1	On semi-empirical decomposition of multidecadal climate variability into forced
2	and intrinsic components
3	
4	by
5	
6	Sergey Kravtsov ¹ and David Callicutt
7	
8	University of Wisconsin-Milwaukee
9	
10	
11	22 February 2016
12	
13	Submitted to the Journal of Climate
14	
15	
16	
17	
18	
19	
20	
21	

¹*Corresponding author address*: Dept. of Mathematical Sciences, Atmospheric Science Group, University of Wisconsin-Milwaukee, P. O. Box 413, Milwaukee, WI 53201. *E-mail*: <u>kravtsov@uwm.edu</u>.

1 Abstract

2 This paper combines CMIP5 historical simulations and observations of sea-surface 3 temperature (SST) and sea-level pressure (SLP) to investigate relative contributions of 4 forced and intrinsic climate variability to long-term (decadal+) climate trends. 5 Climate model simulations match the non-uniform warming of Northern Hemisphere 6 mean surface temperature very well, but are overly sensitive to forcing in the North 7 Atlantic and North Pacific regions, where the models' historical simulations have to 8 be scaled back to match the observed trends. On the other hand, the simulated 9 intrinsic variability in SST and SLP is strongly damped and much weaker than 10 observed, with the exception of variability associated with the Pacific Decadal 11 Oscillation. There are also substantial differences in the spatiotemporal structure of 12 the observed and simulated intrinsic variability. These discrepancies suggest that a 13 contribution of multidecadal intrinsic climate variability to the observed climate 14 change is distorted in the CMIP5 simulations; hence, our ability to attribute and 15 predict climate change using the current generation of climate models is limited.

1. Introduction

2 Analysis of scientific literature suggests that most researchers agree that a 3 sizable fraction of the observed twentieth-century climate warming may be due to 4 human activity (Cook et al. 2013). It is less clear though whether there is such a 5 consensus on the issue of exactly how much of the observed warming over the last 6 few decades has been anthropogenic (see, for example, Ghil and Vautard 1991; 7 DelSole et al. 2011; Wu et al. 2011, among others). The IPCC (2013) fifth 8 assessment report states that: "It is extremely likely that human influence has been the 9 dominant cause of the observed warming since the mid-20th century." Mann et al. 10 (2016) provided quantitative estimates of this (high) likelihood using a semi-empirical 11 approach involving model simulations and observations of surface temperature 12 (Steinman et al. 2015a; Frankcombe et al. 2015). Yet, such quantitative statements 13 necessarily rely on how skillful the state-of-the-art climate-system models are in 14 simulating the observed climatic variability; of particular importance is to assess the 15 models' potential to simulate intrinsic low-frequency (multidecadal) climate 16 variability that may arise in the climate system in the absence of changes in the 17 external (anthropogenic and natural) forcing (Jolliffe and Stephenson 2003; DelSole 18 and Shukla 2010). In this paper, we adopt a modified version of the Steinman et al.'s 19 semi-empirical approach to estimate and compare the intrinsic component of the 20 observed multidecadal climate variability with the one simulated by the CMIP5 21 climate models (Taylor et al. 2015).

Separating the forced climate signal from intrinsic climate variability generally relies on the assumption that these two types of variations possess their own distinct spatiotemporal signatures. Various statistical methods can then be applied for signal detection and attribution in the observed or model simulated climatic time

1 series: the standard empirical orthogonal function analysis (EOF: Preisendorfer 1988; 2 Monahan et al. 2009), singular spectrum analysis (SSA; Ghil and Vautard 1991; 3 Elsner and Tsonis 2006) and its multivariate extension M-SSA (Moron et al. 1998; 4 Ghil et al. 2002; Jamison and Kravtsov 2010; Wyatt et al. 2012; Kravtsov et al. 2014; 5 Groth and Ghil 2015; Groth et al. 2016), multi-taper spectral domain approach (Mann 6 and Park 1994, 1999), empirical mode decomposition (Huang and Wu 2008; Wu et al. 7 2011); discriminant analysis (Schneider and Held 2001; DelSole and Tippett 2007); 8 optimal persistence analysis (DelSole 2001, 2006), and others. Comparison of the 9 observed and simulated space/time patterns detected by these methods serves to 10 assess the models' performance in simulating the observed climate signals and 11 provides clues about dynamical sources of the observed climate variability.

12 In contrast to purely data based signal processing techniques described above, 13 the other class of detection and attribution methods makes a more extensive and immediate use of different climate-model simulations. For example, DelSole et al. 14 15 (2011) derived the "internal multidecadal pattern" (IMP) of climate variability from 16 CMIP3 control runs by maximizing its average predictability time (APT: DelSole and 17 Tippett 2009). They also estimated the forced signal's pattern as the leading 18 discriminant that maximizes the ratio of variance in the forced simulations to that in 19 the control runs of CMIP3 models (see also Ting et al. 2009). The two patterns 20 derived from models were finally combined in a fingerprinting procedure 21 (Hasselmann 1997; Hegerl et al. 1997; Allen and Tett 1999; Tett et al. 1999) to 22 determine the forced and intrinsic components of the observed climate variability, 23 with the IMP component shown to contribute substantially to decadal climate trends 24 and to exhibit decadal predictability (DelSole et al. 2013). Note that applying model-25 derived patterns to the observed data assumes, once again, that models adequately

represent dynamics governing the observed climate variability, which, while certainly hoped for, is not guaranteed. In a sense, the consistency between the models and observations diagnosed in such detection and attribution studies might be a consequence of the procedural design, whereas possible model/data discrepancies may be effectively masked.

6 A more intuitive and easily interpretable way of addressing a mixture of the 7 forced signals and intrinsic climate variability in observations and model simulations is to exploit model ensembles and focus on the observed and simulated large-scale 8 9 low-frequency patterns to achieve requisite reduction of effective degrees of freedom. 10 In particular, to get a naturally unbiased estimate of the forced signal (and the residual 11 intrinsic variability) in any given model, one is to run multiple simulations of this 12 model under the identical history of external forcings and for an ensemble of 13 perturbed initial conditions. These simulations would share the same forced signal, 14 but would have statistically independent, uncorrelated realizations of the intrinsic 15 variability. Single-model ensemble-mean time series (hereafter, SMEM) would thus 16 be dominated by the forced signal, since different realizations of the intrinsic variability would tend to cancel in taking the SMEM. The standard uncertainty of the 17 forced signal estimation in this case is σ/\sqrt{M} , where M is the number of simulations 18 19 in the model ensemble and σ is the standard deviation of the model's intrinsic 20 variability. Since M is typically not that large, one can make use of the fact that the 21 forcing time series and ensuing forced signals in the models are dominated by the 22 response to slowly varying forcing and further reduce the uncertainty of estimated 23 forced signal via smoothing of the individual models' SMEMs, which would thus 24 minimize the fingerprint of the interannual intrinsic variability in the forced signal 25 estimates.

1 Steinman et al. (2015a) and Frankcombe et al. (2015), following Kravtsov et 2 al. (2008), Knight (2009) and Terray (2012), among others, used the CMIP5 multi-3 model ensemble mean (hereafter, MMEM) to estimate the simulated forced signal. In 4 general, this is a very good idea, since the total number of simulations in the multi-5 model ensemble $(M \sim 100)$ is much larger than that in individual-model ensembles 6 $(M\sim5)$, which reduces the MMEM-based forced signal uncertainty substantially as 7 compared to SMEM. The caveat here is that in taking the MMEM, one also averages 8 out, along with intrinsic variability, the uncertainty associated with different forcing 9 subsets and different physical parameterizations used in the models (hereafter, model 10 uncertainty). Frankcombe et al. (2015) derived estimates of this uncertainty in 11 synthetic data sets designed to mimic CMIP5 twentieth century runs, but stopped 12 short of combining these uncertainty estimates with the actual estimated forced 13 signals in CMIP5 simulations and observations. Meanwhile, Kravtsov et al. (2015) 14 showed that the model uncertainty dominates the inferred 'intrinsic variability' in the 15 CMIP5 individual model simulations when the MMEM is used to define the forced 16 signal. They further argued that the forced signals based on the smoothed SMEMs 17 provide more accurate estimates of the true forced signals and residual intrinsic 18 variability in individual model ensembles.

19 The novel aspect of Steinman et al.'s attribution methodology compared to 20 previous studies is in rescaling the MMEM signal using linear regression to best 21 match the observations of a given climatic time series; hence, Steinman et al. (2015a) 22 termed their approach to estimating the forced signal "semi-empirical." Physically, 23 the rescaling is meant to correct for biases in the models' climate sensitivity. **Figure 1** 24 illustrates this methodology using the output of the multi-model ensemble (**Table 1**) 25 for the three climate indices considered by Steinman et al. (2015a): the Atlantic

1 Multidecadal Oscillation index (AMO: Kerr 2000; Enfield et al. 2001) defined as the 2 sea-surface temperature (SST) averaged over the North Atlantic, an analogous SST index for the Pacific - Pacific Multidecadal Oscillation (PMO), as well as the 3 4 Northern Hemisphere mean surface temperature (called NMO in Steinman et al. 5 2015a, HMO hereafter); see further details in section 2. The non-scaled 5-yr low-pass 6 filtered individual-model SMEM time series (multiple gray lines in the left panels) 7 have a large spread characterizing the model uncertainty. Their ensemble mean for 8 either AMO or PMO index exhibits a warm bias after the year 2000 with respect to 9 the rescaled version of the ensemble mean that minimizes its root-mean-square (rms) 10 distance from the observed time series; the bias is slight ($\sim 0.1^{\circ}$ C) in AMO and is more 11 substantial (~0.3°C) in PMO. The simulated HMO signals exhibit essentially no bias. 12 These biases or lack thereof are consistent with the scaling factors (relative 13 sensitivities) of individual models listed in Table 1. Hence, the models considered do, 14 on average, a fairly good job of simulating Northern Hemisphere climate change, but 15 exhibit a wide range of climate responses to forcing and tend, in general, to 16 overestimate the observed regional climate sensitivity over the North Atlantic and 17 North Pacific.

18 The gray lines in the right panels of Fig. 1 show the results of applying the 19 regression-based rescaling minimizing the distance between estimated forced signal 20 and the full observed time series to individual smoothed SMEMs, for each climate 21 index considered. Note that the spread among the rescaled individual forced-signal 22 estimates is naturally narrower than the spread among the non-scaled signals in the 23 corresponding left panels, due to the individual estimates being nudged to the 24 common observed time series. Substantial shrinkage of the raw (non-scaled) spread 25 upon rescaling indicates that a sizable fraction of this original spread is due to

1 different climate sensitivities of individual models in the multi-model ensemble. Still, 2 the remaining spread in the rescaled versions of the forced-signal estimates based on 3 individual models' SMEM (right panels), or the *sensitivity adjusted model error*, is 4 quite large. Note also that the MMEM derived forced signal based on ensembleaveraging the scaled smoothed SMEMs of individual models (thick colored lines in 5 6 the right panels of Fig. 1) is very close to the scaled MMEM of non-scaled smoothed 7 SMEMs (thick black lines in the corresponding left panels of Fig. 1); hence, interchanging the order of scaling and ensemble averaging does not affect much the 8 9 final estimate of the forced signal.

10 Steinman et al. (2015a) and Frankcombe et al. (2015) computed the intrinsic 11 component of observed climate variability by subtracting this rescaled MMEM signal 12 from the raw observed time series of each climate index they considered (that is, by 13 forming the difference between magenta and black curves in left panels of Fig. 1 or, 14 equivalently, between magenta and thick colored curves in the right panels of Fig. 1). 15 They further isolated the multidecadal component of the observed intrinsic variability 16 so defined via low-pass filtering and used the results to interpret recent climate trends. 17 The forced signal estimates in Steinman et al. (2015a) were associated with narrow 18 bootstrap-based errorbars, which gave an impression of a striking (and 19 counterintuitive, given a wide spread of forced signals in Fig. 1) consistency between 20 the forced signals simulated by different CMIP5 models. If it were real, this 21 consistency would also translate into high confidence of the corresponding estimates 22 of the observed intrinsic variability.

In this paper, we revisit the Steinman et al.'s analysis and develop a Monte Carlo method to estimate the uncertainties of the MMEM and SMEM based forced signal inference using the observational and model simulated data sets, as described in

1 section 2. In section 3, we identify the procedure leading to the least uncertainty and 2 estimate its inherent biases. The bias corrected procedure is used in section 4 to 3 isolate forced signals and residual intrinsic variability, along with requisite 4 uncertainties, in observations and CMIP5 model simulations, for five widely used 5 climate indices. We then compare the magnitudes and spatiotemporal structures of the 6 observed and simulated intrinsic variability. Section 5 contains summary of our 7 results and a discussion of their implications. The methodology for constructing 8 stochastic models used to produce synthetic realizations of intrinsic variability is 9 spelled out in Appendix A. In Appendix B, we analyze properties of the MMEM time 10 series of the residual intrinsic variability and, in particular, its small variance 11 purported by Steinman et al. (2015a,b) to indicate statistical independence of the 12 simulated residuals.

