AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER

10.1002/2016GL068159

Key Points:
- The recent slowdown in surface warming likely was not predictable using statistical methods
- The Pacific Multidecadal Oscillation does not appear to exhibit any predictability
- The Atlantic Multidecadal Oscillation does appear to exhibit some predictability

Supporting Information:
- Supporting Information S1

Correspondence to:
B. A. Steinman,
bsteinma@d.umn.edu

Citation:
Mann, M. E., B. A. Steinman, S. K. Miller, L. M. Frankcombe, M. H. England, and A. H. Cheung (2016), Predictability of the recent slowdown and subsequent recovery of large-scale surface warming using statistical methods, *Geophys. Res. Lett.*, *43*, 3459–3467, doi:10.1002/2016GL068159.

Received 6 FEB 2016
Accepted 13 MAR 2016
Accepted article online 18 MAR 2016
Published online 8 APR 2016

Predictability of the recent slowdown and subsequent recovery of large-scale surface warming using statistical methods

Michael E. Mann[1], Byron A. Steinman[2], Sonya K. Miller[1], Leela M. Frankcombe[3], Matthew H. England[3], and Anson H. Cheung[4]

[1]Department of Meteorology and Earth and Environmental Systems Institute, Pennsylvania State University, University Park, Pennsylvania, USA, [2]Department of Earth and Environmental Sciences and Large Lakes Observatory, University of Minnesota Duluth, Duluth, Minnesota, USA, [3]ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, New South Wales, Australia, [4]Department of Geosciences, University of Arizona, Tucson, Arizona, USA

Abstract The temporary slowdown in large-scale surface warming during the early 2000s has been attributed to both external and internal sources of climate variability. Using semiempirical estimates of the internal low-frequency variability component in Northern Hemisphere, Atlantic, and Pacific surface temperatures in concert with statistical hindcast experiments, we investigate whether the slowdown and its recent recovery were predictable. We conclude that the internal variability of the North Pacific, which played a critical role in the slowdown, does not appear to have been predictable using statistical forecast methods. An additional minor contribution from the North Atlantic, by contrast, appears to exhibit some predictability. While our analyses focus on combining semiempirical estimates of internal climatic variability with statistical hindcast experiments, possible implications for initialized model predictions are also discussed.

1. Introduction

Considerable attention has been paid to a temporary slowdown in large-scale surface warming that began during the early 2000s and persisted into the early 2010s. Some studies have attributed the slowdown at least in part to recent changes in natural and anthropogenic external forcing [*Foster and Rahmstorf*, 2011; *Santer et al.*, 2014; *Schmidt et al.*, 2014], while other studies have implicated internal variability [*Trenberth and Fasullo*, 2013; *Mann et al.*, 2014] involving sustained La Niña-like conditions, strengthened Pacific trade winds and/or lowered sea surface temperatures (SSTs) in the tropical Pacific [*Kosaka and Xie*, 2013; *England et al.*, 2014; *Risbey et al.*, 2014; *Dai et al.*, 2015; *Steinman et al.*, 2015], and the negative phase of the so-called "Pacific Decadal Oscillation" or PDO [*Mantua and Hare*, 2002].

Though less pronounced if data through 2014 are used [*Cowtan et al.*, 2015; *Karl et al.*, 2015], the slowdown in Northern Hemisphere (NH) mean warming is evident through at least the early part of this decade. *Steinman et al.* [2015] associate the slowdown with the negative phase of what they term the "Northern Multidecadal Oscillation" (NMO), defined as the multidecadal internal variability in NH mean temperature. The recent NMO trend appears to have been dominated by multidecadal internal variability in North Pacific (NP) mean surface temperature, which *Steinman et al.* [2015] term the "Pacific Multidecadal Oscillation" (PMO) (essentially, the multidecadal component of the PDO). The so-called Atlantic Multidecadal Oscillation (AMO), defined as the corresponding low-frequency internal variability component in North Atlantic (NA) mean surface temperature [*Delworth and Mann*, 2000; *Kerr*, 2000; *Knight et al.*, 2005], appears to have played a more modest role. *Steinman et al.* [2015] speculate that the rate of warming may increase in the decade ahead as the PMO; and hence, NMO begins to reverse direction. Subsequent work [*Dai et al.*, 2015] has reached a similar conclusion.

An intriguing issue that has received somewhat less attention is whether or not such behavior—both the slowdown and now its potential recovery—exhibits any predictability. Initialized climate models, which incorporate some information about the history of internal variability, are able to outperform uninitialized models [*Kirtman et al.*, 2013; *Meehl et al.*, 2014b], suggesting some skill in decadal predictability in large-scale mean temperature. Several recent studies [e.g., *Guemas et al.*, 2013; *Meehl et al.*, 2014a, 2014b] propose that the early 2000s slowdown period was indeed potentially predictable several years in advance. There nonetheless remain questions about the putative forecasting skill; for example, whether the initialized climate state is the source of any apparent predictability or whether it is instead in part due to the climate response to the hindcast forcing (e.g., volcanoes—discussed below).

Newman [2013] compares skill across a range of statistical and model-based forecasting approaches. While finding some evidence of decadal predictability in large-scale (global) mean temperature (attributable essentially to the forced climate change signal), they find little evidence in any of the forecasting schemes for decadal predictability in the PDO or AMO (i.e., for the internal variability), as measured by the root-mean-square error (RMSE) of the forecasts.

Some recent studies, by contrast, have argued for predictability of interdecadal Pacific and Atlantic variability modes closely related to the AMO and PDO/PMO [e.g., *Doblas-Reyes et al*., 2013; *García-Serrano et al*., 2015]. However, such studies have typically employed definitions of these modes using linear detrending (or removal of the global mean SST from the time series)—methods that have been argued to suffer from the potential leakage of forced variability [*Mann et al*., 2014]. The use of such methods could lead to false apparent predictability, since what is inferred to represent predictable internal variability in these cases may instead simply be residual externally forced variability [*Mann et al*., 2014].

