

Reply to “Comment on ‘Comparison of Low-Frequency Internal Climate Variability in CMIP5 Models and Observations’”

ANSON H. CHEUNG^a

Department of Geosciences, The University of Arizona, Tucson, Arizona

MICHAEL E. MANN

Department of Meteorology and Atmospheric Science, and Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, Pennsylvania

BYRON A. STEINMAN

Department of Earth and Environmental Sciences, and Large Lakes Observatory, University of Minnesota Duluth, Duluth, Minnesota

LEELA M. FRANKCOMBE AND MATTHEW H. ENGLAND

ARC Centre of Excellence for Climate System Science, and Climate Change Research Centre, University of New South Wales, Sydney, New South Wales, Australia

SONYA K. MILLER

Department of Meteorology and Atmospheric Science, and Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, Pennsylvania

(Manuscript received 9 August 2017, in final form 28 September 2017)

ABSTRACT

In a comment on a 2017 paper by Cheung et al., Kravtsov states that the results of Cheung et al. are invalidated by errors in the method used to estimate internal variability in historical surface temperatures, which involves using the ensemble mean of simulations from phase 5 of the Coupled Model Intercomparison Project (CMIP5) to estimate the forced signal. Kravtsov claims that differences between the forced signals in the individual models and as defined by the multimodel ensemble mean lead to errors in the assessment of internal variability in both model simulations and the instrumental record. Kravtsov proposes a different method, which instead uses CMIP5 models with at least four realizations to define the forced component. Here, it is shown that the conclusions of Cheung et al. are valid regardless of whether the method of Cheung et al. or that of Kravtsov is applied. Furthermore, many of the points raised by Kravtsov are discussed in Cheung et al., and the disagreements of Kravtsov appear to be mainly due to a misunderstanding of the aims of Cheung et al.

1. Introduction

In our original article (Cheung et al. 2017, hereafter C2017), we applied a semiempirical method [referred to hereafter as the multimodel ensemble mean (MEM) method; Steinman et al. 2015] to the instrumental record

and the ensemble of models from phase 5 of the Coupled Model Intercomparison Project to isolate the surface temperature internal climate variability (ICV) in the Northern Hemisphere (NH), North Atlantic (NA), and North Pacific (NP)—referred as the Northern Hemisphere multidecadal oscillation (NMO), Atlantic multidecadal oscillation (AMO), and Pacific multidecadal oscillation (PMO), respectively—and compared results from the instrumental record and the CMIP5 historical simulations. The MEM method was originally developed by Mann et al. (2014), Steinman et al. (2015)

^a Current affiliation: Department of Earth, Environmental and Planetary Sciences, Brown University, Providence, Rhode Island.

Corresponding author: Anson H. Cheung, anson_cheung@brown.edu

and Frankcombe et al. (2015), who built upon the work of several prior studies (e.g., Knight 2009; Terray 2012), in order 1) to provide improved estimates of the internal variability signals in instrumental surface temperature data, and 2) to demonstrate that simple linear detrending is a highly flawed method for isolating internal variability in both simulated and observed surface temperature series, such that prior studies relying upon this method (e.g., Wyatt et al. 2012; Wyatt and Curry 2014) are erroneous as a result [see Mann et al. (2014) and Steinman et al. (2015) for additional discussion on these points]. We have therefore arrived at this discussion about the merits and limitations of the MMEM and SMEM methods as a direct result of the failures of older methods, most notably the linear detrending method (e.g., Wyatt et al. 2012), to produce valid estimates of the internal variability signal in climate time series.

In response to the work of Steinman et al. 2015 and Mann et al. 2014, Kravtsov (2017a, hereafter K2017) as well as Kravtsov et al. (2015), Kravtsov (2017b), and Kravtsov and Callicutt (2017, hereafter KC2017) have proposed that using the ensemble means of individual models—that is, the single-model ensemble mean (SMEM) method (which was first applied by Steinman et al. 2015; see supplemental material therein)—produces more accurate assessments of internal variability in instrumental and simulated surface temperature series. Through analysis of the Community Earth System Model (CESM) Large Ensemble Project (LENS; Kay et al. 2015), K2017 first argues that the method of KC2017 (the SMEM) should be used instead of the MMEM to isolate internal climate variability. Based on this result, K2017 further applies the SMEM method to isolate ICV in observations and CMIP5 historical simulations. K2017 highlights that the ICV difference, in particular the low-frequency component (>40 yr), between CMIP5 historical simulations and observations is larger than reported in C2017. Last, K2017 suggests 1) that it is more appropriate to correlate the index to the internal component of the SST field instead of the raw SST field, 2) that there are major discrepancies between the spatial patterns when using different methods, and 3) that analyzing the multimodel mean spatial pattern reduces the difference between the methods (i.e., MMEM vs SMEM). Here we show that the evidence presented in K2017 does not invalidate any of the conclusions presented in C2017 and instead assert that results from the MMEM and SMEM methods do not yield substantial differences. We also contend that K2017 misunderstood the aims and disregarded the discussion in C2017 on uncertainties

TABLE 1. List of CMIP5 climate model simulations used in this study. A single asterisk indicates that one realization from this model is not included in spatial pattern analysis. (Acronym expansions are available online at <http://www.ametsoc.org/PubsAcronymList>. The p_n following the GISS model acronyms refers to version n of the physics model used.)

Climate models	No. of realizations
CanESM2	5
CCSM4	6
CNRM-CM5	10
CSIRO Mk3.6.0	10
GFDL CM2.1	10
GFDL CM3	5
GISS-E2-Hp1	6
GISS-E2-Hp2*	6
GISS-E2-Hp3	6
GISS-E2-Rp1	6
GISS-E2-Rp2	6
GISS-E2-Rp3	6
HadCM3	10
HadGEM-ES	5
IPSL-CM5A-LR	6
MIROC5*	5
MRI-CGCM3	3

and potential errors associated with application of the MMEM. Therefore, K2017 has not raised any points that were not addressed at least to some extent by C2017.

