that the issue is important, and although we find many weaknesses in this paper, we have no desire to prevent it from being published, provided our comments are published as part of the public discussion. We find similar weaknesses in other of Shaun Lovejoy's publications, and it may be better that these issues are publicly discussed.

Thus, our recommendation is that the editor reads our comment and use it as a background for the decision.

The subadditivity in the ZC model

In the original manuscript the analysis leading to the subadditivity in the ZC-model was based on an incorrect neglect of a cross-term leading to Eq. (5). The authors were asked to use the full expression and correct the result in Fig. 3 according to this. In their reply the authors acknowledge the error, state that they have corrected it in Fig. 3, but claim that the correction does not alter the conclusion. However, the actual revision they have made is rather disturbing. In the text the incorrect Eq. (5) and the incorrect argument for it is maintained, and so is the conclusion that the "theoretical additive fluctuation level is reduced by a factor ≈ 2.5." Then they add the following paragraph:

"It should be noted that the latter holds assuming independence (pink curve in Fig. 3c) of the solar and volcanic forcing. For comparison, the purple curve in Fig. 3c illustrates the results obtained when analyzing the series constructed by directly summing the two response series (instead of assuming statistical independence). It is clearly seen that the basic result still holds but it is a little less strong (a factor of $\approx$ 2). The reason for the difference is that the cancellation of the cross terms assumed by statistical independence is only approximately valid on simple realizations, especially at the lower frequencies where the statistics are worse."

In Fig. 3b they keep the incorrect curves for additive response and the ratio, and do not plot the correct curve, and in Fig. 3b they also keep the incorrect curve for the ratio, but here they also add the correct curve. For unknown reasons they plot this ratio as a logarithm, so the reader will have to use a calculator to find out what the correct ratio is. The mean $\log_{10}$ ratio in the narrow scale range 250-1000 yr the ratio is ≈ 1.4. In the revised text in Section 3.4 (line 339) the authors claim that this corrected ratio is ≈ 2, and in the caption of Fig. 3 that it is ≈ 1.6. In the concluding section (line 522) it has again risen to "a factor $\approx 2-2.5$."

This way of "correcting" a flawed analysis is unacceptable. As a minimum the authors should take the following actions:

1. Eq. (5) and the arguments leading to it should be taken out of the manuscript. The text gives the reader the false impression that linearity implies additivity of *variances*. The truth is that linearity implies additivity of the *responses*, and this is what should be tested, and nothing else. Moreover, making this approximation does not simplify anything, so there is no scientific reason to make it.

2. In Fig. 3b the incorrect curve for the additive response and the ratio should be replaced by the correct one, and in Fig. 3c the incorrect curve should be removed. In all panels in Fig. 3 it would be easier to grasp the content if the vertical axes are linear, not logarithmic.

3. The authors must refrain from cheating with numbers. The true result of the correct analysis is that the rms-ratio is ≈ 1.4; not 2.5, not 2, and not 1.6. The paper should present only the correct number 1.4.

A linear framework cannot ignore internal noise from stochastic forcing

But the problems do not end here, because the output data from the ZC model contains a noise contribution from internal variability in addition to the model response to forcing. Ideally, if the ensemble size is infinite, the ensemble mean could be interpreted as a "deterministic" response, but as is clearly shown in Fig. 1 of *Mann et al.* (2005) the magnitude of the high-frequency noise depends on the number of realisations averaged over (they show results for 5, 20, and 100 realisations), and even after averaging over 100 realisations the RMS of the noise is as high as the deterministic response. For instance, the 40 yr moving average in Mann et al., Fig. 1c (the maroon curve) can be interpreted as the deterministic response to solar forcing (the blue curve), and the difference between the red and the maroon curve is the noise that remains after averaging the internal variabilty over 100 realisations.

This noise turns out to contribute to the subadditivity and completely explains the factor 1.4, even if the deterministic response is strictly linear!

Proof:

We start from the hypothesis that the model responds linearly to external forcing, but that there is an internal noise level which is independent of the forcing. The temperature after averaging over 100 realisations is $T(t) = T^{\text{det}}(t) + \epsilon(t)$, where $T^{\text{det}}(t) = \hat{L}[F(t)]$ is the deterministic response to the forcing $F(t)$ and $\epsilon(t)$ is the noise that remains after averaging the internal fluctuations over those realisations. Since the response operator $\hat{L}$ is linear we have $T^{\text{det}}_{v+s}(t) = \hat{L}[F_s(t) + F_v(t)] = \hat{L}[F_s(t)] + \hat{L}[F_v(t)] = T^{\text{det}}_s(t) + T^{\text{det}}_v(t)$, where F_s and F_v represent solar and volcanic forcing and T^{det}_s and T^{det}_v the response to these forcings. T^{det}_{v+s} is the response to the combined forcing $F_s + F_v$. The next step is to produce a fluctuation $\Delta T(t, \Delta t)$ by means of a linear wavelet operation, e.g., the Haar wavelet. This yields

$$\Delta T_s(t, \Delta t) = \Delta T^{\text{det}}_s(t, \Delta t) + \epsilon_s(t, \Delta t),$$

$$\Delta T_v(t, \Delta t) = \Delta T^{\text{det}}_v(t, \Delta t) + \epsilon_v(t, \Delta t),$$

$$\Delta T_{s+v}(t,\Delta t) = \Delta T^{\text{det}}_{s+v}(t,\Delta t) + \epsilon_{s+v}(t,\Delta t) = \Delta T^{\text{det}}_{s}(t,\Delta t) + \Delta T^{\text{det}}_{v}(t,\Delta t) + \epsilon_{s+v}(t,\Delta t).$$