A procedure that faithfully estimates the forced climate signal and removes it from the observations (e.g., the so-called "target region regression" method first proposed by *Steinman et al*. [2015] and elaborated upon by *Frankcombe et al*. [2015]) is required to properly identify the internal variability component and to therefore provide a more robust assessment of its predictability. Here we make use of the target region regression approach in predictions of future temperature changes. While we focus on NH mean temperature, parallel analyses are performed for NP and NA mean temperatures. We employ a forecasting scheme that uses estimates of both the forced climate signal and "oscillatory" (i.e., NMO/PMO/AMO) internal variability derived from that approach; the forced signal is estimated using the Coupled Model Intercomparison Project Phase 5 (CMIP5) historical simulation multimodel mean series, while the internal variability is estimated as the residual variation in the observational series once the forced signal is removed.

The NMO, PMO, and AMO series are defined as multidecadally smoothed versions of the corresponding residual series (see supporting information for details of data and methods). The resulting indices display spatial patterns (Figure 1) consistent with previous studies. For example, *Mann and Park* [1994, 1999] find a North Pacific-centered oscillation with a dominant time scale of ~20 years and a pattern similar to that of the PMO found here; while *Delworth and Mann* [2000] identify a North Atlantic multidecadal oscillation with a time scale of ~50–70 years and a pattern similar to that of the AMO evinced here. The PMO displays a classic "horseshoe" pattern of temperature anomalies in the North Pacific and a symmetric pattern of anomalies about the equator, with a dominant interdecadal (>20 year) time scale of variation. The AMO displays a pattern of large positive SST anomalies in the North Atlantic and negative SST anomalies in the South Atlantic and southern ocean, with a dominant multidecadal (>50 year) time scale of variation. The NMO shows spatiotemporal features of both modes. The post-2000 NH mean warming slowdown (i.e., the decline of the NMO during that interval), as noted elsewhere [*Steinman et al*., 2015], is seen (Figure 1) to have arisen primarily from a corresponding decline in the PMO, which appears to have reached peak negative values and recently begun a recovery. A more modest, recent decline in the AMO appears to have delayed slightly the corresponding recovery in the NMO.

A prediction of the overall temperature change requires forecasting both the externally forced and internal variability components. We consider two possible approaches for the prediction of the forced component (which, in turn, leads to two different corresponding estimates of the internal variability component). In the first approach, we assume simple persistence of the climatological (20 year) forced trend, a reasonable assumption given the nearly linear nature of anthropogenic forcing during the late twentieth century interval wherein greenhouse forcing dominates. In the second approach, we use a more sophisticated scheme for projecting the anthropogenic-forced trend which makes use of anthropogenic-only forcing experiments (supporting information). The two schemes are compared in Figure 2 for one particular hindcast experiment.

A second-order autoregressive process is used to project the oscillatory component forward in time. Using sets of analyses performed over the intervals $1880 - X$ (where $2013 \geq X \geq 1980$), we performed hindcast experiments from year $\underline{X}+1$ through 2014. This process was repeated for all choices of X over the range of decadal/interdecadal smoothing time scales for defining an oscillation. (i.e., varying the low-pass filter cutoff frequency from 0.1 to 0.02 cycle/yr corresponding to a 10–50 year range in period; the latter cutoff is a generous estimate of the lowest frequency oscillatory component that can be identified in roughly a century of data).

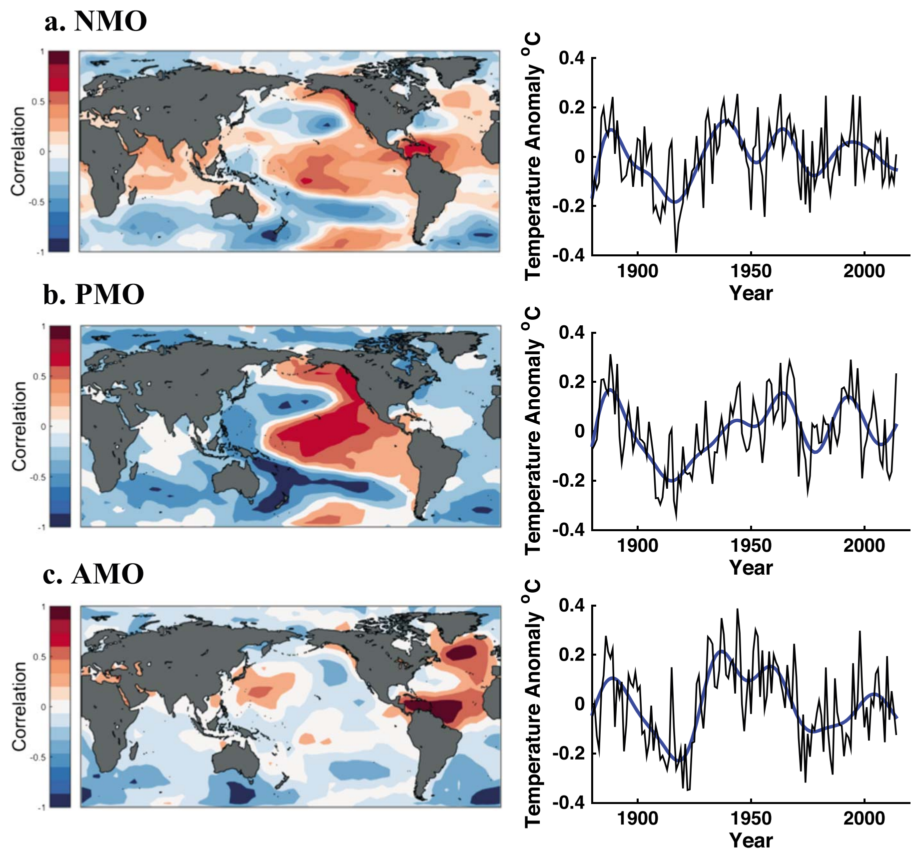

Figure 1. (left) Spatial patterns and (right) time series associated with the (a) NMO, (b) PMO, and (c) AMO. The spatial patterns indicate the correlation of the annual series (black curves) shown on the right with gridded global SST data (1950–2014). The annual series indicate the estimated internal variability series for the NH, NP, and NA surface temperature series, while the corresponding filtered (20 year low pass) NMO, PMO, and AMO series are shown by the blue curves.

