Pronounced differences between observed and CMIP5 simulated multidecadal climate variability in the twentieth century

Sergey Kravtsov[1*] and David Callicutt[1]

Keywords: global warming, forced vs. internal climate variability, stadium wave

Index Terms: 0550, 1616, 1620, 1626, 3305

Key points:

- CMIP5 ensembles can be used to infer forced signal and internal variability
- Observed internal variability has a much larger variance vs. that in models
- The model–data differences are dominated by a hemispheric multidecadal mode

[1] Department of Mathematical Sciences, Atmospheric Science group, University of Wisconsin-Milwaukee, P. O. Box 413, Milwaukee, WI 53201

[*] Corresponding author email: kravtsov@uwm.edu

Abstract: Identification and dynamical attribution of multidecadal climate undulations to either variations in external forcings or to internal sources is one of the most important topics of modern climate science, especially in conjunction with the issue of human induced global warming. Here we utilize ensembles of 20th century climate simulations to isolate forced signal and residual internal variability in a network of observed and modeled climate indices. The observed internal variability so estimated exhibits a pronounced multidecadal mode with a distinctive spatiotemporal signature, which is altogether absent in model simulations. This single mode explains a major fraction of model–data differences over the entire climate-index network considered; it may reflect either biases in the models' forced response or models' lack of requisite internal dynamics, or a combination of both.

1. Introduction

Dynamical inferences about the causes of observed climate change must necessarily rely on analyses of global coupled climate models of the type used in the Coupled Model Intercomparison Project Phase 5 [CMIP5; *Taylor et al.*, 2012]. However, these models are not entirely based on first principles, and can potentially misrepresent true climate dynamics, which may further complicate the interpretation of model simulations. For example, *Zhang et al.* [2007] showed that multidecadal variations — on top of a linear warming trend — of the mean Northern Hemisphere surface atmospheric temperature in the twentieth century can equally well be rationalized as a radiatively forced signal in the fully coupled version of the GFDL2.1 model, or as a response to

multidecadal variability of the North Atlantic sea-surface temperature (SST) in the model simulation with constant radiative forcing but variable Q-fluxes designed to mimic the observed SST evolution in the North Atlantic mixed layer. The caveat of the latter result is that the fully coupled GFDL model is by itself unable to produce multidecadal variations in the basin-wide North Atlantic SSTs — known as the Atlantic Multidecadal Oscillation (AMO) — comparable in magnitude with that of the observed AMO variability. Hence, either this model lacks dynamics required to generate realistic levels of internal multidecadal SST variability in the North Atlantic or, alternatively, it underestimates multidecadal variations in the forced SST response there. This is to some extent true for many other CMIP5 models [see *Zhang and Wang*, 2013]. The perceived fundamental role of AMO dynamics in the climate change problem fueled a vivid (and ongoing) scientific debate around this issue (see *Murphy et al.* [2017] and references therein).

The differences between climate models and observations are, however, not confined to the North Atlantic. *Wyatt et al.* [2012], *Wyatt and Peters* [2012] and *Kravtsov et al.* [2014] detected considerable mismatches in both the magnitude and spatiotemporal structure of the twentieth century's multidecadal variability — defined via deviations from the long-term linear trend — in a network of observed and CMIP5 simulated indices characterizing the Northern Hemisphere climate. Assuming that multidecadal discrepancies between the models and observations reflect differences in the observed and simulated internal climate variability, *Mann et al.* [2014] and *Steinman et al.* [2015a,b] challenged these finding and interpreted them to be an artifact of the linear detrending. In this study, we utilize CMIP5 ensembles of historical 20th century

simulations to isolate (nonlinear) forced signals and internally generated (residual) climate variability in models and observations. We then expand the analyses of *Kravtsov et al.* [2014] to compare the resulting estimated internal components of the observed and CMIP5 simulated climate variability, and find pronounced differences in their magnitude, time scale and spatial patterns, thus reconfirming earlier conclusions of these authors.

2. Data and methods

2.1 Data sources

We utilized the output from CMIP5 historical twentieth-century simulations for models with four or more ensemble members (**Table 1**) to analyze the simulated sea-surface temperature (SST) based climatic indices representing Atlantic Multidecadal Oscillation [AMO: *Kerr*, 2000; *Enfield et al*., 2001], Pacific Multidecadal Oscillation [PMO: *Steinman et al.*, 2015a], and Northern Hemisphere mean surface air temperature [NMO: *Steinman et al.*, 2015a]. The simulated SST indices, as well as their observed counterparts, were the same as used by *Steinman et al.* [2015a] and were downloaded from that manuscript's supplementary website (www.meteo.psu.edu/holocene/public_html/supplements/Science2015). The AMO and PMO indices were based on the SST averaged over the regions (0ºN–60ºN, 80ºW–0º) and (0ºN–60ºN, 120ºE–100ºW), respectively. The NMO index was computed as the mean surface temperature (ocean+land) over the 0ºN–60ºN region. The observed AMO and PMO indices were computed as the average of three SST products: the Hadley Centre Global Sea Ice and Sea Surface Temperature [HadISST: *Rayner et al.*, 2003], National Oceanic and Atmospheric Administration (NOAA) Extended Reconstructed Sea Surface Temperature

(ERSST) [*Xue et al.*, 2003; *Smith et al*., 2008], and Kaplan SSTs [*Kaplan et al.*, 1998; *Parker et al.*, 1994; *Reynolds and Smith*, 1994]. The NMO index was based on Goddard Institute for Space Studies (GISS) Surface Temperature [GISTEMP: *Hansen et al*., 2010; *Cowtan and Way*, 2014]. All indices considered were annual-mean anomalies with respect to the 1880–2005 climatology.