In C2017, we show that 1) the low-frequency ICV spatiotemporal patterns in models are inconsistent with observations, and that this is likely due to a combination of forcing uncertainties in climate models, the relatively short length of the instrumental data, inconsistency between modeled and real-world spatial expressions of internal variability, and underestimation of low-frequency internal variability by the models; 2) the spatial and amplitude disagreement between models and observations increases as the smoothing time scale becomes longer; and 3) modeled and observed internal climate variability in the North Pacific, North Atlantic, and Northern Hemisphere are inconsistent.

2. Data and results

a. Comparisons between observations, MMEM, and SMEM

To be consistent with K2017, we used the data provided by K2017 and reanalyzed the spatial pattern and amplitude on different smoothing time scales. The only difference between the simulation ensemble used in K2017 and that applied here is the omission of two CMIP5 historical simulations for the spatial pattern

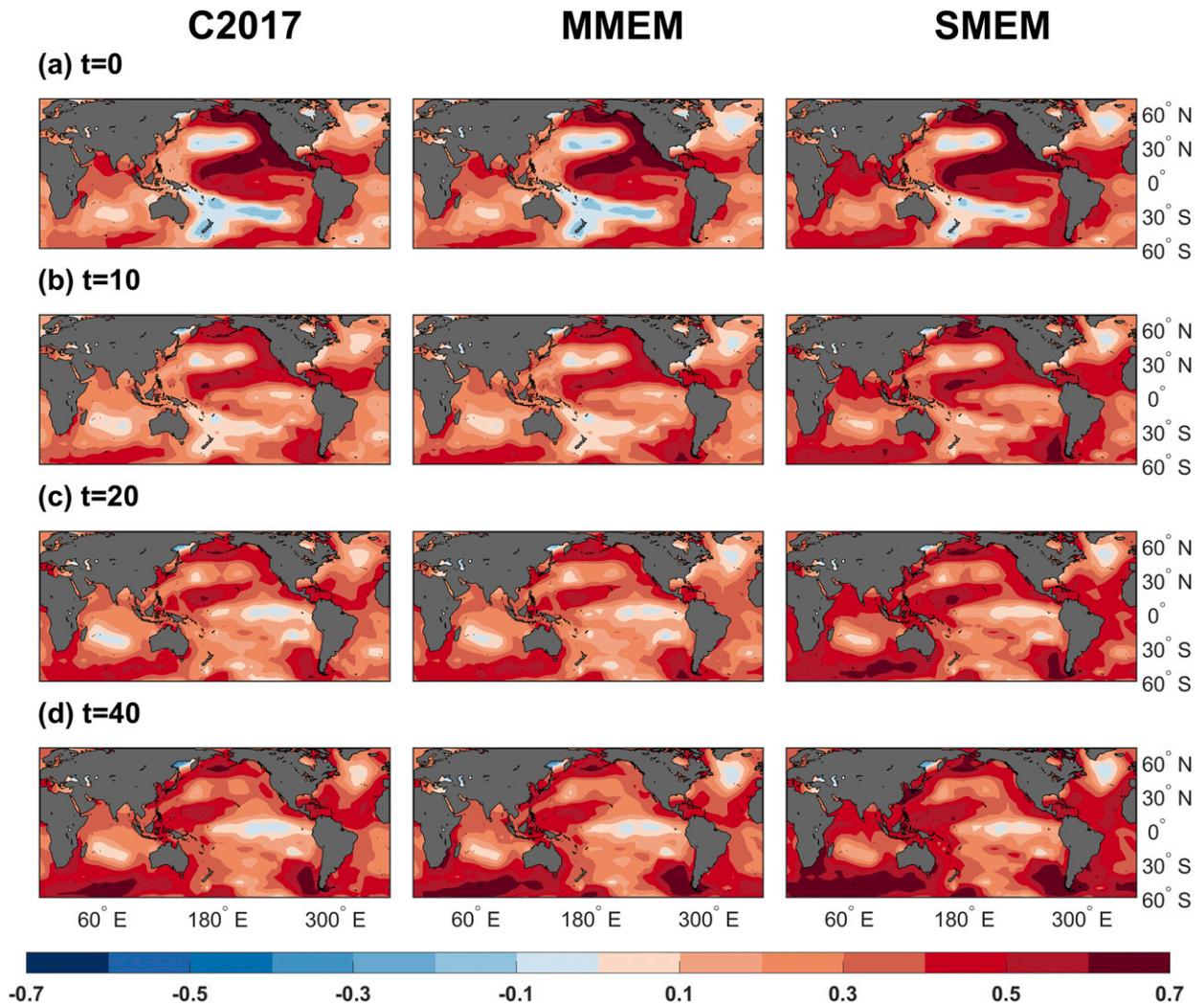


FIG. 1. Spatial patterns of the observed PMO (left) from C2017, (center) using the MMEM method, and (right) using the SMEM method. Shown are the patterns after applying a low-pass filter of (a) 0, (b) 10, (c) 20, and (d) 40 yr.

analysis, as these realizations were not used in C2017 (Table 1). For SMEM-based observations, we analyzed the mean estimated PMO, AMO, and NMO by averaging the 1700 estimates of PMO, AMO, and NMO presented in KC2017.

Comparisons between MMEM-based and SMEM-based observed PMO, AMO, and NMO spatial patterns across four different smoothing time scales (0, 10, 20, and 40 yr) do not yield any substantial difference (Figs. 1–3). Comparisons of MMEM-based and SMEM-based CMIP5 historical PMO, AMO, and NMO spatial patterns also do not yield any notable differences (Figs. 4–6). Therefore, the inconsistency between observed and simulated low-frequency ICV spatial patterns discussed by C2017 remains despite the use of a different method of estimating the forced signal

(cf. Figs. 1–6). K2017 compares observed and simulated PMO, AMO, and NMO amplitude using multiple approaches and shows that the amplitudes of the simulated PMO, AMO, and NMO are lower than the observed amplitudes. Even though a larger difference between the amplitudes is obtained when using SMEM rather than MMEM, the conclusion is the same: the simulated low-frequency ICV amplitude is lower than the observed amplitude. Therefore, the conclusion of C2017 that simulated spatial patterns and amplitudes are inconsistent with observed spatial patterns and amplitudes remains robust.