The smoothing filter [*Mann*, 2008] employs optimal boundary constraints based on the combination of boundary constraints (minimization of norm, slope, or roughness at edges) that minimizes the mean-square-error with respect to the raw series. No future information from the time series is used in implementing boundary constraints, so that subsequent hindcasts represent truly blind predictions of future behavior.

We compared skill for block hindcasts ranging in length from 1 to 15 years. Note that the lead time in these hindcasts is variable (e.g., 1 year for the year 1 hindcast and 15 years for the year 15 hindcast). The earliest hindcasts in these experiments were conducted over the period 1980–1995, and the latest hindcasts over the period 1999–2014. If there is a decadal/multidecadal oscillatory signal that exhibits predictability (i.e., that is distinguishable from simple low-frequency "red noise"), we expect to find *some* time scale (i.e., choice of filtering frequency) wherein the decadal hindcast skill, which we measure via RMSE (while noting that other metrics are worthy of consideration) [see e.g., *Jolliffe and Stephenson*, 2011; *Weigel and Mason*, 2011], consistently outperforms both the prediction of the forced signal alone and all null forecasts for predicting the internal variability, including damped persistence (the null forecast for AR1 red noise) as well as climatology, simple persistence, and extended persistence.

Using the first of the two schemes (i.e., full forced trend persistence; Figure 2a) for forecasting the forced component, we appear to find evidence of such skill (Figure 2c). The average RMSE across all hindcasts for the decadal hindcast of NH mean temperature (e.g., a 14 year forecast which reflects an average lead time of 7.5 years) appears lower than for predictions using the forced component alone or using any of the null forecasting schemes for the internal component (Figure 2e) for any smoothing time scale greater than ~18 years ($f < 0.054$ cycle/yr). A RMSE minimum is obtained at an optimal smoothing time scale of 30 years ($f = 0.033$ cycle/yr). Using this smoothing time scale (which defines our "optimal" prediction), the mean

