Geophysical Research Letters

Supporting Information for

Global mean surface temperature response to large-scale patterns of variability in observations and CMIP5

Jules B. Kajtar[1], Matthew Collins[1], Leela. M. Frankcombe[2,3], Matthew H. England[2,3], Timothy J. Osborn[4], Marcus Juniper[1]

[1] College of Engineering, Mathematics, and Physical Sciences, University of Exeter, Exeter, UK.
[2] Australian Research Council's Centre of Excellence for Climate Extremes, Australia.
[3] Climate Change Research Centre, University of New South Wales, NSW, Australia.
[4] Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK.
Corresponding author: Jules B. Kajtar (j.kajtar@exeter.ac.uk)

Contents of this file

Introduction

This document contains the list of CMIP5 models and ensemble members analyzed in this study (Table S1). Additional details regarding the forced response removal method (Text S1), constructing running trend time-series (Text S2), and computing the statistical significance of correlations are also provided. Supporting figures follow the text.

Model	Ensemble Members	Institution
ACCESS1-0	**r1i1p1**	Commonwealth Scientific and Industrial Research Organisation, and Bureau of Meteorology, Australia
ACCESS1-3	**r1i1p1**	
bcc-csm1-1	**r1i1p1**	Beijing Climate Center, China Meteorological Administration
bcc-csm1-1-m	**r1i1p1**	
CanESM2	**r1i1p1**, r2i1p1, r3i1p1, r4i1p1, r5i1p1	Canadian Centre for Climate Modelling and Analysis
CCSM4	**r1i1p1**, **r2i1p1**, r3i1p1, r4i1p1, r5i1p1, r6i1p1	National Center for Atmospheric Research
CESM1-BGC	**r1i1p1**	National Science Foundation, Department of Energy, and National Center for Atmospheric Research
CESM1-CAM5	**r1i1p1**, r2i1p1, r3i1p1	
CMCC-CESM	**r1i1p1**	Centro Euro-Mediterraneo per I Cambiamenti Climatici
CMCC-CM	**r1i1p1**	
CMCC-CMS	**r1i1p1**	
CNRM-CM5	**r1i1p1**, r2i1p1, r4i1p1, r6i1p1, r10i1p1	Centre National de Recherches Meteorologiques
CSIRO-Mk3-6-0	**r1i1p1**, r2i1p1, r3i1p1, r4i1p1, r5i1p1, r6i1p1, r7i1p1, r8i1p1, r9i1p1, r10i1p1	Commonwealth Scientific and Industrial Research Organisation
EC-EARTH	r1i1p1, r2i1p1, r6i1p1, r8i1p1, r9i1p1, r10i1p1	EC-EARTH consortium
FIO-ESM	**r1i1p1**, r2i1p1, r3i1p1	The First Institute of Oceanography, SOA, China
GFDL-CM3	**r1i1p1**	Geophysical Fluid Dynamics Laboratory
GFDL-ESM2G	**r1i1p1**	
GFDL-ESM2M	**r1i1p1**	
GISS-E2-H	**r1i1p1**, **r1i1p2**, **r1i1p3**, r2i1p3	NASA Goddard Institute for Space Studies
GISS-E2-H-CC	**r1i1p1**	
GISS-E2-R	**r1i1p1**, **r1i1p2**, r2i1p1, r2i1p3	
GISS-E2-R-CC	**r1i1p1**	
HadGEM2-AO	**r1i1p1**	NIMR / Korea Meteorological Administration
HadGEM2-CC	**r1i1p1**	Met Office Hadley Centre
HadGEM2-ES	**r1i1p1**, r2i1p1, r3i1p1, r4i1p1	
inmcm4	r1i1p1	Institute for Numerical Mathematics
IPSL-CM5A-LR	**r1i1p1**, r2i1p1, r3i1p1, r4i1p1	Institut Pierre-Simon Laplace
IPSL-CM5A-MR	**r1i1p1**	
IPSL-CM5B-LR	**r1i1p1**	
MIROC-ESM	**r1i1p1**	Japan Agency for Marine-Earth Science and Technology, and Atmosphere and Ocean Research Institute (U. Tokyo), and National Institute for Environmental Studies
MIROC-ESM-CHEM	r1i1p1	
MIROC5	**r1i1p1**, r2i1p1, r3i1p1, r4i1p1, r5i1p1	
MPI-ESM-LR	**r1i1p1**, r2i1p1, r3i1p1	Max Planck Institute for Meteorology (MPI-M)
MPI-ESM-MR	**r1i1p1**	
MRI-CGCM3	**r1i1p1**	Meteorological Research Institute
MRI-ESM1	r1i1p1	
NorESM1-M	**r1i1p1**	Norwegian Climate Centre
NorESM1-ME	**r1i1p1**	