13

14 **2.** Data sets and analysis methodology

15 *a)* Data sets and procedures

16 We utilized the output from CMIP5 historical twentieth-century simulations 17 for models with four or more ensemble members (Table 1) to analyze the simulated 18 Atlantic Multidecadal Oscillation (AMO: Kerr 2000; Enfield et al. 2001), Pacific 19 Multidecadal Oscillation (PMO: Steinman et al. 2015a), and Northern Hemisphere 20 mean surface temperature (HMO; "NMO" in Steinman et al. 2015a). These indices, as 21 well as their observed counterparts, were the same as used by Steinman et al. (2015a) 22 downloaded from that manuscript's supplementary and were website 23 (www.meteo.psu.edu /holocene/public_html/supplements/Science2015). The AMO 24 and PMO indices were based on SST averaged over the regions (0°N-60°N, 80°W-0°)

1 and (0°N-60°N, 120°E-100°W), respectively. The HMO index was computed as the 2 mean surface temperature (ocean+land) over the 0°N-60°N region. The observed 3 AMO and PMO indices were computed as the average of three SST products: the 4 Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST: Rayner et al. 2003), National Oceanic and Atmospheric Administration (NOAA) Extended 5 6 Reconstructed Sea Surface Temperature (ERSST) (Xue et al. 2003; Smith et al. 2008), 7 and Kaplan SSTs (Kaplan et al. 1998; Parker et al. 1994; Reynolds and Smith 1994). 8 The HMO index was based on Goddard Institute for Space Studies (GISS) Surface 9 Temperature (GISTEMP: Hansen et al. 2010; Cowtan and Way 2014).

10 We also used a subset of model simulations from Table 1 to consider a 11 climate-index network that included two additional indices: the Pacific Decadal 12 Oscillation index (PDO: Mantua et al. 1997; Zhang et al. 1997) and the North Atlantic 13 Oscillation index (NAO: Hurrell 1995; Hurrell and Deser 2009). We computed the 14 PDO as the leading principal component of SST in the region (20°N-60°N, 130°E-15 120°W) after linearly removing the global temperature signal from the raw monthly 16 SST data. The resulting PDO index based on ERSST data is very similar to the one 17 provided by Nate Mantua at http://research.jisao.washington.edu/pdo/ (not shown). 18 The NAO index was computed as the leading principal component of the monthly 19 sea-level pressure (SLP) in the region (15°N-75°N, 90°W-10°W). We used NOAA 20 twentieth-century SLP reanalysis product (20CR: Compo et al. 2011) to define the 21 observed NAO. In addition, we used the station based NAO index 22 (https://climatedataguide.ucar.edu) as an alternative NAO estimate. We normalized the observed and modeled monthly PDO and NAO indices to have the unit standard 23 24 deviation, and then formed and analyzed their annual-mean time series. Note that we 25 also repeated all of the original AMO, PMO and HMO analyses based on the full

model set using our alternative smaller subset of models and ERSST data set for observations; in this case the HMO data was replaced by the mean Northern Hemisphere SST south of 60°N. The latter results (not shown) were completely analogous to those based on the full set of model simulations and different SST products, thus further confirming the robustness of our conclusions.

6

b) Methodology

7 We estimated the forced signals in the 18 individual model ensembles of Table 8 1 as their respective 5-yr low-pass filtered SMEMs, using the data adaptive filter by 9 Mann (2008) available from www.meteo.psu.edu/holocene/public html/smoothing08/. 10 Kravtsov et al. (2015) showed that this procedure results in the uncorrelated 11 realizations of the residual intrinsic variability within each model ensemble. The 12 intrinsic variability so estimated has, however, slightly reduced amplitude due to the 13 method's failing to average out completely the intrinsic variability from the SMEM 14 based forced signal (Steinman et al. 2015b). We will take necessary precautions to 15 account for this bias in our analyses (see section 4).

16 We then fitted a stochastic model to the simulated intrinsic variability 17 (Appendix A) and used it to produce synthetic versions of the CMIP5 multi-model 18 ensemble simulations. We formed synthetic 'CMIP5 simulated' time series by 19 combining the estimated (smoothed SMEM based) forced signals with stochastic 20 realizations of the intrinsic variability. In section 3, we consider 100 synthetic multi-21 model AMO ensembles of 20 'models' and 100 'simulations' (five synthetic 22 simulations per 'model'); here the estimated forced signals from models 1 and 2 of 23 Table 1 enter each multi-model ensemble twice (since we only have 18 different 24 forced signal estimates available). Appendix B analyzes analogous, but even larger synthetic ensembles of 200 'models' and total of 1000 simulations per 'multi-model' ensemble to analyze the dependence of MMEM residuals on the number of simulations considered. Note that in the latter ensemble, we still only have 18 independent forced signals summed up with 1000 independent realizations of intrinsic variability. In section 4, we produce and analyze 100 synthetic multi-model ensembles duplicating the structure of the ensembles in Table 1 in terms of the number of models and individual model simulations.

8 The advantage of working with synthetic time series mimicking the actual 9 CMIP5 data is that the forced signals and intrinsic variability in these time series are 10 known exactly by construction. Therefore, the accuracy and biases in the SMEM and 11 MMEM based forced-signal inference methods can be assessed directly (section 3) 12 and used to compute the uncertainties in both the estimated intrinsic variability of 13 CMIP5 models and in semi-empirical estimates of the observed intrinsic variability 14 (section 4).

15

3. The uncertainties associated with SMEM and MMEM methods to estimate forced and intrinsic variability

18

a) General procedure

We considered 100 synthetic multi-model ensembles of AMO simulations (20 'models' and 5 'simulations' per model), with known forced signals and known stochastic realizations of intrinsic variability (section 2b). We then estimated the forced signals and intrinsic residuals for each synthetic model using: (1) the 5-yr lowpass filtered SMEMs, and (2) a modified version of Steinman et al.'s (2015a,b) MMEM regression method (see below), and compared the methods (1) and (2) in terms of how well they represent the actual forced signals and intrinsic variability
 across all simulations.

3 We introduced two modifications to the original Steinman et al.'s 4 methodology. First, when considering a given model ensemble, we computed the 5 MMEM based on all other ensembles; that is, we excluded all of the simulations of a 6 given model from computing the MMEM. In contrast, Steinman et al. (2015a) defined 7 the MMEM based on all multi-model simulations available, and Steinman et al. 8 (2015b) only excluded the single simulation considered from computation of the 9 MMEM. The second modification was in how we rescale the raw MMEM time series 10 to match a given model's sensitivity. Steinman et al. (2015a,b) defined the rescaling 11 coefficient via linear regression of the MMEM signal against the time series of each 12 individual simulation. Here we used instead a single rescaling coefficient — the 13 average of rescaling coefficients based on individual simulations of this model — per 14 model ensemble.

15

b) Quantifying errors of SMEM and MMEM methods

16 We first computed the spectrum of the actual and estimated intrinsic 17 variability based on SMEM and MMEM methods (Fig. 2a). We used a version of the 18 spectrum that plots the variance of the raw and running-mean boxcar filtered time 19 series as a function of the filter's window size. Naturally, this variance decreases with 20 the window size, as we apply more and more smoothing to the original raw time 21 series. The SMEM based intrinsic residuals have a spectrum that closely matches the 22 observed spectrum, and similar error bars; the estimated variance, however, is slightly 23 lower than the observed variance due to aliasing of some of the actual intrinsic 24 variance into the forced signal, as noted above in section 2b. In contrast, the MMEM-

1 based residuals have a much larger variance compared to the actual intrinsic variance 2 and wide error bars. The lower error bar extends almost as low as the error bars of the 3 SMEM method, due to the actual forced signal in one of the 18 models being close to 4 the MMEM, thus giving, in this particular case, an adequate decomposition of the 5 corresponding model simulation into the forced and intrinsic components. We will see 6 below that the general inflation of residual intrinsic variance in the MMEM regression 7 method in Fig. 2a is dominated by the model error, that is, by the differences between 8 the models' actual and MMEM-regression-estimated forced signals, consistent with 9 the analysis of Kravtsov et al. (2015).

10 The error variance of the SMEM and MMEM based estimates of intrinsic 11 variability (Fig. 2b) is consistent with the spectra in Fig. 2a; once again, the SMEM 12 method gives a much more accurate representation of the actual intrinsic variability. 13 In relative terms, the SMEM method's variance error is uniform in frequency domain, 14 reflecting a reduction of less than 25% compared to the variance of the actual low-15 pass filtered intrinsic variability for averaging window sizes exceeding 5 yr (Figs. 2c, 16 d). On the other hand, the relative error of the MMEM-regression based estimates of 17 intrinsic variability becomes progressively larger at lower frequencies and already 18 exceeds 100% (that is, becomes twice as large as the actual intrinsic variance) for the 19 15-yr low-pass filtered variability. The MMEM based residuals are also characterized 20 by substantially lower correlations with the actual realizations of intrinsic variability 21 than the SMEM based residuals (Fig. 2e), especially at low frequencies, becoming, in 22 fact, statistically uncorrelated with the actual intrinsic variability for the 25-yr low-23 pass filtered and lower-frequency data (where its lower error bar in Fig. 2e straddles 24 zero correlation). Frankcombe et al. (2015) noted a positive bias in the amplitude of 25 the intrinsic variability estimated from CMIP5 twentieth-century runs using the 1 MMEM based forced signal relative to this amplitude in control simulations. Here we 2 show that this bias is frequency dependent and increases substantially for 3 multidecadal variability.

4 Since the estimated forced signal and intrinsic variability add up to the same 5 time series as the sum of actual forced and intrinsic signals by construction, the errors 6 in the intrinsic variability illustrated in Fig. 2 also characterize the errors in forced-7 signal estimation. For our synthetic ensembles considered in this section, we end up 8 with 100 estimates of the forced signal time series for each of 20 models comprising 9 the ensembles. We can further decompose the forced-signal errors into two parts. The 10 model errors arise due to the inferred forced signal being systematically different from 11 the actual forced signal of a given model across all of its 100 available realizations. 12 To compute the model error, we thus averaged the forced-signal difference time series 13 (inferred minus actual) over the 100 realizations, and then computed its root-mean-14 square (rms) average over time and across all 20 models. The remaining "intrinsic" 15 errors are the root-mean-square complement to model errors. They arise due to 16 insufficient cancellation/smoothing of intrinsic variability in the ensemble-mean 17 computation of the forced signal estimate.

18 The forced-signal estimation errors in the SMEM method are dominated by 19 the intrinsic errors, and those in the MMEM method — by the model errors (**Fig. 3**). 20 Smaller intrinsic errors of the MMEM based forced signal estimation are, of course, 21 expected, due to a much larger number of independent intrinsic realizations being 22 averaged in forming the ensemble mean compared to the SMEM method. What's 23 important though is that the MMEM method's model errors drastically exceed the 24 intrinsic errors of the SMEM method, which makes the latter the method of choice for

1 estimating forced signals in the CMIP5 multi-model ensemble.

2 The dominance of the model errors in the MMEM-regression estimates of the 3 forced signal and intrinsic variability also indicates that the intrinsic variability so 4 inferred contains a substantial common bias in the individual model simulations, due 5 to systematic differences between the true and estimated forced signal in each 6 simulation. Indeed, Kravtsov et al. (2015) demonstrated that intrinsic residuals within 7 individual-model ensembles exhibit high statistically significant correlations when the 8 MMEM regression method is used to define the forced signal in these models. This 9 fact is apparently at odds with Steinman et al.'s (2015a,b) claim that the MMEM 10 based intrinsic residuals are statistically independent. They reached this conclusion by 11 forming the grand ensemble-mean time series of intrinsic residuals and analyzing its 12 variance, with small variance supposedly indicating statistical independence. In 13 reality, these arguments are flawed, and small variance of the ensemble-mean 14 residuals instead reflects an algebraic constraint rooted in the definition of the 15 MMEM forced signal (Kravtsov et al. 2015; Appendix B).

16

c) Uncertainties of forced-signal estimates based on multi-model ensemble

17 We can now make use of the surrogate forced-signal estimates computed 18 using SMEM and MMEM methods to construct the most likely forced signal and 19 compute its uncertainty (Fig. 4). The ensemble mean of the SMEM based forced 20 signals (Fig. 4a, red curve) is statistically indistinguishable from the MMEM (Fig. 4b, 21 red curve). The spread of the individual SMEM forced signals (Fig. 4a, gray curves) 22 reflects both model and intrinsic uncertainties, and is characterized by the standard 23 deviation of about 0.1°C. In contrast, the bootstrap-based spread of the MMEM 24 estimates (Fig. 4b, gray curves) — used by Steinman et al. (2015a) to measure the

1 uncertainty of the MMEM-regression inferred forced signal and intrinsic variability-2 is much smaller (0.015°C). Note that the smallness of the MMEM uncertainty does 3 not indicate that the MMEM provides a more robust estimate of the forced signal. 4 Instead, it simply is a consequence of the fact that surrogate bootstrap ensembles, 5 which typically contain about 2/3 of different model simulations from the all-model 6 ensemble, effectively sample the simulations (and forced signals) from all of the 7 models considered, thus resulting in a similar estimate of the MMEM every time. 8 Hence, the narrow ranges of uncertainty in semi-empirical estimates of the forced 9 signal and intrinsic variability in Steinman et al. (2015a) are misleading.