We also analyzed two well-known atmospheric indices based on the observed and simulated sea-level pressure (SLP) data; we used NOAA's twentieth-century SLP reanalysis product [20CR: *Compo et al.*, 2011] to define the observed SLP indices. These indices were the North Atlantic Oscillation index (NAO: *Hurrell*, 1995; *Hurrell and Deser*, 2009), as well as an analog of the Aleutian Low Pressure index ALPI [*Beamish et al.*, 1997]. We computed the NAO and ALPI as the leading principal components of monthly wintertime (DJFM) SLP over the Atlantic (15ºN–75ºN, 90ºW–10ºW) and Pacific (15ºN–75ºN, 130ºE–120ºW) regions, respectively [see *Kravtsov et al.*, 2014 for the resulting loading patterns]. In addition, we used the station based NAO index (https://climatedataguide.ucar.edu) as an alternative NAO estimate. We normalized the observed and simulated monthly NAO and ALPI indices to have the unit standard deviation, and then formed and analyzed their DJFM-mean 1880–2005 time series.

2.2 Methodology

Following *Kravtsov and Callicutt* [2017], we first estimated forced signals for each of the 17 individual models considered as the 5-yr low-pass filtered ensemble mean over the set of a given model's 20th-century simulations (for each index), using the data adaptive filter of *Mann* [2008]. Subtracting the forced signals of a given model from this model's

individual simulations provides multiple realizations of this model's simulated internal variability. The resulting estimates of internal variability are, however, necessarily characterized by smaller variance relative to that of the true internal variability, since the forced signal used to define them in fact contains some of the true internal variability due to insufficient averaging over a small number of individual model runs [*Steinman et al.*, 2015b; *Frankcombe et al.*, 2015]. To alleviate this problem, we derived model-dependent and frequency-dependent inflation factors (**Fig. S1** of *Supporting Information*), which allowed us to obtain variance-bias-corrected estimates of the simulated internal variability for all of the 108 model simulations considered (see **section S1** of *Supporting Information*). These inflation factors were computed from synthetic stochastic versions of the individual model simulations designed to mimic the statistical characteristics of these simulations (see **section S2** of *Supporting Information*).

As a byproduct of the latter procedure, we also obtained 100 estimates of the synthetic forced signals for each of the 17 models available, totaling 1700 forced-signal estimates. Following *Steinman et al.* [2015a,b], *Frankcombe et al.* [2015], *Kravtsov et al.* [2015] and *Kravtsov and Callicutt* [2017], we then rescaled each of these forced-signal estimates via least squares to best match the full observed time series of the climate index considered; this rescaling is meant to correct for climate sensitivity biases of individual models. **Figure S2** of the *Supporting Information* displays the multi-model ensemble-mean forced signals so estimated and their uncertainty — defined as the spread among the 1700 available estimates of the forced signal, — along with the original (raw) observed indices. Subtracting these forced signals from the raw indices provides 1700 estimates of the internal variability in each of the five observed indices considered.

We then analyzed the resulting 108 (variance-bias-corrected) time series of the simulated internal variability, as well as 1700 estimates of the internal component of the observed internal variability for our network of five climate indices using the multi-channel version of the Singular Spectrum Analysis [SSA: *Broomhead and King*, 1986; *Elsner and Tsonis*, 1996] called M-SSA [*Moron et al.*, 1998; *Ghil et al.*, 2002]. M-SSA is an extended variant of a widely used Empirical Orthogonal Function (EOF) analysis technique [*Monahan et al.*, 2009], which looks for the space–time patterns that maximize lagged covariance for a given multivariate time series within a range of M lags; M is referred to as the embedding dimension of M-SSA. The original raw time series can be fully recovered as the sum, over all modes, of the so-called reconstructed components (RCs) associated with each M-SSA mode. Prior to M-SSA analysis, we smoothed the internal components of the NAO and ALPI indices using the 7-yr running-mean boxcar filter to focus on multidecadal time scales. In addition, each index of the five-index network — whether observed or simulated — was normalized to have the unit standard deviation. We performed M-SSA analysis using embedding dimensions from M=20 to 40; all of the results reported below are robust with respect to the choice of M.