Table S1. The 38 climate models, and 87 ensemble members, analyzed in this study. Ensemble members were only retained if the historical experiment, and matching RCP8.5 extension, covered the entire analysis period of 1880 to 2017. The 38 ensemble members in bold denote those that had corresponding piControl experiments.

Text S1. Forced Response Removal

For an individual model realization (or observational dataset) i, the time-series of a quantity $y = y(t)$ can be decomposed into a forced response component, ϕ, and an unforced (or internal variability) component, v, such that

$$y_i(t) = \phi_i(t) + v_i(t). \quad (1)$$

The term $y(t)$ can be taken to represent any time-varying quantity (for example, global mean surface temperature, the AMV index, or SST at an individual grid-point). The forced response component, ϕ_i, depends upon both the forcing and the feedbacks in that model. Here the forced response is estimated from an ensemble of model experiments following the single-factor scaling method of Frankcombe et al. (2015). The premise of using a multi-model mean is that, given different initializations, internal variability will not be synchronized across several simulations. If it is assumed that the multi-ensemble and multi-model mean, M, provides a reasonable approximation to the true forced response, then:

$$\phi_i(t) \cong M(t)\beta_i + c_i \quad (2)$$

where β_i is a scaling factor such that $M(t)$ is scaled to give a best fit to quantity $y(t)$ of realization i. In some sense, β_i represents a measure of that model realization's sensitivity to forcing. The c_i is a constant, which is zero where the mean of M and y_i have been subtracted. The scaling factor β_i can then be estimated by regressing $M(t)$ against $y_i(t)$, arriving at an expression for the internal variability:

$$v_i(t) \cong y_i(t) - M(t)\beta_i. \quad (3)$$

The approach here only differs from Frankcombe et al. (2015) in that the forced signal M is always taken to be the multi-model mean of the CMIP5 historical global mean surface temperature. The term y_i can represent any quantity, be it an SST index, grid-point SST data, or global mean surface temperature, and a corresponding β_i is computed. Although the rate of global warming might vary substantially in different regions across the globe, and across models, the multi-model mean signature of forcing is scaled in all instances before subtraction in each model realization. Therefore, in regions where the long-term warming is weak, or furthermore, where there is weak correlation with the multi-model mean, then the scaling factor will likewise be small.

To compute the scaling factor β_i, an ordinary least-squares regression is computed between M and y_i. Allen & Stott (2003) compare ordinary and total least-squares regressions for computing the scaling factor (see their Figure 2). They show that total least-squares regression should be employed where the sample size of simulated members is small, since there will be residual internal variability (i.e. "expected noise variance") in the ensemble mean. Their algorithm includes a step of "scaling up" M so that the noise variance matches that in y_i. In the present study, the ensemble mean is computed from a large number of model realisations and the residual variability is very low. It is unclear how the noise component should be scaled up, and therefore ordinary rather than total least-squares regression is used.

The single-factor approach, rather than multi-factor (Frankcombe et al. 2015), has been adopted here since not all experiments (e.g. 'historicalNat' or 'historicalGHG') are available for every model in the CMIP5 archive. Other approaches may have been to compute an ensemble mean for models with multiple realizations (following the "single model ensemble mean" method of Frankcombe et al. 2018), or to retain the multi-model mean, but compute single

model-mean scaling factors. The seemingly simplest approach in this regard has been adopted here, which is to treat each realization independently. A comparison of a range of forced response removal processes is presented by Frankcombe et al. (2015, 2018).

In the case of the piControl simulations, where forcings are constant, a simple linear detrending is applied. Some of the models suffer from small residual "drift" (Sen Gupta et al. 2013), and thus here a linear detrending is sufficient.

Text S2. Running trend time-series

To construct an *N*-year running-trend time-series, the linear trend is computed over the first *N*-year window of an annual time-series. The window is then shifted by one year, the linear trend is again computed. This process is repeated to the end of the annual time-series. Linear trends in each window are computed using an ordinary least-squares fit.