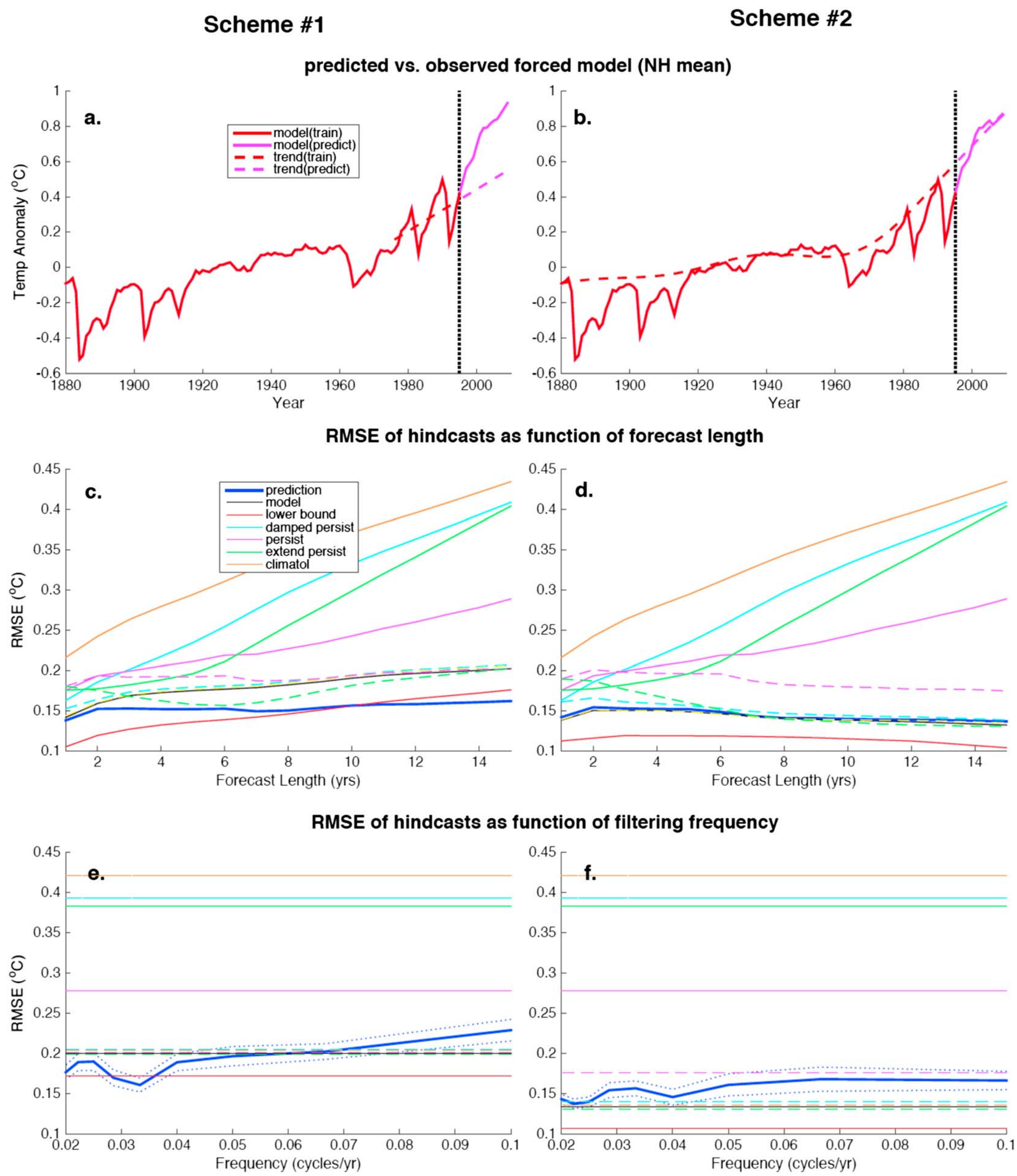

Figure 2. Assessment of hindcast skill for NH mean temperature predictions using the two different schemes: linear forward projection of the (a, c, e) *total forced* trend and linear forward projection of the (b, d, f) estimated *anthropogenic-only* trend. Comparisons of the actual and predicted CMIP5-based forced series during training and prediction intervals based on the two different schemes, for the case of a 14 year prediction made in 1995 (note that the "actual" series are not identical for the two different cases due to the different methods for extending model series to the 2014 boundary) (Figures 2a and 2b). RMSE for optimal prediction (blue; dotted curves indicate one sigma uncertainties) relative to other forecasts as a function of forecast length (Figures 2c and 2d) and filtering frequency (for 14 year predictions) (Figures 2e and 2). Results are shown for (solid line) full forecasts (forced + NMO component) and (dashed line) internal variability components (i.e., "NMO") alone, as well as predictions using forced component of forecast only (black line). Also shown (red line) is the RMSE between the true and predicted forced component, i.e., the lower bound error estimate discussed in the text.

RMSE across hindcasts is found to be lower than for all other forecasting schemes for all forecast lengths (Figure 2c). Indeed, there is a narrow-frequency range over which the RMSE of the total forecast, paradoxically, appears even lower than the "lower bound error" provided by our forecast of the forced component alone. A similar analysis of both NP and NA series (supporting information) yields similar apparent improvements relative to null forecasting schemes.

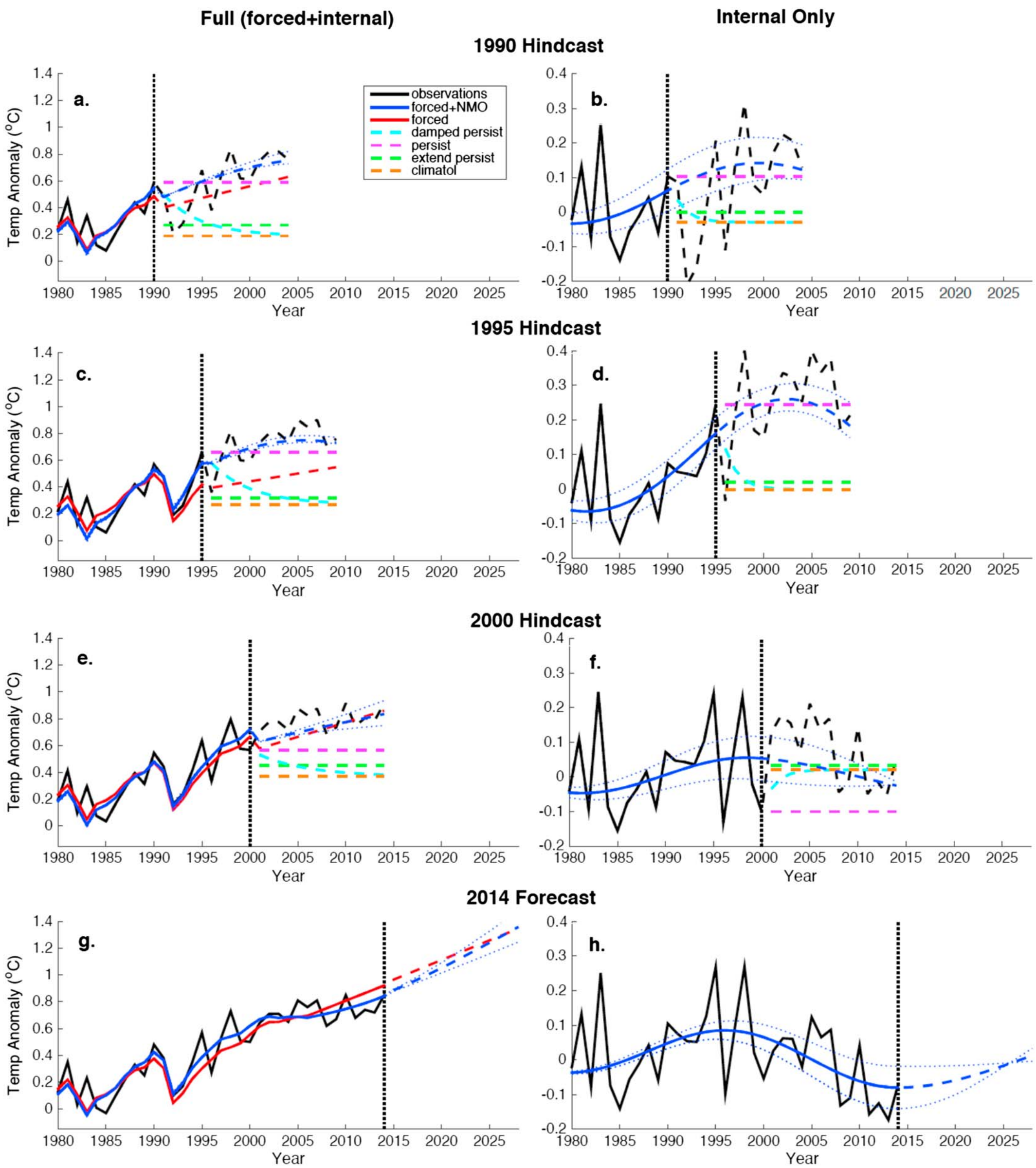

Figure 3. Hindcasts/forecasts of (a, c, e, g) NH mean temperature and (b, d, f, h) NMO based on the total forced trend scheme. Shown are results for 14 year hindcasts in 1990 (Figure 3a and 3b), 1995 (Figures 3c and 3d), and 2000 (Figures 3e and 3f), along with a 14 year prediction in 2014 (Figures 3g and 3). Shown are the annual observations (or in Figures 3b, 3d, and 3e, the observations minus the projected forced trend (black line) both preprediction (solid line) and postprediction (dashed line)), along with model-estimated forced series (red solid line), sum of forced and internal variability series (blue line), hindcast of forced component (red dashed line), and total hindcast (blue dashed line), along with various null forecasts. Upper and lower one sigma uncertainty estimates (blue dotted line) are shown for hindcasts and forecasts. Vertical dashed line indicates time of hindcast/forecast.