10 Enter now the semi-empirical forced-signal estimation, in which we correct 11 for different sensitivities of individual models and rescale their forced signals to 12 match the 'observed' sensitivity. To illustrate this concept, we will take here the 13 MMEM signal as the proxy for the observed time series; when working with the real 14 observed data in section 4, we will use raw observed time series of climate indices considered instead. If x represents the time series of a single simulation, f_0 and x'_0 are 15 the true forced and intrinsic signals, respectively, and \mathbf{f} and \mathbf{x}' are the estimated forced 16 17 and intrinsic signals, then

$$x = f_0 + x'_0 = f + x',$$
 (1a)

19
$$\Delta \mathbf{f} = \mathbf{f} - \mathbf{f}_0; \ \Delta \mathbf{x} = \mathbf{x}' - \mathbf{x}'_0; \ \Delta \mathbf{f} = -\Delta \mathbf{x}. \tag{1b}$$

20 The last equality simply means, as we mentioned earlier, that the error in 21 determination of the forced signal $\Delta \mathbf{f}$ has the same magnitude as that in determination 22 of intrinsic variability $\Delta \mathbf{x}$. Rearranging the first equation in (1b) we have

$$f = \mathbf{f_0} + \Delta \mathbf{f} , \qquad (2)$$

1 which states, trivially, that each of the individual forced-signal estimates in Fig. 4a 2 (gray lines) is the sum of the true forced signal (which we do know in our synthetic 3 samples) and its error. We now do the model sensitivity adjustment and redefine the 4 forced signal by rescaling f_0 with the coefficient α to best match the MMEM, but 5 *keeping the forced signal estimation error intact*:

$$6 \mathbf{f_1} = \alpha \mathbf{f_0} + \Delta \mathbf{f} \ . (3)$$

7 The newly defined forced signal f_1 is, therefore, corrected for the individual model 8 sensitivity bias, but still possesses all other sources of error. Furthermore, the new 9 semi-empirical time series of a given model realization

$$10 \mathbf{x_1} = \mathbf{f_1} + \mathbf{x}' (4)$$

11 is still characterized by the same uncertainty $|\Delta \mathbf{f}| = |\Delta \mathbf{x}|$ in the rescaled forced signal 12 and the (original) intrinsic variability.

13 We created synthetic estimates of the rescaled forcing (3) by randomly pulling 14 the time series of $\Delta \mathbf{f}$ from the ensemble of error time series generated by the SMEM and MMEM methods. The spaghetti plots of rescaled estimated forcing $\mathbf{f_1}$ based on 15 16 Δf ensemble from the SMEM procedure (Fig. 3c) has a narrower spread (0.076°C) 17 than that in Fig. 3a due to eliminating the uncertainty associated with the individual 18 models' different climate sensitivities. The forced signal uncertainty is uniformly 19 distributed in time. The corresponding uncertainty in the forced signal estimate based 20 on the Δf ensemble from the MMEM procedure is larger (with the time-mean value of 21 0.093°C; this amounts to about 50% increase in the error variance compared to Fig. 22 3c) and has time-dependent structure elucidating the forced-response biases in 23 individual models. These results, once again, indicate that the SMEM-based method

of the forced signal inference provides more robust and structurally stable uncertainty
 estimates of the forced signal and residual intrinsic variability compared to the
 MMEM regression method of Steinman et al. (2015a,b) and Frankcombe et al. (2015).

- 4
- 5

6

4. Observed and CMIP5 simulated intrinsic variability

a) Estimating forced signal and its uncertainty from CMIP5 simulations

7 In this section, we apply the Monte Carlo method developed in section 3 to the 8 CMIP5 multi-model ensemble of Table 1. To summarize, this method consists of the 9 following steps: (i) compute the initial forced signal estimates of the (18) individual 10 models as the 5-yr low-pass filtered SMEMs; (ii) generate 100 surrogate multi-model 11 ensembles by combining the forced signal estimates from step (i) with random 12 realizations of intrinsic variability from the stochastic model described in Appendix 13 A; (iii) compute 100 estimates of the forced signal for each model based on surrogate 14 data from step (ii); (iv) derive the best estimate of the forced signal and its uncertainty 15 based on the ensemble mean and spread over all of the (1800) estimated surrogate 16 forced signals from step (iii); (v) correct for biases in the estimated uncertainty by 17 comparing the spreads of the (known) actual and inferred intrinsic variability in 18 surrogate stochastic samples.

The model simulations span the period 1861–2005, and the observed time series extend through 2012. Following Steinman et al. (2015a), we extrapolated the synthetic forced signals from step (iii) above through year 2012 using the slope of 15yr (1991–2005) trend. Steinman et al. used the 30-yr trend for the same purpose, but the (cross-validated retroactive hindcast) skills of the 15-yr and 30-yr trend extrapolation are similar (not shown). In step (iv), we computed three versions of the

1 forced signals and their uncertainties. In the first version (Figs. 5a,d,g), we used the 2 non-scaled forced signals. The uncertainty of these estimates includes that due to 3 different climate sensitivities of the individual models. In the second version (Figs. 4 5b,e,h), we rescaled the linearly extrapolated, 1861–2012 forced signals of version 1 5 via linear regression against the raw observed time series to match the observed 6 climate sensitivity using Eqs. (3) and (4). Finally, we computed the third version 7 (Figs. 5c,f,i) by first rescaling the 1861–2005 forced signals, and then linearly 8 extrapolating the rescaled signals through year 2012. Versions 2 and 3 thus differed 9 by the treatment of extrapolated forced signals; they should give us an idea about the 10 sensitivity of the 2005-2012 forced-signal estimation to the details of linear 11 extrapolation procedure. We reduced the 95% spread of the synthetic forced signals 12 (dashed lines in Fig. 5) by 0.032°C for AMO, 0.03°C for PMO, and by 0.027°C for 13 HMO, based on the smoothed SMEM estimation biases computed in step (v).

14 The spread of the non-scaled forced-signal estimates is, naturally, the largest 15 due to its including the model-sensitivity errors. The semi-empirical, sensitivity 16 corrected estimates of the forced signal are more concentrated around the grand 17 ensemble-mean estimate, but still have a much wider spread than bootstrap based 18 error bars in Steinman et al. (2015a), consistent with the discussion of Fig. 4 in 19 section 3c. Note also that the forced signal uncertainty grows linearly over the 2005-20 2012 period in the version-3 forced signal, reflecting the prediction uncertainty of the 21 linear extrapolation. This uncertainty is effectively erased by rescaling in the version-22 2 forced signal, due to using the information about the future SST behavior, which is 23 in fact unavailable in 2005. Hence, version-3 provides the most accurate 24 representation of the forced signal and its uncertainty.

b) Semi-empirical estimates of observed intrinsic variability

2 Differencing the observed time series (purple lines in Fig. 5) and our surrogate 3 forced-signal estimates (gray lines in Fig. 5) produces the corresponding surrogate 4 estimates of the observed intrinsic variability. Following Steinman et al. (2015a), we 5 concentrated on its multidecadal component by applying, to the estimated time series 6 characterizing the observed intrinsic variability, the 40-yr data-adaptive low-pass 7 filter of Mann (2008). First we note, however, that the direct application of this filter 8 results in large end effects, as also noted by Frankcombe et al. (2015). To demonstrate 9 this, we considered surrogate samples of the intrinsic variability generated by our 10 stationary stochastic model of Appendix A; the spread of the raw, unfiltered 11 variability is uniform throughout the entire 1861–2005 simulation period (not shown). 12 In contrast, the 40-yr low-pass filtered variability exhibits substantial increase of 13 variance at both ends of the time series (Fig. 6a). To alleviate these end effects, we 14 have derived tapers (Fig. 6b), which, when multiplying the filtered time series, 15 produce the uniform (in time) spread of the low-pass filtered variability (not shown).

16 The resulting estimates of the tapered 40-yr low-pass filtered observed intrinsic variability are shown in Fig. 7 based on the version-2 (left) and version-3 17 (right) of the rescaled forced signals. The ensemble-mean estimates of the 18 19 multidecadal intrinsic variability in AMO, PMO and HMO in Fig. 7 are fairly similar 20 to those in Steinman et al. (2015a) (see their Fig. 3c), but the estimates of uncertainty 21 are drastically different. Steinman et al.'s (2015a) narrow error bars are merely an 22 artifact of bootstrap samples generation (see section 3c and Fig. 4), whereas the error 23 bars in Fig. 7 reflect the actual (large) uncertainty associated with the model and 24 intrinsic errors of the forced-signal estimation. In particular, these errors are 25 sufficiently large to render the attribution of the recent cool down of the PMO (Fig.

1 7d) and HMO (Fig. 7f) to the intrinsic variability barely statistically significant if at 2 all (cf. Steinman et al. 2015a,b). For example, statistical significance, at 5% level, would correspond to the difference between the 1995 PMO maximum and 2012 PMO 3 minimum exceeding $2\sqrt{2\sigma^2}$ (if PMO is normally distributed with the standard 4 5 deviation σ), effectively meaning that the lower error bar at 1995 shortened by about a quarter of its length (since the unscaled error bar corresponds to $\sim 2\sigma$, so $3/4 \times$ 6 $2\sigma \sim \sqrt{2}\sigma$) should not overlap with the upper error bar at 2012 shortened in the same 7 8 way (by 1/4 of its length).

9

c) Magnitudes of observed vs. simulated intrinsic variability

10 We now have estimates of both observed (section 4b) and CMIP5 simulated 11 intrinsic variability (obtained by subtracting the 5-yr low-pass filtered SMEMs from 12 individual model simulations). In the latter, we need to correct for the amplitude bias 13 due to forced-signal smoothing (see section 3b). The frequency dependent amplitude 14 correction factors amount to about 6% for raw annual data and saturate at about 9% for the low-pass filtered data (Fig. 8a). The amplitude corrected spectra of the CMIP5 15 16 intrinsic variability in the AMO, PMO and HMO indices (Figs. 8c-d, respectively) 17 still have standard deviations which are substantially smaller than the standard 18 deviations of the observed intrinsic variability (as also noted, for AMO, by 19 Frankcombe et al. 2015). In particular, the observed ensemble-mean amplitudes exceed the 97.5th percentile of the simulated amplitudes for a wide range of time 20 21 scales, including the multidecadal variability. When taking into account the 22 uncertainty of the observed intrinsic variability, the 95% error bars for both the 23 observed and simulated amplitudes have to be reduced by about a quarter of their 24 length (see above), in which case the non-overlapping error bars would indicate that the null hypothesis of the observed and model generated intrinsic variability have the same amplitude. It is clear from Fig. 8 that intrinsic variability simulated by the CMIP5 models is significantly weaker than the observed intrinsic variability inferred by subtracting the CMIP5 derived forced signals from the full observed climatic time series.

6

d) Semi-empirical decomposition of PDO and NAO indices

7 All of the above results for AMO, PMO and HMO indices remain valid for a 8 sub-ensemble of CMIP5 model simulation defined in Table 1 (not shown). For this 9 sub-ensemble, we also computed the PDO and NAO indices (section 2a) and applied 10 our Monte-Carlo procedure to infer the observed and simulated forced signals and 11 intrinsic variability in these indices (Fig. 9). The forced signal in either PDO or NAO 12 is essentially non-existent (Figs. 9a,d), and both indices are dominated by the intrinsic 13 variability characterized by a pronounced multidecadal oscillation shown in Figs. 14 8b,e. Furthermore, both of these indices stand out in terms of how the amplitudes of 15 their intrinsic variations in CMIP5 simulations compare with the amplitude of the 16 observed intrinsic variability. In particular, the PDO is the only index from the five 17 indices considered for which the observed and simulated variations have a consistent 18 magnitude (Figs. 9c). On the other hand, the observed low-frequency variability of the 19 NAO index is strikingly larger than that in the CMIP5 simulations (Fig. 9f). This 20 result holds for both the 20CR based observational NAO estimate and for an 21 alternative, station based version of the NAO index, thus suggesting that the large 22 observed amplitude is not an artifact of observational data set. The results reported in 23 sections 4c,d regarding the comparison between the magnitudes of observed and 24 CMIP5 simulated intrinsic variability are consistent with earlier analyses of Kravtsov 25 et al. (2014, 2015).

2 The observed (40-yr low-pass filtered) intrinsic variability in our network of 3 five climate indices exhibits non-trivial lagged correlations among its members (Fig. 4 10). In terms of the location of maximum statistically significant (non-zero) 5 correlations, -AMO leads PDO by 17 yr and NAO by 7 yr (consistent with estimates 6 by Wyatt et al. 2012 and Kravtsov et al. 2014). The AMO and PMO are nearly in 7 phase, with AMO leading slightly. In contrast, the CMIP5 simulated (40-yr low-pass 8 filtered) intrinsic variability in the 89 CMIP5 runs considered, aside from having 9 much smaller than observed amplitude in all indices but PDO, exhibits no definitive 10 lead-lag relationships among the indices on multidecadal time scale (Fig. 11).

