

The Rypdal and Rypdal manuscript (R&R) raises several critical points about the Lovejoy and Vartsos (L & V) paper, and my recommendation is that it should be published with minor revisions.

General comments:

The original L&V paper broaches important points about the scaling of forcing and response over long time scales. However, it suffers from major flaws both of substance and style, the latter being beyond the scope of this review. The R&R comment addresses some of the substantive flaws, with their critique seemingly centered on three issues:

- (1) A hypothesis test fails to reject the null hypothesis of linearity.
- (2) Properly accounting for natural variability reduces the perceived degree of non-linearity
- (3) Distinct intermitency of forcing and response may still be consistent with a linear system.

In my assessment, points (2) and (3) are well founded, with some specific comments detailed below. Point (1) however merits a more careful discussion.

[1] The response functions of the CZ and GISS models, as well as the climate system, are non-linear in the strictest mathematical sense (e.g. the Black Body/Plank Feedback is non-linear). Thus, the appropriate question to ask is whether a linear approximation is valid for a certain purpose. L&V seem to ask the question as to whether a linear approximation is valid when assessing the mean temperature response to small perturbations in radiative forcing, of the type expected to arise from historical changes in volcanic aerosol emissions and solar variability. This is an important question for purposes of inferring climate sensitivity from palaeorecords, and L&V seem to reach the conclusion that the linear approximation is not appropriate, as the response is markedly sub-additive, by a factor of $R \sim 1.5$. L&V do not set up the problem as such a null-hypothesis, and I agree with R&R that this is a major flaw. They also fail to quantify the uncertainty range of their factor R . Although unlikely, it might be that 1.5 is consistent with a linear approximation in the presence of noise.

However, issues of presentation and robustness aside, the L&V paper can be easily interpreted as a rejection of the null hypothesis of linearity. While R&R perform a more properly set-up hypothesis test, it is a somewhat different hypothesis test, examining the second order statistics of the residuals of a specific model fit, as opposed to the responses themselves. Thus, between L&V and R&R, a null hypothesis has been proposed, and two different tests have

been performed, one of which rejects it, while the other one fails to reject it. This issue needs to be addressed before the R&R test can be interpreted as a rebuttal of the L&V result. If the R&R test implies that the L&V test should have also rejected linearity, then the question arises where is the error in L&V? (perhaps not including internal variability?). Otherwise, the alternative test is not by itself a rebuttal.

[2] At times the language used is not appropriate for a scientific paper. I recommend that the critique sticks to the sciences and does not become personal and subjective.

- Line 164: “invalid (and completely unnecessary) approximation”. ‘Completely unnecessary’ seems a bit hyperbolic.

- Line 164-165: “The authors admit in the published paper that this analysis is wrong”. I think R&R misrepresent L&V when claiming that the latter explicitly admit to a faulty analysis.

- Line 318: “but L&V are blind to this fact”. This phrasing is inappropriate and, I dare say, completely unnecessary.

More Specific Comments:

[2] Line 31: Geoffroy et al (2013) finds that the linear approximation is appropriate when considering two different forcing scenarios with the same type of forcing (CO₂). Even then, in many of the models a small but robust overestimation of the temperature response in the 1pctCO₂ scenario can be observed. There is also evidence (Merlis et al 2014) that there is a small but noticeable difference in the response to volcanic forcing as opposed to CO₂ forcing.

[3] Line 37: The ‘nonlinearity’ that Andrews et al (2012) find refers to the fact that the relation between global temperature and global radiation imbalance is “not a line” in the transient regime. This may be entirely consistent with linear dynamics (Armour et al 2013).

[4] Line 115 (equation 8). An AR(1) type transfer function would be easy to fit and more consistent with a dynamical system than using a constant lag. Additionally, as per Geoffroy et al (2013), using at least two time-scales to represent a fast and slow component seems to be a minimal requirement for a decent fit. Such a fit should be easy enough to perform given knowledge of the forcing (e.g. Castruccio et al 2014). However, for the author’s purposes equation (8) should be sufficient, and I would not list this as a required improvement.

[5] Along the same lines: How different are the transfer functions fit to the solar only, volcanic only, and solar+volcanic scenarios? This could be construed as yet another hypothesis test.

[6] Lines 153-155: A visual comparison seems unsatisfactory as a hypothesis test. A more rigorous p-value test would be nice. However, as L&V do not provide one either, I do not think that the R&R comment should be held to a different standard.

[7] Section 2.4: Lines 157-180. Confusing. What is the invalid and completely unnecessary approximation? An alternative estimate is given, but no explanation for why the original estimate was wrong when variability is not taken into account.

[8] Lines 224-226: In support of this, the authors can cite the Geoffroy et al (2013) also cited in the introduction, which provides estimates of the transfer functions for CMIP5 gcms (appendix of part I). Additionally, MacMynowski et al. (2011) explicitly compute the transfer function for solar forcing, showing that it is not a simple power law.

[9]Line 331: A damped harmonic oscillator is an unusual choice for a transfer function. The response to volcanic forcing is generally assumed to follow a standard decaying exponential form. One could use a standard transfer function such as those mentioned in [8].

Armour, K. C., Bitz, C. M. & Roe, G. H. (2013). Time-Varying Climate Sensitivity from Regional Feedbacks. *J. Climate* 26, 4518–4534 (2013).

Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L., & Moyer, E. J. (2014). Statistical Emulation of Climate Model Projections Based on Precomputed GCM runs. *Journal of Climate*.

Merlis, T.M., Held, I. M., Stenchikov, I. M., Zeng, F. and Horowitz, L. W. (2014). Constraining Transient Climate Sensitivity Using Coupled Climate Model Simulations of Volcanic Eruptions. *Journal of Climate*.

MacMynowski, D. G., Shin, H.-J., and Caldeira, K., The Frequency Response of Temperature and Precipitation in a Climate Model. *Geophysical Research Letters*.