A set of 14 year hindcasts appear to skillfully predict the warming slowdown as early as 1990 (Figure 3). This slowdown is observed to result from a predicted apparent peak in the NMO near 2000 followed by a steep subsequent decline (Figures 3a and 3b). Hindcasts from 1995 and 2000 (Figures 3c and 3d and 3e and 3f, respectively) appear to predict the continued evolution of the slowdown and recovery through the present. Successful prediction of the slowdown at decadal lead times would represent a substantial advance relative to past efforts [*Newman*, 2013]. The analysis predicts (Figures 3g and 3h) an increase in the NMO (and associated acceleration of NH mean warming) over the next decade.

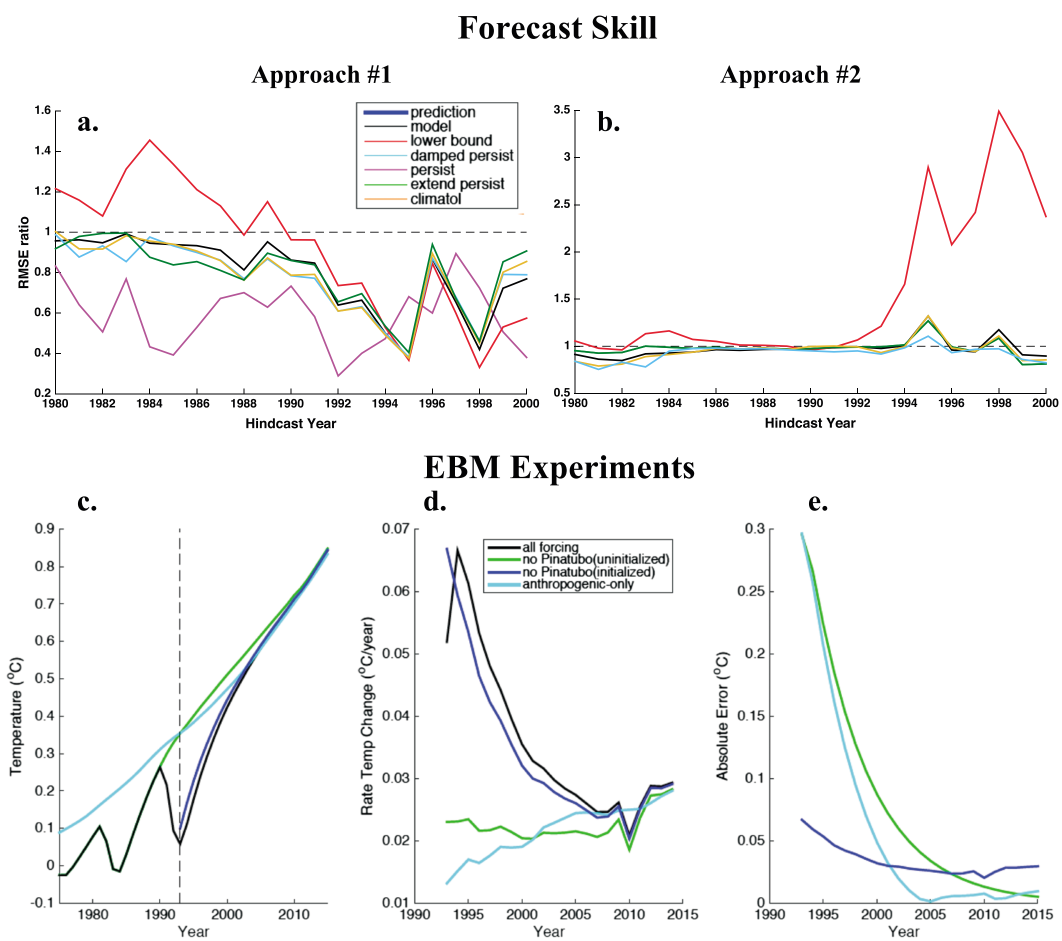

Figure 4. Impact of Mount Pinatubo response on assessment of NH mean forecasting skill. Shown are (a and b) ratio of RMSE of optimal NMO forecast to that for other forecasts as a function of hindcast year using the two different schemes for predicting forced component: projection of climatological total forced trend (Figure 4a) and projection of estimated anthropogenic-only trend (Figure 4b); dashed black line denotes a ratio of unity. Also shown (c) are EBM simulations of NH mean temperature based on full (anthropogenic + natural) radiative forcing (black line), full forcing minus the 1991 Mount Pinatubo eruption, i.e., no Pinatubo (green line), and prediction based on initializing the no Pinatubo simulation with the approximate 1993 temperature value (blue line). Timing of the 1993 prediction is marked by vertical dashed line. Shown also for comparison is an anthropogenic forcing-only simulation (cyan line). (d and e) The rate of temperature change and absolute error of each simulation relative to the full radiative forcing simulation.

Unfortunately, the apparent skill in these forecasts turns out to be an artifact of misprediction of the forced signal. This artifact is apparent in the stability of the RMSE of hindcasts, with increasing forecast length in the face of increasing RMSE in the forced model-only component of the hindcast (Figure 2c). The simplest interpretation is that we are doing a relatively poor job predicting the future forced trend.

