Comment on "Scaling regimes and linear/nonlinear responses of last millennium climate to volcanic and solar forcing" by S. Lovejoy and C. Varotsos

Kristoffer Rypdal[1] and Martin Rypdal[1]

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

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

Abstract. *Lovejoy and Varotsos* (2016) (L&V) analyse the temperature response to solar, volcanic, and solar plus volcanic, forcing in the Zebiak-Cane (ZC) model, and to solar and solar plus volcanic forcing in the GISS-E2-R model. By a simple wavelet filtering technique they conclude that the responses in the ZC model combine subadditively on time scales from 50 to 1000 yr. Nonlinear
5 response on shorter time scales is claimed by analysis of intermittencies in the forcing and the temperature signal for both models. The analysis of additivity in the ZC model suffers from a confusing presentation of results based on an invalid approximation, and from ignoring the effect of internal variability. We present a test without this approximation which is not able to reject the linear response hypothesis, even without accounting for internal variability. We also demonstrate that internal vari-
10 ability will appear as subadditivity if it is not accounted for. The analysis of intermittencies is based on a mathematical corollary stating that the intermittencies of forcing and response is the same if the response is linear. We argue that there are at least three different factors that may invalidate the application of this corollary for these data. First, the corollary is valid only for a power-law response function. This implies a strong response on centennial time scales, which the authors claim does not
15 take place in these models. Second, it assumes power-law scaling of structure functions of forcing as well as temperature signal, which is not the case for these data. And third, the internal variability, which is strong at least on the short time scales, will exert an influence temperature intermittence which is independent of the forcing. We demonstrate by a synthetic example that the differences in intermittencies observed by L&V easily can be accounted for by these effects under the assumption
20 of a linear response. Our conclusion is that the analysis performed by L&V does not present valid evidence for a nonlinear response in the global temperature in these climate models.

1 Introduction

The issue of linearity in the global temperature responses of modern General Circulation Models (GCMs) and Earth System Models (ESMs) is important because the prospect of predicting global
25 aspects of the climate under different forcing scenarios is considerably brighter if the response is

reasonably linear. Linear response models with two characteristic response times or a long-memory power-law response have had considerable success in describing global temperature response in GCM data, instrumental data and in multiproxy reconstructions (*Geoffroy et al.*, 2013; *Rypdal and Rypdal*, 2014; *Østvand et al.*, 2014; *Rypdal et al.*, 2015; *Lovejoy et al.*, 2013; *Fredriksen and Rypdal*,

30 2015). The credibility of these results depends crucially on the validity of the linear approximation in the global response. Particularly relevant is *Geoffroy et al.* (2013), who estimate the parameters of a linear two-box energy balance model by data from runs of a large number of CMIP5 ESMs with step-function forcing and linearly increasing forcing, respectively. Very good fits to the simulated global temperature are found in this study, with the same values of the two-box model parameters

35 for the two different forcing scenarios. This is a very clear demonstration of the linearity of the global temperature response in the CMIP5 ensemble, and there are others that find only weak nonlinearities in some models (see *Andrews et al.* (2012) and references therein).

The paper by *Lovejoy and Varotsos* (2016) (in the following denoted L&V) is a research paper, but has the character of a review of earlier papers of Shaun Lovejoy and coworkers. The review

40 style has the unfortunate effect of masking the substance of the new results presented, which is an analysis of the responses in two different climate models to solar and volcanic forcing, and to combinations of these forcings. The actual analysis is made in Section 3.4 of the L&V paper, where the authors test the additivity of responses to solar and volcanic forcing in the Zebiak-Cane (ZC) model, and Section 4.2, where they study the intermittency of forcing and response and conclude

45 that difference in their intermittency implies nonlinearity of the response. In Sect. 2 of this comment we present a critical examination of the methods leading L&V to conclude that combined solar and volcanic forcing leads to a weaker response than the sum of the solar and volcanic responses in the ZC model. Sect. 3 examines the intermittency analysis and demonstrates that L&V's results for the ZC model can be reproduced in the response of a simple linear response model.

2 Additivity of global response in the ZC model

2.1 The linear response null hypothesis

There is an infinity of ways the response can be nonlinear, so the only reasonable approach is to formulate a test that may, or may not, reject the null hypothesis that the response is linear. This linear hypothesis, however, must be formulated with some care. Hydrodynamic flow models like

55 the ZC and GCMs are inherently nonlinear. "Unforced" control simulations are of course driven by the constant solar energy flux, and this results in a turbulent, nonlinear cascade that forms the "internal variability" of the model. In a linear model for the global response this internal variability is represented as a noise process $\epsilon(t)$ in a global variable $T(t)$. Forcing $F(t)$ in the model means a variation of the global energy flux around the flux that drives such a turbulent equilibrium state.

60 After these remarks we are ready to formulate the *linear response hypothesis*:

(i) For realistic strength of the global forcing the internal variability $\epsilon(t)$ is unaffected by the forcing.

(ii) The global temperature can be expressed as a sum of this internal variability and a linear response to the forcing, i.e.,

$$T(t) = T^{\mathrm{det}}(t) + \epsilon(t), \ \ T^{\mathrm{det}}(t) = \hat{L}[F(t)], \tag{1}$$

where $T(t)$ is the global surface temperature, $T^{\mathrm{det}}(t)$ is the deterministic, linear response to the global forcing $F(t)$, and $\hat{L}$ is the linear response operator.

2.2 Internal noise and response additivity

The data used from the ZC model is the temperature (more precisely; the Niño3 index) after averaging over 100 simulations with the same forcing. If the internal variability is a persistent noise, averaging over N independent runs will reduce the standard deviation by a factor $N^{-1/2} = 0.1$, but the correlation structure of the noise will be preserved. In the following, $\epsilon(t)$ is the noise that remains after averaging the internal noise over those N realisations.