3. Results

3.1 Variance of observed and simulated climate variability

Figure 1 shows standard deviations of the observed and simulated internal variability for raw and low-pass filtered data, as a function of the filter time scale. Naturally, the more smoothing is applied to a time series, the less is the resulting standard

deviation. Left panels display the standard deviations of the full internal variability for the five indices considered, with the mean values and the spread computed over the 108 samples of simulated internal variability and 1700 samples of the internal component of the observed internal variability (see section 2.2). Strikingly, the observed internal variability has a large variance at decadal and longer time scales unmatched by that in the models, for both the oceanic indices [compare with *Kravtsov and Callicutt*, 2017] and the atmospheric indices. Quantitatively, the standard deviations of all indices except for the NMO index have a range of decadal-to-interdecadal time scales at which the error bars based on the observed and simulated data do not overlap. These results are qualitatively consistent with earlier work of *Kravtsov et al.* [2014], the main difference between the two analyses being that the latter study compared the anomalies with respect to the corresponding linear trends, whereas the present study uses CMIP5 based forced signals to define 'internal' anomalies.

Right panels of the same figure show the standard deviations of the observed and simulated data from which we subtracted the RCs associated with their respective leading M-SSA pairs (see section 3.2 below). This time, while the simulated internal variability still lacks some amplitude compared to the observed variability, the difference between the two is much reduced, thereby indicating that the dominant M-SSA pairs account for a major fraction of this difference. The present results were obtained with the M-SSA embedding dimension of M=20; the results with M=40 (and the values of M in between) are analogous (**Fig. S3** of *Supporting Information*).

3.2 Dominant multidecadal variability

Consistent with the results of section 3.1, M-SSA spectra of the observed and simulated climate networks display pronounced differences dominated by the leading M-SSA pair (**Fig. 2**, upper panels). The spectra shown here are the 'space–time' variances associated with the EOFs of the data trajectory matrix obtained by augmenting the original five-index time series by its *M* lagged copies; these 'space–time' (ST-) EOFs represent the M-SSA modes. In the observed spectra, the leading two M-SSA modes stand out with respect to the background spectrum of the remaining (trailing) M-SSA modes; by contrast, the simulated spectra are much more flat, and their dominant M-SSA modes account for a relatively smaller fraction of the total variance. Unlike the variances of the leading two modes, the variances of the trailing M-SSA modes based on the observed and simulated data are consistent, as reflected by their overlapping error bars.

Not only the leading M-SSA pair of the observed internal variability has a much larger variance relative to that of the simulated variability, but it is also characterized by a completely different spatiotemporal structure absent from model simulations. Indeed, the variances associated with the projection of the trajectory matrices of simulated data onto the observed leading ST-EOFs (green error bars in the upper panels of Fig. 2) are small relative to the corresponding variances of both observed and simulated M-SSA modes (blue and red error bars, respectively).

The dominant spatiotemporal pattern of the observed internal variability is best visualized by plotting the sum of two leading observed RCs (Fig. 2, bottom panels). This pattern represents a coherent multidecadal signal propagating across the entire climate

index network; *Wyatt et al.* (2012) termed this signal the "stadium wave." The propagation characteristics of the stadium wave — that is, the time lags between the multidecadal signals in individual indices, — are clearly sensitive to subtraction of the background trend, as the present stadium wave has a different structure compared to that found by *Wyatt et al.* [2012] and *Kravtsov et al.* [2014]. (Once again, these authors used anomalies with respect to long-term linear trend in their analysis, in contrast to the anomalies with respect to the CMIP5 based forced signals here.) However, none of the CMIP5 models are able to mimic neither the observed spatial coherence in the internal variability within the climate network considered, nor the dominant — multidecadal — time scale of this variability (see **Figs. S4–S7** of *Supporting Information*).

4. Discussion

Our study documents pronounced differences between the observed and CMIP5 simulated climate variability in the twentieth century. These differences are dominated by a coherent multidecadal hemispheric-scale signal present in the observed SST and SLP fields, but completely missing in the CMIP5 simulations. Our results are consistent with an earlier study by *Kravtsov et al.* [2014], but generalize it to the entirety of the CMIP5 ensemble. Furthermore, we here focused on the (estimated) internal component of the observed and simulated climate variability, defined in terms of deviations from the CMIP5 based forced-signal estimates, rather than on the deviations from linear trend as in *Kravtsov et al.* [2014].

Despite our explicit decomposition of the climate variability into the forced and internally generated components, dynamical attribution of the multidecadal model–data differences still remains uncertain. On one hand, if our derived CMIP5 based forced signals are realistic, these differences must arise from internal climate-system dynamics presumably misrepresented in CMIP5 models, such as sea-ice dynamics [*Wyatt and Curry*, 2014], oceanic mesoscale eddies [*Siqueira and Kirtman*, 2016], positive cloud and dust feedbacks [*Evan et al.*, 2013; *Martin et al.*, 2014; *Brown et al.*, 2016; *Yuan et al.*, 2016], or SST forced NAO response [*Kushnir et al.*, 2002; *Eade et al.*, 2014; *Stockdale et al.*, 2015; *Siegert et al.*, 2016]. On the other hand, however, it is possible that CMIP5 models underestimate multidecadal variations in the true response of the climate system to external forcing or misrepresent the forcing itself [*Booth et al.*, 2012; *Murphy et al.*, 2017]; if this is true, the model–data differences reflect the mismatch between the actual and CMIP5 simulated forced signals, whereas the real-world's internal climate variability may be consistent with that simulated by the models. In either case, we strongly believe that model development activities should strive to alleviate the present large discrepancies between the observed and simulated multidecadal climate variability, as these discrepancies hinder our fundamental understanding of the observed climate change.