We also reanalyzed the power spectra of observed and simulated PMO, AMO, and NMO produced using the two different methods (SMEM and MMEM) to determine whether the amplitude derived from the SMEM

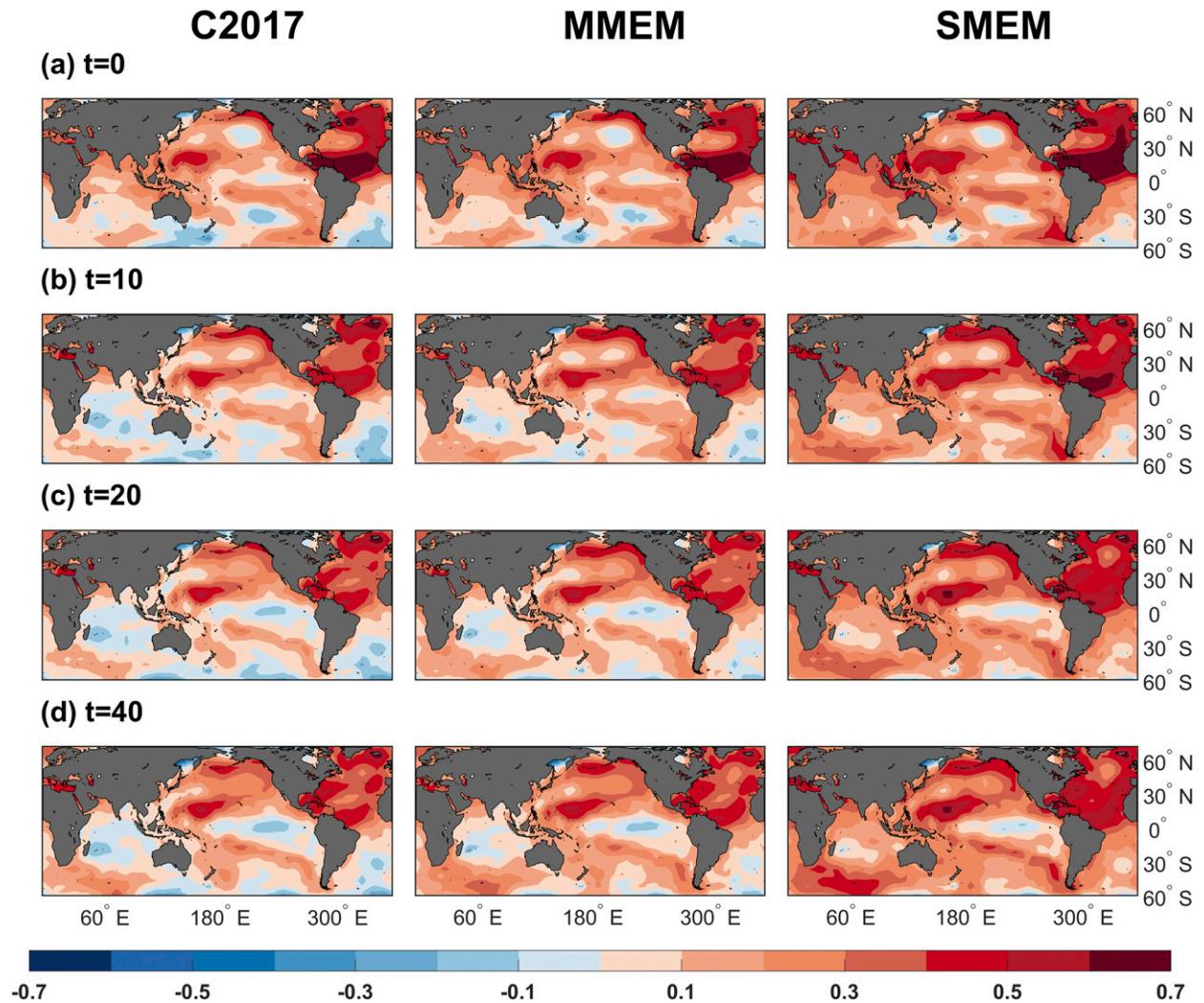


FIG. 2. As in Fig. 1, but displaying the AMO patterns.

approach is different from the MMEM approach. The power spectra are slightly different, with simulated ICV having less power, while observed ICV has more power at low frequencies when using the SMEM method in comparison to the MMEM method (Fig. 7). However, both results show that the difference between observations and historical simulations becomes substantial when focusing on low-frequency variations. We further analyzed the spatial properties of ICVs on different smoothing time scales based on the two different methods. Both methods show that discrepancies increase at longer smoothing time scales (Figs. 1–6). These results therefore support the conclusion of C2017 that the spatial and temporal disagreement between ICV in the historical simulations and observations increases as the smoothing time scale becomes longer.

Finally, we regressed the PMO and AMO onto the NMO with different smoothing time scales in order to examine the relative roles of the NP and NA in influencing NH mean temperatures. We find that regression results are not sensitive to the method (Figs. 8 and 9); however, they are sensitive to the choice of datasets, both for the observations and the models (i.e., which models are included in the CMIP5 ensemble). We note that the observational results for MMEM and SMEM presented here (orange and purple bars in Figs. 8 and 9) differ from the original study (green bars) even when applying the same method, whereas results using the K2017 data show the dominant role of the Pacific over all time scales. We suggest that this difference could be caused by the slightly different boundary constraints (e.g., different ensemble size to estimate the forced component or the application of a different NH temperature

