

TECHNICAL RESPONSE

CLIMATE CHANGE

Response to Comment on “Atlantic and Pacific multidecadal oscillations and Northern Hemisphere temperatures”

B. A. Steinman,^{1*}† L. M. Frankcombe,^{2†} M. E. Mann,³ S. K. Miller,³ M. H. England²

Kravtsov *et al.* claim that we incorrectly assess the statistical independence of simulated samples of internal climate variability and that we underestimate uncertainty in our calculations of observed internal variability. Their analysis is fundamentally flawed, owing to the use of model ensembles with too few realizations and the fact that no one model can adequately represent the forced signal.

Kravtsov *et al.* (*1*) wrongly claim (i) that our assertion of statistical independence among estimates of the internal variability in regional temperature change in climate models (*2*) is an artifact of a flawed procedure. They further suggest (ii) that robust assessments of simulated internal variability cannot rely on a multimodel ensemble mean (MMEM) that differs from the true forced signal of the individual models, because the residuals of the two forced signals masquerade as low-frequency internal variability, which leads to correlation among ensemble members. Finally, they claim (iii) that we substantially underestimate the uncertainty in the semi-empirical estimate of internal variability derived using the MMEM to approximate the forced signal.

Regarding their first point, Kravtsov *et al.* assert (in their second reference/note) that the standard deviation of the mean of internal variability is not exactly zero only because the data were filtered before analysis [figure 2 and figures S2 to S4 in (*2*)]. This is true, however, only if the full ensemble of realizations, N , is used in the calculation. If $(N - 1)$ realizations are instead used, as is the case in our analysis, the standard deviation of the mean is not zero (regardless of whether the data are filtered beforehand) (*3*). Our method for assessing statistical independence is valid.

Regarding their second point, Kravtsov *et al.* claim that internal variability calculated using the regional regression method and a scaled MMEM

introduces errors because each model has a different forced response. Instead, they use a single-model ensemble mean (SMEM) for models with four or more historical runs and show that the internal variability calculated in this manner has a lower variance and lower intramodel correlation than that determined using a MMEM [figure 1, A to F, in (*1*)]. They assert that this occurs because the difference between the MMEM and the true forced signal for individual models introduces extra internal variability at low frequencies. Although this could be true in principle, this point is irrelevant because the ensembles of simulated internal variability determined using regional regression nevertheless universally satisfy the requirements for statistical independence [figure 2 in (*2*)].

Moreover, we show that more than four (indeed, more than 10) ensemble members are required for a robust estimate of the forced signal from a SMEM, and thus the lower variance of the internal variability estimates is due to the small ensemble size, which leads to the removal of too much of the internal variability. We demonstrate this point using synthetic autoregressive-1 time series (Fig. 1, A and B), each with the same forced signal, and a different realization of red noise (*4*). We divide the 160-member ensemble into subsets of four- or ten-member ensembles, which were smoothed using a 5-year low-pass filter. Using the code provided by Kravtsov *et al.*, we show that the application of small ensemble SMEMs results in lower variance of internal variability (Fig. 1A), even though in this case we know that the large ensemble yields a far more accurate estimate of the forced signal (Fig. 1B). This idealized example reveals that the higher variance obtained when using the MMEM relative to using the SMEM is in fact due to the removal of too much internal variability when using the SMEM.

Regarding their third point, Kravtsov *et al.* claim that there is a wide range of possible semi-

empirical estimates of observed internal variability resulting from the range of possible SMEM estimates of the forced signal. They attempt to assess the uncertainty in the Atlantic Multidecadal Oscillation (AMO), Pacific Multidecadal Oscillation (PMO), and Northern Hemisphere Multidecadal Oscillation (NMO) by applying the mean of individual model ensembles with four or more realizations and claim that the resulting spread of these separate estimates defines uncertainty inherent in the regional regression method (*5*). However, as we have shown, none of these SMEMs is in itself a robust estimate of the forced signal, because none of the models have enough ensemble members for suitable cancellation of the different realizations of internal variability [which also explains the narrow 2σ range in their figure 1, G to I (*1*)]. More importantly, it is unreasonable to expect an individual model to have a better representation of the forced signal than the multimodel ensemble. In fact, the spread in the internal variability estimates of Kravtsov *et al.* [figure 1, G to I in (*1*)] supports this assertion, indicating that regional regression-based estimates of internal variability and their uncertainties are intrinsically dependent on the choice of the forced signal and therefore that only robust estimates derived from a large number of models or realizations should be used for this assessment. The MMEMs from the Coupled Model Intercomparison Project Phase 5 [CMIP5-All (all models) and CMIP5-AIE (models that include the first and second aerosol indirect effects)] applied in our study fulfill these requirements, whereas the SMEMs do not. By this rationale, the CMIP5-All mean provides the best overall estimate of the forced signal, with an uncertainty that can be estimated (among other methods) using bootstrap resampling [figure 3C in (*2*)].