Acknowledgements. We acknowledge the World Climate Research Programme's Working Group on Coupled Modeling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. **Section S6** of the *Supporting Information* PDF document contains the link to all data and MATLAB © scripts for this paper. This research was supported by the NSF grants OCE-1243158 and AGS-1408897.

References

Beamish R. J., et al. (1997), Production of Fraser River sockeye salmon (Oncorhynchus nerka) in relation to decadal-scale changes in the climate and the ocean, Can. J. Fish. Aquat. Sci., 54, 543–554, 10.1139/f96-310.

Booth, B. B. B., et al. (2012), Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability, Nature, 484, 228–232, doi:10.1038/nature10946.

Brohan, P., et al. (2006), Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850, J. Geophys. Res., 111, D12106, doi:10.1029/2005JD006548.

Broomhead D. S., and G. P. King (1986), Extracting qualitative dynamics from experimental data, Phys. D, 20, 217–236, http://dx.doi.org/10.1016/0167-2789(86)90031-X.

Brown, P.T., et al. (2016), The necessity of cloud feedback for a basin-scale Atlantic Multidecadal Oscillation, Geophys. Res. Lett., 43, 3955–3963.

Compo, G. P., et al. (2011), The twentieth century reanalysis project, Quart. J. Royal Meteor. Soc., 654, 1–28, doi: 10.1002/qj.776.

Cowtan, K., and R. G. Way (2014), Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends, Quart. J. Royal Meteor. Soc., 140, 1935–1944.

Eade, R., et al. (2014), Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? Geophys. Res. Lett., 41, 5620–5628.

1 Elsner, J. B., and A. A. Tsonis (1996), Singular spectrum analysis: A new tool in time
2 series analysis, Springer, New York.
3 Enfield, D. B., et al. (2001), The Atlantic multidecadal oscillation and its relation to
4 rainfall and river flows in the continental U. S., Geophys. Res. Lett., 28, 277–280.
5 Evan, A. T., et al. (2013), The modification of sea surface temperature anomaly linear
6 damping time scales by stratocumulus clouds, J. Climate, 26, 3619–3630.
7 Frankcombe, L. M., et al. (2015), Separating internal variability from the externally
8 forced climate response, J. Climate, 28, 8184–8202, doi:
9 http://dx.doi.org/10.1175/JCLI-D-15-0069.1.
10 Ghil M., et al. (2002), Advanced spectral methods for climatic time series, Rev. Geophys.,
11 40(1), 3.1–3.41, doi:10.1029/2000GR000092.
12 Hansen, J., et al. (2010), Global surface temperature change, Rev. Geophys., 48, RG4004,
13 doi:10.1029/2010RG000345.
14 Hurrell, J. W. (1995), Decadal trends in the North Atlantic Oscillation: Regional
15 temperatures and precipitation, Science, 269, 676–679, doi:
16 10.1126/science.269.5224.676.
17 Hurrell, J. W., and C. Deser (2009), North Atlantic climate variability: The role of the
18 North Atlantic Oscillation, J. Marine Sys., 78, 28–41.
19 Kaplan A., et al. (1998), Analyses of global sea surface temperature 1856–1991, J.
20 Geophys. Res., 103, 18,567–18,589, doi:10.1029/97JC01736.
21 Kerr, R. A. (2000), A North Atlantic climate pacemaker for the centuries, Science, 288,
22 1984–1985, doi: 10.1126/science.288.5473.1984.

Kravtsov, S., et al. (2014), Two contrasting views of multidecadal climate variability in the twentieth century, Geophys. Res. Lett., 41, 1326, 6881–6888, doi: 10.1002/2014GL061416.

Kravtsov, S., et al. (2015), Comment on "Atlantic and Pacific multidecadal oscillations and Northern Hemisphere temperatures," Science, 350, 1326, doi: 10.1126/science.aab3570.

Kravtsov, S., and D. Callicutt (2017), On semi-empirical decomposition of multidecadal climate variability into forced and internally generated components, Inter. J. Climatol., accepted. Available from https://people.uwm.edu/kravtsov/publications/.

Kushnir Y., et al. (2002), Atmospheric GCM response to extratropical SST anomalies: Synthesis and evaluation, J. Climate, 15, 2233–2256, doi: http://dx.doi.org/10.1175/1520-0442(2002)015<2233:AGRTES>2.0.CO;2.

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

Mann, M. E., et al. (2014), On forced temperature changes, internal variability, and the AMO, Geophys. Res. Lett., 41, 3211–3219, doi:10.1002/2014GL059233.

Martin, E. R., et al. (2014), The multidecadal Atlantic SST—Sahel rainfall teleconnection in CMIP5 simulations, J. Climate, 27, 784–806.

Monahan, A. H., et al. (2009), Empirical Orthogonal Functions: The medium is the message, J. Climate, 22, 6501–6514, doi: http://dx.doi.org/10.1175/2009JCLI3062.1.

Moron, V., et al. (1998), Trends, interdecadal and interannual oscillations in global sea-surface temperatures, Climate Dyn., 14, 545–569, doi: 10.1007/s003820050241.