The next step is to produce a fluctuation $\Delta T(t, \Delta t)$ by means of a linear low-pass filtering operation. It could for example be a simple moving average over a window Δt, or the Haar wavelet smoothing employed by L&V. In the following we shall for notational simplicity omit the arguments $(t, \Delta t)$. The results presented hold for the temperature signal itself ($\Delta t = 0$) as well as for any degree Δt of filtering. Since the response operator $\hat{L}$ is linear we have

$$\Delta T_{v+s}^{\mathrm{det}} = \Delta T_s^{\mathrm{det}} + \Delta T_v^{\mathrm{det}}, \tag{2}$$

where $\Delta T_s^{\mathrm{det}}$ and $\Delta T_v^{\mathrm{det}}$ are the responses to the solar and volcanic forcings, ΔF_s and ΔF_v, respectively, and $\Delta T_{v+s}^{\mathrm{det}}$ is the response to the combined forcing $\Delta F_s + \Delta F_v$. This yields

$$\Delta T_s = \Delta T_s^{\mathrm{det}} + \Delta \epsilon_s, \tag{3}$$
$$\Delta T_v = \Delta T_v^{\mathrm{det}} + \Delta \epsilon_v, \tag{4}$$
$$\Delta T_{s+v} = \Delta T_s^{\mathrm{det}} + \Delta T_v^{\mathrm{det}} + \Delta \epsilon_{s+v}. \tag{5}$$

Here $\Delta \epsilon_s$, $\Delta \epsilon_v$, and $\Delta \epsilon_{s+v}$ are the filtered fluctuations of independent realisations of the same noise process $\epsilon(t)$ (here $\epsilon(t)$ is the average over 100 realisations of internal variability). By subtracting Eqs. (3) and (4) from Eq. (5), and using Eq. (2), we find

$$\Delta T_{s+v} - \Delta T_s - \Delta T_v = \Delta \epsilon_{v+s} - \Delta \epsilon_s - \Delta \epsilon_v \equiv \Delta \varepsilon. \tag{6}$$

Here, $\Delta \varepsilon$ is the sum of three independent realisations of the same noise process $\Delta \epsilon \stackrel{d}{=} \Delta \epsilon_s \stackrel{d}{=} \Delta \epsilon_v \stackrel{d}{=} \Delta \epsilon_{v+s}$, where $\stackrel{d}{=}$ is identity in distribution. This imples that

$$\Delta \varepsilon \stackrel{d}{=} \sqrt{3} \Delta \epsilon. \tag{7}$$

Hence, a prediction based on the linear response hypothesis is that the difference between the temperature driven by combined solar and volcanic forcing and the sum of the temperatures driven by solar and volcanic forcing is realisation of a noise process which is $\sqrt{3}$ times the internal variability
95 process. In the next subsection we shall test this prediction on the data from the ZC model. If the prediction is inconsistent with the data the linear response hypothesis is rejected for this model. If the prediction is confirmed by the data the linear hypothesis stands stronger.

2.3 Alternative test of additivity in the ZC model

Fig. 1 (a) shows time series of the solar and volcanic forcing for the last millennium used in the
100 simulations of the ZC model. Unfortunately we do not have available control runs on millennial scale from this model. This would have been very useful in establishing directly the statistical properties of the internal noise $\epsilon(t)$. The approach we will use as an alternative, is to assume the validity of the linear response hypothesis, which will allow us to extract the internal noise from the simulation with solar forcing only. Then we will formulate a test by which the hypothesis can be rejected by
105 the data for volcanic forcing only and for volcanic plus solar forcing. Assuming the validity of the linear hypothesis from the start may seem like circular reasoning, but it is not. Any valid hypothesis testing makes predictions based on the null hypothesis, which are then tested against observation.

If the linear response hypothesis is true we can determine $\epsilon(t)$ from the solar forcing signal and the corresponding temperature signal. The solar forcing signal in Fig. 1a has a smooth appearance,
110 in particular for the first 750 years of the record, when no sunspot counts were available. As a contrast, the corresponding temperature signal shown as the thin orange curve in Fig. 1b is noisy on all scales down to the annual scale. This appearance of the temperature signal under the smooth solar forcing already lends support to the assumption that the variability up to century time scale is internal. However, according to L&V the subadditivity is most prominent on time scales longer than
115 50 yr, so we have to pay special attention to the slow components of the noise spectrum. We now write a linear response to the solar forcing in the form;

$$\Delta T_s^{\text{det}}(t, \Delta t) = -S \Delta F_s(t - \tau, \Delta t).\tag{8}$$

Here $\Delta t = 50$ yr over which we have performed a moving average of the temperature and forcing. The time lag τ of the response is estimated to be ≈ 25 yr from inspection of the filtered time series.
120 The climate sensitivity S is chosen to give the best least-square fit of $\Delta T_s^{\text{det}}(t, \Delta t)$ (the black curve in Fig. 1b) to the filtered temperature signal $\Delta T_s(t, \Delta t)$ (the orange, thick curve).

Because of the smooth character of the solar forcing signal in the first 750 yr of the record, the 50 yr filtering of this signal has almost no effect, and we can therefore interpret the black curve in Fig. 1b as the linear, deterministic response to the solar forcing, and the difference between the
125 orange, thin curve and the black curve as the internal noise, i.e.,

$$\epsilon(t) = T_s(t) - \Delta T_s^{\text{det}}(t, \Delta t).\tag{9}$$

This difference is plotted as the brown, thin curve at the bottom of Fig. 1b, and the thick brown curve
is the 50 yr moving average.