Although the use of SMEMs based on a small number of realizations is an inherently flawed method, we nonetheless show that both the MMEM and SMEM regression-based methods provide better estimates of the internal signal than simple linear detrending, which produces a large overestimation of the variance of the internal variability, especially in recent decades [magenta curves in Fig. 1, A and C, and figure 1, G to I in (*1*)]. Indeed, this is one of the key points of our original Report and is a principal focus of Frankcombe *et al.* (*6*). This recent study concludes that linear detrending introduces large biases in both the amplitude and phase of the internal variability and that regression-based approaches that rely on large historical simulation ensembles to estimate the forced signal produce less biased estimates of internal variability.

The use of different MMEM and large ensemble SMEM estimates of the forced series (*5*) produces internal variability trends that are generally consistent [see figure 3 and figure S6 in (*2*)]. For example, the AMO, PMO, and NMO behavior over the most recent two decades are in each case largely similar to one another (e.g., in the most recent decade: PMO decreasing, NMO decreasing, AMO flat) (*7*) and inconsistent with estimates of

¹Large Lakes Observatory and Department of Earth and Environmental Sciences, University of Minnesota Duluth, Duluth, MN, USA. ²Australian Research Council Centre of Excellence for Climate System Science and Climate Change Research Centre, University of New South Wales, Sydney, New South Wales, Australia. ³Department of Meteorology and Earth and Environmental Systems Institute, Pennsylvania State University, University Park, PA, USA.

*Corresponding author. E-mail: bsteinma@d.umn.edu

†These authors contributed equally to this work.

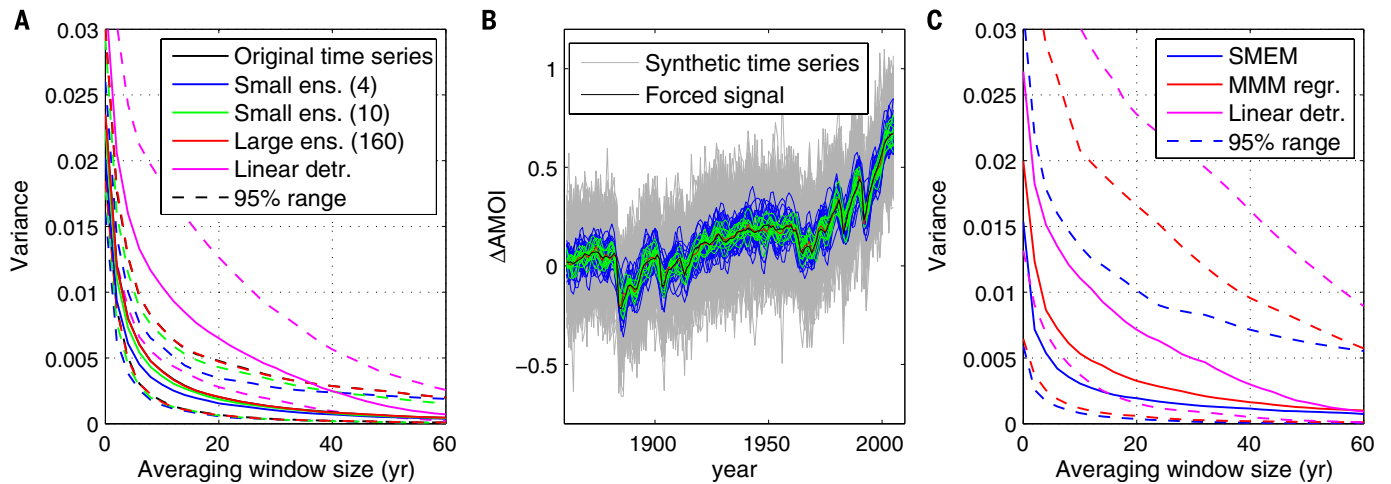


Fig. 1. Estimates of AMO variance based on ensemble size and smoothing window. (A) The variance of internal variability in synthetic AMO time series produced using the same forced signal and different realizations of red noise. The mean variance (solid lines) and 2σ range (dashed lines) of the large ensemble (red, 160 members) and small ensembles (green, 10 members; blue, 4 members) are shown as a function of the time-averaging window size. The small ensembles consist of random groupings taken from the large ensemble. Also shown are the variance of the original synthetic time series (not including the forced signal; black, almost coinciding with the red curve) and the variance obtained after linear detrending (magenta). (B) Individual estimates of the

forced signal from the large and small ensembles means [colors as in (A)] shown with the true forced signal (black). Note the extremely high level of agreement between the large ensemble mean and the true forced signal relative to that of the small ensemble means. The synthetic time series (including the forced signal) are shown in gray. (C) The variance of semi-empirical estimates of AMO variability from the 18 CMIP5 models considered in Kravtsov *et al.* using multimodel ensemble mean regression (red), individual model ensemble mean subtraction (blue), and linear detrending (magenta). Ensemble mean variance (solid lines) and 2σ ranges (dashed lines) are shown as a function of the time-averaging window size.

internal variability derived from simple statistical methods (i.e., detrending, with the exception of, perhaps, the PMO). Furthermore, the principal conclusions of Steinman *et al.* (2) (regarding the recent slowdown in surface warming) have been supported by at least eight other prominent studies (8–15), of which the most recent (15) uses a semi-empirical method (in which the forced signal is estimated using a MMEM) that is very similar to the target region regression method of Steinman *et al.* (2).