Murphy, L. N., et al. (2017), The role of historical forcings in simulating the observed Atlantic Multidecadal Oscillation. Geophys. Res. Lett.

Parker, D. E., et al. (1994), Interdecadal changes of surface temperature since the late nineteenth century, J. Geophys. Res., 99, 14,373–14,399.

Rayner, N.A., et al. (2003), Globally complete analyses of sea surface temperature, sea ice and night marine air temperature, 1871–2000, J. Geophys. Res., 108, 4407, doi:10.1029/2002JD002670.

Reynolds, R. W., and T. M. Smith (1994), Improved global sea-surface temperature analysis using optimum interpolation, J. Climate, 7, 929–948.

Siegert, S., et al. (2016), A Bayesian Framework for Verification and Recalibration of Ensemble Forecasts: How Uncertain is NAO Predictability? J. Climate, 29, 995–1012, doi: http://dx.doi.org/10.1175/JCLI-D-15-0196.1

Siqueira, L., and B. P. Kirtman (2016), Atlantic near-term climate variability and the role of a resolved Gulf Stream, Geophys. Res. Lett., 43, 3964–3972, doi:10.1002/2016GL068694.

Smith, T. M., et al. (2008), Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006), J. Climate, 21, 2283–2296, doi: http://dx.doi.org/10.1175/2007JCLI2100.1.

Steinman, B. A., et al. (2015a), Atlantic and Pacific multidecadal oscillations and Northern Hemisphere temperatures, Science, 347, 988, doi:

10.1126/science.1257856.

Steinman, B. A., et al. (2015b), Response to comment on "Atlantic and Pacific multidecadal oscillations and Northern Hemisphere temperatures," Science, 350, 1326, doi: 10.1126/science.aac5208.

Stockdale, T. N., et al. (2015) Atmospheric initial conditions and the predictability of the Arcitc Oscillation, Geophys. Res. Lett., 42, 1173–1179.

Taylor, K. E., et al. (2012) An Overview of CMIP5 and the experiment design, Bull. Am. Meteorol. Soc., 93, 485–498.

Wyatt, M. G., et al. (2012), Atlantic Multidecadal Oscillation and Northern Hemisphere's climate variability, Climate Dyn., 38(5–6), 929–949, doi:10.1007/s00382-011-1071-8.

Wyatt, M. G., and J. M. Peters (2012), A secularly varying hemispheric climate-signal propagation previously detected in instrumental and proxy data not detected in CMIP3 data base, SpringerPlus, 1–68, doi:10.1186/2193-1801-1-68.

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, doi: 10.1007/s00382-013-1950-2.

Xue, Y., et al. (2003), Interdecadal changes of 30-yr SST normals during 1871–2000, J. Climate, 16, 1601–1612.

Yuan, T., et al. (2016), Positive low cloud and dust feedbacks amplify tropical North Atlantic Multidecadal Oscillation, Geophys. Res. Lett., 43, 1349–1356, doi:10.1002/2016GL067679.

Zhang, R., et al. (2007), Can the Atlantic Ocean drive the observed multidecadal

variability in Northern Hemisphere mean temperature? Geophys. Res. Letts., 34, L02709, doi:10.1029/2006GL028683.

Zhang, L., and C. Wang (2013), Multidecadal North Atlantic sea surface temperature and Atlantic Meridional Overturning Circulation variability in CMIP5 historical simulations, J. Geophys. Res. Oceans, 118 (10), 5772–5791, doi:10.1002/jgrc.20390.

Table captions

Table 1: CMIP5 twentieth century simulations used in this study. We selected the models with four or more realizations, and analyzed the runs for which sea-surface temperature (SST), surface-air temperature (SAT) and sea-level pressure (SLP) outputs were all available.

Figure captions

Figure 1: Standard deviations (STDs) of the estimated observed (blue) and CMIP5 simulated (red) internal variability for the five indices considered; top-to-bottom rows correspond to the results for the AMO, PMO, NMO, NAO and ALPI indices, respectively. The STDs were computed for raw and boxcar running-mean low-pass filtered time series using different window sizes of $2\times K+1$ yr, $K = 0, 1, \ldots, 30$ (shown on the horizontal axis); $K = 0$ corresponds to raw annual data, $K = 1$ — to 3-yr low-pass filtered data and so on. Error bars show the 70% spread of the STDs, between 15th and 85th percentiles of the available estimates of internal variability (see text for details). The NAO plot also includes the results (black curve) based on an alternative, station based observed NAO index (https://climatedataguide.ucar.edu /climate-data/ hurrell-north- atlantic-oscillation-nao-index-station-based). Left column: the results based on the full annual data; right column: the results based on the anomalies with respect to the leading M-SSA pair of the corresponding observed or simulated realization of internal variability (see text for details); the M-SSA embedding dimension M=20. Comments: (i) *The simulated multidecadal variability is much weaker than observed* (left column). (ii) *Much of this model–data difference is rationalized by the leading M-SSA pair* (right column).