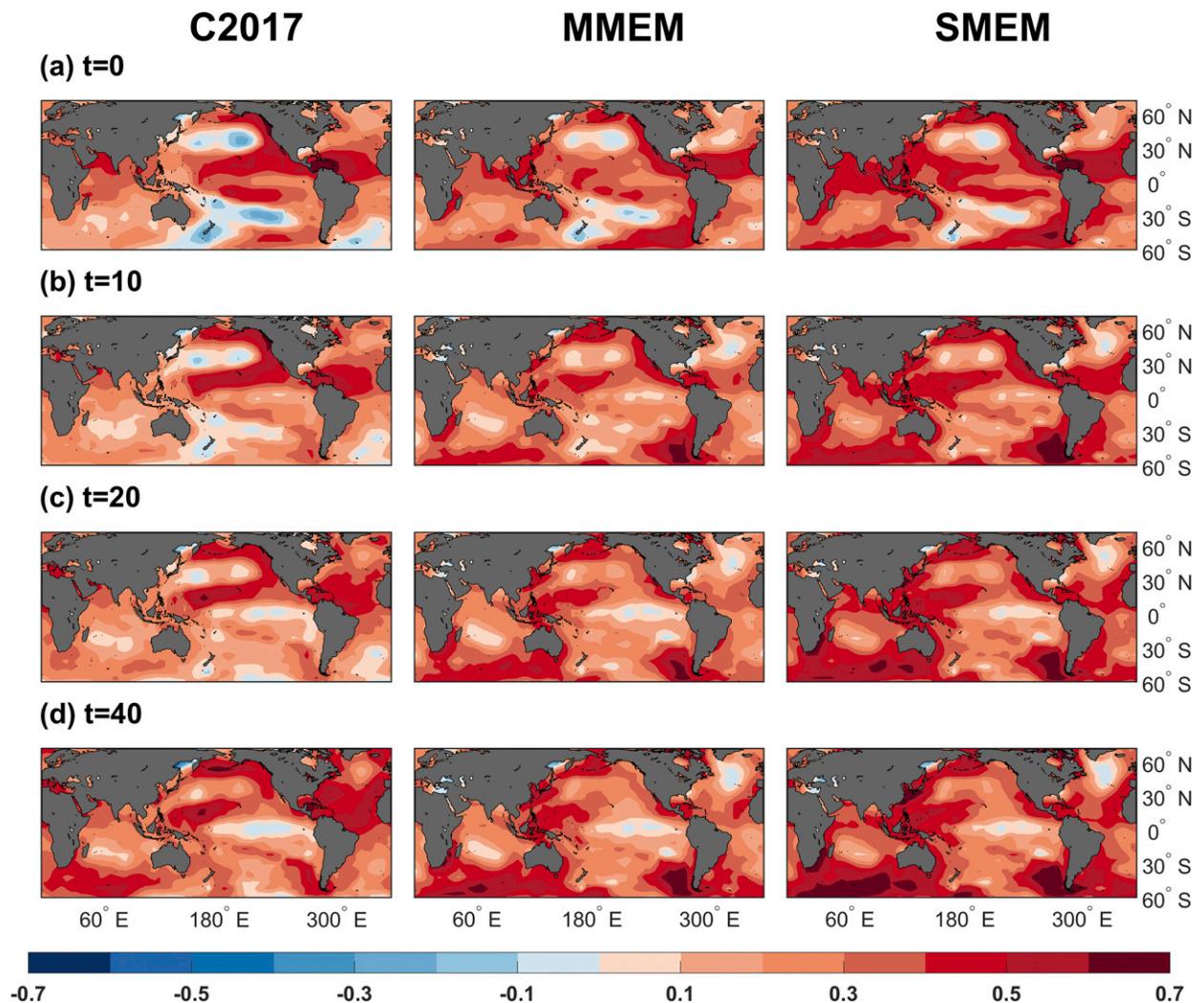


FIG. 3. As in Fig. 1, but displaying the NMO patterns.

dataset), making the results from our original study subject to uncertainty. Nevertheless, regression results based on the two different approaches are the same, showing that the uncertainty does not result from the choice of method.

b. Aims of C2017

Most of the arguments put forth in K2017 appear to result from a misunderstanding of the aims of C2017. The primary objective of our original article is to better understand internal variability in the observational record and to compare it to internal variability in the CMIP5 model ensemble. The basis of the MMEM approach is that we only have one realization of the observational record, and that the forced component of the observational record can be best estimated by the multimodel mean of the CMIP5 historical simulations. We agree that it is more suitable to use realizations from the

same model to characterize the internal variability of an individual model; however, this gives us no guidance as to the best way of applying the method to *observations*, which may be treated as a model with only a single ensemble member. While there have been developments on how to weigh different climate models when studying various characteristics of the climate system (e.g., Knutti et al. 2017), presently there is no consensus on which model(s) may be best at simulating the forced signal. Therefore, there is no justification for the application of a particular individual model ensemble in this capacity. We maintain that the multimodel ensemble mean remains the most sensible choice for estimating the forced signal in the observational record.

The second aim of C2017 is to generally understand the ICV of CMIP5 models. We agree that some models do better in simulating certain aspects of the