11 Note that the values of maximum CMIP5 correlations in Fig. 11 are similar to 12 the observed maximum correlations in Fig. 10, indicating that the latter (observed 13 lagged) correlations among individual pairs of climate indices considered are quite 14 likely to arise by chance. The likelihood of capturing the strength of the observed 15 relationship between the climate indices in CMIP5 models quickly drops, however, 16 when multiple indices are considered simultaneously in the single climate network to 17 identify the leading mode of its spatiotemporal variability via MSSA (Wyatt and 18 Peters 2012; Kravtsov et al. 2014). These authors removed linear trends to study the 19 resulting networks of the observed and simulated climate-index anomalies. We plan to 20 extend their analyses by applying MSSA to the present paper's semi-empirical 21 estimates of the observed and simulated intrinsic variability.

22

5. Summary and discussion

2 We considered five indices characterizing Northern Hemisphere climate 3 variability (AMO, PMO, HMO, PDO, NAO) and developed a Monte Carlo approach 4 to estimate uncertainties of their forced and intrinsic components using the multi-5 model ensemble of CMIP5 twentieth-century simulations. We showed that using the 6 multi-model ensemble mean (MMEM) to infer forced signals in individual models 7 (Frankcombe et al. 2015; Steinman et al. 2015a,b) results in substantial overestimation of the residual intrinsic variability, especially on multidecadal time 8 9 scales, due to the differences between MMEM and true forced response of individual 10 models masquerading as their intrinsic variability (Kravtsov et al. 2015). On the other 11 hand, intrinsic variability inferred by subtracting smoothed single-model ensemble-12 mean (SMEM) forced signals from individual model simulations approximates true 13 simulated intrinsic variability much more accurately, albeit with a slight negative 14 amplitude bias resulting from lack of cancelation among different realizations of 15 intrinsic variability in the SMEM forced-signal estimates (Steinman et al. 2015b).

16 One source of uncertainty in the SMEM based forced signals of individual 17 CMIP5 models comes from their different climate sensitivities. The models capture 18 non-uniform warming trends in Northern Hemisphere mean surface temperature 19 (HMO) very well, whereas they tend to overestimate the rate of the observed regional 20 climate change in the AMO and PMO indices. On the other hand, the CMIP5 21 simulated PDO and NAO indices do not exhibit any pronounced long-term trends. 22 Following Steinman et al. (2015a,b) and Frankcombe et al. (2015), we rescaled the 23 model-derived forced signals to best match the observed climate sensitivity; in doing 24 so, we, however, preserved the bias-corrected uncertainty associated with (SMEM 25 estimated) forced signals of individual models.

1 We used the resulting semi-empirical (sensitivity adjusted) forced signals to 2 isolate intrinsic variability in the observed network of climate indices and compared 3 its spatiotemporal characteristics with those of CMIP5 simulated networks. The 4 observed intrinsic variability is characterized by a pronounced multidecadal 5 oscillation in all of the indices, and distinctive lead-lag relationships among them 6 consistent with earlier analyses (Wyatt et al. 2012; Wyatt and Curry 2014; Kravtsov et 7 al. 2014). In contrast, the CMIP5 simulated indices exhibit a much weaker (by a 8 factor of 5–10 in terms of variance) multidecadal intrinsic variability in all of them 9 except PDO, and no robust coherence in time.

10 Frankcombe et al. (2015) [see their Fig. 5c] and Trenary and DelSole (2016) 11 previously noted a striking lack of 'intrinsic' AMO variance in CMIP5 model 12 simulations. Trenary and DelSole (2016) went on to conclude that this property must 13 imply that the observed multidecadal excursions of AMO, unmatched by most of the 14 CMIP5 models, are largely externally forced (see also Mann and Emanuel 2006; 15 Mann et al. 2014). In this interpretation, more pronounced multidecadal undulations 16 of the observed surface temperatures would be due to models' underestimating the 17 multidecadal component of the true forced climate response, while the true intrinsic 18 variability in observations would be consistent with the simulated intrinsic variability. 19 Indeed, Booth et al. (2012) proof-of-concept results demonstrated quantitative 20 feasibility of this scenario; in their model (HadGEM2-ES), indirect aerosol effects 21 drove multidecadal AMO variability in SST that closely matched observations. 22 Similarly, Golaz et al. (2013) documented large sensitivity of GFDL CM3 model to 23 small changes in the cloud formulation parameters (via ensuing large changes in the 24 magnitude of aerosol indirect effects on climate), implying that the magnitude of 25 forced variations in simulated climates is tunable. On the other hand, Zhang et al. (2013) identified a number of major discrepancies between the observed climate
changes and those simulated in the HadGEM2-ES model, and attributed them to
excessively strong aerosol effects. More generally, Stevens (2015) argued that the
majority of CMIP5 models already have excessive aerosol forcing, which thus cannot
fully explain the multidecadal downturns of the observed climate warming.

6 An alternative (or complementary) explanation of excessively damped 7 multidecadal climate variability in CMIP5 models involves the possibility that these 8 models misrepresent some of the dynamical feedbacks at work in the real climate 9 system, in which case the model-data differences would reflect a lack or distortion of 10 multidecadal intrinsic dynamics in climate models. Multidecadal climate variations in 11 the North Atlantic region have been associated with Atlantic Meridional Overturning 12 Circulation (AMOC: Delworth et al. 1993; Timmermann et al. 1998; Delworth and 13 Mann 2000; Latif et al. 2004; Knight et al. 2005, 2006). The dynamics of the AMOC 14 and North Atlantic SST signals are still not fully understood, and a large number of 15 different theories are available (Delworth et al. 1993; Timmermann et al. 1998; 16 Frankcombe et al. 2009, 2010; Frankcombe and Dijkstra 2011; Clement et al. 2015; 17 Trenary and DelSole 2016). Meanwhile, the simulated signals exhibit a wide spread in 18 their characteristic time scales and amplitudes across CMIP5 models (Zhang and 19 Wang 2013; Ba et al. 2014), and the climate response mostly confined to the North 20 Atlantic region and its immediate surroundings (Enfield et al. 2001; Sutton and 21 Hodson 2005; Knight 2006), perhaps with an in-phase (simultaneous) teleconnection 22 to North Pacific (Kravtsov and Spannagle 2008; DelSole et al. 2011; Wyatt and Peters 23 2012; Kravtsov et al. 2014).

In contrast, Wyatt et al. (2012) and Kravtsov et al. (2014) argued that the spatiotemporal structure of observed multidecadal climate variability in the twentieth

1 century is more complex and involves hemispheric propagation of the AMO 2 multidecadal signal, which they termed the "stadium wave." They further suggested 3 that the absence of the stadium wave in CMIP3 (Wyatt and Peters 2012) and some of 4 the CMIP5 models (Kravtsov et al. 2014) is due to the models' failing to transmit the 5 simulated AMO signal to the overlying atmosphere, which may in turn result from the 6 insufficient atmospheric response to SST or sea-ice anomalies in model simulations; 7 see Kushnir et al. (2002) and Wyatt and Curry (2014). This lack of atmospheric 8 response is reflected in the deficit of multidecadal SLP variability in model 9 simulations. Our results, based on a large CMIP5 multi-model ensemble, corroborate 10 these conclusions.

11 Finally, we venture to speculate about a hypothetical scenario in which climate 12 models have a more pronounced multidecadal intrinsic variability consistent with our 13 semi-empirical estimates. In this case, a combination of a weaker forced warming 14 trend and large multidecadal climate variability could easily produce synthetic 15 realizations of climate variability as close to the observed variations as the current 16 CMIP5 simulations dominated by the strong non-uniform forced trend. Were such a 17 configuration possible, it could introduce significant upward revisions in the Mann et 18 al.'s (2016) estimates of the likelihood of the recent warming to occur in the absence 19 of anthropogenic effects.

To conclude, state-of-the-art climate models are characterized by a substantial model uncertainty, large sensitivity to aerosol and cloud parameterizations and a possible lack of feedbacks that could amplify multidecadal intrinsic variability, which impedes clear attribution of the observed twentieth century climate change.

24

1 Acknowledgements. We acknowledge the World Climate Research Programme's 2 Working Group on Coupled Modeling, which is responsible for CMIP, and we thank 3 the climate modeling groups for producing and making available their model output. 4 ERSST and 20CR SLP data were provided by NOAA: www.ncdc.noaa.gov/data-5 access/marineocean-data/extended-reconstructed-sea-surface-temperature-ersst-v3b, 6 https://climatedataguide.ucar.edu/climate-data/noaa-20th-century-reanalysis-version-7 2-and-2c. Kaplan SST V2 data were provided by the NOAA/ Office of Oceanic and 8 Atmospheric Research/Earth System Research Laboratory Physical Sciences Division, 9 Boulder, Colorado, USA: www.esrl.noaa.gov/psd. HadISST data were provided by 10 the Met Office Hadley Centre: www.metoffice.gov.uk/hadobs. Station based NAO 11 index was retrieved from https://climatedataguide.ucar.edu/climate-data/hurrell-north-12 atlantic-oscillation-nao-index-station-based. We also thank the authors of Steinman et 13 al. (2015) for making their data publicly available. This research was supported by 14 NSF grants OCE-1243158 and AGS-1408897. All data and MATLAB © scripts for 15 this paper are available for downloading from http://atmo.math.uwm.edu:8181/ \rightarrow S. 16 Kravtsov Data \rightarrow KC2016.

17

18 Appendix A: Stochastic model for CMIP5-simulated intrinsic variability

To produce independent Monte Carlo realizations of intrinsic variability that best mimic CMIP5-simulated variability, we worked, for each climate index xconsidered, with residual time series obtained by subtracting the respective 5-yr lowpass filtered SMEMs from the individual model simulations. For the full set of simulations in Table 1, this resulted in 116 time series of estimated intrinsic variability, 145-yr-long each. Next, we concatenated all these multiple time series into a single 145×116-yr-long time series and fitted, to this extended time series, a threelevel stochastic model following the multi-level methodology of Kravtsov et al.
(2005, 2009). The model had the following form:

$$x^{n+1} = a_1 x^n + r^n, (A1a)$$

5
$$r^{n+1} = a_2 x^n + b_2 r^n + q^n$$
, (A1b)

6
$$q^{n+1} = a_3 x^n + b_3 r^n + c_3 q^n + d_3 \xi$$
, (A1c)

7 where *n* is the time index, *r* and *q* are residual time series for the first (A1a) and 8 second (A1b) model levels, and the coefficients a, b, c were found sequentially for 9 each model level, from top to bottom, by multiple linear regression. To produce synthetic realizations of the intrinsic variability x, model (A1) was randomly 10 11 initialized and driven, at the third model level (A1c), by a Gaussian white noise ξ with amplitude d_3 inferred from that of the actual third-level residual. The model (A1) is a 12 slightly extended version of the lowest-order auto-regressive moving-average 13 14 (ARMA) models used by Mann et al. (2016) to generate synthetic Monte Carlo 15 realizations of intrinsic climate variability. Kravtsov et al. (2015) used the models 16 (A1) fitted separately to CMIP5 individual-model ensembles, rather than to 17 concatenated multi-model time series as in here, and produced synthetic multi-model 18 intrinsic samples statistically indistinguishable from the realizations of the single 19 empirical model (A1) here (not shown). This indicates that different CMIP5 models 20 are consistent in terms of the intrinsic variability they generate, and that fitting the 21 single model (A1) to the CMIP5 multi-model ensemble is sufficient to capture the 22 salient characteristics of the simulated intrinsic variability.

23

Appendix B: Variance of ensemble-mean time series of residual intrinsic variability

Following the notations of Kravtsov et al. (2015), we consider *M* time series of length *T*, corresponding to *M* different climate simulations: $x_m^{(t)}$; m = 1, ..., M; t =1, ..., *T*. Let the bar denote averaging across the time dimension (*t*), and square brackets denote averaging across the ensemble-member dimension (*m*). For example, the time mean of each ensemble member $\overline{x_m}$ and the ensemble-average time series $\begin{bmatrix} x^{(t)} \end{bmatrix}$ are defined as follows:

$$\overline{x_m} = \frac{1}{T} \sum_{t=1}^T x_m^{(t)},\tag{B1}$$

$$[x^{(t)}] = \frac{1}{M} \sum_{m=1}^{M} x_m^{(t)} .$$
(B2)

9 Consider a decomposition of $x_m^{(t)}$ into the forced signal $f_m^{(t)}$ and residual 10 intrinsic variability $\epsilon_m^{(t)}$:

$$x_m^{(t)} = f_m^{(t)} + \epsilon_m^{(t)}.$$
 (B3)

11 Without loss of generality, we can assume $\overline{x_m} = \overline{f_m} = 0$, hence $\overline{\epsilon_m} = 0$. For unbiased 12 forced signal $f_m^{(t)}$ and the distribution of $\epsilon_m^{(t)}$ with variance σ^2 , the *ensemble-mean* 13 residual time series $[\epsilon^{(t)}]$ will have the variance σ^2/M . Hence, one can quantitatively 14 assess the statistical independence of different realizations of simulated intrinsic 15 variability by comparing the actual dispersion $\overline{[\epsilon]^2}$ of the ensemble-mean time series 16 $[\epsilon^{(t)}]$ with its theoretical prediction $[\overline{\epsilon^2}]/M$, where we estimated $\sigma^2 \sim [\overline{\epsilon^2}]$.