Text S3. Statistical significance of correlations

The statistical significance of correlations between running-trend time-series were tested based on a one-tailed Student's *t*-distribution with adjusted degrees of freedom. One-tailed tests were chosen in these cases because models and observations are largely in agreement on the sign of the correlation coefficients (positive), over most time-scales. The degrees of freedom in the statistical tests were adjusted to account for non-independence of overlapping periods. The effective degrees of freedom were specified by the number of data-points in the running trend time-series divided by *N*/2. This choice is analogous to that of Wang et al. (2017), who scale the degrees of freedom by half the filter frequency used in the smoothing of their AMV time-series. The significance test for the model data correlation incorporates the number of available model realizations in the adjusted degrees of freedom.

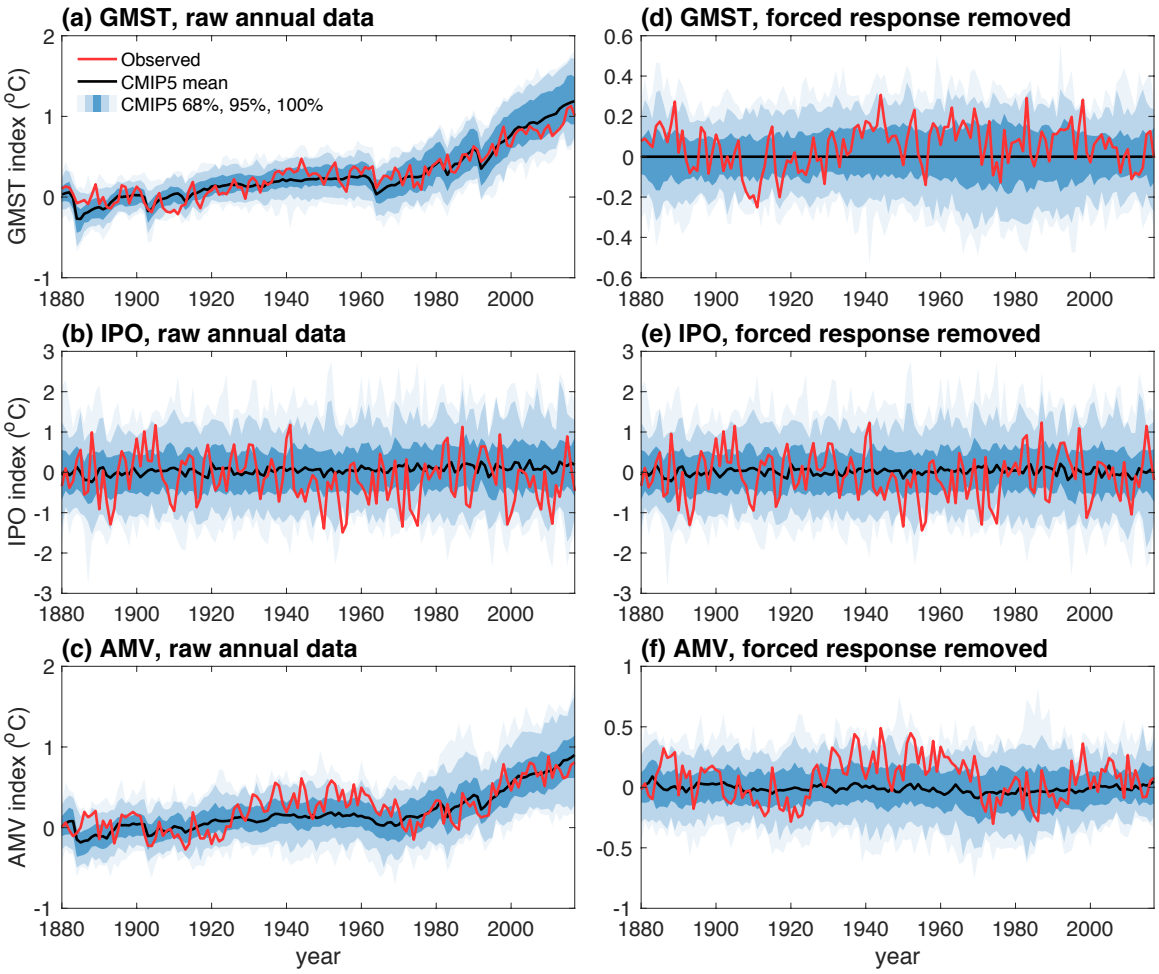

Figure S1. Annual time-series of global mean surface temperature (GMST), the Interdecadal Pacific Oscillation (IPO), and the Atlantic Multidecadal Variability (AMV) indices, in observations and CMIP5 historical simulations. **(a-c)** Raw annual data, and **(d-f)** with the forcing response removed. In each case, anomalies with respect to the 1880-1930 baseline period are shown. The shaded blue regions denote the central 68%, 95%, and 100% of the CMIP5 realization ensemble.