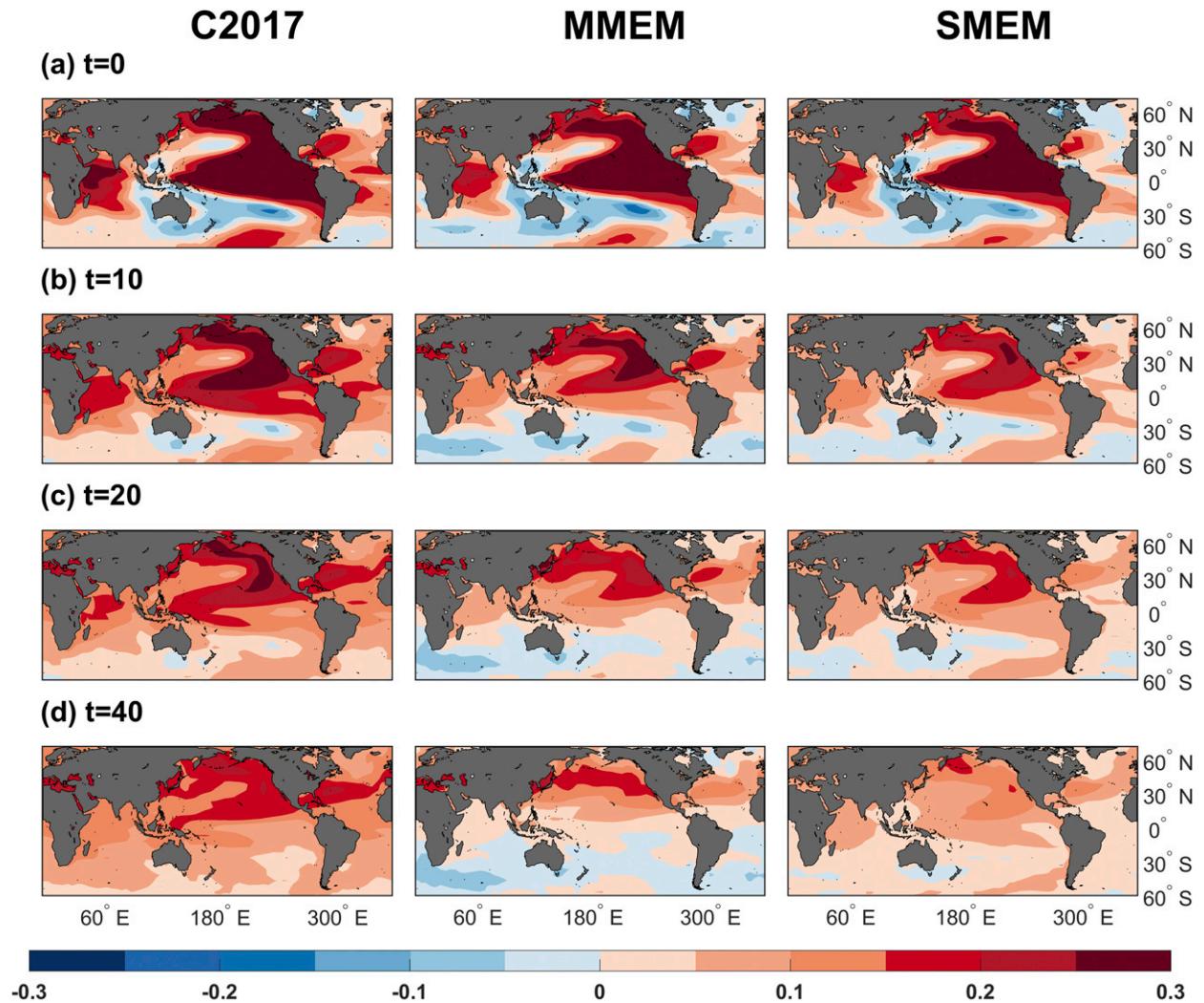


FIG. 4. Spatial patterns of the historical PMO (left) from C2017, (center) using the MMEM method, and (right) using the SMEM method. Shown are the patterns after applying a low-pass filter of (a) 0, (b) 10, (c) 20, and (d) 40 yr.

climate system than others. In fact, C2017 showed that models exhibit a wide range of spatial patterns, amplitude, and spectral characteristics (see Figs. 4, 5, and 7 in C2017). However, it is noteworthy that the aim of C2017 is not to find the models that best simulate the ICV of each target region, but to understand the behavior of the CMIP5 ensemble in a general sense. To this end, C2017 compares ICV in the observational record to that of the CMIP5 ensemble, instead of to results from individual models.

K2017 argues that the MMEM approach mischaracterizes the forced component of individual climate models. However, section 3 of C2017 discusses the possible effects of inadequate removal of external forcing, the uncertainty of external forcings used in the CMIP5 ensemble, and the fact that differences in

model physics could lead to model–model and model–observation discrepancies. K2017 also disagrees with the method used to analyze the spatial pattern of ICV, arguing that the forced component of each grid point should be removed before computing the spatial correlation pattern. We agree that this method has its advantages theoretically; however, in practice the forced component might not be sufficiently removed because of larger variability at individual locations in comparison to regional averages, which could lead to substantial errors when analyzing the spatial patterns. Here we demonstrate that when assessing the spatial patterns of ICV, the choice of method for removing the forced signal does not substantially change the results when analyzing the multimodel ensemble average (see Fig. 5 in K2017; see also Figs. 4–6 herein). Therefore, the