Steinman et al. (2015a) considered the following two methods for estimating
 the forced signal, both based on the multi-model ensemble-mean time series [x^(t)],
 the differencing method and regression method:

4
$$f_m^{(t)} = [x^{(t)}],$$
 (B4a)

5
$$f_m^{(t)} = a_m[x^{(t)}],$$
 (B4b)

6 where a_m is found via least squares to minimize $\overline{\epsilon_m^2}$ in (B3).

7 With both of these choices of forced signal, the ensemble-mean residual time 8 series $[\epsilon^{(t)}]$ is identically zero

9
$$[\epsilon^{(t)}]=0; t = 1, ..., T,$$
 (B5)

and so is its variance $\overline{[\epsilon]^2} = 0$. The identity (B5) is trivial to prove by taking the ensemble average of (B3) and using the definition of forcing (B4a). For the forcing given by (B4b),

$$a_m = \frac{\overline{x_m[x]}}{[x]^2} ; \tag{B6}$$

13 [a] = 1, so the ensemble average of (B3) also vanishes. Hence, the extreme smallness 14 of the dispersion of ensemble-average intrinsic variability attributed by Steinman et 15 al. to the statistical independence of its different realizations is actually an artifact of 16 the algebraic constraint (B5); see Kravtsov et al. (2015).

Steinman et al. (2015b) repeated the calculations of Steinman et al. (2015a),
but using *M*-1 simulations of the multi-model ensemble to compute the MMEM and
defined it as the forced signal for the *M*-th simulation. In doing so, they obtained

1 similar small values of $\overline{[\epsilon]^2}$. This is to be expected, however, since excluding only one 2 model simulation from the multi-model ensemble cannot significantly affect the 3 MMEM (Kravtsov et al. 2015) and the algebraic constraint (B4) would still be the 4 root cause of cancellations within the MMEM of residual time series.

5 To show this explicitly, consider first a simpler case of the differencing 6 method (B4a). The estimate of the *m*-th model's forcing in this case is

$$f_m^{(t)} = \frac{M}{M-1} \left[x^{(t)} \right] - \frac{1}{M-1} x_m^{(t)} \,. \tag{B7}$$

7 Plugging (B7) in (B3) and taking the ensemble average gives

$$\left[\epsilon^{(t)}\right] = \left[x^{(t)}\right] - \frac{M}{M-1}\left[x^{(t)}\right] + \frac{1}{M-1}\left[x^{(t)}\right] = 0,$$
(B8)

- 8 so the original algebraic constraint (B5) is recovered exactly.
 - For the forced signal estimate (B4b),

$$a_{m} = \frac{(M-1)}{M} \frac{\left(\overline{x_{m}[x]} - \overline{x_{m}^{2}} / M\right)}{([x] - x_{m} / M)^{2}} \approx$$

$$\approx \frac{(M-1)}{M} \frac{\left(\overline{x_{m}[x]} - \overline{x_{m}^{2}} / M\right)}{[x]^{2}} \left(1 + \frac{2}{M} \frac{\overline{x_{m}[x]}}{[x]^{2}}\right), \tag{B9}$$

10 where the last line uses the assumption of large number of simulations $M \gg 1$.

11 Ensemble averaging (B9), we get

$$\frac{M}{M-1}[a] = \frac{\overline{[x]^2} - \left[\overline{x^2}\right]/M}{\overline{[x]^2}} + \frac{2}{M} \frac{\left[(\overline{x_m[x]})^2\right]}{\overline{[x]^2}} + O\left(\frac{1}{M^2}\right) =$$

12

$$= 1 - \frac{(1 - 2c^2)}{M} \frac{\left[\overline{x^2}\right]}{[x]^2},$$
(B10)

1 with $0 < c^2 < 1$ representing the averaged squared correlation of individual 2 simulation time series with the MMEM.

3

Now, taking the ensemble average of (B3) using (B7) and (B10), we can write

$$\left[\epsilon^{(t)}\right] = \left[x^{(t)}\right] \left(1 - \frac{M}{M-1}[a]\right) + \frac{1}{M-1}\left[ax^{(t)}\right] =$$
$$= \frac{\left[x^{(t)}\right]}{M} \left(1 - 2c^2 + \frac{M}{M-1}\tilde{a}\right),$$
(B11)

where $\tilde{a} = [ax^{(t)}]/[x^{(t)}]$, close to unity, is the representative sensitivity of the 4 individual model simulations with respect to MMEM. Eq. (B11) states that the 5 MMEM time series of the residual intrinsic variability $[\epsilon^{(t)}]$ computed by subtracting, 6 7 from each model simulation, the MMEM computed over all other M-1 simulations, is proportional to the grand MMEM $[x^{(t)}]$ divided by *M*. Hence, the variance $\overline{[\epsilon]^2}$ scales 8 as σ^2/M^2 , rather than σ^2/M as expected from the cancellation of independent 9 10 random realizations of intrinsic variability. Since we never made any assumptions 11 about the independence of intrinsic residuals in deriving the constraint (B11), this constraint, valid for any collection of time series $x_m^{(t)}$, reflects an algebraic property of 12 13 the ensemble mean for the specific definition of the forced signal (B7).

14 This behavior is illustrated in **Fig. 12**a for 1000 synthetic multi-model 15 ensembles, each consisting of 200 models and five simulations per model, totaling 16 1000 model simulations per ensemble (section 2b). We computed the distribution of 17 the standard deviations of ensemble-mean residuals $\Sigma = \sqrt{[\epsilon]^2}$ for multi-model sub1 ensembles of size M=100, 200, ..., 1000. We also computed the parameters σ , c^2 and 2 $\tilde{\alpha}$ to obtain the theoretical estimate of the ensemble-mean residual time series $[\epsilon^{(t)}]$ 3 via (B11). The theoretical results match very well the actual computed Σ , which scales 4 as σ/M due to algebraic constraint (B11) and is thus much smaller than the expected 5 standard deviation $\Sigma = \sigma/\sqrt{M}$ based on the sum of independent residuals.

6 The same conclusions hold for the case in which not one, but all of a given 7 model's simulations are excluded from computing the MMEM to estimate its forced 8 signal (section 3a); see Fig. 12b. Once again, this is to be expected since the single 9 model only has a slight influence on the MMEM computed over a large ensemble of 10 models. We also included in Fig. 12b the estimates of Σ based on the smoothed 11 SMEM forced signals, which are closer to the theoretical expectation, but still smaller 12 (note that they are non-zero only here because of the smoothing of individual 13 SMEMs).

Finally, **Figure 13** complements the results in Fig. 12b by showing that the entire algebraic constraint (B11) originally derived for the case with individual simulations excluded from the computation of MMEM, is in fact approximately valid for a more general case that leaves out the individual model ensembles: the ensemblemean residual time series is proportional to the grand MMEM and its amplitude scales as 1/M.

In summary, the small variance of the grand-mean of residual time series based on regressing out the MMEM from individual model simulations (Steinman et al. 2015a,b) has nothing to do with the independence of the intrinsic residuals so estimated, but is rather a reflection of the algebraic constrained rooted in the definition of the forced signal via the MMEM.

1 References

2	Allen, M. R., and S. F. B. Tett, 1999: Checking for model consistency in optimal					
3	fingerprinting. Climate Dyn., 15, 419–434.					
4	Ba, J., Keenlyside, N. S., M. Latif, et al., 2014: A multi-model comparison of Atlantic					
5	multidecadal variability. Climate Dynamics, 43, 2333-2348,					
6	http://dx.doi.org/10.1007/s00382-014-2056-1.					
7	Booth, B. B. B., N. J. Dunstone, P. R. Halloran, R. Andrews, and N. Bellouin, 2012:					
8	Aerosols implicated as a prime driver of twentieth-century North Atlantic					
9	climate variability. Nature, 484, 228-232.					
10	Clement, A., Bellomo, K., Murphy, L., Cane, M., Mauritsen, T., Rädel, G., and					
11	Stevens, B., 2015: The Atlantic Multidecadal Oscillation without a role for					
12	ocean circulation. Science, 340, 320-324, doi:10.1126/science.aab3980					
13	Cook, J., D. Nuccitelli, S. A. Green, et al., 2013: Quantifying the consensus on					
14	anthropogenic global warming in the scientific literature. Environ. Res. Lett.,					
15	8, 024024, doi:10.1088/1748-9326/8/2/024024.					
16	Compo, G. P., J. S. Whitaker, P. D. Sardeshmukh, et al. 2011: The twentieth century					
17	reanalysis project. Quart. J. Royal Meteor. Soc., 654, 1-28,					
18	doi:10.1002/qj.776.					
19	Cowtan, K., and R. G. Way, 2014: Coverage bias in the HadCRUT4 temperature					
20	series and its impact on recent temperature trends. Quart. J. Royal Meteor.					
21	<i>Soc.</i> , 140 , 1935–1944.					
22	DelSole, T., 2001: Optimally persistent patterns in time-varying fields. J. Atmos. Sci.,					
23	58 , 1341–1356.					
24	DelSole, T., 2006: Low-frequency variations of surface temperature in observations					
25	and simulations. J. Climate, 19, 4487–4507.					

1	DelSole, T., and M. K. Tippett, 2007: Predictability: Recent insights from information				
2	theory. Rev. Geophys., 45, RG4002, doi:10.1029/2006RG000202.				
3	DelSole, T., and M. K. Tippett, 2009: Average predictability time: Part II: Seamless				
4	diagnosis of predictability on multiple time scales. J. Atmos. Sci., 66, 1188-				
5	1204.				
6	DelSole, T., and J. Shukla, 2010: Model fidelity versus skill in seasonal forecasting. J.				
7	<i>Climate</i> , 18 , 4794–4806.				
8	DelSole, T., M. K. Tippett, and J. Shukla, 2011: A significant component of unforced				
9	multidecadal variability in the recent acceleration of global warming. J .				
10	<i>Climate</i> , 24 , 909–926.				
11	DelSole, T., L. Jia, and M. K. Tippett, 2013: Decadal prediction of observed and				
12	simulated sea surface temperatures. Geophys. Res. Lett., 40, 2773-2778.				
13	Delworth, T. L., and M. E. Mann, 2000: Observed and simulated multidecadal				
14	variability in the Northern Hemisphere. Climate Dyn., 16, 661-676.				
15	Delworth, T. L., S. Manabe, and R. Stouffer, 1993. Interdecadal variability of the				
16	thermohaline circulation in a coupled ocean-atmosphere model. J. Climate. 6,				
17	1993–2011.				
18	Elsner, J. B., and A. A. Tsonis, 2006: Singular Spectrum Analysis: A New Tool in				
19	Time Series Analysis. Springer US, 164 pp., doi:10.1007/978-1-4757-2514-8.				
20	Enfield, D. B., A. M. Mestas-Nuñez, and P. J. Trimble, 2001: The Atlantic				
21	multidecadal oscillation and its relation to rainfall and river flows in the				
22	continental U. S., Geophys. Res. Lett., 28, 277-280.				
23	Frankcombe, L. M., and H. A. Dijkstra, 2011: The role of Atlantic-Arctic exchange in				
24	North Atlantic multidecadal climate variability. Geophys. Res. Lett., 38,				
25	L16603, doi:10.1029/2011GL048158.				