Here $\Delta\epsilon_s(t,\Delta t)$, $\Delta\epsilon_v(t,\Delta t)$, and $\Delta\epsilon_{s+v}(t,\Delta t)$ are the fluctuations of independent, realisations of the same noise process $\epsilon(t)$ = the average over 100 realisations of internal variability. The final step is to form the variance $\langle \Delta T(t,\Delta t)^2\rangle$, where $\langle\ldots\rangle$ denotes averaging over disjoint time intervals of length Δt. From the equations above, using that ϵ_s, ϵ_v and ϵ_{s+v} are independent, we find

$$\langle \Delta T_{s+v}(t,\Delta t)^2\rangle = \langle \Delta T^{\text{det}}_{s+v}(t,\Delta t)^2\rangle + \langle \Delta\epsilon(t,\Delta t)^2\rangle,$$

$$\langle \Delta T_{s}(t,\Delta t)^2 + \Delta T_{v}(t,\Delta t)^2\rangle = \langle \Delta T^{\text{det}}_{s+v}(t,\Delta t)^2\rangle + 2\langle \Delta\epsilon(t,\Delta t)^2\rangle.$$

The RMS-ratio computed by the authors, and displayed in Fig. 3c is

$$\text{RMS-ratio} = \sqrt{\frac{\langle \Delta T_{s}(t,\Delta t)^2 + \Delta T_{v}(t,\Delta t)^2\rangle}{\langle \Delta T_{s+v}(t,\Delta t)^2\rangle}} = \sqrt{1 + \frac{\langle \Delta\epsilon(t,\Delta t)^2\rangle}{\langle \Delta T^{\text{det}}_{s+v}(t,\Delta t)^2\rangle + \langle \Delta\epsilon(t,\Delta t)^2\rangle}}.$$

Since, after averaging over 100 realisations, the RMS of the noise is as high as the deterministic response, the second term under the root-sign on the right is typically of the order unity. If the internal variability after averaging over 100-realisations is close to 1/f noise, then the numerator $\langle \Delta\epsilon(t,\Delta t)^2\rangle$ may be relatively independent of the scale Δt, while the denominator $\langle \Delta T^{\text{det}}_{s+v}(t,\Delta t)^2\rangle$ will be larger on the small scales where volcanoes contribute strongly. Hence this term arising from the noise may explain the entire deviation from unity of the RMS-ratio shown by the magenta curve in Fig. 3c.

Our conclusion is that the results shown in Fig. 3 are completely consistent with a linear response in the ZC model.

Multifractality and linearity in responses

The equations that constitute the climate models studied in this paper are nonlinear. The noise (internal variability) in the models (as appearing in control runs) is a result of nonlinear dynamical processes. However, the issue studied in this paper is whether the responses to external forcing add up. It is not easy for us to see the relevance of all the multifractal formalism presented in Section 4, but the essence seems to be a mathematical corollary claiming that linearity in the response implies that the intermittency function $K(q)$ is the same for forcing and response (lines 400-410). However, as was pointed out in the first referee report, this results depends on a particular form of the linear response function, namely that it is a power law, which is necessarily an idealisation, since it leads to infinite responses on long time scales. In their reply the authors claim that there is no cut-off of the power-law tail of the response function. This is in direct contradiction to their claim that GCMs do not reproduce low-frequency (multicentennial) variability. If the linear response function has a power-law tail up to multicentennial scales, then a white-noise stochastic forcing would produce an internal variability with a power-law spectrum up to these scales, and this would be reflected in the spectrum of control runs. This is what the authors claim does *not* happen in GCMs on lines 82-83. In the referee report the authors were asked to test their results by using also other response functions, but they have refused to do that.

The lack of accounting for internal variability also arises in the multifractal analysis. If the data output from the climate model is one realisation (or the average over limited number of realisations), and this output is modeled as a linear response to a forcing, then it is necessary to assume the existence of a stochastic forcing. Otherwise, the linear model cannot account for the internal noise in control runs of GCMs and the noisy output in the ZC model with smooth solar forcing observed in Fig. 1c of *Mann et al.* (2005). Even more striking is Fig. 1a of *Mann et al.* (2005) for volcanic forcing, where the 100-realisation mean is completely dominated by this noise for all times except 1-2 years after major eruptions. But this stochastic forcing is not included in the computation of structure functions of the forcing in Fig. 6. Thus, the linear model put to test is a model that cannot produce internal variability, but it is tested against data that is known to exhibit such variability.