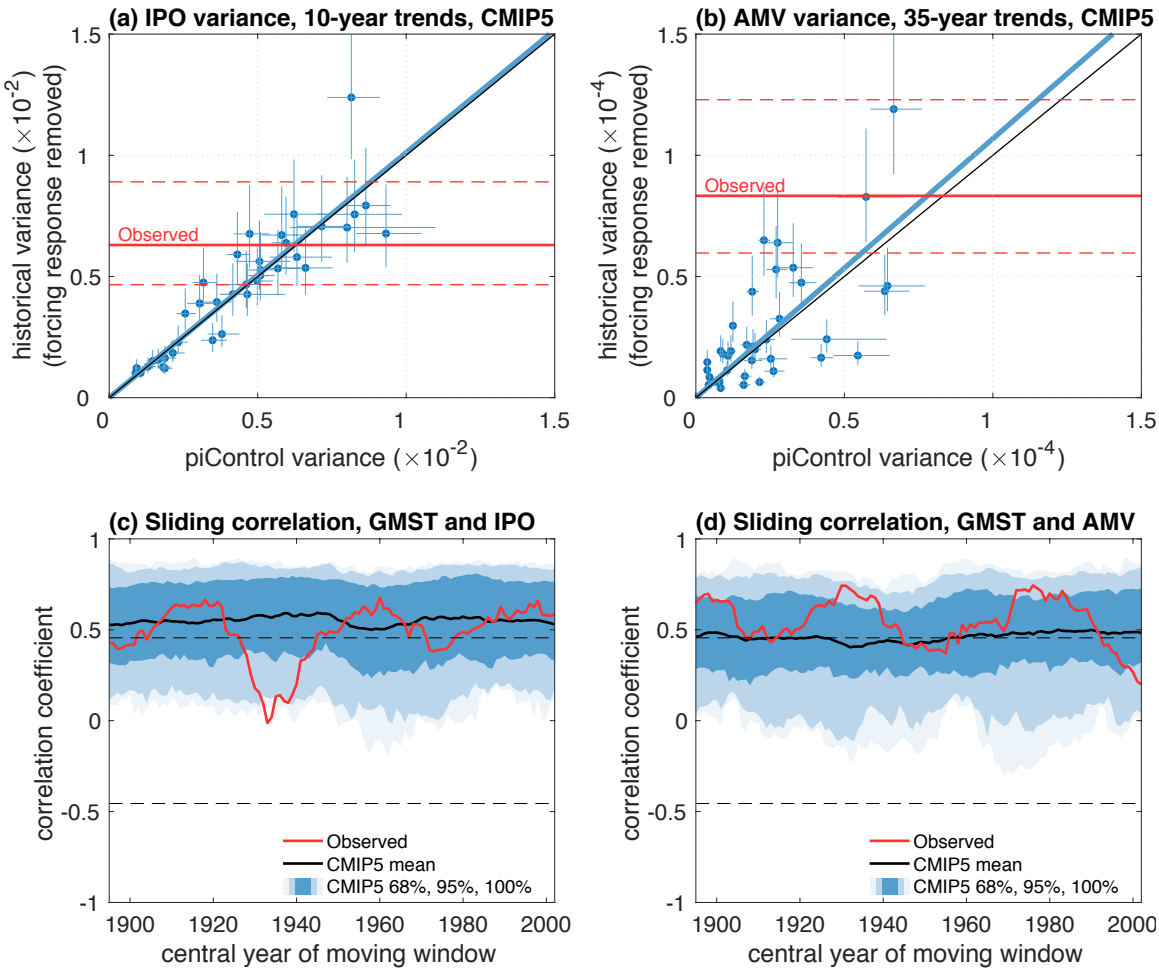

Figure S2. (a,b) Variances of **(a)** 10-year running trends of the Interdecadal Pacific Oscillation (IPO) index, and **(b)** 35-year running trends of the Atlantic Multidecadal Variability (AMV) index. The variance in piControl is plotted against the variance in historical (with forcing response removed) for each available ensemble member (Table S1). The one-to-one line is plotted in black, and blue line denotes the inter-model least-squares fit regressed through the origin. Whiskers denote the 99% confidence interval of the computed variance using a chi-squared distribution, with degrees of freedom based on the number of data-points in the running-trend time-series. The red dashed lines denoted the 99% confidence interval for the observed data. **(c,d)** Correlations of global mean