1	Frankcombe, L. M., H. A. Dijkstra, and A. von der Heydt, 2009: Noise-induced
2	multidecadal variability in the North Atlantic: Excitation of normal modes. J.
3	<i>Phys. Oceanogr.</i> , 39 , 220–233.
4	Frankcombe, L. M., A. von der Heydt, and H. A. Dijkstra, 2010: North Atlantic
5	multidecadal climate variability: An investigation of dominant time scales and
6	processes. J. Climate, 23, 3626–3638.
7	Frankcombe, L. M., M. H. England, M. E. Mann, and B. A. Steinman, 2015:
8	Separating internal variability from the externally forced climate response. J.
9	<i>Climate</i> , 28 , 8184–8202. doi: <u>http://dx.doi.org/10.1175/JCLI-D-15-0069.1</u>
10	Ghil M., and R. Vautard, 1991: Interdecadal oscillations and the warming trend in
11	global temperature time series. Nature, 350, 324-327 (28 March 1991);
12	doi:10.1038/350324a0
13	Ghil M., R. M. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A.
14	Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou, 2002: Advanced
15	spectral methods for climatic time series. Rev. Geophys., 40(1), 1003,
16	doi:10.1029/2000RG000092
17	GISTEMP Team, 2016: GISS Surface Temperature Analysis (GISTEMP). NASA
18	Goddard Institute for Space Studies. Dataset accessed 2016-02-16 at
19	http://data.giss.nasa.gov/gistemp/.
20	Golaz, JC., L. W. Horowitz, and H. Levy II, 2013: Cloud tuning in a coupled climate
21	model: impact on 20th century warming Geophys. Res. Lett., 40, 2246-2251.
22	Groth, A., and M. Ghil, 2015: Monte Carlo SSA revisited: Detecting oscillator
23	clusters in multivariate data sets. J. Climate, 28, 7873-7893. doi:
24	10.1175/JCLI-D-15-0100.1.
25	Groth, A., Y. Feliks, D. Kondrashov and M. Ghil, 2016: Interannual variability in the

1	North Atlantic temperature field and its association with the wind-stress						
2	forcing. J. Climate, submitted.						
3	Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change.						
4	Rev. Geophys., 48, RG4004, doi:10.1029/2010RG000345.						
5	Hasselmann, 1997: Multi-pattern fingerprint method for detection and attribution of						
6	climate change. Climate Dyn., 13, 601-611.						
7	Hegerl, G. C., K. Hasselmann, U. Cubasch, J. F. B. Mitchell, E. Roeckner, R. Voss,						
8	and J. Waszkewitz, 1997: Multi-fingerprint detection and attribution of						
9	greenhouse gas and aerosol forced climate change. Climate Dyn. 13, 613-634.						
10	Huang, N. E., and Z. Wu, 2008: A review on Hilbert-Huang transform: Method and						
11	its applications to geophysical studies. Rev. Geophys., 46, RG2006,						
12	doi:10.1029/2007RG000228.						
13	Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: Regional						
14	temperatures and precipitation. Science, 269, 676-679.						
15	Hurrell, J. W., and C. Deser, 2009: North Atlantic climate variability: The role of the						
16	North Atlantic Oscillation. J. Mar. Syst., 78, 28-41.						
17	IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of						
18	Working Group I to the Fifth Assessment Report of the Intergovernmental						
19	Panel on Climate Change [Stocker, T.F., D. Qin, GK. Plattner, M. Tignor,						
20	S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)].						
21	Cambridge University Press, Cambridge, United Kingdom and New York,						
22	NY, USA, 1535 pp.						
23	Jamison, N., and S. Kravtsov, 2010: Decadal variations of North Atlantic sea-surface						
24	temperature in observations and CMIP3 simulations. J. Climate, 23, 4619-						
25	4636.						

1	Jolliffe, I. T., and D. B. Stephenson, Eds., 2003: Forecast Verification: A
2	Practitioner's Guide in Atmospheric Science. Wiley-Interscience, 254 pp.
3	Jones, G. S., P. A. Stott, and N. Christidis, 2013: Attribution of observed historical
4	near-surface temperature variations to anthropogenic and natural causes using
5	CMIP5 simulations, J. Geophys. Res. Atmos., 118, 4001–4024,
6	doi:10.1002/jgrd.50239.
7	Kaplan, A., M. Cane, Y. Kushnir, A. Clement, M. Blumenthal, and B. Rajagopalan,
8	1998: Analyses of global sea surface temperature 1856–1991. J. Geophys.
9	<i>Res.</i> , 103 , 18,567–18,589.
10	Knight, J. R., R. J. Allan, C. K. Folland, M. Vellinga, and M. E. Mann, 2005: A
11	signature of persistent natural thermohaline circulation cycles in observed
12	climate. Geophys. Res. Lett., 32, L20708, doi:10.1029/2005GL024233.
13	Knight, J. R., C. K. Folland, and A. A. Scaife, 2006: Climate impacts of the Atlantic
14	Multidecadal Oscillation. Geophys. Res. Lett., L17706,
15	doi:10.1029/2006GL026242.
16	Knight, J. R., 2009: The Atlantic Multidecadal Oscillation inferred from the forced
17	climate response in coupled general circulation models. J. Climate, 22, 1610-
18	1625, doi:10.1175/2008JCLI2628.1.
19	Kravtsov, S., D. Kondrashov, and M. Ghil, 2005: Multi-level regression modeling of
20	nonlinear processes: Derivation and applications to climatic variability. J.
21	<i>Climate</i> , 18 , 4404–4424.
22	Kravtsov, S., and C. Spannagle, 2008: Multidecadal climate variability in observed
23	and modeled surface temperatures. J. Climate, 21, 1104–1121. doi:
24	http://dx.doi.org/10.1175/2007JCLI1874.1.
25	Kravtsov, S., M. Ghil, and D. Kondrashov, 2010: Empirical Model Reduction and the

1	Modeling Hierarchy in Climate Dynamics and the Geosciences. In: Stochastic						
2	Physics and Climate Modeling, T. Palmer and P. Williams, Eds., Cambridge						
3	University Press, pp. 35–72.						
4	Kravtsov, S., M. G. Wyatt, J. A. Curry, and A. A. Tsonis, 2014: Two contrasting						
5	views of multidecadal climate variability in the twentieth century. Geophys.						
6	Res. Lett., 41, 1326, 6881-6888, doi: 10.1002/2014GL061416.						
7	Kravtsov, S., M. G. Wyatt, J. A. Curry, and A. A. Tsonis, 2015: Comment on						
8	"Atlantic and Pacific multidecadal oscillations and Northern Hemisphere						
9	temperatures." Science, 350 , 1326, doi: 10.1126/science.aab3570.						
10	Kushnir, Y., W. A. Robinson, I. Bladé, N. M. J. Hall, S. Peng, and R. Sutton, 2002:						
11	Atmospheric GCM response to extratropical SST anomalies: Synthesis and						
12	evaluation. J. Climate, 15, 2233-2256, doi:10.1175/1520-						
13	0442(2002)015<2233:AGRTES>2.0.CO;2.						
14	Latif, M., E. Roeckner, M. Botzet, et al., 2004: Reconstructing, monitoring, and						
15	predicting multidecadal-scale changes in the North Atlantic thermohaline						
16	circulation with sea surface temperature. J. Climate, 17, 1605–1614.						
17	Mann, M. E., 2008: Smoothing of climate time series revisited. Geophys. Res. Lett.,						
18	35 , L16708, doi:10.1029/2008GL034716.						
19	Mann, M.E., and K. A. Emanuel, 2006. Atlantic hurricane trends linked to climate						
20	change. EOS: Transactions of the American Geophysical Union, 87, 233–244.						
21	Mann, M. E., and J. Park, 1994: Global scale modes of surface temperature variability						
22	on interannual to century time scales, J. Geophys. Res., 99, 25819–25833.						
23	Mann, M. E., and J. Park, 1999: Oscillatory spatiotemporal signal detection in climate						
24	studies: A multiple-taper spectral domain approach. Advances Geophys., 41,						
25	1–131, 1999.						

1	Mann, M. E., B. A., Steinman and S. K. Miller, 2014: On forced temperature changes
2	internal variability, and the AMO. Geophys. Res. Lett. 41, 3211-3219.
3	Mann, M. E., S. Rahmstorf, B. A. Steinman, M. Tingley and S. K. Miller, 2016: The
4	likelihood of recent record warmth. Scientific Reports, 6, Article
5	number:19831, doi:10.1038/srep19831.
6	Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A
7	Pacific interdecadal climate oscillation with impacts on salmon production.
8	Bull. Amer. Meteor. Soc., 78, 1069-1079, doi: 10.1175/1520-
9	0477(1997)078<1069:APICOW>2.0.CO;2.
10	Monahan, A., J. C. Fyfe, M. H. Ambaum, D. B. Stephenson, and G. R. North, 2009:
11	Empirical orthogonal functions: The medium is the message. J. Climate, 22,
12	6501–6514.
13	Moron, V., R. Vautard, and M. Ghil, 1998: Trends, interdecadal and interannual
14	oscillations in global sea-surface temperatures. Climate Dyn., 14, 545-569.
15	Parker, D. E., P. D. Jones, C. K. Folland, and A. Bevan, 1994: Interdecadal changes
16	of surface temperature since the late nineteenth century. J. Geophys. Res., 99,
17	14,373–14,399.
18	Preisendorfer, R. W., 1988: Principal Component Analysis in Meteorology and
19	Oceanography. Elsevier, New York, 425 pp.
20	Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V, Rowell,
21	D.P., Kent, E.C. and Kaplan, A., 2003: Globally complete analyses of sea
22	surface temperature, sea ice and night marine air temperature, 1871–2000. J.
23	Geophys. Res., 108, 4407, doi:10.1029/2002JD002670.
24	Reynolds, R. W., and T. M. Smith, 1994: Improved global sea-surface temperature
25	analysis using optimum interpolation. J. Climate, 7, 929–948.

1	Schlesinger, M. E., and N. Ramankutty, 1994: An oscillation in the global climate
2	system of period 65-70 years. Nature, 367, 723-726.
3	Schneider, T., and I. M. Held, 2001: Discriminants of twentieth-century changes in
4	earth surface temperatures. J. Climate., 14, 249–254.
5	Smith, T.M., R.W. Reynolds, T.C. Peterson, and J. Lawrimore, 2008: Improvements
6	to NOAA's historical merged land-ocean surface temperature analysis (1880-
7	2006). J. Climate, 21, 2283-2296, doi:
8	http://dx.doi.org/10.1175/2007JCLI2100.1
9	Sutton, R. T., and D. L. R. Hodson, 2005: Atlantic Ocean forcing of North American
10	and European summer climate. Science, 290, 2133–2137.
11	Steinman, B. A., M. E. Mann and S. K. Miller, 2015a: Atlantic and Pacific
12	multidecadal oscillations and Northern Hemisphere temperatures. Science,
13	347 , 988, doi: 10.1126/science.1257856.
14	Steinman, B. A., L. M. Frankcombe, M. E., Mann, S. K. Miller, and M. H. England,
15	2015b: Response to comment on "Atlantic and Pacific multidecadal
16	oscillations and Northern Hemisphere temperatures." Science, 350, 1326, doi:
17	10.1126/science.aac5208.
18	Stevens, B., 2015: Rethinking the lower bound on aerosol radiative forcing. J.
19	Climate, 28, 4794-4819, doi:10.1175/JCLI-D-14-00656.1
20	Taylor, K. E., R. J. Stouffer and G. A. Meehl, 2012: An Overview of CMIP5 and the
21	experiment design. Bull. Am. Meteorol. Soc., 93, 485-498.
22	Terray, L., 2012: Evidence for multiple drivers of North Atlantic multi-decadal
23	climate variability. Geophys. Res. Lett. 39, L19712,
24	doi:10.1029/2012GL053046.
25	Tett, S. F. B., P.A. Stott, M. A. Allen, W. J. Ingram, and J. F. B. Mitchell, 1999:

1	Causes of twentieth century temperature change. Nature, 399, 569-572.						
2	Timmermann, A., M. Latif, R. Voss, and A. Grötzner, 1998: Northern hemispheric						
3	interdecadal variability: A coupled air-sea mode. J. Climate, 11, 1906-1931,						
4	doi:10.1175/1520-0442(1998)011(1906:NHIVAC)2.0.CO;2						
5	Ting, M., Y. Kushnir, R. Seager, and C. Li, 2009: Forced and internal twentieth-						
6	century SST in the North Atlantic. J. Climate, 22, 1469–1481.						
7	Trenary, L., and T. DelSole, 2016: Does the Atlantic Multidecadal Oscillation get its						
8	predictability from the Atlantic Meridional Overturning Circulation. J.						
9	Climate, submitted.						
10	Wu, Z., N. E. Huang, J. M. Wallace, B. Smoliak, and X. Chen, 2011: On the time-						
11	varying trend in global-mean surface temperature. Climate Dyn., 37, 759-773,						
12	doi:10.1007/s00382-011- 1128-8.						
13	Wyatt, M. G., and J. A. Curry, 2014: Role for Eurasian Arctic shelf sea ice in a						
14	secularly varying hemispheric climate signal during the 20th century. Climate						
15	<i>Dyn.</i> , 42 , 2763–2782, doi:10.1007/s00382-013-1950-2.						
16	Wyatt, M. G., and J. M. Peters, 2012: A secularly varying hemispheric climate-signal						
17	propagation previously detected in instrumental and proxy data not detected in						
18	CMIP3 data base. SpringerPlus, 1, 68, doi:10.1186/2193-1801-1-68.						
19	Wyatt, M., S. Kravtsov, and A. A. Tsonis, 2012: Atlantic Multidecadal Oscillation						
20	and Northern Hemisphere's climate variability. Climate Dyn., 38, 929–949,						
21	doi:10.1007/s00382-011-1071-8.						
22	Xue, Y., T. M. Smith, and R. W. Reynolds, 2003: Interdecadal changes of 30-yr SST						
23	normals during 1871–2000. J. Climate, 16, 1601–1612.						
24	Zhang, R., Delworth, T. L., and I. M. Held, 2007. Can the Atlantic Ocean drive the						
25	observed multidecadal variability in Northern Hemisphere mean temperature?						

