Manuscript prepared for Earth Syst. Dynam.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 17 April 2016

# Reply to Lovejoy, Sarlis and Varotsos: “On testing the additivity of Zebiak-Cane model response to volcanic and solar forcing”

**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)

**Abstract.** Lovejoy, Sarlis, and Varotsos (LSV) replicate a test in our comment by using a different time series representing internal variability, and conclude that this test rejects the linear response hypthesis in the Zebiak-Cane model. This time series is the first 195 yr of the volcanic response record, during which there was no volcanic forcing. We demonstrate that the short length of this record creates large finite size uncertainties which render their result statistically insignificant. We also comment on some passages in their reply about physical paradigms, and on faulty statistical reasoning and apparent self-contradictions in L&V’s writings.

## 1 Reply to “Introduction”

As a motivation for conduction these tests in the ZC model LSV write:

> “*To situate the debate, recall that whereas at short enough time scales, when external forcings are small enough, then theoretically we may expect the atmospheric response to be approximately linear, however, at long enough time scales, due to temperature - albedo feedbacks, the response is expected to become nonlinear. At the same time, it is possible that at long enough time scales, due to quite different surface and atmospheric interactions, that solar and volcanic external forcings combine nonlinearly.*”

A qualitative mental picture, a paradigm, of the physical mechanisms that govern the macro dynamics is of course an important guideline, but it can also become a straitjacket that restricts the range of alternatives that one is willing to investigate. Our favorite picture is quite different L&V’s. We se no reason why responses should be more linear on short than on long time scales, in particular not the response to volcanic forcing, which is strong on short time time scales. LSV write about “atmospheric response” which is a vague concept that is impossible to quantify without specifying the physical variable and on which spatial and temporal scale it is measured. The response of local climatic variables on synoptic and seasonal scales to strong volcanic eruptions is certainly nonlinear, since it is dominated by hydrodynamical vortex structures. But on long time scales, the global temperature change will change in proportion to the change in heat content in the upper ocean, which again will change in proportion to the net radiative flux. The response in presence of feedbacks that modify the radiative flux, like albedo feedbacks, are not “expected to become nonlinear” by most climate modelers. They are typically modeled linearly. Consider, for instance, the linear energy balance model $d_tT = S^{-1}T + F$, and assume that the albedo decreases proportionally to $T$, such that the effective radiative flux can be written $F = F_0 - \alpha T$. The resulting equation is $d_tT = S'^{-1}T + F_0$, where the feedback has enhanced the climate sensitivity to $S' = (1-\alpha S)^{-1}S$.

The ENSO phenomenon is probably a nonlinear mode in the climate system, and is part of the internal variability, even though it can be influenced by external forcing. The nonlinear nature of the oscillation makes it likely that the timing of El Niño events can be influenced by external forcing such as strong volcanic eruptions. But we find it less likely that the response on centennial time scales is nonlinear. This is the issue discussed in section 2 of our comment article. But whatever our prejudices are, proper tests is what should settle the issue.

## 2 Reply to "The R+R linear response null hypothesis test fails for the ZC model response "

In section 2 of our comment we presented two tests of linearity of the response in the ZC-model. In Sect. 2.3 we presented an alternative test of response additivity which involved an estimate of internal variability and its effect on the test. This is the only test that is addressed in LSV's reply. In Sect. 2.4 we replicated the test made in the original paper by L&V (which does not take internal variability into account). This replication failed to detect the subadditivity claimed by L&V and presented in a very confusing manner in their paper. It would be interesting to know weather L&V will dispute the correctness of Fig. 2 in our comment article, for instance by replicating that figure.

In our test in Sect. 2.3 we estimated internal variability from the solar forcing and response, using the entire 1000 yr time series. It involved fitting a 25 yr shifted linear response to the observed response signal and interpreting the residue as the internal variability. A weakness of this procedure is that it involves some assumptions about the slow linear response, but an advantage is that finite sample size uncertainty could be kept low because we used the entire time series. The mentioned assumptions reduces the strength of the test, so the only conclusion we could draw is that this relatively weak test could not reject the linearity hypothesis. We also wrote in the comment, that long control runs of the model could give us a better estimate of the internal noise and a stronger test.

In their reply, LSV suggest to use a different estimate of the internal noise, namely the first 195 yr of the volcanic-driven response time series. This is justified, since there was no volcanic forcing in this period. The drawback, however, is that an estimate of the Haar fluctuation from such a short time series is associated with much higher estimation uncertainty (finite-sample size errors). LSV make no attempt to demonstrate that the estimates of the difference $|\sqrt{3}\epsilon(t)-\varepsilon(t)|$ is significantly different from zero in a statistical sense. Such a test is easy to make by creating a Monte Carlo ensemble of time series containing 195 data points with statistical properties similar to those of the observed volcano response. The statistical scatter of the Haar fluctuations within this ensemble will give us information about the finite sample uncertainty of the Haar estimate. This is done in attached Figure 1, where the specifications of the Monte Carlo are desccribed in the caption. The figure shows that the difference between the Haar fluctuations of $\sqrt{3}\epsilon(t)$ and $\varepsilon(t)$ is smaller than this uncertainty in the interesting scale range $\Delta t > 10$ yr.