surface temperature with **(c)** IPO index, and **(d)** the AMV index, in observations and CMIP5 historical models. The correlations are computed from annual data in 31-year sliding windows, where year on the x-axes denotes the central year of the 31-year window. The shaded blue regions denote the central 68%, 95%, and 100% of the CMIP5 model ensemble. Dashed lines denote the 99% levels for statistically significant correlations based on a two-tailed Student's *t*-distribution with 30 degrees of freedom.

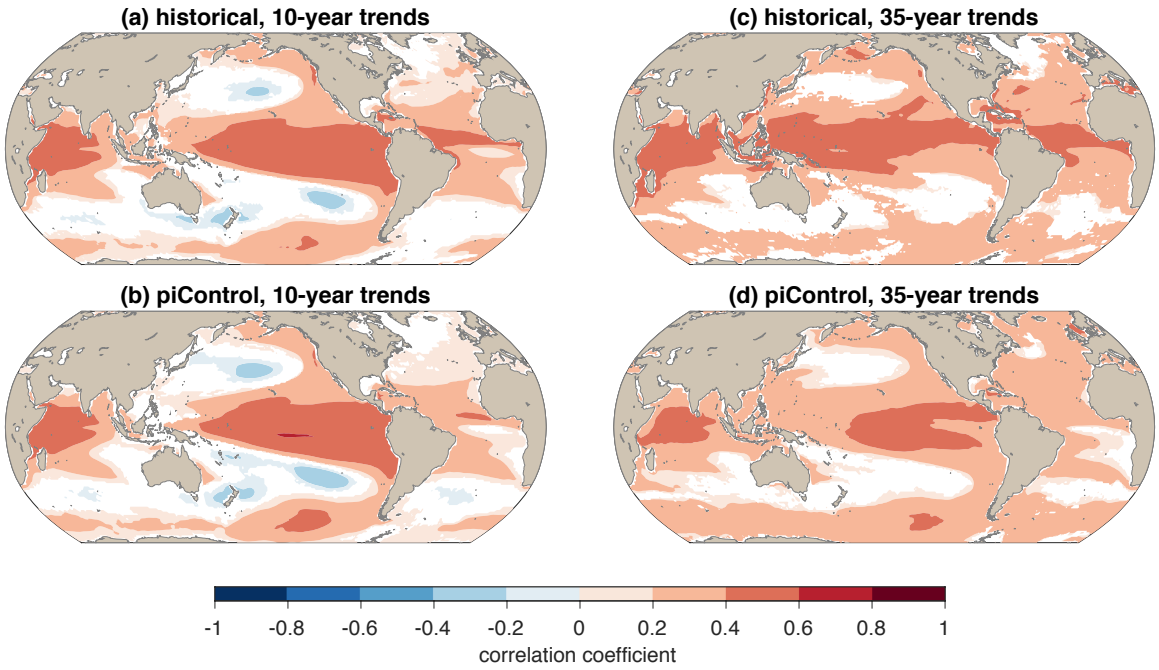

Figure S3. Correlations between running trends of global mean surface temperature (GMST) and grid-point sea surface temperature (SST). As in Figure 3, but only for the multi-model means of the CMIP5 **(a,c)** historical (same panels as Figure 3b,f) and **(b,d)** piControl simulations**.**