1	Geophys. Res. Letts., 34, L02709, doi:10.1029/2006GL028683.
2	Zhang, L., and C. Wang, 2013: Multidecadal North Atlantic sea surface temperature
3	and Atlantic Meridional Overturning Circulation variability in CMIP5
4	historical simulations. J. Geophys. Res. Oceans, 118 (10), 5772-5791,
5	doi:10.1002/jgrc.20390.
6	Zhang, Y., J. M. Wallace, and D. S. Battisti, 1997: ENSO-like interdecadal
7	variability: 1900–93. J. Climate, 10, 1004–1020.
8	Zhang, R., T. L. Delworth, R. Sutton, et al., 2013: Have aerosols caused the observed
9	Atlantic Multidecadal Variability? J. Atmos. Sci., 70, 1135–1144. doi:
10	http://dx.doi.org/10.1175/JAS-D-12-0331.1.
11	

1 Table captions

2 Table 1. CMIP5 twentieth century simulations with four or more realizations, 3 resulting in 18 independent ensembles with the total of 116 simulations. Models with acronyms in bold have the aerosol indirect effects (cloud 4 5 albedo+lifetime) included. We also considered a smaller subset of models, 6 which excluded the models marked by the asterisk. For the MRI-CGCM3 7 model, the latter subset had three of the four original simulations. In total, the 8 second subset consisted of 15 models and 89 simulations. The last six columns 9 list relative model sensitivities in terms of the scaling factors obtained via 10 regression of the multi-model ensemble mean (MMEM) time series against the single-model ensemble mean (SMEM) of the individual models, as well as via 11 12 regression of SMEMs against the observed time series, for the AMO, PMO 13 and HMO climate indices (see text). The numbers in parentheses show the 14 MMEM scaling factors based on the observed time series in the first case, and 15 the multi-model average of the scaling factors in the second.

1 Figure captions

2 Figure 1: Full observed time series and its estimated forced component for the AMO 3 (a, b), PMO (c, d) and HMO (e, f). Left panels show the 1861-2005 multimodel ensemble mean (MMEM) time series (solid colored lines) linearly 4 5 extrapolated through 2012 (dotted lines), as well as its rescaled version (black) 6 that best matches the observed time series (magenta). Thin gray lines plot the 7 5-yr low-pass filtered single-model ensemble mean (SMEM) time series for each of the 18 models considered. In the right panels, gray lines represent 8 9 these SMEM time series individually rescaled to match observations, solid 10 colored lines — their multi-model average, and dotted lines — a linear 11 extrapolation of the resulting estimated forced signal through 2012.

12 Figure 2: Performance of the MMEM-regression (red) and SMEM-subtraction (blue) 13 attribution methods in recovering the known intrinsic variability (black) in 100 14 surrogate multi-model AMO data sets (see text). (a) Variance of actual and inferred intrinsic signals; (b) error variance (variance of the difference 15 16 between inferred and actual intrinsic time series); (c) variance ratio [ratio of 17 the black to the blue and black to the red lines in (a)]; (d) relative error 18 variance [error variance in (b) divided by the actual variance (black lines) in 19 (a) times 100%]; (e) correlation between the actual and inferred intrinsic 20 signals. All of these characteristics were computed for raw and boxcar 21 running-mean low-pass filtered time series using different window sizes of 22 $2 \times K + 1$ yr, K = 0, 1, ..., 30 (shown on the horizontal axis); K = 0corresponds to raw annual data, K = 1 — to 3-yr low-pass filtered data and so 23

2

on. Error bars, where present, show the 95% spread of the quantity displayed across all of the surrogate simulations considered.

3 Figure 3: Decomposition of errors in the MMEM regression (red) and SMEM 4 subtraction (blue) attribution methods, for raw and low-pass filtered time 5 series, based on the same data as in Fig. 2. The horizontal axis related to the running-mean smoother window size is also the same as in Fig. 2. "Model" 6 7 errors (x-symbols) are associated with differences between the actual and 8 inferred forced signals that are present in all of the 100 surrogate multi-model 9 data sets; they were determined by averaging the differences between the inferred and actual forced signals over the 100 multi-model ensembles 10 11 considered, and then computing the root-mean-square error for the resulting 12 multi-model difference time series. "Intrinsic" errors (+-symbols) arise due to 13 leaking a fraction of the intrinsic variance to the forced signal estimate due to 14 insufficient cancellation of independent realizations of intrinsic variability in 15 the MMEM or SMEM ensemble means. The total attribution error (solid lines) 16 is the root-mean-square sum of the model and intrinsic errors.

17 Figure 4: Spaghetti plots of individual-model forced-signal estimates (gray), as well 18 as their grand ensemble-mean (red) for the MMEM and SMEM attribution 19 methods applied to the surrogate AMO data of Figs. 2 and 3. (a) Forced 20 signals defined via 5-yr low-pass filtered SMEM of individual models; (b) 21 forced signals computed as MMEM of bootstrap subsamples of surrogate 22 model simulations; (c) forced signals obtained via rescaling of known 23 synthetic raw forced signals to best match their grand ensemble-mean [red line 24 in (a)], and then adding a randomly chosen time series pulled from an

2

ensemble of the difference time series between the actual and SMEM-inferred forced signals; (d) the same as in (c), but for the MMEM regression method.

3 Figure 5: Estimated forced signals and their uncertainties for the AMO (a-c), PMO 4 (d-f) and HMO (g-i) time series obtained via SMEM-based Monte Carlo 5 method applied to the multi-model ensemble of twentieth-century simulations 6 in Table 1 (see text for details). Magenta lines in all panels show the observed 7 time series. Left column: 1861-2005 5-yr low-pass filtered SMEM based forced signals linearly extrapolated through 2012 using 1986-2006 trend 8 9 slopes (gray lines), their ensemble mean (solid colored lines) and the 10 associated 95% confidence interval (dashed colored lines). Middle column: grav lines show forced signals obtained via rescaling of each individual 11 12 SMEM forced-signal estimate [that is, the signals shown by gray lines in the 13 corresponding left panel] to best match the corresponding observed (magenta) 14 time series, and then adding a randomly chosen time series pulled from an 15 ensemble of the difference time series between the known actual and inferred raw (unscaled) forced signals. The solid and dashed colored lines show the 16 17 ensemble-mean and the 95% spread of the individual forced-signal estimates. 18 Right column: the same as in the middle column, but the rescaling is 19 performed first over the 1861–2005 period, and the resulting rescaled signals 20 are then linearly extrapolated through 2012 using 1986–2006 trend slopes.

Figure 6: End effects of the data-adaptive 40-yr low-pass filter. (a) Gray lines show the 40-yr low-pass filtered time series of synthetic stationary intrinsic AMO signals derived for the multi-model ensemble of Table 1; thick blue lines display these signals' ensemble mean and 95% spread. Note the enhanced spread at both ends of the time series. (b) The tapers derived to alleviate the
end effects in (a) for the AMO, PMO and HMO. The tapers were constructed
so that the low-pass filtered intrinsic AMO, PMO and HMO signals multiplied
by their respective tapers result in the uniform standard deviation of the lowfrequency intrinsic surrogates throughout the entire length of the data record.

Figure 7: Estimates of observed multidecadal intrinsic variability for AMO (a, b), 6 7 PMO (c, d) and HMO (e, f). These estimates were obtained using the two Monte Carlo ensembles of rescaled forced signals in Fig. 5: namely, those in 8 9 Figs. 5b, e, h for panels (a), (c), (e) here, and those in Figs. 5c, f, I for panels 10 (b), (d), (f) here. Rescaled forced signals were subtracted from the corresponding observed time series, 40-vr low-pass filtered and windowed 11 12 using the tapers in Fig. 6. Heavy solid colored lines (AMO: blue, PMO: green, 13 and HMO: red) show the ensemble mean of the resulting intrinsic signal 14 estimates, and error bars — their 95% spread. Each panel also contains for reference the "intrinsic" estimates based on subtracting linear trend from the 15 entire observed time series, as well as the one based on the piecewise linear 16 17 detrending of the observed time series with the break point at 1900.

Figure 8: Spectra of the observed (blue) and CMIP5 simulated (red) intrinsic variability in the AMO (b), PMO (c) and HMO (d) indices. The spectra were computed in terms of the standard deviation (STD) of boxcar low-pass filtered data as a function of the filter's size, the latter shown on the abscissa of each panel (same conventions as in Fig. 2). The input data for the observed intrinsic signals were the same as in Fig. 7. The simulated intrinsic time series were obtained by subtracting the 5-yr low-pass filtered SMEM from individual

1 model simulations (Table 1). To account for the spectral leakage of the 2 intrinsic variance associated with the SMEM subtraction procedure, we 3 computed the STD inflation factors as the ratio of the known actual to the SMEM-inferred standard deviations of the intrinsic variability in the Monte 4 Carlo simulations of each climate index. The filter-size dependent inflation 5 6 factors shown in panel (a) multiplied the standard deviations of the 116 raw 7 CMIP5 intrinsic signals; blue lines in (b-d) show these inflated spectra. The 8 heavy lines in (b–d) show the ensemble-mean, and error bars — the associated 9 95% spread of the individual spectra.

10 Figure 9: Forced and intrinsic variability in the observed and simulated PDO (a-c) 11 and NAO (d-f). The input data used the reduced CMIP5 model ensemble from 12 Table 1 (15 models and 89 simulations), as well as the observations based on 13 ERSST sea-surface temperature and 20CR sea-level pressure. (a, d) Forced 14 signal estimates analogous to those in Figs. 5b, e, h. (b, e) Multidecadal 15 intrinsic variability estimates analogous to those in Figs. 7a, c, e. (c, f) Spectra of the observed and simulated intrinsic variability analogous to those in Figs. 16 17 8b-d. The NAO spectrum panel (f) also includes the spectrum based on an 18 alternative, station based NAO index (https://climatedataguide.ucar.edu 19 /climate-data/ hurrell-north- atlantic-oscillation-nao-index-station-based).

Figure 10: Lagged correlations between the multidecadal intrinsic signals associated with the AMO, PMO, PDO and NAO climate indices (Figs. 7a, c, 9b, e). The semi-empirical estimates of these intrinsic signals were obtained using the same input data as in Fig. 9. Gray lines show lagged correlations between individual Monte Carlo estimates of the intrinsic signals considered, heavy blue lines — the ensemble-mean correlation and the dots — 95% spread of the lagged correlations, all as a function of time lag (yr). The pairs of indices considered in each panel are listed in that panel caption; negative lags correspond to the first member of the index pair leading the second member.

1

2

3

Figure 11: The same as in Fig. 10, but for the 89 simulated intrinsic multidecadal
signals obtained by subtracting the 5-yr low-pass filtered SMEM from the
individual model simulations in the reduced 15-model subset of Table 1.

7 Figure 12: The standard deviation Σ of the MMEM of 'intrinsic' residual time series 8 estimated using the MMEM regression and smoothed SMEM subtraction 9 methods. The input synthetic AMO data were constructed by combining the 10 18 different forced signals and 1000 surrogate intrinsic samples based on the 11 CMIP5 simulations of Table 1 (see text for details). (a) Observed (blue line) 12 and predicted (red line) Σ for a version of MMEM-regression attribution 13 which regresses out, from each simulated AMO ensemble-member time series, 14 the MMEM computed over all other ensemble members; black line shows 15 expected Σ for independent residuals. All Σ estimates are shown as functions 16 of the total number N of simulations in a multi-model ensemble. Error bars 17 show the 95% spread of Σ based on 1000 independent synthetic multi-model 18 ensembles. (b) The estimates of Σ for the true (synthetic) intrinsic signals 19 (black), for intrinsic signals defined using 5-yr low-pass filtered SMEM 20 subtraction (blue), and for those computed using a modified version of the 21 MMEM-regression method which leaves out the individual-model ensembles 22 when estimating the MMEM (see text for details), all as a function of N. Error 23 bars show the 95% spread of Σ based on 100 independent synthetic multi-24 model ensembles.

1 Figure 13: The MMEM of the 'intrinsic' residual AMO time series for different 2 number N of simulations (thin colored lines), using the same synthetic data set 3 as in Fig. 12. The intrinsic residuals were computed using a modified version 4 of the MMEM-regression method that leaves out the individual-model ensembles when estimating the MMEM (see Fig. 12b and text for details). For 5 6 reference, a scaled version of the synthetic data set AMO's MMEM is also 7 shown (heavy black line). The MMEMs of the 'intrinsic' residual AMO time 8 series for different N are almost perfectly correlated; their correlation with the 9 raw AMO's MMEM is ~0.97.

10

1 **Table 1.** CMIP5 twentieth century simulations with four or more realizations,

2 resulting in 18 independent ensembles with the total of 116 simulations. Models with

3 acronyms in bold have the aerosol indirect effects (cloud albedo+lifetime) included.

4 We also considered a smaller subset of models, which excluded the models marked by

- 5 the asterisk. For the MRI-CGCM3 model, the latter subset had three of the four
- 6 original simulations. In total, the second subset consisted of 15 models and 89
- 7 simulations. The last six columns list relative model sensitivities in terms of the
- 8 scaling factors obtained via regression of the multi-model ensemble mean (MMEM)
- 9 time series against the single-model ensemble mean (SMEM) of the individual 10 models, as well as via regression of SMEMs against the observed time series, for the
- 10 models, as well as via regression of SMEMs against the observed time series, for the 11 AMO, PMO and HMO climate indices (see text). The numbers in parentheses show
- 12 the MMEM scaling factors based on the observed time series in the first case, and the
- 12 methyly scaling factors based on the observed time series in the first ca 13 multi-model average of the scaling factors in the second.