In short, we find no merit to the criticisms of Kravtsov *et al.* We once again emphasize that the linear detrending procedure used in their past work (16) leads to extremely biased estimates of internal variability and should not be employed. Our regression-based approach (2, 6), by contrast, yields faithful estimates of the internal variability.

REFERENCES AND NOTES

1. S. Kravtsov, M. G. Wyatt, J. A. Curry, A. A. Tsonis, *Science* **350**, 1326 (2015).
2. B. A. Steinman, M. E. Mann, S. K. Miller, *Science* **347**, 988–991 (2015).

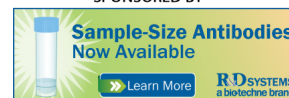
3. In our original analysis, we incorrectly assessed the standard deviation of the mean of the internal variability estimates by including all realizations ($N = 170$, in the case of CMIP5-All) rather than ($N - 1$) realizations in demonstrating that only the regional regression approach satisfies the requirement for statistical independence. However, the correct calculation results in an insignificant difference in the results that has no effect on our findings and conclusions. The manuscript, code, and results have been updated to reflect this correction.
4. The forced signal is the Coupled Model Intercomparison Project Phase 5 all model (CMIP5-All) ensemble mean North Atlantic time series. The realizations of red noise are scaled to match the average autocorrelation of the North Atlantic series in the CMIP5 models.
5. We agree, in part, that the true uncertainty in semi-empirical estimates of internal variability is best assessed using multiple, equally valid estimates of the forced signal based on different models with a large number of realizations. We addressed this specific point in our manuscript through the inclusion of five separate estimates of internal variability based on five individual SMEMs consisting of 10 or more realizations [our ensembles include CMIP5-All (all models)/CMIP5-GISS (GISS model E2 ensemble)/CMIP5-AIE (models that include the first and second aerosol indirect effects)/five different SMEMs with more than 10 members] [figure S6 in (2)].
6. L. M. Frankcombe, M. H. England, M. E. Mann, B. A. Steinman, *J. Clim.* **28**, 8184–8202 (2015).
7. The linear extrapolation method used to extend model series from 2005 through 2012 produces results that are generally

- consistent with simple propagation of the 2005 value. The claim in reference/note 11 in Kravtsov *et al.* (1) that our methodology here is “unfortunate” is therefore inappropriate. This issue was addressed in detail by Mann *et al.* (8).
8. M. E. Mann, B. A. Steinman, S. K. Miller, *Geophys. Res. Lett.* **41**, 3211–3219 (2014).
9. Y. Kosaka, S.-P. Xie, *Nature* **501**, 403–407 (2013).
10. M. H. England *et al.*, *Nature Clim. Change* **4**, 222–227 (2014).
11. K. E. Trenberth, J. T. Fasullo, *Earth’s Future* **1**, 19–32 (2013).
12. X. Chen, K.-K. Tung, *Science* **345**, 897–903 (2014).
13. V. Guemas, F. J. Doblas-Reyes, I. Andreu-Burillo, M. Asif, *Nat. Clim. Chang.* **3**, 649–653 (2013).
14. G. A. Meehl, A. Hu, J. M. Arblaster, J. Fasullo, K. E. Trenberth, *J. Clim.* **26**, 7298–7310 (2013).
15. A. Dai, J. C. Fyfe, S.-P. Xie, X. Dai, *Nat. Clim. Change* **5**, 555–559 (2015).
16. M. G. Wyatt, S. Kravtsov, A. A. Tsonis, *Clim. Dyn.* **38**, 929–949 (2012).

ACKNOWLEDGMENTS

We thank Kravtsov *et al.* for providing the MATLAB code used for their analyses. All raw data, MATLAB code, and results from our analysis are available at the supplementary website www.meteo.psu.edu/holocene/public_html/supplements/Science2015.

14 May 2015; accepted 6 November 2015
10.1126/science.aac5208



Response to Comment on "Atlantic and Pacific multidecadal oscillations and Northern Hemisphere temperatures"

B. A. Steinman *et al.*

Science **350**, 1326 (2015);

DOI: 10.1126/science.aac5208

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of December 10, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/350/6266/1326.3.full.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/350/6266/1326.3.full.html#related>

This article **cites 12 articles**, 3 of which can be accessed free:

<http://www.sciencemag.org/content/350/6266/1326.3.full.html#ref-list-1>

This article appears in the following **subject collections**:

Geochemistry, Geophysics

http://www.sciencemag.org/cgi/collection/geochem_phys