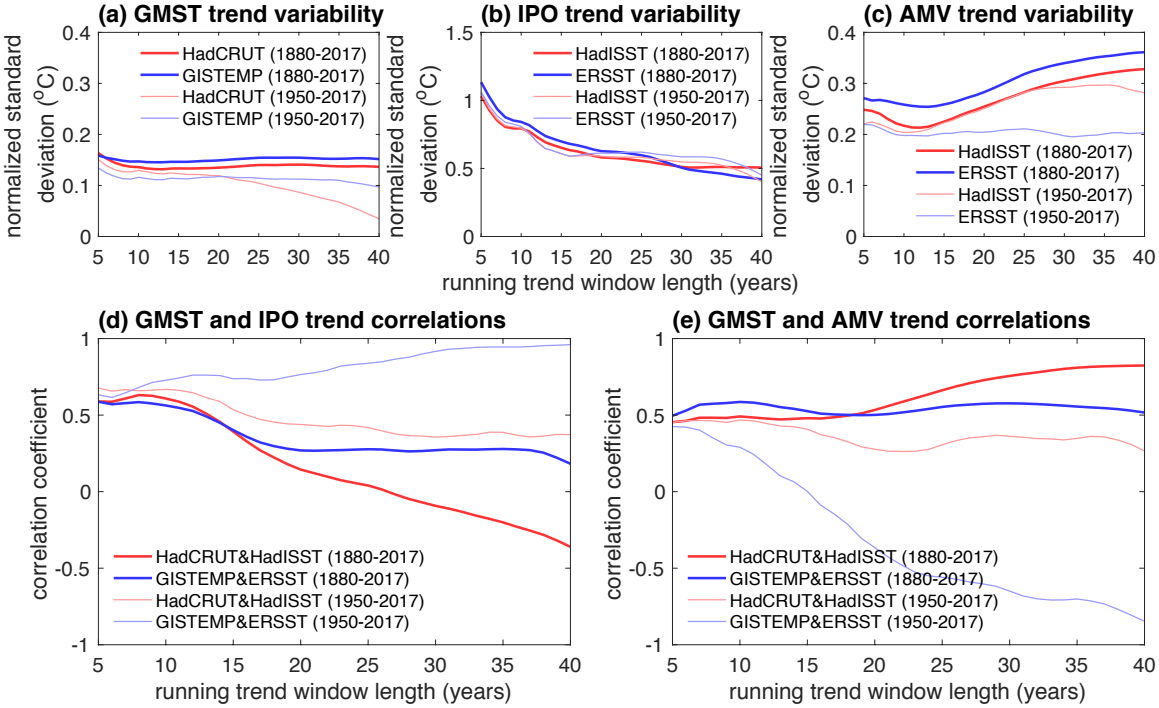

Figure S4. Standard deviation and correlations in the running trends of global mean surface temperature (GMST), the Interdecadal Pacific Oscillation (IPO), and the Atlantic Multidecadal Variability (AMV) indices, over a range of time-scales. As in Figure 1, but comparing different observational data-sets, including GISTEMP (Hansen et al. 2010; GISTEMP Team 2018) and ERSST (Huang et al. 2017). The thicker red curves for HadCRUT and HadISST are the same as

those shown in Figure 1. For the thinner curves in each panel, the data was first trimmed to the period 1950 to 2017 before computing the standard deviations and correlations.

References

Allen MR, Stott PA. Estimating signal amplitudes in optimal fingerprinting, part I: theory. *Clim Dyn* **21**, 477–491 (2003)

Frankcombe LM, England MH, Kajtar JB, et al. On the choice of ensemble mean for estimating the forced signal in the presence of internal variability. *J Clim* **31**, 5681–5693 (2018)

Frankcombe LM, England MH, Mann ME, Steinman BA. Separating internal variability from the externally forced climate response. *J Clim* **28**, 8184–8202 (2015)

GISTEMP Team. GISS Surface Temperature Analysis (GISTEMP). http://www.data.giss.nasa.gov/gistemp/. Dataset accessed 2019-01-15. (2018)

Hansen JE, Ruedy R, Sato M, Lo K. Global surface temperature change. *Rev Geophys* **48**, RG4004 (2010)

Huang B, Thorne PW, Banzon VF, et al. Extended Reconstructed Sea Surface Temperature version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons. *J Clim* **30**, 8179–8205 (2017)

Sen Gupta A, Jourdain NC, Brown JN, Monselesan DP. Climate drift in the CMIP5 models. *J Clim* **26**, 8597–8615 (2013)

Wang J, Yang B, Ljungqvist FC, et al. Internal and external forcing of multidecadal Atlantic climate variability over the past 1,200 years. *Nat Geosci* **10**, 512–517 (2017)