Figure 2: The M-SSA analysis of the observed and simulated internal variability in the AMO–PMO–NMO–NAO–ALPI climate-index network. Left column: the results of M-SSA analysis using the embedding dimension of M=20; right column: the same for M=40. Upper row: M-SSA spectra. The spectra based on the observed

data — ensemble-mean variances and their ±1 STD spread — are in blue, the spectra based on CMIP5 model simulations are in red. The green error-bar plot represents the variances of the time series obtained by projecting the trajectory matrices from model simulations onto the ST-EOFs of the *observed* trajectory matrix. Bottom row: Reconstructed components (RCs) corresponding to the leading M-SSA pair of the observed data. The NAO and ALPI RCs were multiplied by –1. For the purposes of better visualization, the RCs corresponding to different indices (channels) were vertically stacked by adding the constant offsets of +1, +2, +3, and +4 to the RCs for AMO, –NAO, PMO and NMO indices, respectively. Comments: (i) *Model–data differences are dominated by the leading M-SSA pair, whose observed spatiotemporal structure is absent from models* (upper row). (ii) *This mode, in observations, is characterized by a pronounced coherent multidecadal signal across the entire climate network considered* (bottom row).

Table 1: CMIP5 twentieth century simulations used in this study. We selected the models with four or more realizations, and analyzed the runs for which sea-surface temperature (SST), surface-air temperature (SAT) and sea-level pressure (SLP) outputs were all available.

Model #	Model acronym	Ensemble size
1	CanESM2	5
2	CCSM4	6
3	CNRM-CM5	10
4	CSIRO-MK3-6-0	10
5	GFDL-CM2.1	10
6	GFDL-CM3	4
7	GISS-E2-Hp1	6
8	GISS-E2-Hp1	5
9	GISS-E2-Hp3	6
10	GISS-E2-Rp1	6
11	GISS-E2-Rp2	6
12	GISS-E2-Rp3	6
13	HadCM3	10
14	HadGEM2-ES	5
15	IPSL-CM5A-LR	6
16	MIROC5	4
17	MRI-CGSM3	3
Total:	17 models	108 simulations

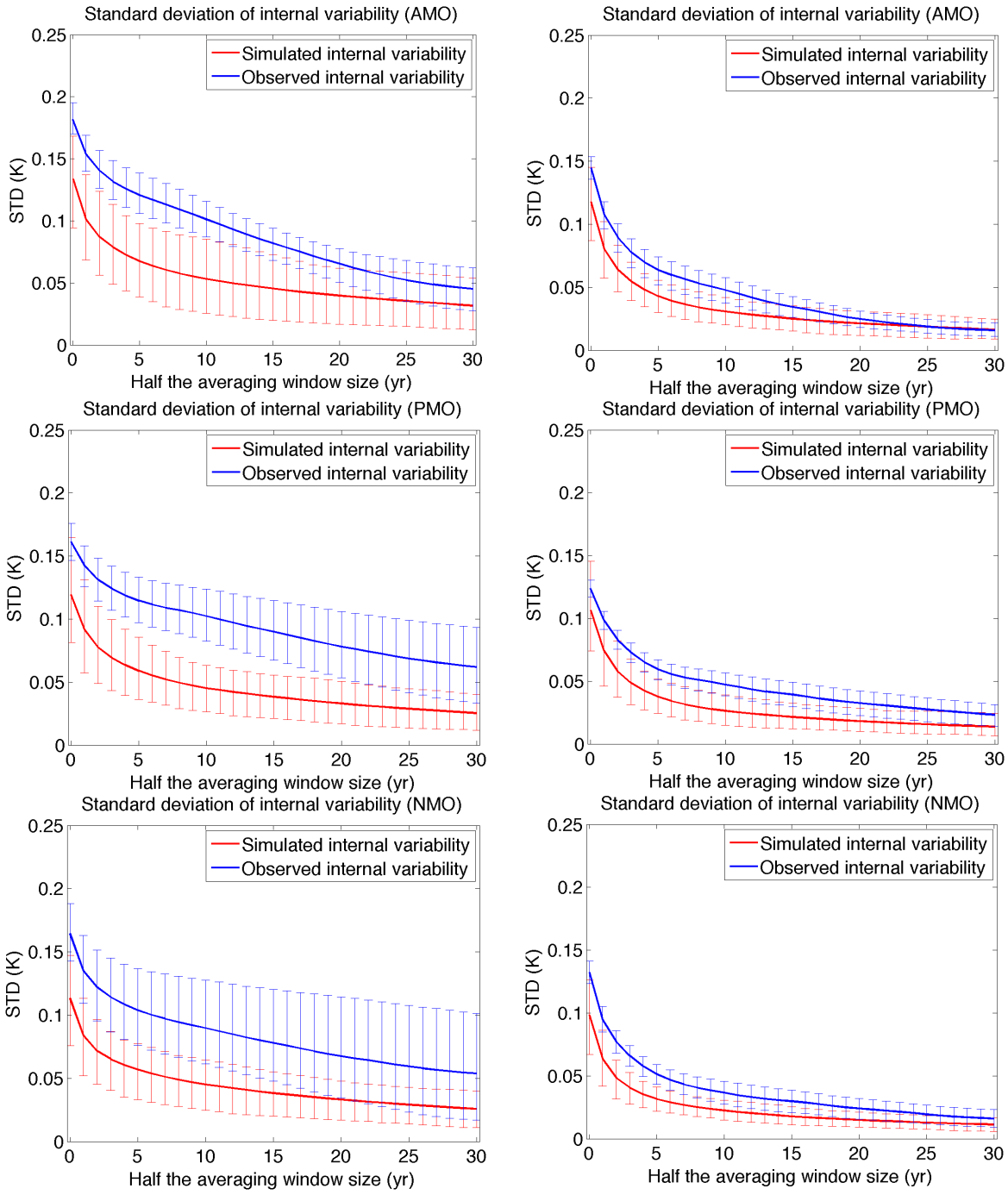

Figure 1: (see the remaining panels and figure caption on the next page)