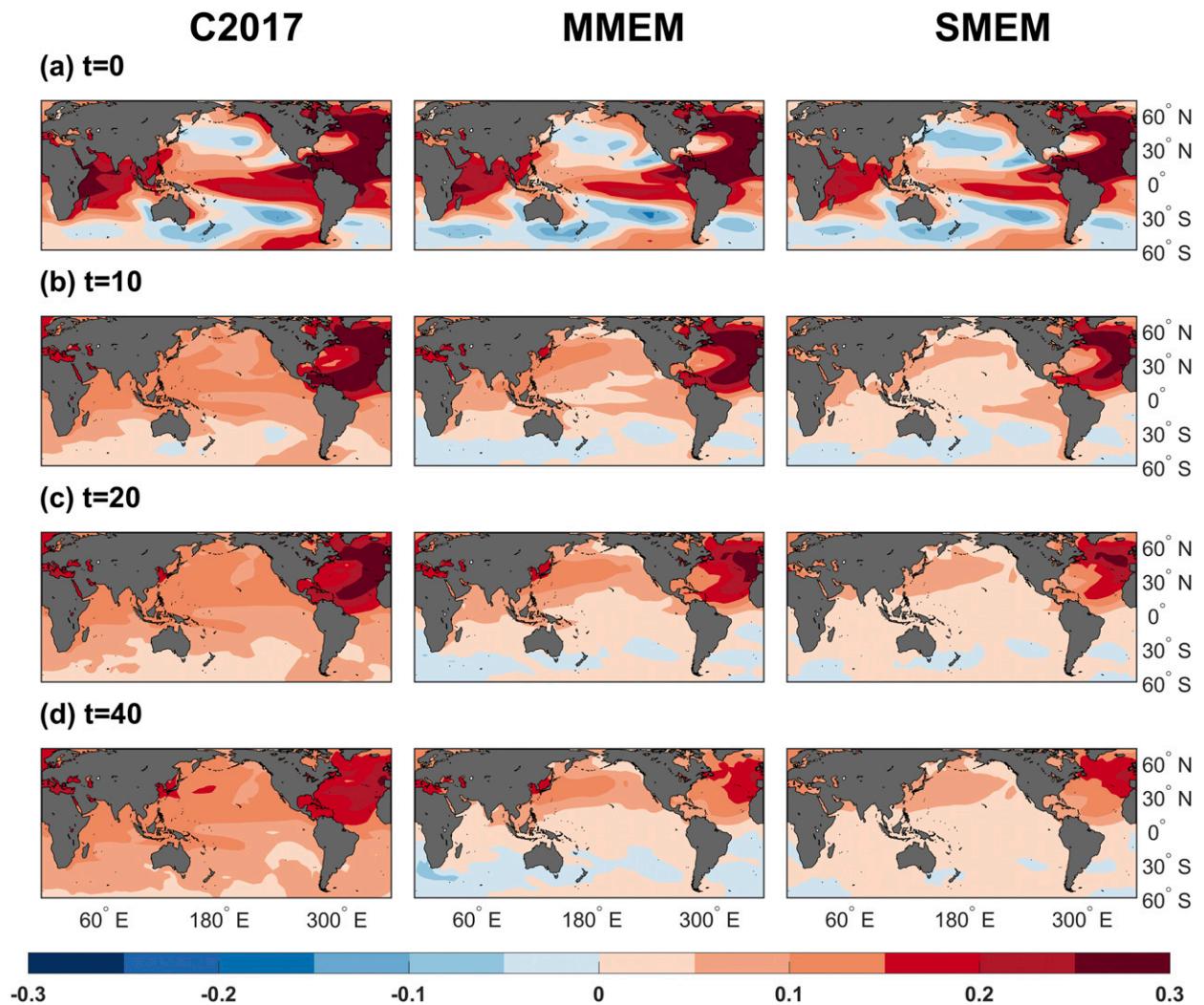


FIG. 5. As in Fig. 4, but displaying the AMO patterns.

spatial patterns of ICV derived from MMEM and SMEM methods are not markedly different, and the conclusions of C2017 remain robust.

3. Conclusions

K2017 argues that the conclusions of C2017 are invalid largely due to methodological errors associated with the MMEM method. We do acknowledge that there are uncertainties and deficiencies in the MMEM method, as discussed in detail in Frankcombe et al. (2015) as well as in C2017, which affect the estimation of internal variability in the instrumental record and in the CMIP5 historical simulations. By comparing results derived from the two different approaches (SMEM and MMEM), we show that the results of C2017 are robust and, furthermore, are strengthened by the fact that they

can be obtained using distinct methods. We further reiterate that the goals of C2017 were to isolate internal climate variability in the observational record and compare it with results from the CMIP5 model ensemble. We did not aim to identify the individual model that can best simulate ICV in targeted regions. Therefore, we find that the arguments and criticisms raised by K2017 are primarily due to a misunderstanding of the aims of C2017, and that the K2017 study is complementary rather than in opposition to C2017. Last, we would like to stress that both the SMEM and MMEM methods are superior to older methods for estimating the internal variability signal in climate data and, in particular, are far more robust methods than the linear detrending procedure applied by many prior studies (e.g., Wyatt et al. 2012; Wyatt and Curry 2014).

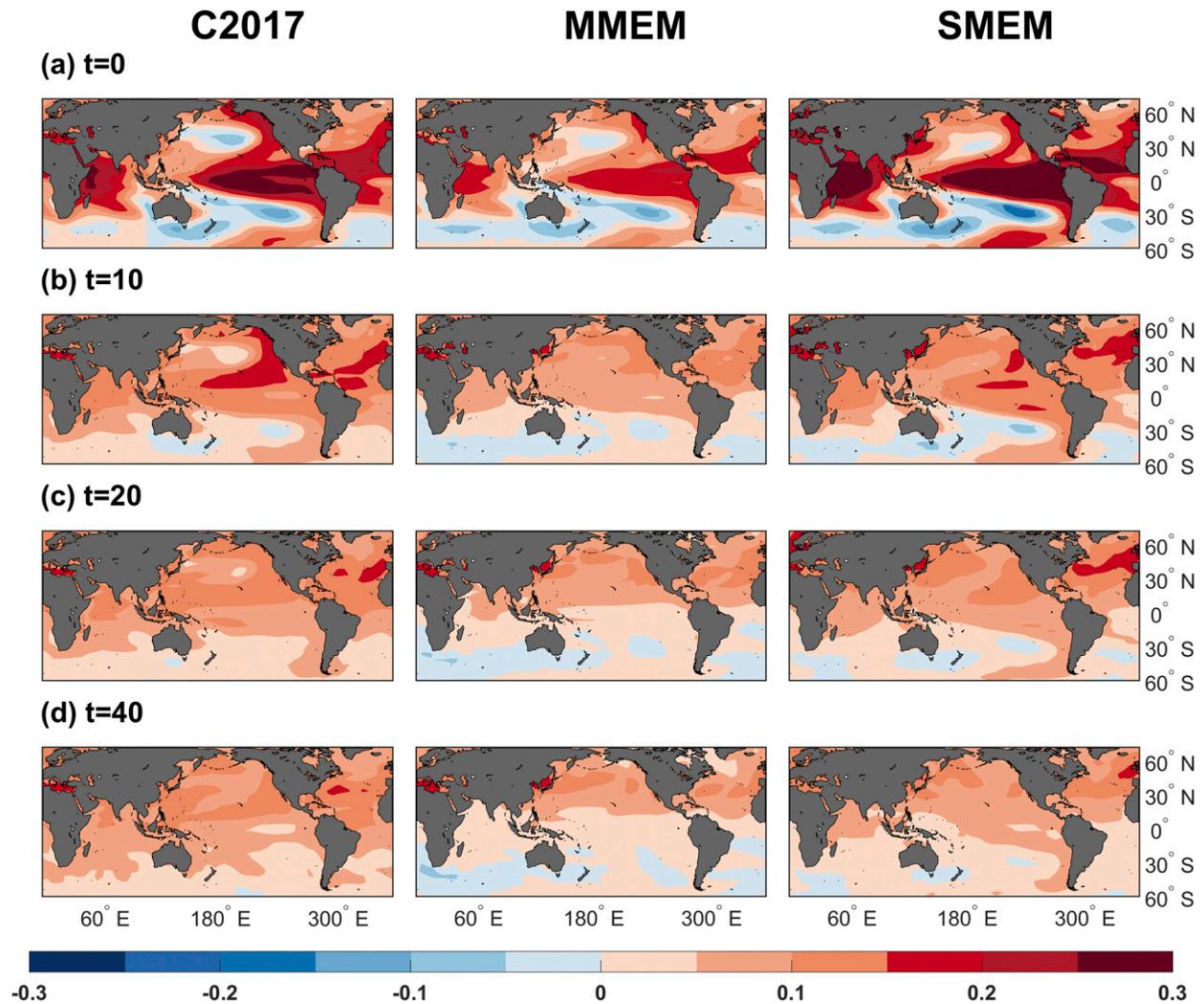


FIG. 6. As in Fig. 4, but displaying the NMO patterns.