Using instead the more sophisticated second scheme (i.e., anthropogenic-only trend persistence; Figure 2b) for predicting the forced component, we arrive at very different conclusions. We no longer observe any evidence of predictability (see Figures 2d and 2f) of the internal variability (i.e., of the NMO). The optimal forecast now yields less skill (i.e., greater error) than either the forced model only or several null forecasting schemes. The lower bound error estimates are now seen to be substantially lower than with the first scheme and lie below all other estimates. The only apparent predictable signal is indeed the forced signal itself.

The naïve use of a linear extension of the climatological trend as in the first scheme is subject to bias if the termination interval lies within the period of recovery from a large volcanic cooling event such as Pinatubo (e.g., the 1995 hindcast). Even 4 years after the eruption, the residual cooling downwardly leverages the recent (and future projected) trend. The large uptick predicted in the mid-1990s NMO hindcast (Figure 3)

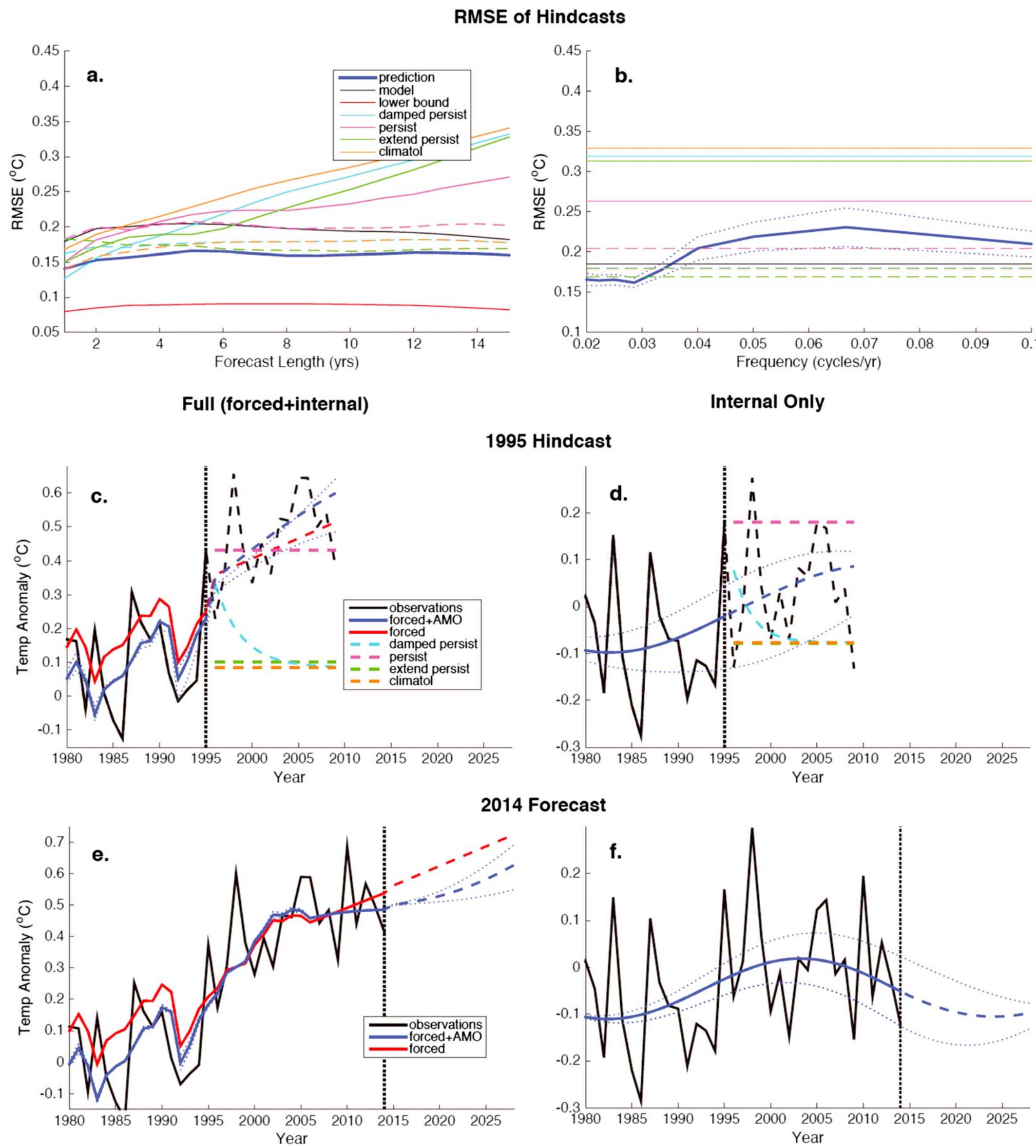

Figure 5. North Atlantic mean temperature predictions. Shown are RMSE as a function of (a) forecast length and (b) filtering frequency, along with 14 year hindcasts in 1995 for (c) North Atlantic SST and (d) AMO component, and a 14 year prediction in 2014 for (e) North Atlantic SST and (f) AMO component. Upper and lower one sigma uncertainty estimates (blue dotted line) are shown for hindcasts and forecasts.

simply reflects the ongoing recovery from the eruption. The more sophisticated second scheme skillfully projects the future trend even in the presence of this complicating signal. Examining the forecast skill as a function of the year of the hindcast (Figure 4a), the first scheme suggests a substantial increase in skill following the 1991 Pinatubo eruption but only as an artifact of an increasingly poor forecast of the forced signal. The second scheme (Figure 4b), by contrast, confirms a consistent absence of forecasting skill over time.

While it might be tempting to dismiss these findings as specific only to the statistical forecasting approaches investigated here, we suspect it has broader implications from the standpoint of decadal predictability. Predictions founded on model-based data assimilation approaches, for example, may similarly embed information about forced responses (e.g., the residual response to the Pinatubo eruption in the mid-1990s in the form of cold SSTs) in the initial conditions imposed. Perhaps, it is that information, rather than any information about the phase of internal decadal/multidecadal variability, that leads to the apparent predictability of the warming slowdown in some hindcast experiments using models initialized in the mid-1990s [*Meehl et al.*, 2014a].