We have now distinguished the internal noise from the solar-driven temperature signal by means
of the linear-response assumption. The orange bullets in Fig. 1d is a characterisation of this noise by
means of the Haar structure function employed by L&V. The definition of this structure function is

$$\sqrt{S_2^{\text{Haar}}(\Delta t)} = \langle |\Delta T(t, \Delta t)|^2 \rangle^{1/2}, \tag{10}$$

where $\langle \ldots \rangle$ denotes averaging over disjoint time intervals of length Δt. It measures the root-mean-
square fluctuation level on the scale Δt. The flat appearance on scales above a decade indicates a
strongly persistent noise process with equally strong fluctuations on scales $\Delta t > 10$ yr. The straight-
line character of the log-log plot in this scale range is symptomatic of a scaling process, and the
corresponding power spectral density has the form $\sim f^{-\beta}$, where $\beta \approx 1$ (sometimes denoted $1/f$-
noise or pink noise). The higher fluctuations for $\Delta t < 10$ yr is characteristic for the El Niño Southern
Oscillation (ENSO). This mode is particularly strong in the ZC model, which is designed specifically
for the study of ENSO, and the global output $T(t)$ is the so-called Niño3 index.

If the characterisation we have made of the internal noise is correct, and the linear hypothesis is
true, then Eq. (7) must be true. But ε in Eq. (7) must be computed from Eq. (6), which requires the
temperature signals T_v and T_{u+v}, in addition to T_s. The characterisation of ϵ only used T_s, so if the
linear hypothesis is false, it is very unlikely that the estimated ε and ϵ will give good agreement with
Eq. (6). This means that we should have a strong test.

In Fig. 1c the thin, blue curve represents $T_{s+v}(t)$, the thin, red curve is $T_s(t) + T_v(t)$, and the thin,
black curve is their difference $\varepsilon(t) = T_s(t) + T_v(t) - T_{u+v}$. Note that the narrow spikes from the fast
responses to the volcanic eruptions are completely absent in the difference signal $\varepsilon(t)$, demonstrating
that the addition of solar forcing does not influence the response to the volcanic eruptions on the
short time scales up to a few years. The thick curves in Fig. 1c are the corresponding 50 yr moving
averages. The Haar structure function of the signal $\varepsilon(t)$ is shown as the red bullets in Fig. 1d. The
brown bullets are $\sqrt{3}\epsilon(t)$, i.e., the orange bullets multiplied by $\sqrt{3}$. We observe that the red and
brown bullets are more or less on top of each other, which means that the second-order statistics of
the noise processes $\varepsilon(t)$ and $\sqrt{3}\epsilon$ are the same, in agreement with Eq. (7). Thus, this test is not able
to reject the linear response hypothesis.

2.4 Examination of L&V's test of response additivity

The L&V test of additivity is simpler, but ignores internal variability. Here we shall demonstrate
that their test also fails to reject the linearity hypothesis, even when this variability is not taken
into account. Their main conclusion concerning additivity of responses in the ZC model is that for

160 $\Delta t > 50$ yr the rms-ratio,

$$R \equiv \sqrt{\frac{\langle |\Delta T_s + \Delta T_v|^2 \rangle}{\langle |\Delta T_{u+v}|^2 \rangle}}, \tag{11}$$

is found to be $R \approx 1.5$. As will be shown below, our analysis yields a number closer to unity. But the authors also make attempts in their Figure 3 to inflate this ratio further by presenting approximate results for the numerator based an invalid (and completely unnecessary) approximation. The authors

165 admit in the published paper that this analysis is wrong, but still they print the old graphs based on the faulty analysis in their Figure 3. For instance, in Figure 3b they show the flawed graph of $\sqrt{\langle |\Delta T_s + \Delta T_v|^2 \rangle}$ together with the graph of $\sqrt{\langle |\Delta T_{u+v}|^2 \rangle}$, which appears to show that the former is larger than the latter by a factor ≈ 2.5 for $\Delta t > 50$ yr. In Fig. 2 we show the results that we obtain without the faulty approximation. This figure should be compared to Figure 3b in the L&V

170 paper, where the faulty result for $\sqrt{\langle |\Delta T_{u+v}|^2 \rangle}$ is plotted rather than the correct one, and gives a false impression of subadditivity. When the correct curve is plotted, we cannot find any significant difference between the two graphs (red and blue bullets) for $\Delta t < 300$ yr. For $\Delta t > 300$ yr the statistics is so poor that the observed differences seem statistically significant.

An alternative, and very simple, estimate for this ratio can be obtained from the data for the red

175 and blue, thick curves in Fig. 1c, by computing ΔT's as 50 yr moving averages rather than Haar fluctuations. The standard deviation of $\Delta T_s + \Delta T_v$ is 0.072 K, and of ΔT_{s+v} it is 0.060 K, which yields $R \approx 1.20$. This ratio is slightly greater than unity due to the higher fluctuations in the red graph compared to the blue graph in Fig. 2 for $\Delta t > 300$ yr. Since this difference on the longest time scales appears to be a statistical error due to limited sample size, $R = 1$ may be within the

180 error bars of the estimated R. If such an error test were crucial, we could have computed the error bars via a Monte Carlo ensemble of the $1/f$ noise process. However, as will be shown in the next subsection, internal variability gives an additional positive contribution to R which exceeds the error that is required to explain the estimate $R \approx 1.2$ under the linear response hypothesis.