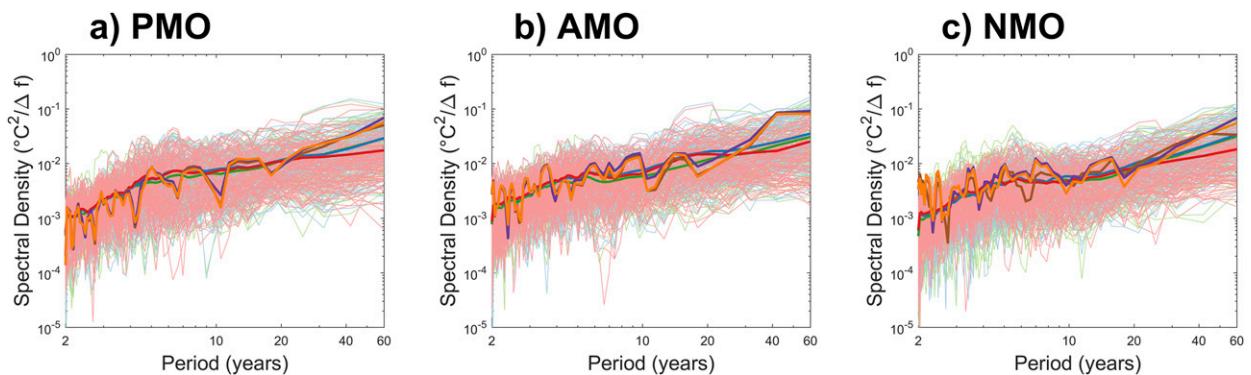


FIG. 7. Power spectra of the (a) PMO, (b) AMO, and (c) NMO from observations (brown) and historical simulations (green) used in C2017, observations (purple) and historical simulations (blue) derived from the MMEM method, and observations (orange) and historical simulations (red) derived from the SMEM method. The thick line for historical simulations represents the mean spectral power whereas the thin lines represent each realization's spectral power.

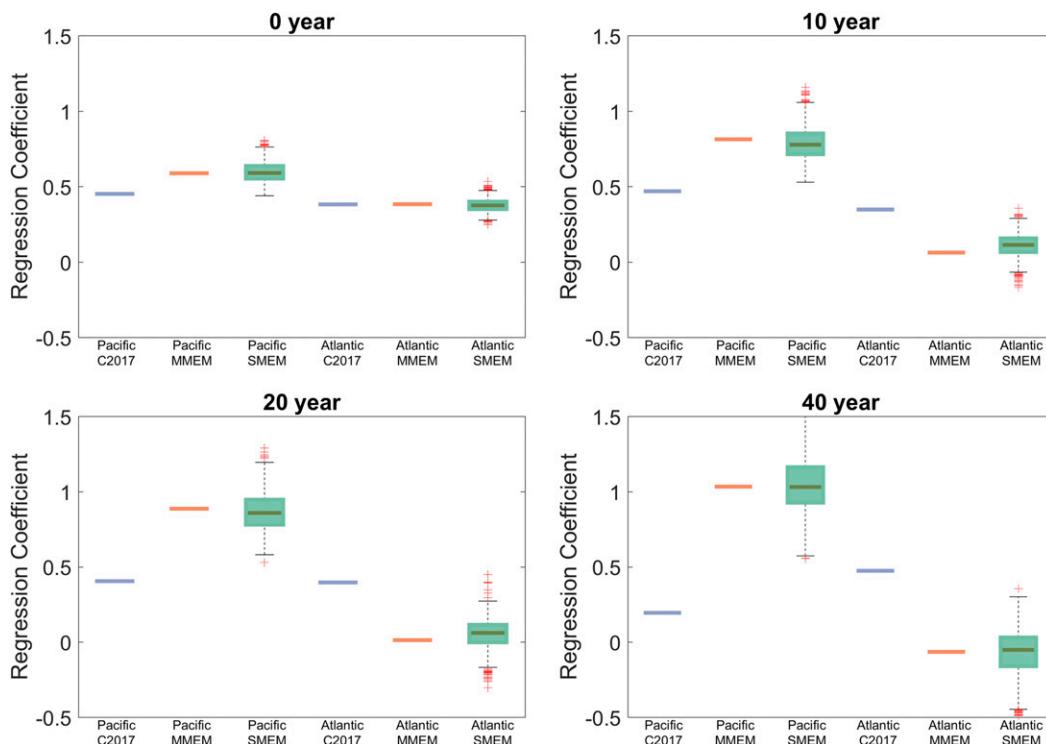


FIG. 8. Regression coefficients of observed PMO and AMO from the original C2017 study (purple) and using the K2017 data analyzed with the MMEM (orange) and SMEM (green) method. Red plus symbols represent outliers (e.g., $x > 2.7\sigma$, where σ is the standard deviation) and bars depict median values.