We illustrate the issue using an Energy Balance Model (EBM) simulation of NH mean temperature driven with estimated historical natural and anthropogenic radiative forcing (see *Mann et al*. [2014] and supporting information for further details). Our reference simulation (i.e., the "truth" in the context of this example) employs the full (natural + anthropogenic) historical radiative forcing. An additional simulation entirely removes the radiative forcing due to the 1991 Mount Pinatubo eruption, while yet another simulation removes the Pinatubo forcing but *initializes* the EBM at the approximate (cold) observed temperature in 1993 (Figures 4c–4e). The uninitialized "no Pinatubo" simulation, unsurprisingly, fails to capture the Pinatubo cooling and recovery. The *initialized* no Pinatubo simulation, however, closely reproduces the feature and appears to predict the subsequent "slowdown" in warming through 2010 (which in reality arises through a relaxation back to the pre-Pinatubo state). In this example, the apparent skill of the initialized simulation relative to the uninitialized simulation, including the ability to "predict" the warming slowdown, arises purely as an artifact of the misspecification of the radiative forcing (in this case, the absence of the Pinatubo forcing), the effect of which is largely corrected by initialization. It does not indicate skillful prediction of internal variability (indeed, in the example provided, there is, by design, *no internal variability*). While this may be an extreme example, even a more modest misspecification of forcing (e.g., an error in the estimated amplitude of the applied Pinatubo forcing) will lead to spurious apparent improvements in skill through initialization. Additional work will be necessary to assess the impact that this phenomenon might have on estimates of forecasting skill in initialized model simulation-based forecasts.

Our findings are not entirely negative. While we find no predictability in the PMO (supporting information), we do find evidence of a predictable multidecadal AMO signal in our hindcast experiments (Figure 5) consistent with past work identifying specific physical mechanisms that support oscillatory AMO behavior with a well-defined 50–70 year time scale [*Delworth and Mann*, 2000; *Kerr*, 2000; *Knight et al*., 2005]. In the past (e.g., midtwentieth century), the AMO played a more dominant role in the behavior of the NMO (Figure 1) and was to play a similarly dominant role in the future, our results suggest the possibility of enhanced predictability in larger-scale temperature changes, with the caveat that the long time scale of this signal means that the assessed skill is based on just one "cycle" of the putative oscillation.

Some forecasts [*Keenlyside et al*., 2008] have indeed focused on the role of the North Atlantic in decadal predictability, forecasting large-scale cooling as a result of multidecadal fluctuations in the Atlantic Meridional Overturning Circulation. One recent study [*McCarthy et al*., 2015] predicted a decline in the AMO over the next several decades large enough to offset global warming. That study, however, employed a linear detrending procedure that has been heavily criticized for yielding a biased climate signal [*Mann and Emanuel*, 2006; *Trenberth and Shea*, 2006; *Mann et al*., 2014; *Steinman et al*., 2015; *Frankcombe et al*., 2015]. We see no evidence for such a conclusion: our predictions (Figure 5) indicate a very slight (~0.04°C) decrease in the AMO over the next decade, implying less than a 0.02°C negative impact on NH mean temperature, an amount which is dwarfed by the ongoing ~0.15°C/decade pace of anthropogenic warming. Beyond a decade from now, the AMO is predicted in our analysis to begin rising again.

Steinman et al. [2015] argued, based on data through 2012, that the recent NH warming slowdown was associated with a negative trend in the NMO, reflecting a combination of a relatively flat, modestly positive AMO and a sharply negative-trending PMO. Given the historical pattern, they speculated that the trend in the PMO (and hence NMO) would likely reverse with internal variability instead adding to anthropogenic warming in the coming decades. While data from 2013 to 2014 suggest that such a reversal may indeed already be underway, we find little evidence for a truly predictable Pacific-centered PMO oscillation. Our findings are consistent with the PMO largely or entirely reflecting climatic red noise. An Atlantic-centered AMO oscillation, by contrast, appears to exhibit true predictability, but its amplitude is currently modest, and it is projected to have little influence on large-scale temperatures over the next one-to-two decades.

Acknowledgments
All raw data, ©Matlab code, and results from our analysis are available at the supplementary website: http://www.meteo.psu.edu/~mann/supplements/GRL2016. 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. B.A.S. acknowledges support by the U.S. National Science Foundation (EAR-1447048). M.H.E. and L.M.F. acknowledge support from the Australian Research Council (FL100100214). A.H.C. acknowledges support from the U.S. National Science Foundation (AGS-1263225). Kaplan SST V2 data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/. HadISST data were provided by the Met Office Hadley Centre: www.metoffice.gov.uk/hadobs. ERSST data were provided by NOAA: www.ncdc.noaa.gov/data-access/marineocean-data/extended-reconstructed-sea-surface-temperature-ersst-v3b.

References

Cowtan, K., Z. Hausfather, E. Hawkins, P. Jacobs, M. E. Mann, S. K. Miller, B. A. Steinman, M. B. Stolpe, and R. G. Way (2015), Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures, *Geophys. Res. Lett.*, *42*, 6526–6534, doi:10.1002/2015GL064888.

Dai, A., J. C. Fyfe, S.-P. Xie, and X. Dai (2015), Decadal modulation of global surface temperature by internal climate variability, *Nat. Clim. Change*, *5*(6), 555–559, doi:10.1038/nclimate2605.

Delworth, T. L., and M. E. Mann (2000), Observed and simulated multidecadal variability in the Northern Hemisphere, *Clim. Dyn.*, *16*, 661–676, doi:10.1007/s003820000075.

Doblas-Reyes, F. J., I. Andreu-Burillo, Y. Chikamoto, J. García-Serrano, V. Guemas, M. Kimoto, T. Mochizuki, L. R. L. Rodrigues, and G. J. van Oldenborgh (2013), Initialized near-term regional climate change prediction, *Nat. Commun.*, *4*, 1715, doi:10.1038/ncomms2704.