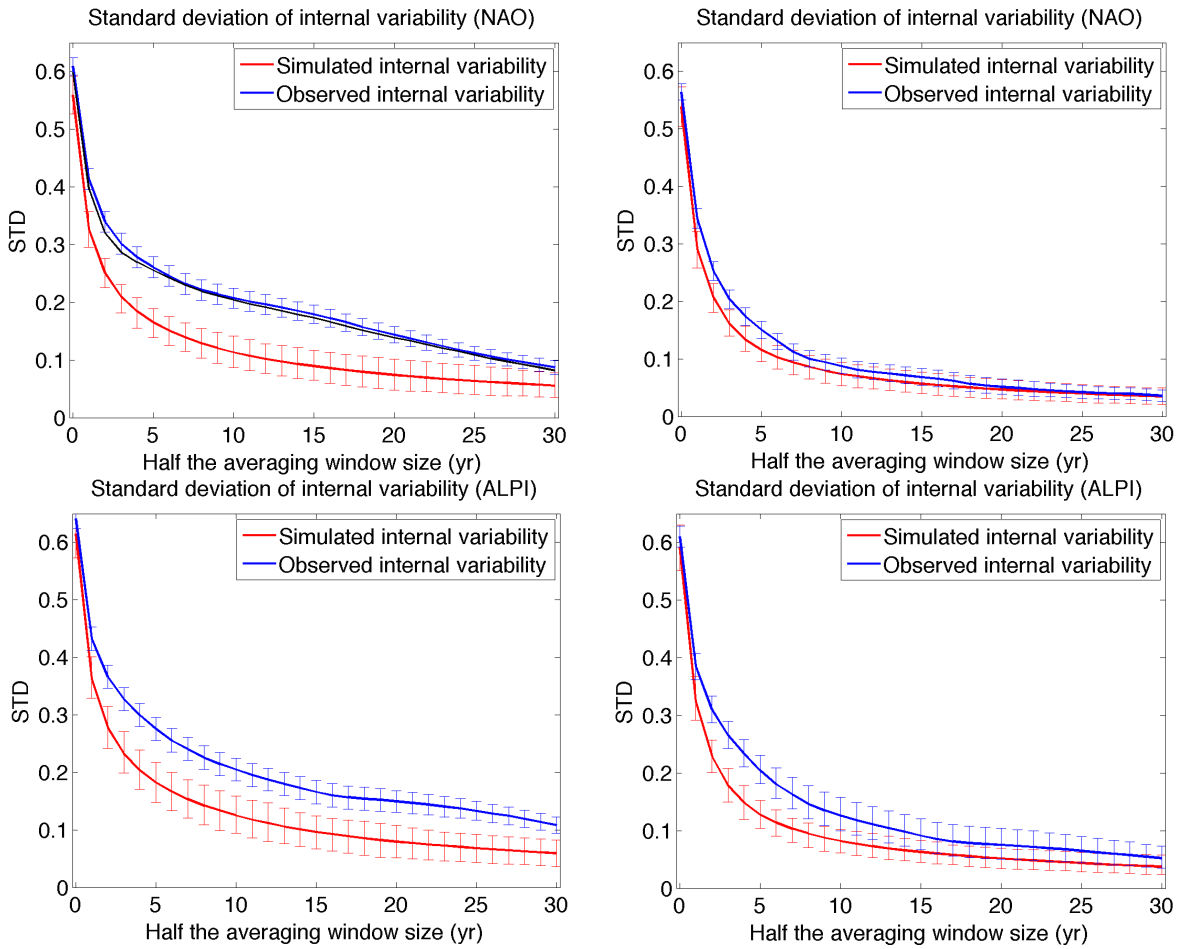

Figure 1 (cont'd): Standard deviations (STDs) of the estimated observed (blue) and CMIP5 simulated (red) internal variability for the five indices considered; top-to-bottom rows correspond to the results for the AMO, PMO, NMO, NAO and ALPI indices, respectively. The STDs were computed for raw and boxcar running-mean low-pass filtered time series using different window sizes of $2\times K+1$ yr, $K = 0, 1, \ldots, 30$ (shown on the horizontal axis); $K = 0$ corresponds to raw annual data, $K = 1$ — to 3-yr low-pass filtered data and so on. Error bars show the 70% spread of the STDs, between 15th and 85th percentiles of the available estimates of internal variability (see text for details). The NAO plot also includes the results (black curve) based on an alternative, station based observed NAO index (https://climatedataguide.ucar.edu /climate-data/ hurrell-north- atlantic-oscillation-nao-index-station-based). Left column: the results based on the full annual data; right column: the results based on the anomalies with respect to the leading M-SSA pair of the corresponding observed or simulated realization of internal variability (see text for details); the M-SSA embedding dimension M=20. Comments: (i) *The simulated multidecadal variability is much weaker than observed* (left column). (ii) *Much of this model–data difference is rationalized by the leading M-SSA pair* (right column).