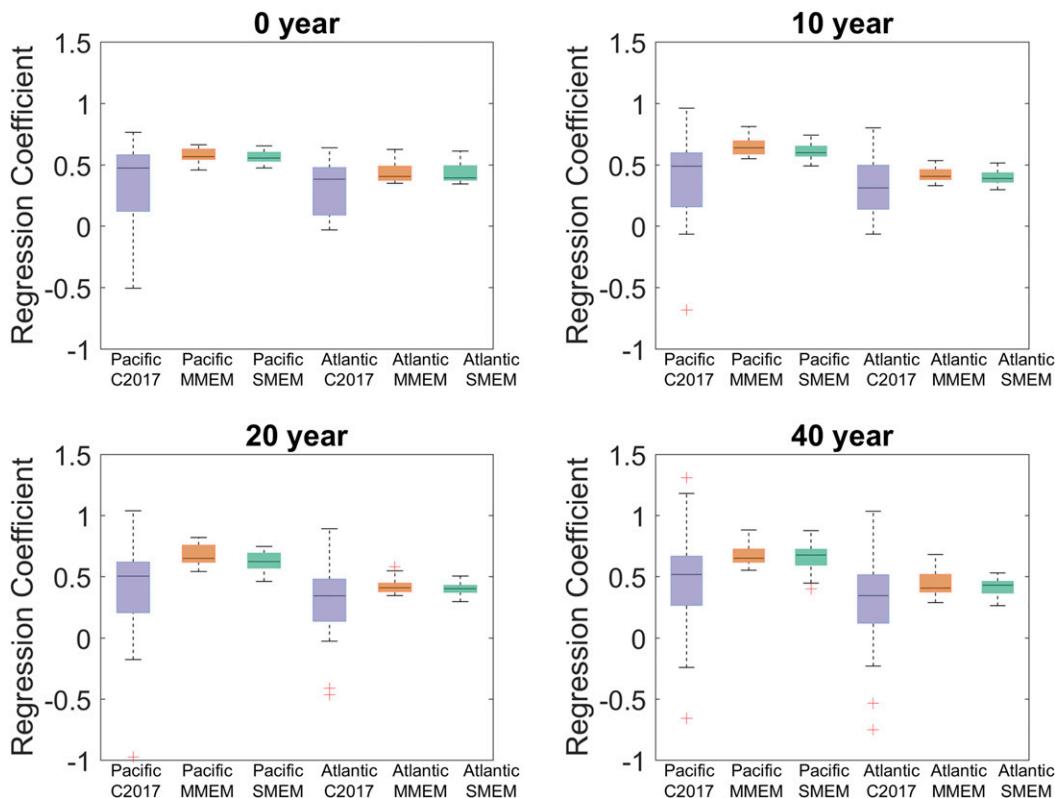


FIG. 9. As in Fig. 8, but using the CMIP5 historical simulations. Orange plus symbols represent outliers (e.g., $x > 2.7\sigma$, where σ is the standard deviation) and bars depict median values.

Acknowledgments. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. AHC acknowledges support from the U.S. National Science Foundation (AGS-1263225). MHE and LMF are supported by the Australian Research Council. We thank S. Kravtsov for making his data available. Kaplan SST version 2 data are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, from their website at <http://www.esrl.noaa.gov/psd/>. HadISST data are provided by the Met Office Hadley Centre (online at <http://www.metoffice.gov.uk/hadobs>). ERSST data are provided by NOAA (online at <http://www.ncdc.noaa.gov/data-access/marineocean-data/extended-reconstructed-sea-surface-temperature-ersst-v3b>).

REFERENCES

- Cheung, A. H., M. E. Mann, B. A. Steinman, L. M. Frankcombe, M. H. England, and S. K. Miller, 2017: Comparison of low-frequency internal climate variability in CMIP5 models and observations. *J. Climate*, **30**, 4763–4776, <https://doi.org/10.1175/JCLI-D-16-0712.1>.
- Frankcombe, L. M., M. H. England, M. E. Mann, and B. A. Steinman, 2015: Separating internal variability from the externally forced climate response. *J. Climate*, **28**, 8184–8202, <https://doi.org/10.1175/JCLI-D-15-0069.1>.
- Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) Large Ensemble Project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, <https://doi.org/10.1175/BAMS-D-13-00255.1>.
- Knight, J. R., 2009: The Atlantic multidecadal oscillation inferred from the forced climate response in coupled general circulation models. *J. Climate*, **22**, 1610–1625, <https://doi.org/10.1175/2008JCLI2628.1>.
- Knutti, R., J. Sedlacek, B. M. Sanders, R. Lorenz, E. M. Fischer, and V. Eyring, 2017: A climate model projection weighting scheme accounting for performance and interdependence. *Geophys. Res. Lett.*, **44**, 1909–1908, <https://doi.org/10.1002/2016GL072012>.
- Kravtsov, S., 2017a: Comment on “Comparison of low-frequency internal climate variability in CMIP5 models and observations.” *J. Climate*, **30**, 9763–9772, <https://doi.org/10.1175/JCLI-D-17-0438.1>.
- , 2017b: Pronounced differences between observed and CMIP5-simulated multidecadal climate variability in the twentieth century. *Geophys. Res. Lett.*, **44**, 5749–5757, <https://doi.org/10.1002/2017GL074016>.
- , and D. Callicutt, 2017: On semi-empirical decomposition of multidecadal climate variability into forced and internally generated components. *Int. J. Climatol.*, **37**, 4417–4433, <https://doi.org/10.1002/joc.5096>.
- , M. G. Wyatt, J. A. Curry, and A. A. Tsonis, 2015: Comment on “Atlantic and Pacific multidecadal oscillations and Northern Hemisphere temperatures.” *Science*, **350**, 1326, <https://doi.org/10.1126/science.aab3570>.
- Mann, M. E., B. A. Steinman, and S. K. Miller, 2014: On forced temperature changes, internal variability, and the AMO. *Geophys. Res. Lett.*, **41**, 3211–3219, <https://doi.org/10.1002/2014GL059233>.
- Steinman, B. A., M. E. Mann, and S. K. Miller, 2015: Atlantic and Pacific multidecadal oscillations and Northern Hemisphere temperatures. *Science*, **347**, 988–991, <https://doi.org/10.1126/science.1257856>.
- Terray, L., 2012: Evidence for multiple drivers of North Atlantic multi-decadal climate variability. *Geophys. Res. Lett.*, **39**, L19712, <https://doi.org/10.1029/2012GL053046>.
- Wyatt, M. G., and J. A. Curry, 2014: Role for Eurasian Arctic shelf sea ice in a secularly varying hemispheric climate signal during the 20th century. *Climate Dyn.*, **42**, 2763–2782, <https://doi.org/10.1007/s00382-013-1950-2>.
- , S. Kravtsov, and A. A. Tsonis, 2012: Atlantic multidecadal oscillation and Northern Hemisphere's climate variability. *Climate Dyn.*, **38**, 929–949, <https://doi.org/10.1007/s00382-011-1071-8>.