England, M. H., S. McGregor, P. Spence, G. A. Meehl, A. Timmermann, W. Cai, A. Sen Gupta, M. J. McPhaden, A. Purich, and A. Santoso (2014), Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus, *Nat. Clim. Change*, *4*(3), 222–227, doi:10.1038/nclimate2106.

Foster, G., and S. Rahmstorf (2011), Global temperature evolution 1979–2010, *Environ. Res. Lett.*, *6*(4), 044022, doi:10.1088/1748-9326/6/4/044022.

Frankcombe, L. M., M. H. England, M. E. Mann, and B. A. Steinman (2015), Separating internal variability from the externally forced climate response, *J. Clim.*, *28*, 8184–8202.

García-Serrano, J., V. Guemas, and F. J. Doblas-Reyes (2015), Added-value from initialization in predictions of Atlantic multi-decadal variability, *Clim. Dyn.*, *44*, 2539–2555.

Guemas, V., F. J. Doblas-Reyes, I. Andreu-Burillo, and M. Asif (2013), Retrospective prediction of the global warming slowdown in the past decade, *Nat. Clim. Change*, *3*, 649–653.

Jolliffe, I. T., and D. B. Stephenson (2011), *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, 274 pp., Wiley, West Sussex, U. K.

Karl, T. R., A. Arguez, B. Huang, J. H. Lawrimore, J. R. McMahon, M. J. Menne, T. C. Peterson, R. S. Vose, and H.-M. Zhang (2015), Possible artifacts of data biases in the recent global surface warming hiatus, *Science*, *348*(6242), 1469–1472, doi:10.1126/science.aaa5632.

Keenlyside, N. S., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner (2008), Advancing decadal-scale climate prediction in the North Atlantic sector, *Nature*, *453*(7191), 84–8, doi:10.1038/nature06921.

Kerr, R. A. (2000), A North Atlantic climate pacemaker for the centuries, *Science*, *288*, 1984–1985, doi:10.1126/science.288.5473.1984.

Kirtman, B., et al (2013), Near-term climate change: Projections and predictability, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker et al., pp. 962–968, Cambridge Univ. Press, Cambridge, U. K., and New York.

Knight, J. R., R. J. Allan, C. K. Folland, M. Vellinga, and M. E. Mann (2005), A signature of persistent natural thermohaline circulation cycles in observed climate, *Geophys. Res. Lett.*, *32*, 1–4, doi:10.1029/2005GL024233.

Kosaka, Y., and S.-P. Xie (2013), Recent global-warming hiatus tied to equatorial Pacific surface cooling, *Nature*, *501*(7467), 403–407, doi:10.1038/nature12534.

Mann, M. E. (2008), Smoothing of climate time series revisited, *Geophys. Res. Lett.*, *35*, L16708, doi:10.1029/2008GL034716.

Mann, M. E., and J. Park (1994), Global-scale modes of surface temperature variability on interannual to century timescales, *J. Geophys. Res.*, *99*(D12), 25,819–25,834, doi:10.1029/94JD02396.

Mann, M. E., and J. Park (1999), Advances in geophysics, *Adv. Geophys*., *41*, 1–131.

Mann, M. E., and K. A. Emanuel (2006), Atlantic hurricane trends linked to climate change, *Eos Trans. AGU*, *87*, 233, doi:10.1029/2006EO240001.

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, doi:10.1002/2014GL059233.

Mantua, N. J., and S. R. Hare (2002), The Pacific Decadal Oscillation, *J. Oceanogr.*, *58*, 35–44, doi:10.1023/A:1015820616384.

McCarthy, G. D., I. D. Haigh, J. J.-M. Hirschi, J. P. Grist, and D. A. Smeed (2015), Ocean impact on decadal Atlantic climate variability revealed by sea-level observations, *Nature*, *521*(7553), 508–510, doi:10.1038/nature14491.

Meehl, G. A., H. Teng, and J. M. Arblaster (2014a), Climate model simulations of the observed early-2000s hiatus of global warming, *Nat. Clim. Change*, *4*(10), 898–902, doi:10.1038/nclimate2357.

Meehl, G. A., et al. (2014b), Decadal climate prediction: An update from the trenches, *Bull. Am. Meteorol. Soc.*, *95*(2), 243–267, doi:10.1175/BAMS-D-12-00241.1.

Newman, M. (2013), An empirical benchmark for decadal forecasts of global surface temperature anomalies, *J. Clim.*, *26*(14), 5260–5269, doi:10.1175/JCLI-D-12-00590.1.

Risbey, J. S., S. Lewandowsky, C. Langlais, D. P. Monselesan, T. J. O. Kane, and N. Oreskes (2014), Well-estimated global surface warming in climate projections selected for ENSO phase, *Nat. Clim. Change*, 1–7, doi:10.1038/NCLIMATE2310.

Santer, B. D., et al. (2014), Volcanic contribution to decadal changes in tropospheric temperature, *Nat. Geosci.*, *7*, 185–189.

Schmidt, G. a., D. T. Shindell, and K. Tsigaridis (2014), Reconciling warming trends, *Nat. Geosci.*, *7*(3), 158–160, doi:10.1038/ngeo2105.

Steinman, B. A., M. E. Mann, and S. K. Miller (2015), Atlantic and Pacific multidecadal oscillations and Northern Hemisphere temperatures, *Science*, *347*(6225), 988–91, doi:10.1126/science.1257856.

Trenberth, K. E., and D. J. Shea (2006), Atlantic hurricanes and natural variability in 2005, *Geophys. Res. Lett.*, *33*, L12704, doi:10.1029/2006GL026894.

Trenberth, K. E., and J. T. Fasullo (2013), An apparent hiatus in global warming? *Earth's Future*, *1*(1), 19–32, doi:10.1002/2013EF000165.

Weigel, A. P., and S. J. Mason (2011), The generalized discrimination score for ensemble forecasts, *Mon. Weather Rev.*, *139*, 3069–3074, doi:10.1175/MWR-D-10-05069.1.