2.5   The effect of internal variability on the L&V test

185 The ratio R defined in Eq. (11) only measures the ratio of responses if the internal noise is negligible. Hence, even if R were significantly (in statistical sense) greater than unity, this increase might be caused by the internal variability in a model whose response to forcing is perfectly linear. By using Eq. (6), which is valid for a linear response model, Eq. (11) can be written as

$$R = \sqrt{1 + \frac{\langle |\Delta \varepsilon|^2 \rangle}{\langle |\Delta T_{s+v}|^2 \rangle}}. \tag{12}$$

190 This shows that internal noise can increase the rms-ratio computed by L&V even if the response is linear. From the data for the thick, brown curve in Fig. 1b we have that the standard deviation for the internal noise $\Delta \epsilon$ is 0.03, and hence for $\Delta \varepsilon$ a factor $\sqrt{3}$ larger. The standard deviation of

ΔT_{s+v} can be estimated from the data for the thick, blue curve in Fig. 1c and is 0.06. This yields $\langle|\Delta\varepsilon|^2\rangle/\langle|\Delta T_{s+v}|^2\rangle \approx 0.75$, and hence $R \approx 1.32$ is the estimate of the rms-ratio based on the linear response hypothesis.

2.6 L&V's arguments against high internal variability

In the first and second drafts of the L&V discussion paper internal variability was not mentioned. After this problem was raised by us in the interactive discussion, L&V have in the final paper presented two arguments against the presence of sufficiently high internal fluctuations on the centennial time scales to explain the raised rms-ratio R.

The first argument uses the internal variability of the GISS model as an estimate of the centennial scale internal variability of the ZC model, and concludes that this estimate is less than 20% of the total variability in the ZC model. The authors overlook the fact that the output of the ZC model is the Niño3 index (temperature anomalies in the tropical pacific), while the GISS model output is the average over the northern hemisphere land. In Figure 4 of the L&V paper, fluctuation levels versus scale for ZC and GISS are plotted in the same panel. For $\Delta t > 10$ yr they almost overlap. However, the ZC model data are averaged over 100 model runs, so the actual fluctuation level for the stochastic component is ten times greater than for the output from GISS control simulations.

The second argument assumes that the internal noise must have a scaling exponent $\beta \approx 0.6$, which would yield a negative slope $H = (\beta - 1)/2 \approx -0.2$ of the structure-function plot (see Fig. 1d). The actual plot of the structure function of the solar residual (the yellow circles in Fig. 1d) has a weakly positive slope, and hence the authors conclude that the latter is dominated by forced fluctuations on the centennial to millennium scale. The weakness of this argument is that it takes as assumption what the authors want to prove, namely that internal fluctuations on long time scales are small. It seems that only long control runs of the ZC model can settle this issue.

3 Linearity and intermittencies

The essence of Section 4 in the L&V paper is a mathematical corollary claiming that linearity in the response implies that the intermittency (the curvature of the scaling function) is the same for forcing and response. We have a number of reservations against the application of this result to the data and the climate models studied in this paper.

3.1 The essence of our critique

There are at least three possible sources of different intermittencies of the forcing and temperatures that are missed in the L&V paper:

(I) The mentioned corollary depends on a power-law form of the linear response function. On the long time scales this assumption is in direct contradiction to L&V's own claim that GCMs do not reproduce low-frequency (multicentennial) variability (see also *Lovejoy et al.* (2013)).

(II) It depends on the perfect power-law scaling of the structure functions of forcing and response, i.e., that these processes belong to the multifractal class. This is not true for any of the signals analysed in the L&V paper.

(III) The analysis does not account for the internal variabiliy. The authors have argued that internal variability may be smaller than forced variability on the longest time scales (se our discussion in Sect. 2.6). But in analysis of intermittency the emphasis is on the smallest time scales. The intermittency of the temperature signal will be strongly influenced by, or even dominated by, the internal noise, and hence there is no reason there should be a strong similarity between intermittencies of forcing and temperature in a linear response model.

3.2 Effect of imperfect power laws on intermittencies

Here we present some theoretical considerations which demonstrate that imperfect scaling (power laws) of the response kernel and the structure functions can lead to different intermittency of forcing and response in a linear response model. In Section 3.3 we demonstrate this by an example, so the present subsection can be skipped by readers who are only interested in such a demonstration. The general, linear response model Eq. (1) can be written as a convolution of the forcing $F(t)$ with a response kernel $G(t)$;

$$\hat{L}[F(t)] = \int_{-\infty}^{\infty} G(t - t')F(t')\,dt'. \tag{13}$$

For a general analysis of moments it is convenient to formulate the moments in the frequency domain rather than the time domain. Thus, we Fourier transform Eq. (13) to write

$$\mathcal{T}(f) = \mathcal{G}(f)\mathcal{F}(f), \tag{14}$$

where $\mathcal{T}(f)$, $\mathcal{F}(f)$, and $\mathcal{G}(f)$, are the Fourier transforms of $T(t)$, $F(t)$, and $G(t)$, respectively. By defining structure functions in frequency domain, $\mathcal{S}_q^T(f) \equiv \langle |\mathcal{T}(f)|^q \rangle$, $\mathcal{S}_q^F(f) \equiv \langle |\mathcal{F}(f)|^q \rangle$, we have the general, linear response model formulated as a linear relation between forcing and response structure functions of order q in the frequency domain, with the ensemble average of the q'th power of the transfer function $|\mathcal{G}(f)|$ as a constant of proportionality;

$$\mathcal{S}_q^T(f) = \langle |\mathcal{G}(f)|^q \rangle \mathcal{S}_q^F(f). \tag{15}$$

L&V assume a power-law, linear response. This corresponds to a response function of the form

$$G(t) = \xi (t/\mu)^{H-1/2}\theta(t), \tag{16}$$

where $\xi = 1\ \mathrm{Km}^2\mathrm{J}^{-1}$, μ is a constant in units of time which characterises the strength of the response, H is the scaling exponent for the response used by L&V, and $\theta(t)$ is the unit step function. The Fourier transform of this response function yields (see *Rypdal and Rypdal* (2014))

$$|\mathcal{G}(f)| = \left(\frac{f}{f_0}\right)^{-(H+1/2)}, \tag{17}$$

where

$$f_0 = \frac{[\xi\mu\Gamma(H+1/2)]^{\frac{1}{H+1/2}}}{2\pi\mu}, \tag{}$$

and $\Gamma(x)$ is the Euler Gamma function. Hence the L&V special case of Eq. (15) is

$$\mathcal{S}_q^T(f) = \left(\frac{f}{f_0}\right)^{-q(H+1/2)} \mathcal{S}_q^F(f). \tag{18}$$

The next assumption made by L&V is that both forcing and response exhibit multifractal scaling. If we write the structure functions on the form (dropping the superscripts T and F);

$$\mathcal{S}_q(f) = C_q(f)\, f^{-\eta(q)}, \tag{19}$$

the multifractal scaling assumption is that the multiplicative factor $C_q(f)$ is independent of the frequency f, such that the structure functions are perfect power laws in f. This is a very restrictive assumption that is not satisfied by any of the data in this study. If Eq. (19) holds true a plot of $\log\mathcal{S}_q(f)$ vs. $\log f$ is linear with slope $-\eta(q)$. The essence of the L&V approach (although some technicalities differ) corresponds to fitting the $\log\mathcal{S}_q(f)$ vs. $\log f$ curves with straight lines at the highest frequencies, or in other words, to draw tangent lines to the curves at the Nyquist frequency f_N. The negative slopes of these lines are interpreted as the scaling functions $\eta(q)$. This corresponds to defining the scaling functions by

$$\eta(q) = \left[\frac{d\,\mathcal{S}_q(f)}{d\,(\log f)}\right]_{f=f_N}, \tag{20}$$

and from Eq. (19) we then find the f-dependence of $C_q(f)$ which represents the deviation from multifractal scaling. The L&V approach includes normalizing the signals $T(t)$ and $F(t)$ such that they have the same power at the lowest frequency $f = 1$, i.e., $S_2^T(1) = S_2^F(1)$. If $H \neq -1/2$ Eq. (18) then implies that $f_0 = 1$, and putting $f = 1$ in Eqs. (18) and (19) we find,

$$\mathcal{S}_q^T(1) = \mathcal{S}_q^F(1) = C_q^T(1) = C_q^F(1) \tag{21}$$

for all all q. From the logarithm of Eqs. (18) and (19) we find for $f > 1$,

$$\eta_F(q) - \eta_T(q) + q(H+1/2) = \frac{\log\left[C_q^F(f))/C_q^T(f)\right]}{\log f}. \tag{22}$$

If $T(t)$ and $F(t)$ exhibit perfect multifractal scaling we have $C_q(f) = C_q(1)$, and from Eq. (21) the right hand side of Eq. (22) vanishes. Hence, for this case we have the L&V results that the curves

$\eta^T(q)$ and $\eta^F(q)$ have the same curvature, i.e., the response and forcing exhibit the same multifractal intermittency. However, the term $q(H + 1/2)$ on the left hand side arises from the particular power-law form of the linear response function shown in Eq. (17). With another form of the linear response kernel this term might not be linear in q, and this could introduce different curvature of $\eta^T(q)$ and $\eta^F(q)$. Different curvature is also introduced if the structure functions are not perfect power laws. Then the term on the right of Eq. (22) will in general not vanish, and it may have a non-zero second derivative. This may give rise to different curvatures of $\eta^T(q)$ and $\eta^F(q)$ even if the response is linear with the power-law response kernel given by Eq. (17).

3.3 Response to volcanic forcing

An important point in L&V is that intermittency in volcanic forcing and the corresponding temperature response are different, and that this is a signature of nonlinearity in the response. In this subsection we shall first demonstrate that the intermittency in the volcanic forcing is not multifractal, i.e., all the structure functions are not power laws. This is a symptom of the lack of correlations between bursts that characterises a multiplicative cascade. Next, we shall show by using L&V's trace moment analysis on a simple, linear response model that we can reproduce the intermittency observed in the response to volcanic forcing in the ZC model. This linear response exhibits a similar power spectrum, similar trace moments, and almost identical intermittency parameters as the ZC response. It demonstrates that these results obtained from the ZC model is not a signature of nonlinearity in the response.

Let us first build some intuition on the nature of the volcanic forcing. In Fig. 3a we have zoomed in on the volcanic forcing signal used in the ZC model. Each volcanic eruption is represented by 2-3 data points (years) different from zero (some large eruptions are represented by a few more points). If the eruptions are distributed randomly in time (Poisson distributed) the autocorrelation function (ACF) will vanish after a time lag of a few years. This is exactly what we observe in Fig. 3b. The spectral structure functions used in Sect. 3.2 are convenient for theoretical studies, but not for estimation based on short and spiky time series. Here it is better to use the standard structure functions which are computed the from empirical moments;

$$\hat{S}_q(\Delta t) = (N - \Delta t)^{-1} \sum_{t=1}^{N-\Delta t} |Y(t + \Delta t) - Y(t)|^q \tag{23}$$

where $Y(t) = \sum_{t'=0}^{t} F(t')$ is the cumulative sum of the forcing time series. These are shown in Fig. 3c. The steeper slopes (slope $\approx q$) for $\Delta t \le 4$ is due to the smoothness of the forcing signal on these short time scales, signified by the ACF in Fig. 3b. For $q = 2$ the structure function looks quite straight and with slope close to 1 in the log-log plot for the scale range 4-100 yr. For smaller q the plots become more curved. This is symptomatic for a stationary, uncorrelated process (Lévy process) which is non-Gaussian on short time scales, although the central limit theorem requires that it converges to a Gaussian on the longer scales. Such a process is not multifractal, but L&V are blind

to this fact. In practice, their approach corresponds to assuming that the moments can be written in