> *This means that the deviation from linearity observed by LSV is statistically insignificant, and hence does not reject the linear response hypothesis.*

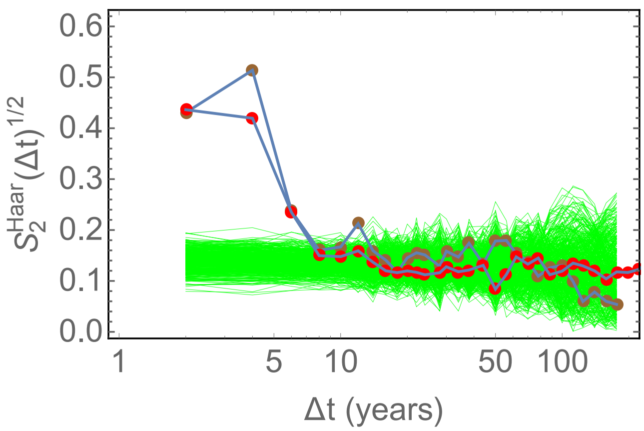


**Figure 1.** Brown bullets: Haar fluctuation function of $\sqrt{3}\epsilon(t)$, where $\epsilon(t)$ is the first 195 yr of the volcanic forcing record. Red bullets is Haar-fluctuation of $\varepsilon(t)$ as defined in the comment article of R&R. These two curves look similar to the corresponding curves in Fig. 2 in LSV. The crucial issue is whether the difference between these two curves is statistically significant. The green curves is a 500 member ensemble of fractional Gaussian noises (fGn's) with $H=-0.01$ ($\beta=2H+1=0.98$) and Haar fluctuation equal to that of $\sqrt{3}\epsilon(t)$ on 100 yr time scale. On time scale less than 10 yr the fGn is not a good model for the internal noise because of the ENSO dynamics, but on longer time scales the flat Haar-fluctuation curve suggests that an fGn with $\beta\approx 1$ is a crude statistical model of the internal variability. The scatter of the Haar fluctuation in this ensemble gives an idea about the statistical uncertainty of this volcano estimate of internal variability. This uncertainty is considerably larger than the estimate of $|\sqrt{3}\epsilon(t)-\varepsilon(t)|$ (the difference between the brown and the red curves), hence this difference is not statistically significant.

## 3 Elementary errors of statistical reasoning and persistent self-contradicion in L&V's writings

The last paragraph in the LSV reply reads:

> "Other deficiencies of R+R are:
>
> a) they make use of statistical independence in their Eq. (12) -thus increasing the effect of the internal variability E to the ratio R - while they criticize L+V for doing so (see lines 162-164) and
>
> b) they claim (in lines 164-166) that L+V admit that their analysis is wrong, which of course is not the case."

Eq. (12) in our comment reads

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

No assumption of statistical independence was used up to that point. Note that $R=R(\Delta t)$, i.e., it is a generalisation of the $R(\Delta t)$-curve plotted in Figure 3 in the L&V paper, including internal noise. If we had set out to estimate this curve we should not estimate $\langle|\Delta\varepsilon|^2\rangle$ by using $\sqrt{3}\epsilon(t)$ as an estimator, but rather the exact definition of $\Delta\varepsilon$ given by Eq. (6); $\Delta\varepsilon = \Delta T_{v+s} - \Delta T_v - \Delta T_s$. We did not plot such a curve, however, because our ambition was only to demonstrate that internal variability increase $R$, and to indicate the order of magnitude of that increase.

What L&V did in their paper was different. They used an approximation analogous to, but different from, ours to estimate Haar fluctuation curves, and this turned out to give a systematic bias in favour of subadditivity. Our approximation was concerned a noise process (internal variability), and the error is a finite size error, not a bias. The L&V approximation treated the volcanic and solar forcing as independent stochastic processes, while the appropriate way of dealing with these 1000 yr historical records (in particular the solar) is to treat them as deterministic signals. In that case omission of cross-terms have little justification.

After being pushed on this point by us in the ESDD discussion of their paper, L&V kept their original approximation in the text and in the caption of Figure 3, but included a revision of the $R$-curve in the figure, so that the figure became a hybrid that contains results both based on the invalid approximation and results that are not. They acknowledged that the approximation makes a difference, and in the text they added the following paragraph:

> "The reason for the difference is that the cancellation of the cross terms assumed by statistical independence is only approximately valid on single realizations, especially at the lower frequencies where the statistics are worse (even on a single realization, at any given scale - except the very longest - there are several fluctuations, so that there is still some averaging).

We interpreted this as an admission that the approximation is wrong. However, in (b) above they now state that this is not the case. They believe both results are correct, and thereby transcend the trivial realm of logic. . .