#	Model	Number of	Scaling wrt MMEM			Scaling wrt observed		
	acronym	realizations	AMO	PMO	HMO	AMO	PMO	HMO
			(0.89)	(0.58)	(1.05)	(0.8)	(0.57)	(1.02)
1.	CCSM4	6	1.18	1.50	1.31	0.68	0.47	0.78
2.	CNRM-CM5	10	0.68	0.86	0.84	1.15	0.76	1.22
3.	CSIRO- Mk3-6-0	10	0.68	0.64	0.72	1.10	0.57	1.21
4.	CanESM2	5	0.97	1.02	1.00	0.81	0.53	0.95
5.	GFDL- CM2p1*	10	1.26	1.41	1.31	0.61	0.48	0.78
6.	GFDL-CM3	5	0.67	0.60	0.79	0.80	0.21	0.91
7.	GISS-E2-Hp1	6	1.0	0.95	1.00	0.82	0.70	1.04
8.	GISS-E2-Hp2	5	0.83	0.82	0.84	1.03	0.72	1.21
9.	GISS-E2-Hp3	6	1.23	1.17	1.15	0.72	0.61	0.92
10.	GISS-E2-Rp1	6	1.03	0.82	0.93	0.8	0.70	1.11
11.	GISS-E2-Rp2	6	0.76	0.67	0.73	1.1	0.71	1.30
12.	GISS-E2-Rp3	6	1.53	0.93	1.11	0.48	0.64	0.93
13.	GISS-E2-Rp4*	6	1.46	1.42	1.42	0.56	0.50	0.71
14.	HadCM3*	10	0.66	1.09	0.90	0.83	0.57	1.11
15.	HadGEM2- ES	5	0.86	0.34	0.64	0.84	0.33	1.21
16.	IPSL-CM5A- LR	6	1.66	1.72	1.44	0.48	0.42	0.72
17.	MIROC5	4	0.92	0.67	0.77	0.93	0.64	1.13
18.	MRI- CGCM3*	4 (3)	0.75	0.73	0.73	0.77	0.70	1.20

14



1 Figure 1: Full observed time series and its estimated forced component for the AMO 2 (a, b), PMO (c, d) and HMO (e, f). Left panels show the 1861-2005 multi-3 model ensemble mean (MMEM) time series (solid colored lines) linearly 4 extrapolated through 2012 (dotted lines), as well as its rescaled version (black) 5 that best matches the observed time series (magenta). Thin gray lines plot the 5-yr low-pass filtered single-model ensemble mean (SMEM) time series for 6 7 each of the 18 models considered. In the right panels, gray lines represent 8 these SMEM time series individually rescaled to match observations, solid 9 colored lines — their multi-model average, and dotted lines — a linear 10 extrapolation of the resulting estimated forced signal through 2012. 11



2 Figure 2: Performance of the MMEM-regression (red) and SMEM-subtraction (blue) 3 attribution methods in recovering the known intrinsic variability (black) in 100 4 surrogate multi-model AMO data sets (see text). (a) Variance of actual and 5 inferred intrinsic signals; (b) error variance (variance of the difference 6 between inferred and actual intrinsic time series); (c) variance ratio [ratio of 7 the black to the blue and black to the red lines in (a)]; (d) relative error 8 variance [error variance in (b) divided by the actual variance (black lines) in 9 (a) times 100%]; (e) correlation between the actual and inferred intrinsic 10 signals. All of these characteristics were computed for raw and boxcar 11 running-mean low-pass filtered time series using different window sizes of 12 $2 \times K + 1$ yr, K = 0, 1, ..., 30 (shown on the horizontal axis); K = 013 corresponds to raw annual data, K = 1 — to 3-yr low-pass filtered data and so 14 on. Error bars, where present, show the 95% spread of the quantity displayed 15 across all of the surrogate simulations considered.





3 Figure 3: Decomposition of errors in the MMEM regression (red) and SMEM 4 subtraction (blue) attribution methods, for raw and low-pass filtered time 5 series, based on the same data as in Fig. 2. The horizontal axis related to the 6 running-mean smoother window size is also the same as in Fig. 2. "Model" 7 errors (x-symbols) are associated with differences between the actual and 8 inferred forced signals that are present in all of the 100 surrogate multi-model 9 data sets; they were determined by averaging the differences between the inferred and actual forced signals over the 100 multi-model ensembles 10 considered, and then computing the root-mean-square error for the resulting 11 12 multi-model difference time series. "Intrinsic" errors (+-symbols) arise due to 13 leaking a fraction of the intrinsic variance to the forced signal estimate due to insufficient cancellation of independent realizations of intrinsic variability in 14 15 the MMEM or SMEM ensemble means. The total attribution error (solid lines) 16 is the root-mean-square sum of the model and intrinsic errors. 17



1 Figure 4: Spaghetti plots of individual-model forced-signal estimates (gray), as well 2 as their grand ensemble-mean (red) for the MMEM and SMEM attribution 3 methods applied to the surrogate AMO data of Figs. 2 and 3. (a) Forced 4 signals defined via 5-yr low-pass filtered SMEM of individual models; (b) 5 forced signals computed as MMEM of bootstrap subsamples of surrogate model simulations; (c) forced signals obtained via rescaling of known 6 7 synthetic raw forced signals to best match their grand ensemble-mean [red line 8 in (a)], and then adding a randomly chosen time series pulled from an 9 ensemble of the difference time series between the actual and SMEM-inferred 10 forced signals; (d) the same as in (c), but for the MMEM regression method. 11



2 Figure 5: Estimated forced signals and their uncertainties for the AMO (a-c), PMO 3 (d-f) and HMO (g-i) time series obtained via SMEM-based Monte Carlo 4 method applied to the multi-model ensemble of twentieth-century simulations 5 in Table 1 (see text for details). Magenta lines in all panels show the observed time series. Left column: 1861-2005 5-yr low-pass filtered SMEM based 6 7 forced signals linearly extrapolated through 2012 using 1986-2006 trend 8 slopes (gray lines), their ensemble mean (solid colored lines) and the 9 associated 95% confidence interval (dashed colored lines). Middle column: 10 gray lines show forced signals obtained via rescaling of each individual SMEM forced-signal estimate [that is, the signals shown by gray lines in the 11 12 corresponding left panel] to best match the corresponding observed (magenta) 13 time series, and then adding a randomly chosen time series pulled from an 14 ensemble of the difference time series between the known actual and inferred 15 raw (unscaled) forced signals. The solid and dashed colored lines show the ensemble-mean and the 95% spread of the individual forced-signal estimates. 16 17 Right column: the same as in the middle column, but the rescaling is 18 performed first over the 1861–2005 period, and the resulting rescaled signals 19 are then linearly extrapolated through 2012 using 1986–2006 trend slopes.



1 Figure 6: End effects of the data-adaptive 40-yr low-pass filter. (a) Gray lines show 2 the 40-yr low-pass filtered time series of synthetic stationary intrinsic AMO 3 signals derived for the multi-model ensemble of Table 1; thick blue lines 4 display these signals' ensemble mean and 95% spread. Note the enhanced 5 spread at both ends of the time series. (b) The tapers derived to alleviate the 6 end effects in (a) for the AMO, PMO and HMO. The tapers were constructed 7 so that the low-pass filtered intrinsic AMO, PMO and HMO signals multiplied 8 by their respective tapers result in the uniform standard deviation of the low-9 frequency intrinsic surrogates throughout the entire length of the data record.

11



1 Figure 7: Estimates of observed multidecadal intrinsic variability for AMO (a, b), 2 PMO (c, d) and HMO (e, f). These estimates were obtained using the two 3 Monte Carlo ensembles of rescaled forced signals in Fig. 5: namely, those in 4 Figs. 5b, e, h for panels (a), (c), (e) here, and those in Figs. 5c, f, I for panels 5 (b), (d), (f) here. Rescaled forced signals were subtracted from the corresponding observed time series, 40-yr low-pass filtered and windowed 6 7 using the tapers in Fig. 6. Heavy solid colored lines (AMO: blue, PMO: green, and HMO: red) show the ensemble mean of the resulting intrinsic signal 8 9 estimates, and error bars — their 95% spread. Each panel also contains for 10 reference the "intrinsic" estimates based on subtracting linear trend from the 11 entire observed time series, as well as the one based on the piecewise linear 12 detrending of the observed time series with the break point at 1900.



2 Figure 8: Spectra of the observed (blue) and CMIP5 simulated (red) intrinsic 3 variability in the AMO (b), PMO (c) and HMO (d) indices. The spectra were 4 computed in terms of the standard deviation (STD) of boxcar low-pass filtered 5 data as a function of the filter's size, the latter shown on the abscissa of each 6 panel (same conventions as in Fig. 2). The input data for the observed intrinsic 7 signals were the same as in Fig. 7. The simulated intrinsic time series were 8 obtained by subtracting the 5-yr low-pass filtered SMEM from individual 9 model simulations (Table 1). To account for the spectral leakage of the 10 intrinsic variance associated with the SMEM subtraction procedure, we 11 computed the STD inflation factors as the ratio of the known actual to the 12 SMEM-inferred standard deviations of the intrinsic variability in the Monte 13 Carlo simulations of each climate index. The filter-size dependent inflation 14 factors shown in panel (a) multiplied the standard deviations of the 116 raw 15 CMIP5 intrinsic signals; blue lines in (b-d) show these inflated spectra. The 16 heavy lines in (b–d) show the ensemble-mean, and error bars — the associated 17 95% spread of the individual spectra.



Figure 9: Forced and intrinsic variability in the observed and simulated PDO (a-c) 1 2 and NAO (d-f). The input data used the reduced CMIP5 model ensemble from 3 Table 1 (15 models and 89 simulations), as well as the observations based on ERSST sea-surface temperature and 20CR sea-level pressure. (a, d) Forced 4 5 signal estimates analogous to those in Figs. 5b, e, h. (b, e) Multidecadal intrinsic variability estimates analogous to those in Figs. 7a, c, e. (c, f) Spectra 6 7 of the observed and simulated intrinsic variability analogous to those in Figs. 8 8b-d. The NAO spectrum panel (f) also includes the spectrum based on an 9 alternative, station based NAO index (https://climatedataguide.ucar.edu 10 /climate-data/ hurrell-north- atlantic-oscillation-nao-index-station-based).



1 2 with the AMO, PMO, PDO and NAO climate indices (Figs. 7a, c, 9b, e). The 3 semi-empirical estimates of these intrinsic signals were obtained using the 4 same input data as in Fig. 9. Gray lines show lagged correlations between 5 individual Monte Carlo estimates of the intrinsic signals considered, heavy blue lines — the ensemble-mean correlation and the dots — 95% spread of the 6 7 lagged correlations, all as a function of time lag (yr). The pairs of indices 8 considered in each panel are listed in that panel caption; negative lags 9 correspond to the first member of the index pair leading the second member.



Figure 11: The same as in Fig. 10, but for the 89 simulated intrinsic multidecadal signals obtained by subtracting the 5-yr low-pass filtered SMEM from the individual model simulations in the reduced 15-model subset of Table 1.



1 Figure 12: The standard deviation Σ of the MMEM of 'intrinsic' residual time series 2 estimated using the MMEM regression and smoothed SMEM subtraction 3 methods. The input synthetic AMO data were constructed by combining the 4 18 different forced signals and 1000 surrogate intrinsic samples based on the 5 CMIP5 simulations of Table 1 (see text for details). (a) Observed (blue line) 6 and predicted (red line) Σ for a version of MMEM-regression attribution 7 which regresses out, from each simulated AMO ensemble-member time series, 8 the MMEM computed over all other ensemble members; black line shows 9 expected σ for independent residuals. All Σ estimates are shown as functions 10 of the total number N of simulations in a multi-model ensemble. Error bars 11 show the 95% spread of Σ based on 1000 independent synthetic multi-model 12 ensembles. (b) The estimates of Σ for the true (synthetic) intrinsic signals 13 (black), for intrinsic signals defined using 5-yr low-pass filtered SMEM 14 subtraction (blue), and for those computed using a modified version of the 15 MMEM-regression method which leaves out the individual-model ensembles 16 when estimating the MMEM (see text for details), all as a function of N. Error 17 bars show the 95% spread of Σ based on 100 independent synthetic multi-18 model ensembles.



2 Figure 13: The MMEM of the 'intrinsic' residual AMO time series for different 3 number N of simulations (thin colored lines), using the same synthetic data set 4 as in Fig. 12. The intrinsic residuals were computed using a modified version 5 of the MMEM-regression method that leaves out the individual-model ensembles when estimating the MMEM (see Fig. 12b and text for details). For 6 7 reference, a scaled version of the synthetic data set AMO's MMEM is also 8 shown (heavy black line). The MMEMs of the 'intrinsic' residual AMO time 9 series for different N are almost perfectly correlated; their correlation with the 10 raw AMO's MMEM is ~0.97.