320 the power-law form $\hat{S}_q(\Delta t) \sim \Delta t^{\zeta(q)}$, where the scaling function $\zeta(q)$ is estimated by fitting straight lines to the structure functions in the log-log plot in the range 4-100 yr. This has been done in Fig. 3d. The curved scaling function is incorrectly interpreted by L&V as a signature of multifractality. But this interpretation is correct only if all structure functions are power laws (straight lines in log-log plots). It is easily demonstrated that very similar results are obtained by random shuffling of the

325 onset times of the volcanic spikes, which would convert a multifractal signal into a realisation of a Lévy process. If the original signal were a multifractal, the result should be quite different after shuffling.

Our main focus in this comment, however, is not on the incorrect multifractal interpretation of the scaling analysis, but on the incorrect conclusions drawn from this analysis when it comes to non-

330 linearity in the response. As a means to investigate this point we construct a linear response model that mimics the ZC response to the volcanic forcing. The ZC response is shown by the blue curve in Fig. 4a. We observe that every volcanic spike seems to be succeeded by a damped oscillation. Thus, we construct a linear, damped harmonic oscillator response model and select the parameters to produce a response signal that looks similar to that of the ZC response to the volcanic forcing

335 when we drive the model with stochastic forcing in addition to the volcanic forcing. We make no attempts to fine-tune the model parameters, since this extremely simple model obviously is not an accurate substitute for the ZC model. The purpose of devising this model is only to demonstrate that a linear model can produce a response with intermittency parameters very different from those of the forcing. These are results which L&V contend can only arise from nonlinearity of the response.

340 The response according to the linear model is shown by the red curve in Fig. 4a, and we compute the trace moments and intermittency coefficients for this linear response signal. We have used the Mathematica routines downloaded from Shaun Lovejoy's web page for these computations to make sure that the results are comparable to those presented by L&V. Fig. 4b is a reproduction of Figure 6a, top right, in L&V for the volcanic forcing. L&V interpret the wide spread in slopes of the

345 trace-moment curves as signature of multifractal intermittency, and they compute the intermittency coefficients $C_1 = 0.48$ and $\alpha = 0.31$ (their Table 1). The results depend on the exact fitting range chosen, so we cannot expect to get exactly the same results for these parameters. We find $C_1 = 0.52$ and $\alpha = 0.13$ (which makes us wonder if $\alpha = 0.31$ in L&V is a misprint). In Fig. 4c we have computed the trace moments for the linear response model. This figure is very similar to their Figure 6a,

350 bottom right, for the ZC response. The intermittency parameters computed by L&V for this case are $C_1 = 0.054$ and $\alpha = 2.0$, while our results for the linear model are $C_1 = 0.059$ and $\alpha = 1.97$. L&V interpret the difference in intermittency parameters between forcing and response as a signature of nonlinearity, but our exercise shows that such a difference can be obtained from a simple linear response model.

355 4   Conclusions

A correct treatment, without unjustified approximations, of the issue of additivity in the Zebiak-Cane
model gives no reason for rejection of a linear response model (see Fig. 2). This conclusion holds
even without accounting for internal variability, but is enforced by the inclusion of this effect.

L&V's analysis of intermittencies is based on a corollary which states that if the response is linear
360 the intermittency computed through trace moment analysis must be the same in forcing and response.
However, this corollary holds only if both forcing and response belong to the class of multifractals,
i.e., if all structure functions are power-laws, and in addition that the response function is a power law
on all scales. These assumptions do not hold for any of the time series in question and for realistic
linear response functions. These issues will be discussed in depth in a forthcoming paper; here it has
365 been sufficient to demonstrate by an example that the intermittencies can be very different in forcing
and response produced by a linear response model. Hence, our conclusion is that the intermittency
analysis of L&V does not constitute a valid test for rejecting the linear response hypothesis.

Acknowledgements. This paper was supported by the the Norwegian Research Council (KLIMAFORSK pro-
gramme) under grant no. 229754. We are grateful to Shaun Lovejoy for providing Mathematica codes for esti-
370 mation of Haar structure functions and trace moments on his web page: http://www.physics.mcgill.ca/ gang/software/index.html,
and to Hege-Beate Fredriksen for debugging these codes.

References

Andrews, T., Gregory, J. M, Webb, M. J., and Taylor, K. E.: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models, Geophys. Res. Lett., 39, L09712, doi:10.1029/2012GL051607, 2012.

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.

Geoffroy, O., Saint-Martin, D., Olivié, D. J. L., Voldoire, A., Bellon, G., and Tytcá, S.: Transient Climate Response in a Two-Layer Energy-Balance Model. Part I: Analytical Solution and Parameter Calibration Using CMIP5 AOGCM Experiments, *J. Climate*, 6, 1841-1857, doi:10.1175/JCLI-D-12-00195.1, 2013.

Lovejoy, S., and Schertzer, D., and Varon, D.: Do GCMs predict the climate··· or macroweather?, Earth Syst. Dyn., 4, 439-454, doi:10.5194/esd-4-439-2013, 2013.

Lovejoy, S., del Rio Amador, L., and Hébert, R.: The ScaLIng Macroweather Model (SLIMM): using scaling to forecast global-scale macroweather from months to decades, Earth. Syst. Dynam.., 6, 637-658. doi:10.5194/esd-6-637-2015, 2015.

Lovejoy, S., and Varotsos, C.: Scaling regimes and linear/nonlinear responses of last millennium climate to volcanic and solar forcings, Earth Syst. Dyn., 7, 133-150, doi:10.5194/esd-7-1-2016, 2016.

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.

Rypdal, M., and Rypdal, K.: Long-memory effects in linear-response models of Earth's temperature and implications for future global warming, J. Climate, 27, 5240-5258, doi:10.1175/JCLI-D-13-00296.1, 2014.

Rypdal, K., Rypdal, M., and H.-B. Fredriksen: Spatiotemporal Long-Range Persistence in Earth's Temperature Field: Analysis of Stochastic-Diffusive Energy Balance Models, J. Climate, 28, 8379- 8395, 2015, 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.
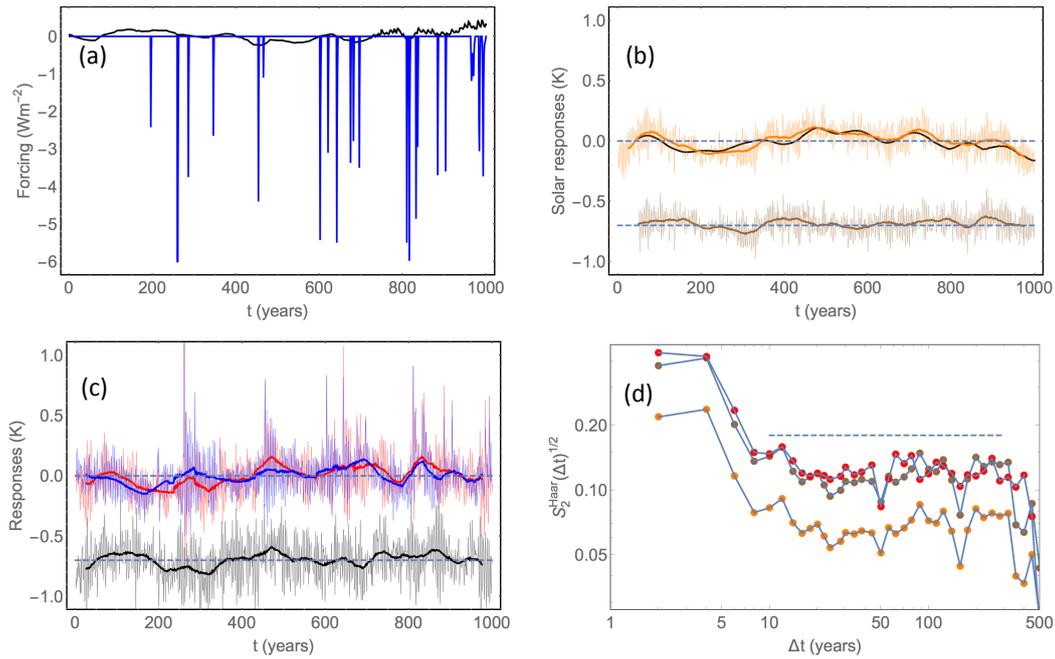
Figure 1. (a): Time series of the solar (black) and volcanic forcing (blue) for the last millennium used in the simulations of the ZC model. (b): Responses after averaging over 100 realisations. Thin, orange curve is response to solar forcing and the thick, orange is filtered by a 50 yr moving average. The thick, black curve is the filtered and shifted solar forcing signal $\Delta T_s^{\text{det}}(t, \Delta t)$ given by Eq. (8). The brown, thin curve is the internal noise $\epsilon(t)$ defined in Eq. (9), and the thick brown is the filtered time series. (c): Thin, blue curve represents $T_{s+v}(t)$, the thin, red curve is $T_s(t) + T_v(t)$, and the thin, black curve is their difference $\varepsilon(t) = T_s(t) + T_v(t) - T_{u+v}$. Thick curves are the corresponding filtered series. (d): Haar structure function of $\epsilon(t)$ (orange bullets), of $\varepsilon(t)$ (red bullets), and of $\sqrt{3}\epsilon(t)$ (brown bullets).
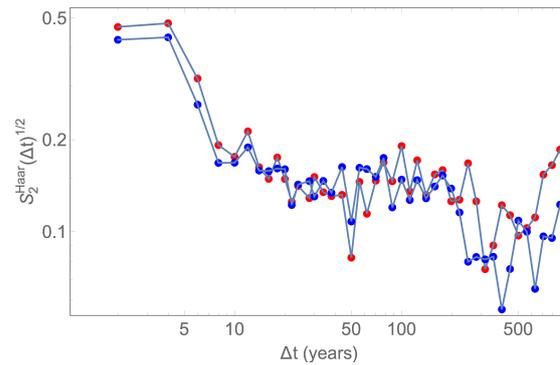
Figure 2. Haar structure functions $\sqrt{\langle |\Delta T_s + \Delta T_v|^2 \rangle}$ (red bullets) and $\langle |\Delta T_{s+v}|^2 \rangle$ (blue bullets).
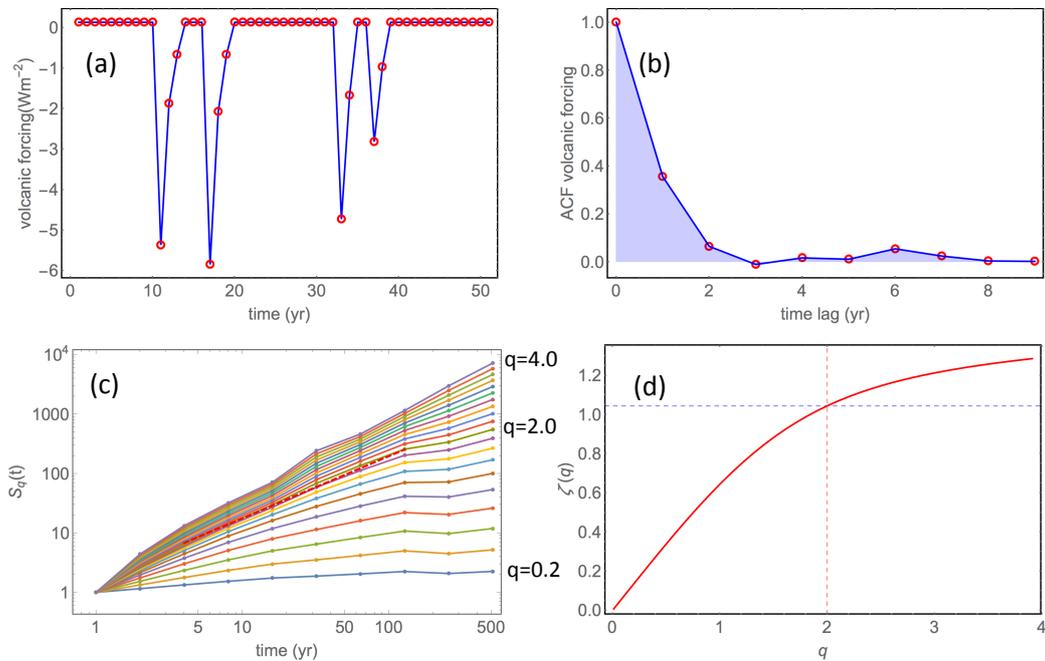
Figure 3. (a): A zoom-in on the volcanic forcing signal shown in Fig. 1a. (b): The ACF estimated for the volcanic forcing signal. (c): The structure functions (empirical moments) $\hat{S}_q(\Delta t)$ for the volcanic forcing signal estimated for $q = (0.2, 0.4, \ldots, 4.0)$ The red, dashed line is a linear fit to the log-log plot of $\hat{S}_2(\Delta t)$. (d): The scaling function $\zeta(q)$ computed from linear fits to the $\hat{S}_q(\Delta t)$'s over the interval $\Delta t \in (4, 128)$. The observation that $\zeta(2) \approx 1$ suggests that the process is uncorrelated on these time scales.
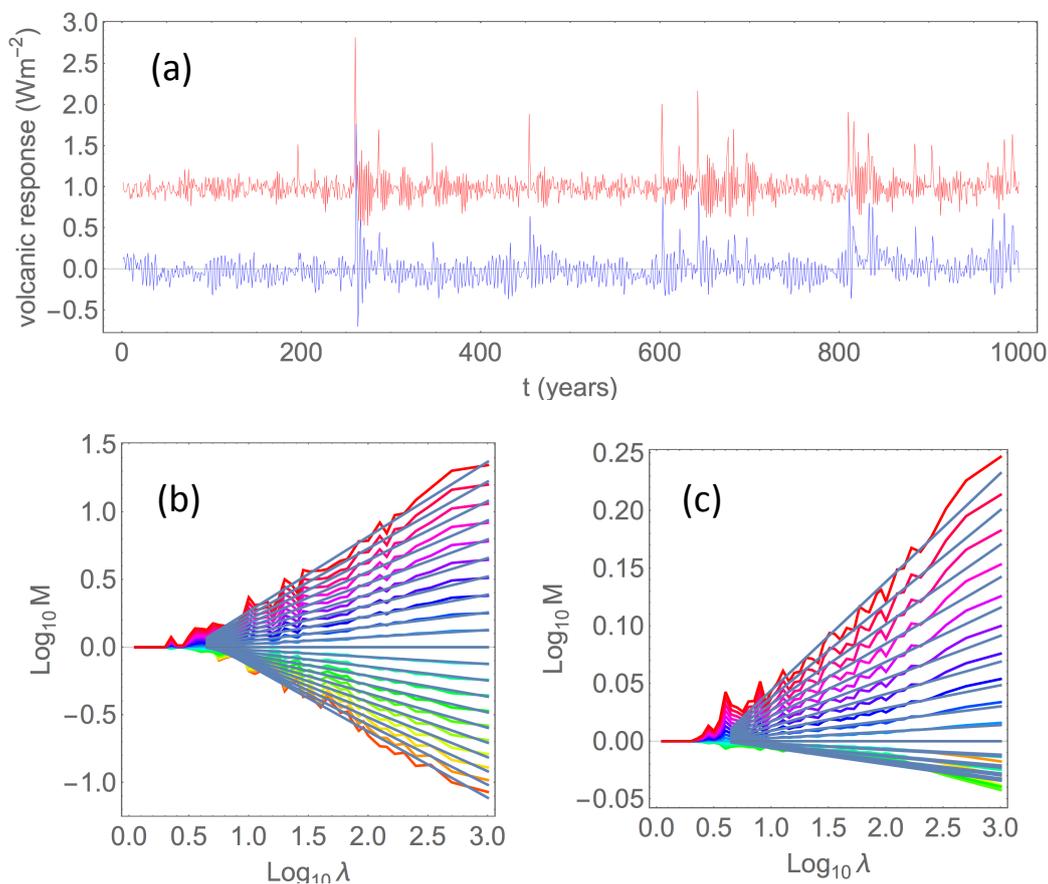
Figure 4. (a): Blue curve is the average over 100 realisation of the response to volcanic forcing in the ZC model, and the red curve is the response to this forcing plus a stochastic Gaussian white noise forcing in a linear, damped harmonic oscillator model. (b): Result of trace moment analysis of the volcanic forcing signal. It is very similar to the corresponding panel in Figure 6 of L&V. (c): Result of trace moment analysis of the harmonic oscillator response shown by the red curve in panel (a). It is very similar to the corresponding panel for the ZC response to volcanic forcing in Figure 6 of L&V.