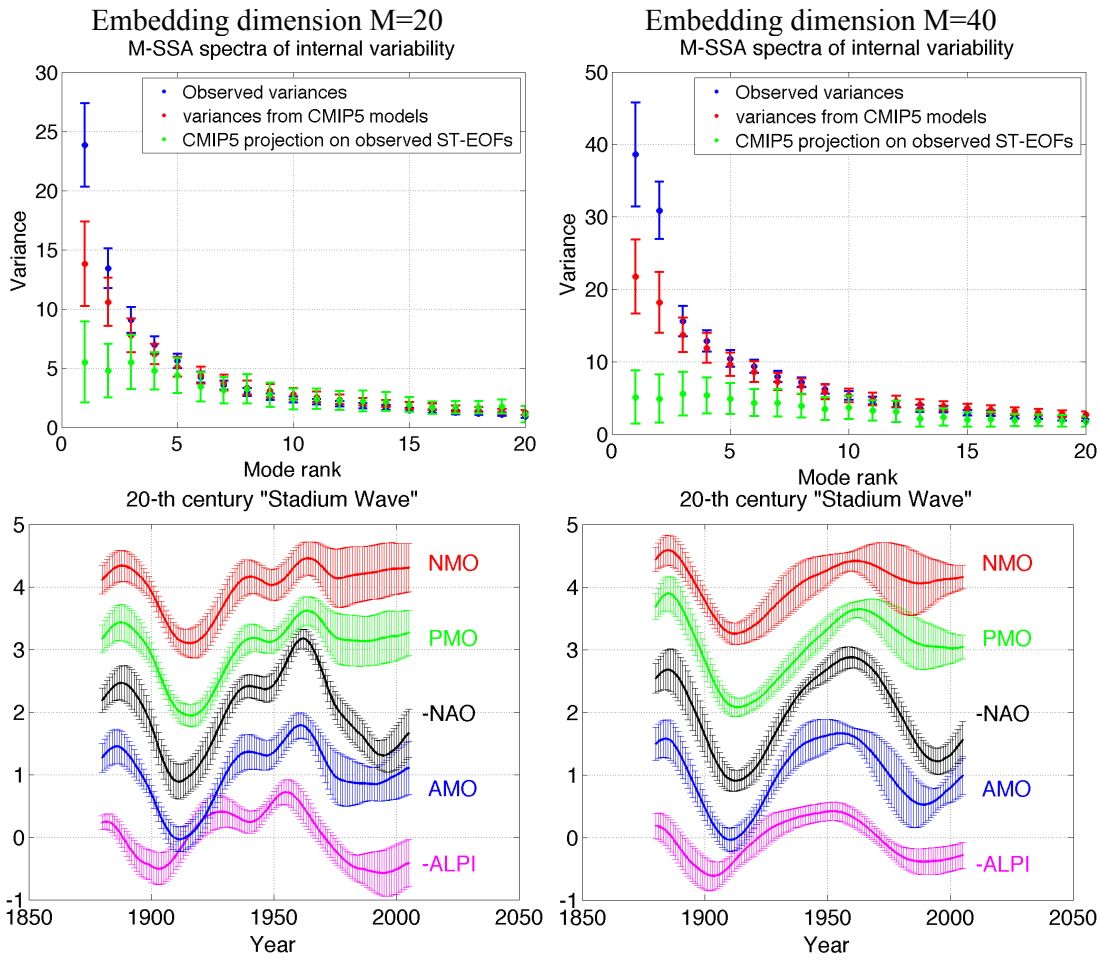

Figure 2: The M-SSA analysis of the observed and simulated internal variability in the AMO–PMO–NMO–NAO–ALPI climate-index network. Left column: the results of M-SSA analysis using the embedding dimension of M=20; right column: the same for M=40. Upper row: M-SSA spectra. The spectra based on the observed data — ensemble-mean variances and their ±1 STD spread — are in blue, the spectra based on CMIP5 model simulations are in red. The green error-bar plot represents the variances of the time series obtained by projecting the trajectory matrices from model simulations onto the ST-EOFs of the *observed* trajectory matrix. Bottom row: Reconstructed components (RCs) corresponding to the leading M-SSA pair of the observed data. The NAO and ALPI RCs were multiplied by –1. For the purposes of better visualization, the RCs corresponding to different indices (channels) were vertically stacked by adding the constant offsets of +1, +2, +3, and +4 to the RCs for AMO, –NAO, PMO and NMO indices, respectively. Comments: (i) *Model–data differences are dominated by the leading M-SSA pair, whose observed spatiotemporal structure is absent from models* (upper row). (ii) *This mode, in observations, is characterized by a pronounced coherent multidecadal signal across the entire climate network considered* (bottom row).