The multifractal analysis of the climate model output is mostly an analysis of the internal noise, and hence there is no reason to find a linear connection between forcing and output.

The weaknesses discussed above are all symptoms of a missing hypothesis-testing strategy. The linearity hypothesis is not clearly formulated, which creates confusion with respect to the rôle of internal variability. There are of course always some deviation from linearity, but is no attempt in this paper to test if the observed deviations are significant. This is particularly problematic for the multifractal analysis. Here results are notoriously unconvincing unless they are followed up by Monte Carlo simulations of linear response models with a range of response functions and stochastic forcing. Such simulations will provide information about biases and statistical errors, and could be used to reject a linear response hypothesis if this hypothesis is false.

Other comments

The first referee-report recommended a drastic shortening of this review-style paper, and therefore refrained from commenting on the extensive discussions of issues not directly related to the new results presented. The authors have not followed this recommendation (which was also made by referee #3), so we feel that we in this final referee report should also make some comments on these parts of the manuscript.

Section 1.1

Line 36-44: Here the authors give the impression that assuming a linear response is synonymous with neglecting feedbacks. This is of course nonsense. Feedbacks, like the ice-albedo feedback, can in many cases, and on different time scales, be accounted for by constant feedback factors, i.e., in a linear framework. But there are of course also situations where this is impossible, in particular at times when the climate system undergoes transitions. The validity of a linear approximation is not primarily a question of time-scale, but of the structural stability of the system. For instance, if the climate system is close to a tipping point, the nonlinearities that are responsible for the bifurcation will be important and detectable in the response to external forcing as well as to "internal" stochastic forcing. The papers by other authors referred to do not assume linearity as a general feature of the climate system, but as reasonable approximation for global responses to realistic forcing in the present stable Holocene climate state.

Section 1.2

This subsection is unnecessary as a background for assessing the results presented in the paper. Our understanding of the different spectral regimes is very different from what is presented here. For instance, in a recent paper just acepted for ESD (where Shaun Lovejoy was a referee) we demonstrate that the spectra of global temperature can be described as a relatively non-intermittent background noise with spectrum of the form $1/f^{\beta}$, $\beta \approx 1$, superposed on a succession of rapid (nonlinear) transitions. The authors end the subsection with stating that opinions diverge on the value of the global transition scale. This is not a correct description of the disagreement, since we don't accept the very notion of a transition scale in the Holocene climate.

Section 1.3

Line 82-83: "From the point of view of the GCMs, the low-frequency (multicentennial) variability *arises exclusively as a response to external forcings*, although potentially - with the addition of (known or currently unknown) slow processes such as land-ice or biogeochemical processes - new internal sources of low-frequency variability could be included."

This is a gross overstatement. In a recent paper (*Fredriksen and Rypdal*, 2015) we analyse an ECHO-G control run and a large number of long control runs in the CMIP5 ensemble. ECHO-G shows near-perfect long-memory power-law scaling up to a millennium, while most CMIP5 models exhibit a flattening tendency of the spectrum on scales larger than a century. *Østvand et al.* (2014) analysed millennium-long control runs and runs with full forcing of the ECHO-G and COSMOS models. Also analysed were forced runs of HadCM3, and LOVECLIM models (no controls were available), and here the residuals after subtracting a forced, deterministic response were also subject to scaling analysis. The results showed no enhanced variability on the multicentennial scales by inclusion of forcing, and the internal variability showed no deviation from power-law scaling with $0.75 < \beta < 1$ on the scales available in these experiments. This demonstrates that GCMs differ among each other when it comes to variability on multicentennial scales. Various techniques are used to deal with phenomenon of model drift, which will influence variability on large time scales. But these issues with the GCMs cannot be summarised as the authors do on lines 82-83.

Section 3.1 lines 187-192

Here the authors suggest to use as a test of model performance to check the equality $\Delta T_{sim} \stackrel{d}{=} \Delta T_{obs}$. We don't disagree with that, but what the authors do not specify is that in order to compare statistics of models with the statistics of observations, ΔT_{sim} should be computed from single realisations of model output, and *not* from ensemble averages over many model runs as they do in their analysis of the ZC model.

References

Fredriksen, H.- B., and Rypdal, K.: Scaling of Atmosphere and Ocean Temperature Correlations in Observations and Climate Models, *J. Climate*, e-view, doi: http://dx.doi.org/10.1175/JCLI-D-15-0457.1, 2015.

Mann, M. E, Cane, M. A., Zebiak, S. E., and Clement, A.: Volcanic and solar forcing of the tropical pacific over the past 1000 years, J. Climate, 18, 447-456, 2005.
DOI: 10.1175/JCLI-D-15-0183.1.

Østvand, L., Nilsen, T., Rypdal, K., Divine, D., and Rypdal, M.: Long-range memory in internal and forced dynamics of millennium-long climate model simulations, Earth. Syst. Dynam., 5, 295-308, doi:10.5194/esd-5-295-2014, 2014.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |