Robust anthropogenic signal identified in the seasonal cycle of tropospheric temperature

Benjamin D. Santer,[a b] Stephen Po-Chedley[a], Nicole Feldl[c], John C. Fyfe[d], Qiang Fu[e], Susan Solomon[f], Mark England[c], Keith B. Rodgers[g h], Malte F. Stuecker[i], Carl Mears[j], Cheng-Zhi Zou[k], Céline J. W. Bonfils[a], Giuliana Pallotta[a], Mark D. Zelinka[a], Nan Rosenbloom[l], and Jim Edwards[l]

[a] *Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California*

[b] *Joint Institute for Regional Earth System Science & Engineering, University of California at Los Angeles, Los Angeles, California*

[c] *Department of Earth and Planetary Sciences, University of California at Santa Cruz, Santa Cruz, California*

[d] *Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, Victoria, British Columbia, Canada*

[e] *Dept. of Atmospheric Sciences, University of Washington, Seattle, Washington*

[f] *Massachusetts Institute of Technology, Earth, Atmospheric, and Planetary Sciences, Cambridge, Massachusetts*

[g] *Center for Climate Physics, Institute for Basic Science, Busan, South Korea*

[h] *Pusan National University, Busan, South Korea*

[i] *Department of Oceanography and International Pacific Research Center, School of Ocean and Earth Science and Technology, University of Hawai'i at Mānoa, Honolulu*

[j] *Remote Sensing Systems, Santa Rosa, California*

[k] *Center for Satellite Applications and Research, NOAA/NESDIS, Camp Springs, Maryland*

[l] *National Center for Atmospheric Research, Boulder, Colorado.*

Corresponding author: bensanter1289@gmail.com

ABSTRACT: Previous work identified an anthropogenic fingerprint pattern in $T_{AC}(x,t)$, the amplitude of the seasonal cycle of mid- to upper tropospheric temperature (TMT), but did not explicitly consider whether fingerprint identification in satellite $T_{AC}(x,t)$ data could have been influenced by real-world multidecadal internal variability (MIV). We address this question here using large ensembles (LEs) performed with five climate models. LEs provide many different sequences of internal variability noise superimposed on an underlying forced signal. Despite differences in historical external forcings, climate sensitivity, and MIV properties of the five models, their $T_{AC}(x,t)$ fingerprints are similar and statistically identifiable in 239 of the 240 LE realizations of historical climate change. Comparing simulated and observed variability spectra reveals that consistent fingerprint identification is unlikely to be biased by model underestimates of observed MIV. Even in the presence of large (factor of 3-4) inter-model and inter-realization differences in the amplitude of MIV, the anthropogenic fingerprints of seasonal cycle changes are robustly identifiable in models and satellite data. This is primarily due to the fact that the distinctive, global-scale fingerprint patterns are spatially dissimilar to the smaller-scale patterns of internal $T_{AC}(x,t)$ variability associated with the Atlantic Multidecadal Oscillation and the El Niño~Southern Oscillation. The robustness of the seasonal cycle D&A results shown here, taken together with the evidence from idealized aquaplanet simulations, suggest that basic physical processes are dictating a common pattern of forced $T_{AC}(x,t)$ changes in observations and in the five LEs. The key processes involved include GHG-induced expansion of the tropics, lapse-rate changes, land surface drying, and sea ice decrease.

1. Introduction

Detection and attribution ("D&A") studies seek to disentangle human and natural influences on Earth's climate. This research made a significant contribution to the recent finding that human influence on climate is unequivocal (IPCC 2021). Pattern-based "fingerprint" methods are a key element of D&A research (Hasselmann 1979; North et al. 1995; Hegerl et al. 1996; Santer et al. 1996; Tett et al. 1996; Stott et al. 2000; Barnett et al. 2005).

The initial focus of fingerprint research was on changes in annual- or decadal-mean properties of surface temperature (Hegerl et al. 1996; Stott et al. 2000), atmospheric temperature (Santer et al. 1996; Tett et al. 1996; Thorne et al. 2002; Santer et al. 2003), and ocean heat content (Barnett et al. 2005). Examination of the hydrological cycle, cryosphere, and atmospheric circulation followed, targeting surface specific humidity and water vapor (Willett et al. 2007; Santer et al. 2009), rainfall (Zhang et al. 2007; Marvel and Bonfils 2013), salinity (Pierce et al. 2012), sea-level pressure (Gillett et al. 2003), and Arctic sea ice (Min et al. 2008). Model-predicted patterns of mean changes in these and many other variables were detectable in observations and attributable to human influences (Santer et al. 1995; Mitchell and Karoly 2001; Hegerl et al. 2007).

After comprehensive interrogation of the causes of historical changes in average climate, the attention of D&A analysts shifted to aspects of climate change that are more directly relevant to societal impacts (Bindoff et al. 2013). Research began to examine extreme rainfall and heat (Min et al. 2009; Stott et al. 2016), the likelihood and severity of individual extreme events (Stott et al. 2004; Risser and Wehner 2017), and the seasonality of precipitation (Marvel et al. 2017) and temperature (Santer et al. 2018; Duan et al. 2019).

It is changes in the amplitude of the seasonal cycle that are of interest here. They have the potential to impact water availability, hydropower production, energy demand, agriculture, fire weather, vector-borne diseases, and many other aspects of society, the economy, and human health. Seasonality also influences animal and plant distributions and abundances (Parmesan and Yohe 2003; Root et al. 2005; Cohen et al. 2018). It is critically important to understand how this seasonal pacemaker may have been modulated by historical changes in anthropogenic forcing – and how seasonality may change over the 21st century (Dwyer et al. 2012; Stine and Huybers 2012; Donohoe and Battisti 2013; Qian and Zhang 2015; Yettella and England 2018).

A previous study by Santer et al. (2018) reported that satellite temperature records contained a fingerprint of human-caused changes in $T_{\mathrm{AC}}(x,t)$, the amplitude of the annual cycle of mid- to upper tropospheric temperature (TMT).[i] Related work showed that internal climate variability affected observed annual-mean TMT changes over the satellite era (Kamae et al. 2015; Suárez-Gutiérrez et al. 2017; Po-Chedley et al. 2021). The relationship between changes in annual-mean TMT and changes in $T_{\mathrm{AC}}(x,t)$ is unclear. It is conceivable, however, that multidecadal internal variability (MIV) may have influenced the identification of a human fingerprint in satellite $T_{\mathrm{AC}}(x,t)$ data.

We explore this possibility here using output from large initial condition ensembles (LEs) performed with five different Earth System Models (ESMs; Deser et al. 2012; Fyfe et al. 2017, 2021; Tatebe et al. 2019; Rodgers et al. 2021). In total, these five LEs provide 240 different plausible realizations of historical climate change, each with a unique sequence of internal variability ("noise") superimposed on the response to anthropogenic and natural external forcing ("signal"). With such information, we can assess how frequently fingerprint detection occurs in model realizations of $T_{\mathrm{AC}}(x,t)$. If fingerprint detection is a robust result in the 240 realizations, despite differences in the forcings, climate sensitivity, and MIV properties of the five LEs, it suggests that positive fingerprint detection in real-world $T_{\mathrm{AC}}(x,t)$ data is unlikely to be due to the fortuitous phasing of MIV.

Most fingerprint methods rely on model MIV estimates to assess whether the random action of internal variability could explain a "match" between observed climate change patterns and a model-predicted anthropogenic fingerprint. Concerns have been raised about the adequacy of model noise estimates, thus calling into question the reliability of fingerprint results (Curry and Webster 2011; O'Reilly et al. 2021). We address such concerns here by comparing simulated and observed spectra for three key modes of MIV: the Atlantic Multidecadal Oscillation (AMO), the El Niño/Southern Oscillation (ENSO), and the Interdecadal Pacific Oscillation (IPO).

We use information from these spectra as the basis for a number of sensitivity studies. These studies explore whether the positive identification of annual cycle fingerprints in observations and model simulations is robust to large model differences in the amplitude of specific modes of internal variability. A further sensitivity study considers whether fingerprint identification is hampered by removing all information regarding global-mean $T_{\mathrm{AC}}(x,t)$ changes.

---

[i] For each model and satellite data set, and at each grid-point x and year t, there are 12 monthly-mean values of TMT. We use these 12 values to calculate the amplitude of the first harmonic – the annual cycle (Wilks 1995; Yettella and England 2018). Our focus in this study is solely on the amplitude of the first harmonic. Here and throughout, x is an index over the combined latitude and longitude dimensions of the spatial field and t is an index over time in years.

In addition to assessing the robustness of our fingerprint detection results for annual cycle changes, we also seek to improve understanding of the physical mechanisms driving these changes. Some insights are provided by novel aquaplanet simulations with realistic, seasonally varying insolation (Feldl et al. 2017). These experiments were performed under preindustrial and quadrupled CO_2 conditions with two climate models, each with a different representation of the effects of sea-ice on high-latitude climate processes. We compare the two sets of aquaplanet experiments with conventional (land+ocean+ice) ESM simulations to investigate how the annual temperature cycle is affected by the presence or absence of land.

The structure of our paper is as follows. Section 2 introduces the observational and model data sets used here, with additional information available in the Supplementary Materials (SM) and in a previous paper (Santer et al. 2021). Section 3 introduces the spatial patterns of satellite-era $T_{\mathrm{AC}}(x,t)$ trends in four observational data sets and in the average of the five LEs. As a prelude to the signal-to-noise (S/N) analysis of global patterns of annual cycle changes, Section 4 performs a local S/N analysis of $T_{\mathrm{AC}}(x,t)$ trends at individual grid-points in each LE. The fingerprint method applied to discriminate between forced and unforced annual cycle changes is introduced in Section 5 and documented in detail in the SM. Section 6 discusses the S/N ratios and "baseline" fingerprint detection times obtained for the full global pattern of $T_{\mathrm{AC}}(x,t)$ changes. After using the five LEs to estimate and subtract signals of forced SST changes from individual LE realizations and observations, Section 7 compares the simulated and observed variability spectra for the AMO, Niño 3.4 SSTs, and the IPO. Section 8 uses information from the model spectra to repeat the "baseline" fingerprint analysis of Section 6 with subsets of the 240 realizations of internal $T_{\mathrm{AC}}(x,t)$ fluctuations. These subsets comprise realizations with low- and high-amplitude variability of the AMO and ENSO. Annual cycle changes in the aquaplanet simulations performed with two different climate models are analyzed in Section 9. We provide brief conclusions in Section 10.

2. Observational data and model simulations

a. Satellite and reanalysis data

Our focus here is on $T_{\mathrm{AC}}(x,t)$ changes over the satellite era (January 1979 to December 2020). We rely on satellite TMT data from three research groups: Remote Sensing Systems (RSS; Mears and Wentz 2017), the Center for Satellite Applications and Research (STAR; Zou et al. 2018), and the University of Alabama at Huntsville (UAH; Spencer et al. 2017). All three groups analyze microwave emissions from oxygen molecules. Emissions are measured with Microwave Sounding Units (MSU) and Advanced Microwave Sounding Units (AMSU) and depend on the temperature of different broad atmospheric layers. Measurements at different microwave frequencies provide information on temperatures at different heights. In addition to TMT, we use measurements of the temperature of the lower stratosphere (TLS) to adjust TMT for the contribution it receives from stratospheric cooling (Fu et al. 2004; Fu and Johanson 2004; see SM).

Our comparisons of simulated and observed $T_{\mathrm{AC}}(x,t)$ changes also make use of synthetic TMT data from version 5.1 of the state-of-the-art ERA reanalysis of the European Centre for Medium-Range Weather Forecasts (ECMWF; Hersbach et al. 2020; Simmons et al. 2020; see SM). Reanalyses are a retrospective analysis of many different types of observational data using a data assimilation system and numerical weather forecast model that do not change over time (Kalnay et al. 1996).

b. SST data

Section 7 considers three commonly-used indices of modes of SST variability. We use version 4 of the data set developed jointly by the Hadley Centre and the Climatic Research Unit (HadCRUT4; Morice et al. 2012) to compute observational time series of the AMO, Niño 3.4 SSTs, and the IPO. Information regarding calculation of these indices is provided in the SM. Our focus in Section 7 is on the 852 months from January 1950 to December 2020, a period unaffected by potential problems associated with SST measurements during World War II (Thompson et al. 2008).

c. Model simulations

We analyze $T_{\mathrm{AC}}(x,t)$ changes in five different large initial condition ensembles (LEs). Deser et al. (2020) provide a comprehensive introduction to LEs and their many scientific applications. An LE typically consists of between 30 to 100 individual members. The ensemble is generated by repeatedly running the same physical climate model with the same spatio-temporal changes in external forcings. Each ensemble member commences from different initial states of the atmosphere and/or ocean. These are selected in various ways (see SM). Slight differences in initial states result in different sequences of natural variability superimposed on the underlying forced response. The result is an envelope of plausible trajectories of historical and/or future climate change.

Here, we use LEs to explore both the local (Section 4) and global (Sections 5, 6, and 8) S/N characteristics of simulated changes in annual cycle amplitude. Of particular interest is the information LEs provide regarding the robustness of fingerprint detection, the stochastic uncertainty in fingerprint detection time, estimates of externally forced signals in the AMO, Niño 3.4 SSTs, and the IPO, and uncertainties in the internal variability spectra of these three modes.

The LEs considered here rely on both older and newer model versions and estimates of external forcings. Two LEs were generated with models participating in the older phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012). The CMIP5 LEs were performed with version 1 of the Community Earth System Model (CESM1; Kay et al. 2015) and with version 2 of the Canadian Earth System Model (CanESM2; Kirchmeier-Young et al. 2017; Fyfe et al. 2017; Swart et al. 2018). The CESM1 and CanESM2 LEs have 40 and 50 members, respectively. The three LEs produced with models taking part in the newer phase 6 of CMIP (CMIP6; Eyring et al. 2016) relied on version 5 of CanESM (CanESM5; Swart et al. 2019; Fyfe et al. 2021), version 2 of CESM (CESM2; Rodgers et al. 2021), and version 6 of the Model for Interdisciplinary Research on Climate (MIROC6; Tatebe et al. 2019). Each CMIP6 LE had 50 ensemble members.[ii]

The CMIP5 and CMIP6 historical simulations ended in 2005 and 2014, respectively. To facilitate comparison with observational $T_{\mathrm{AC}}(x,t)$ changes over the full 42-year satellite era (1979 to 2020), historical simulations were spliced with scenario integrations initiated from the end of each historical run. The scenario integrations are Representative Concentration Pathway 8.5 for CanESM2 and CESM1 (Meinshausen et al. 2011), Shared Socioeconomic Pathway 5-8.5 (SSP5) for CanESM5 and MIROC6, and SSP 3-7.0 for CESM2 (SSP3; Riahi et al. 2017). Further details of these scenarios are given in the SM.

Our pattern-based fingerprinting method requires model estimates of natural internal variability. We obtain these estimates from two sources: 1) multi-model ensembles of preindustrial control simulations with no year-to-year changes in external forcings; and 2) the between-realization variability of each of the five LEs. In the former case, we use output from preindustrial control runs performed with 36 CMIP5 models and 30 CMIP6 models. In the latter case, we estimate the

[ii] The CESM2 LE described in Rodgers et al. (2021) has 100 ensemble members. The first 50 members were run with CMIP6 SSP 3-7.0 forcing; the remaining 50 members have modified biomass forcing over recent decades (Fasullo et al. 2021). We analyze only the first 50 members here.

between-realization variability in a single model's LE by subtracting the ensemble-mean changes in $T_{\mathrm{AC}}(x,t)$ from each realization in the LE (see Section 5 and SM). Tables S1 and S2 of the SM identify the CMIP5 and CMIP6 models we relied on for our multi-model noise estimates.

Section 9 examines changes in the amplitude of the annual cycle of TMT in aquaplanet simulations performed with two climate models. The first is version 2.1 of the Geophysical Fluid Dynamics Laboratory Atmospheric Model (GFDL-AM2.1). The model was run in a configuration with a 30-meter fixed-depth slab ocean with no meridional ocean heat transport and a realistic seasonal cycle of insolation (Feldl et al. 2017). The simulations explore the impact of large differences in sea-ice albedo under preindustrial and quadrupled CO_2 conditions.

The second model relies on version 6 of the Community Atmospheric Model (CAM6; Rodgers et al. 2021). This is the atmospheric component of CESM2. Like GFDL-AM2.1, CESM2-CAM6 was run with a 30-meter fixed-depth slab ocean, but with a symmetrical annual-mean ocean heat transport (an average of NH and SH conditions) diagnosed from the CESM2 pre-industrial control run. A significant difference in the two models is that GFDL-AM2.1 has no ice thermodynamics, while CESM2-CAM6 includes ice thermodynamics and uses a simple version of the Los Alamos sea-ice model (CICE5; Smith et al. 1992). As we show subsequently, model differences in sea-ice treatment yield different high-latitude changes in $T_{\mathrm{AC}}(x,t)$ in response to CO_2 forcing.

Both sets of aquaplanet simulations allow us to investigate whether large-scale features of the annual cycle fingerprints in full ESMs can be captured without representation of land surface processes and without hemispheric asymmetry in land distribution or land-ocean differences in heat capacity. Further details of the aquaplanet simulations are given in the SM.

3. Changes in annual cycle amplitude in observations and the LE average

Santer et al. (2018) analyzed observed spatial patterns of $T_{\mathrm{AC}}(x,t)$ trends over 1979 to 2016. It is useful to re-examine these patterns given four additional years of corrected TMT data, improved versions of satellite TMT data sets, and results from the state-of-the-art ERA5.1 reanalysis.

Updates and improvements to satellite TMT data have not altered the basic features of the $T_{\mathrm{AC}}(x,t)$ trends. These features include increases in annual cycle amplitude at mid-latitudes in both hemispheres (with larger increases in the NH than the SH), decreases in amplitude over the Arctic, and small changes of either sign in the tropics (Figs. 1a-c). ERA5.1 shows similiar behavior

(Fig. 1d). UAH differs from the other observational data sets at high latitudes in the SH: $T_{AC}(x,t)$ trends are positive in UAH and negative in RSS, STAR, and ERA5.1. The anomalous UAH results appear to be related to the decisions made by the UAH group in merging information from MSU and AMSU during the period of overlap between these different instruments (Santer et al. 2018).

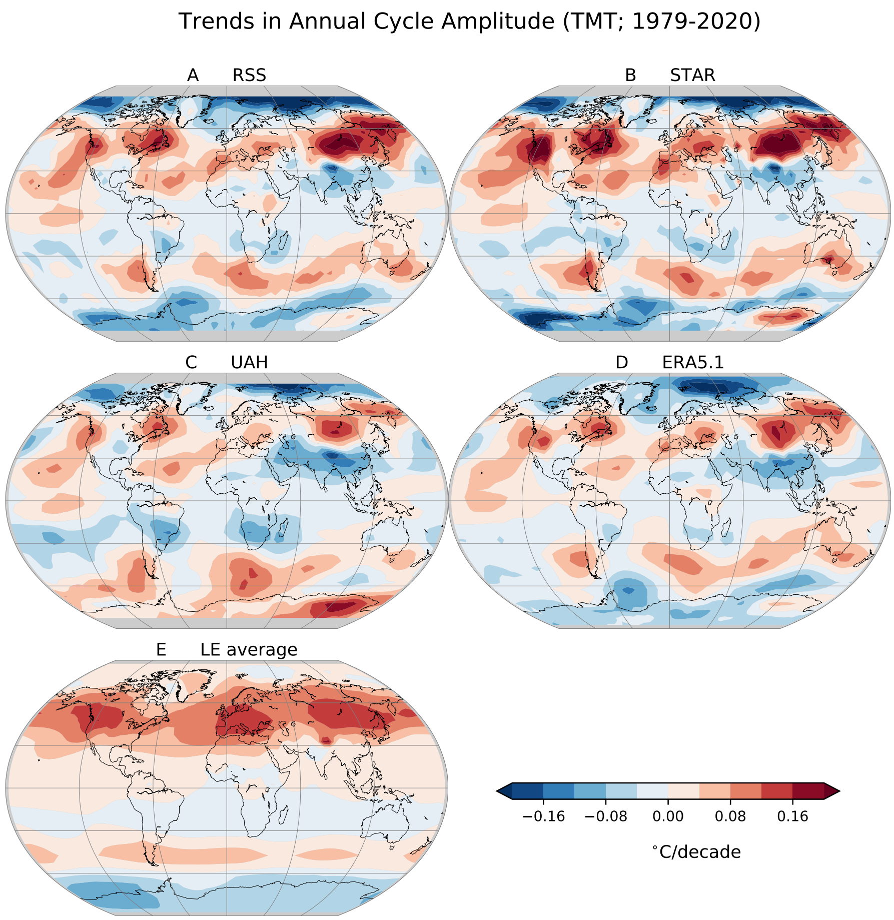

FIG. 1: Least-squares linear trends over 1979 to 2020 in $T_{AC}(x,t)$, the amplitude of the annual cycle of mid- to upper tropospheric temperature (TMT). (a-c) Satellite data from Remote Sensing Systems (RSS), the Center for Satellite Applications and Research (STAR), and the University of Alabama at Huntsville (UAH). (d) Version 5.1 of the reanalysis produced by the European Centre for Medium-Range Weather Forecasts. (e) The average of the ensemble-mean trends in $T_{AC}(x,t)$ in the five LEs analyzed here (see Figs. 2a-e). TMT is adjusted for stratospheric cooling in all satellite, reanalysis, and climate model data sets (see SM).

Figure 1e shows the average of the ensemble-mean $T_{AC}(x,t)$ trends in the five LEs. As expected, simulated changes are smoother than in the observations (Santer et al. 2018; Po-Chedley et al. 2021). This is because the model results have been averaged over individual realizations with different sequences of internal variability, and then averaged over models. Averaging over realizations and models damps internal variability and reduces uncorrelated model biases, more clearly revealing the underlying forced response. Despite the larger spatial noise in observations, there is correspondence between the large-scale features of the simulated and observed $T_{\mathrm{AC}}(x,t)$ changes in Fig. 1. Whether this correspondence is statistically significant is considered in Section 6.

4. Local signal-to-noise ratios

Pattern-based fingerprinting utilizes the signal and noise properties of entire spatial fields (Hasselmann 1979; Santer et al. 1994; Hegerl et al. 1996). It provides an efficient means of discriminating between externally forced climate changes and the complex noise of internal variability. An alternate form of S/N analysis considers forced and unforced climate changes at individual model grid-points (Hawkins and Sutton 2012; Mahlstein et al. 2012; Deser et al. 2014; Rodgers et al. 2015). Local S/N information can help to inform and interpret results from pattern-based fingerprinting (Santer et al. 2019). In this section, we briefly discuss a local S/N analysis before detailed consideration of our fingerprint results in Section 5.

Figures 2a-e show the ensemble-mean $T_{\mathrm{AC}}(x,t)$ trends in the five LEs. Trends are calculated over the same 1979 to 2020 analysis period used for the observations in Fig. 1. Although there are pronounced differences between the LEs in the amplitude of the changes, there are also key common features in the trend patterns. These include the previously noted increases in annual cycle amplitude at mid-latitudes in both hemispheres (with larger increases in the NH than the SH), decreases in $T_{\mathrm{AC}}(x,t)$ at high latitudes in the SH, and small changes with differing signs in the tropics (see Section 3). At high latitudes in the NH, the observations and CanESM5 show pronounced decreases in $T_{\mathrm{AC}}(x,t)$. This feature is absent in the other LEs.

The denominator of the local S/N ratio is the between-realization standard deviation of the 42-year trend in $T_{\mathrm{AC}}(x,t)$, calculated across all members of an ensemble. Patterns of this local noise are similar in the five LEs, with smallest values in the tropics and largest values at high latitudes in both hemispheres (Figs. 2f-j). There is some agreement across LEs in small-scale features of the noise patterns, such as the maxima over Greenland, the Himalayas, and East Antarctica. In all LEs, the local S/N ratio displays highest values at mid-latitudes in the NH, where increases in $T_{\mathrm{AC}}(x,t)$ are largest and noise is relatively low (Figs. 2k-o).

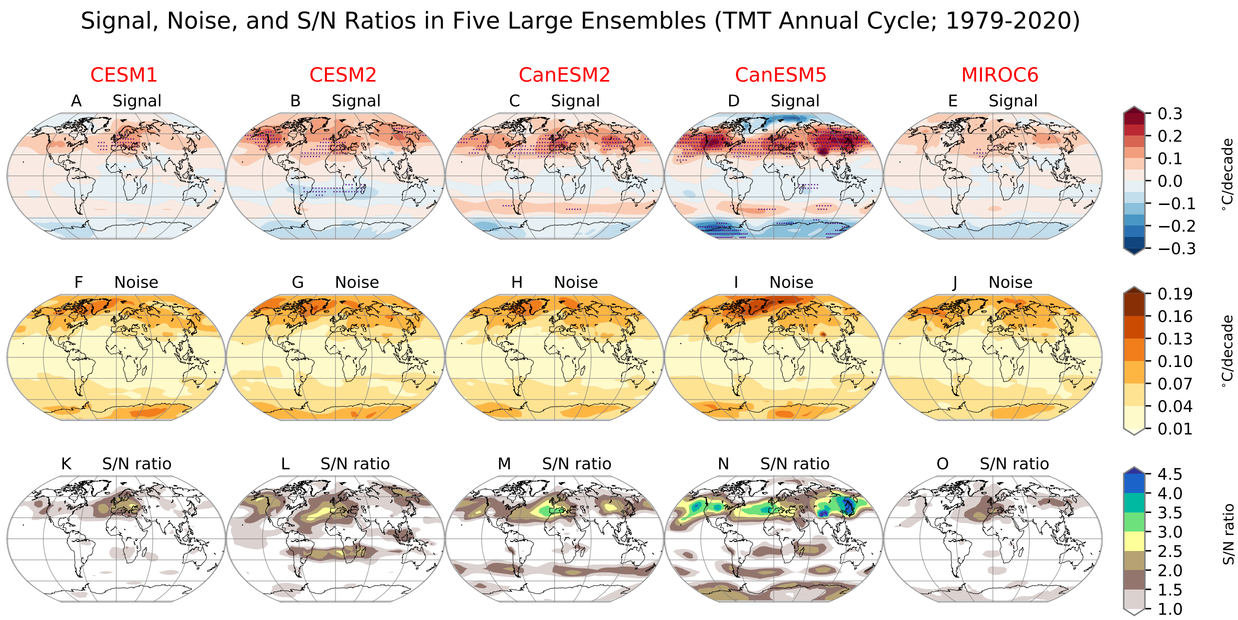

FIG. 2: Local signal-to-noise (S/N) analysis of least-squares linear trends over 1979 to 2020 in $T_{AC}(x,t)$. Results are from five different LEs (columns 1-5). (a-e) Ensemble-mean $T_{AC}(x,t)$ trends. (f-j) Local 1σ standard deviation of the 42-year trends in $T_{AC}(x,t)$ across all members in the LE. (k-o) S/N ratio: the absolute value of the ensemble-mean trend in an LE (the signal) divided by the local standard deviation of trends in the same LE (the noise). Stippling in the top row identifies grid-points where the local S/N ratio for ensemble-mean trends exceeds 2.

It is of interest to compare the annual cycle changes for TMT with those obtained for surface temperature (TS). In the Arctic and Antarctic, there are large reductions in the amplitude of the annual cycle of TS (Figs. 3a-e). These reductions in annual cycle amplitude have been linked to sea-ice loss and associated seasonal feedbacks, ocean-atmosphere energy transfer, and changes in surface heat capacity (Serreze and Barry 2011; Donohoe and Battisti 2013; Bintanja and van der Linden 2013; Taylor et al. 2013; Santer et al. 2018; Feldl et al. 2020; Feldl and Merlis 2021). As for TMT, the amplitude of the annual cycle of TS increases at mid-latitudes in the NH, but TS increases there are smaller, without the well-defined zonal structure of the TMT amplitude increases. Even for TS, however, there are mid-latitude areas of the North Atlantic and North Pacific oceans displaying significant increases in annual cycle amplitude, suggesting that the TS changes are not driven by land surface processes alone. Information on some of the factors driving annual cycle changes in TS and atmospheric temperature is given in Donohoe and Battisti (2013). In addition to the sea ice changes mentioned above, these factors include the shortwave absorption associated with GHG-forced increases in upper tropospheric water vapor.

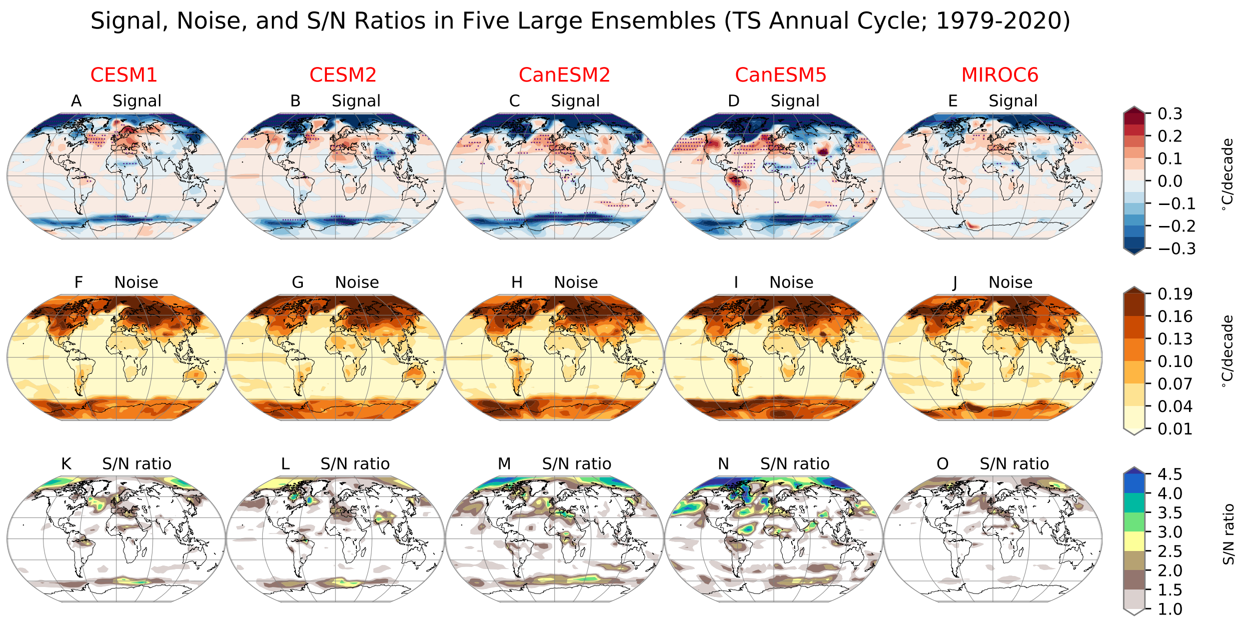

FIG. 3: As for Fig 2 but for the annual cycle of surface skin temperature. To facilitate comparison with TMT results the colorbar ranges are identical to those in Fig. 2.

As expected, the between-realization variability of trends in annual cycle amplitude has a strong land-sea contrast component for TS but not for TMT (compare Figs. 3f-j and Figs. 2f-j). Because of the higher noise over land for TS, few land areas have S/N ratios > 2 for changes in the annual cycle of TS (Figs. 3k-o). A notable exception is the Mediterranean region (Yettella and England 2018). Some of the most extensive areas of high S/N are in the regions of Arctic and Antarctic sea-ice decrease where TS signals are largest.

5. Fingerprint method and results

Next, we seek to determine whether the patterns of forced changes in $T_{\mathrm{AC}}(x,t)$ can be identified in observations and individual realizations of the LEs. The latter provide 240 different trajectories of climate change over the satellite era, each with a different estimate of MIV superimposed on the underlying response to forcing. The LEs allow us to estimate the stochastic uncertainty in t_d, the time required to identify the searched-for fingerprints of forced change (Santer et al. 2019).

We use a standard pattern-based fingerprint method to calculate t_d (Hasselmann 1979). The method has been successfully employed to identify anthropogenic fingerprints in many different independently monitored aspects of climate change (Hegerl et al. 1996; Santer et al. 1996, 2009, 2018; Marvel and Bonfils 2013; Bonfils et al. 2020; Sippel et al. 2020, 2021). The statistical methodology follows Santer et al. (2018); full details are provided in the SM. A brief description of the method is given below.

In the present application, the fingerprint pattern $F_{\mathrm{AC}}(x)$ is an estimate of the response of the amplitude of the annual cycle of TMT to combined anthropogenic and natural forcing. Five different fingerprints are used here. Each is the leading Empirical Orthogonal Function (EOF) of ensemble-mean $T_{\mathrm{AC}}(x,t)$ in an LE, calculated over 1979 to 2020 (Figs. 4a-e). We assume that the spatial pattern of $F_{\mathrm{AC}}(x)$ does not change markedly over time. For changes in the annual cycle of TMT, this assumption has been tested elsewhere and found to be reasonable (see SM).

The five LE estimates of $F_{\mathrm{AC}}(x)$ shown in Figs. 4a-e are searched for in sequences of time-varying $T_{\mathrm{AC}}(x,t)$ patterns derived from satellite data, the ERA5.1 reanalysis, and individual realizations of an LE. In the latter case, a searched-for model fingerprint is always compared with individual realizations of $T_{\mathrm{AC}}(x,t)$ changes generated with the same model – e.g., the CESM1 fingerprint in Fig. 4a is compared with the 40 individual realizations of $T_{\mathrm{AC}}(x,t)$ changes in the CESM1 LE (see Fig. 5a and left box-and-whiskers bar in Figs. 6a,b). In searching for $F_{\mathrm{AC}}(x)$ in observations, each of the five model fingerprints is compared with each observational data set (Fig. 5f).

These comparisons involve computing a measure of pattern similarity (an uncentered spatial covariance). This yields the signal time series $Z(t)$. If the observations or individual LE realizations are exhibiting greater magnitude of $F_{\mathrm{AC}}(x)$ over time, $Z(t)$ will exhibit a trend. To determine whether this trend in $Z(t)$ is significant, we require null distributions of pattern similarity trends in which we know *a priori* that any changes in pattern similarity with time are due to the effects of natural variability only (see SM).

We generate these null distributions by fitting trends to the noise time series $N(t)$, which is calculated by measuring the pattern similarity between $F_{\mathrm{AC}}(x)$ and time-varying patterns of natural internal variability in $T_{\mathrm{AC}}(x,t)$. The latter are obtained from two sources: 1) multiple pre-industrial control runs performed with either CMIP5 or CMIP6 models; and 2) the between-realization variability of $T_{\mathrm{AC}}(x,t)$ changes in each LE. We refer to these subsequently as multi-model and single-model noise estimates, respectively.

In the multi-model noise case there are n_m model control runs, each of length 150 years. These are concatenated into one data set (see SM). The single-model noise is computed by subtracting the ensemble-mean $T_{\mathrm{AC}}(x,t)$ changes in an LE from each realization of the LE. Calculation of the ensemble mean and residuals is over the 42-year satellite era (1979 to 2020). The residuals are then concatenated and have the time dimension $42 \times n_r$, the number of years in the satellite era times the number of realizations in the LE. Differences between single-model and multi-model noise estimates are discussed in Section 6.

Our detection time estimates are based on SN_L, the S/N ratio between b_L, an L-year trend in $Z(t)$, and σ_L, the standard deviation of the sampling distribution of L-year trends in $N(t)$. Here, L varies from 10, 11, ... 42 years. A key aspect of our analysis is that trends in $Z(t)$ and $N(t)$ are always compared on the same timescale. Explicit consideration of the timescale-dependence of S/N ratios is important because noise patterns and amplitude vary as a function of timescale (Tett et al. 1997; Stouffer et al. 2000).

For $L = 10$ years, for example, b_L is calculated over 1979 to 1988 and σ_L is computed from the sampling distribution of overlapping 10-year trends in $N(t)$. For $L = 11$ years, b_L is the trend in $Z(t)$ over the first 11 years (1979 to 1989) and σ_L is calculated from the sampling distribution of overlapping 11-year trends in $N(t)$. The full satellite era (1979 to 2020) is the $L = 42$ case. The detection time t_d is defined as the final year of the L-year period at which SN_L first exceeds some stipulated significance level (generally 5% here) and then remains continuously above this level for all larger values of L. The null hypothesis we are testing is that trends in $Z(t)$ are consistent with internal variability alone and SN_L values are not statistically unusual relative to an assumed Gaussian distribution (see SM for further details).

Before considering t_d results, it is useful to first examine the $F_{\mathrm{AC}}(x)$ patterns and dominant modes of between-realization variability in the five LEs. The fingerprints are spatially similar across the LEs (Figs. 4a-e) and capture the zonally coherent mean changes in annual cycle amplitude described in the local S/N analysis (Section 4). In contrast, the dominant noise modes are characterized by variability at smaller spatial scales. The leading noise EOF displays ENSO-like features (Po-Chedley et al. 2021) which are similar across the five LEs (Figs. 4f-j). The second noise EOF is also similar in the LEs, capturing anticorrelated variability in $T_{\mathrm{AC}}(x,t)$ between North America, Northern Eurasia, and the Indian subcontinent (Figs. 4k-o). The spatial dissimilarity[iii] between

[iii] The centered (spatial mean removed) pattern correlation between the fingerprint and leading noise mode in each LE is very small, ranging from close to zero for CanESM2 to 0.15 for CanESM5.

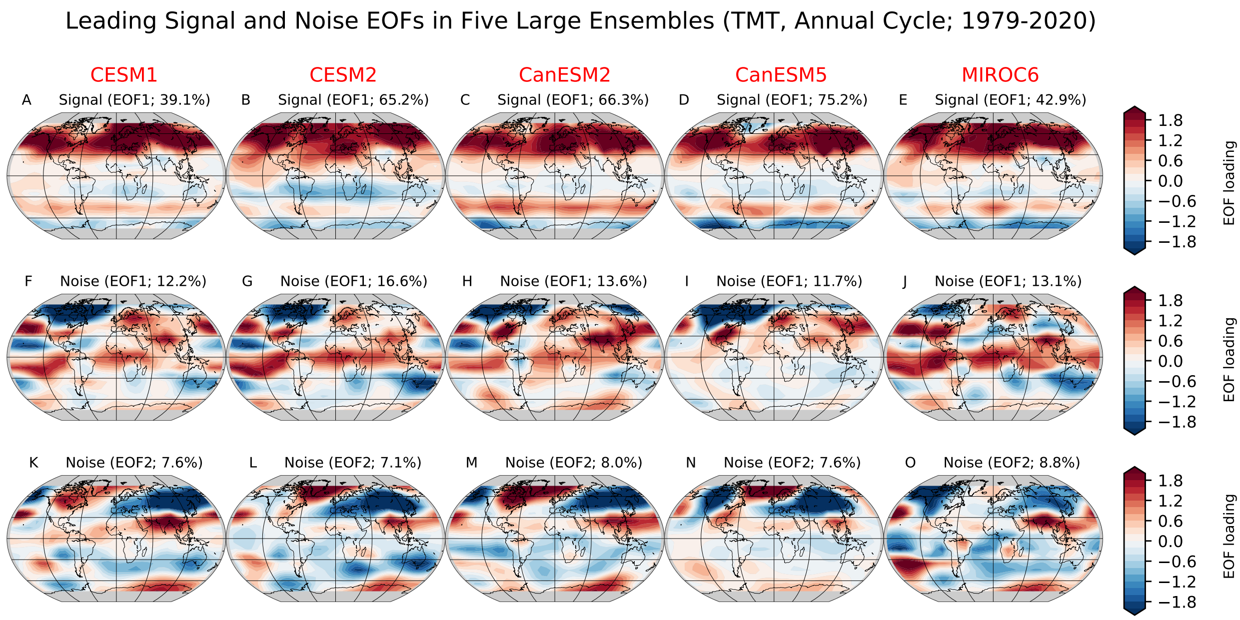

FIG. 4: Leading modes of response to external forcing and natural internal climate variability for changes in the amplitude of the annual cycle of TMT. (a-e) Fingerprints of changes in $T_{\mathrm{AC}}(x,t)$ in five LEs. The fingerprints are the leading EOF of changes in ensemble-mean $T_{\mathrm{AC}}(x,t)$ over the 42-year period from 1979 to 2020. (f-j) First EOF of natural internal climate variability of $T_{\mathrm{AC}}(x,t)$, estimated from the between-realization variability of each LE. (k-o) Second EOF of natural internal variability. The total variance explained by each EOF is listed. The grey shaded regions poleward of 80° arise because of regridding to a 10°×10° grid and masking model simulation output with observational TMT coverage (see SM).

the large-scale, zonally distinctive fingerprints and the smaller-scale noise patterns is important in explaining the fingerprint detection results described in the next section.

6. Fingerprint detection times in LEs and observationally based data

Values of SN_L used for calculating t_d are given in Fig. 5. The 1991 Pinatubo eruption has a clear effect on simulated and observed annual cycle amplitude (Santer et al. 2018), resulting in an initial dip in SN_L for analysis periods ending between 1991 and 1994. Thereafter, SN_L increases linearly with increasing L, except in CESM2 and in observational data, where SN_L exhibits relatively little change or decreases for L-year trends ending after ca. 2012 (Figs. 5d,f).

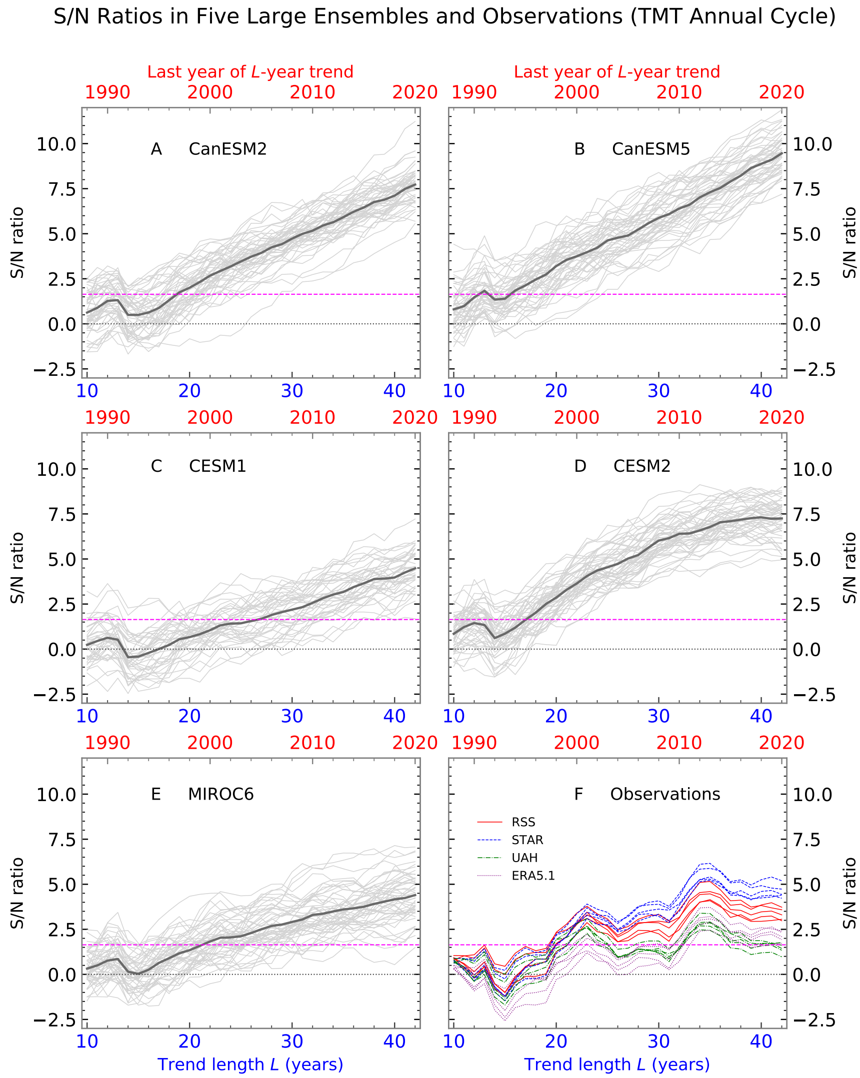

FIG. 5: Signal-to-noise ratio SN_L as a function of the trend length L. (a-e) SN_L for the strength of the model $F_{\mathrm{AC}}(x)$ fingerprints in individual realizations of $T_{\mathrm{AC}}(x,t)$ (thin grey lines) and in ensemble-mean $T_{\mathrm{AC}}(x,t)$ changes (dark grey lines). Results are from five different LEs. Model fingerprints used in panels a-e are shown in the top row of Fig. 4. For CanESM2 and CESM1 (which are both CMIP5 models), the denominator of SN_L was estimated with the unforced variability from 36 different CMIP5 pre-industrial control runs. For the CMIP6 LEs (CanESM5, CESM2, and MIROC6), the denominator of SN_L was computed with the internally generated variability from 30 different CMIP6 control integrations. (f) SN_L ratios for the strength of model fingerprints in satellite and reanalysis $T_{\mathrm{AC}}(x,t)$ data. There are five lines for each observational data set. Each line corresponds to use of a different LE for estimating the fingerprint and noise (see Fig. 4 and SM). SN_L is always plotted on the final year of the L-year analysis period, which is given in red in the upper x-axis. The trend length L is given in blue in the lower x-axis. The first analysis period is over 1979 to 1988; the final analysis period is over 1979 to 2020. The dashed horizontal magenta line is the stipulated 5% significance level used for calculating the t_d values shown in Fig. 6a.

The individual LE realizations cross the stipulated 5% significance threshold at a wide range of L values. When multi-model noise estimates are used to compute the denominator of SN_L, the median detection time in the five LEs, $t_{d\{\mathrm{med}\}}$, ranges from 1994 for CanESM5 to 2005 for CESM1 (Fig. 6a). A similar range of $t_{d\{\mathrm{med}\}}$ results is obtained by calculating the denominator of SN_L with the between-realization variability of an individual LE (Fig. 6b).

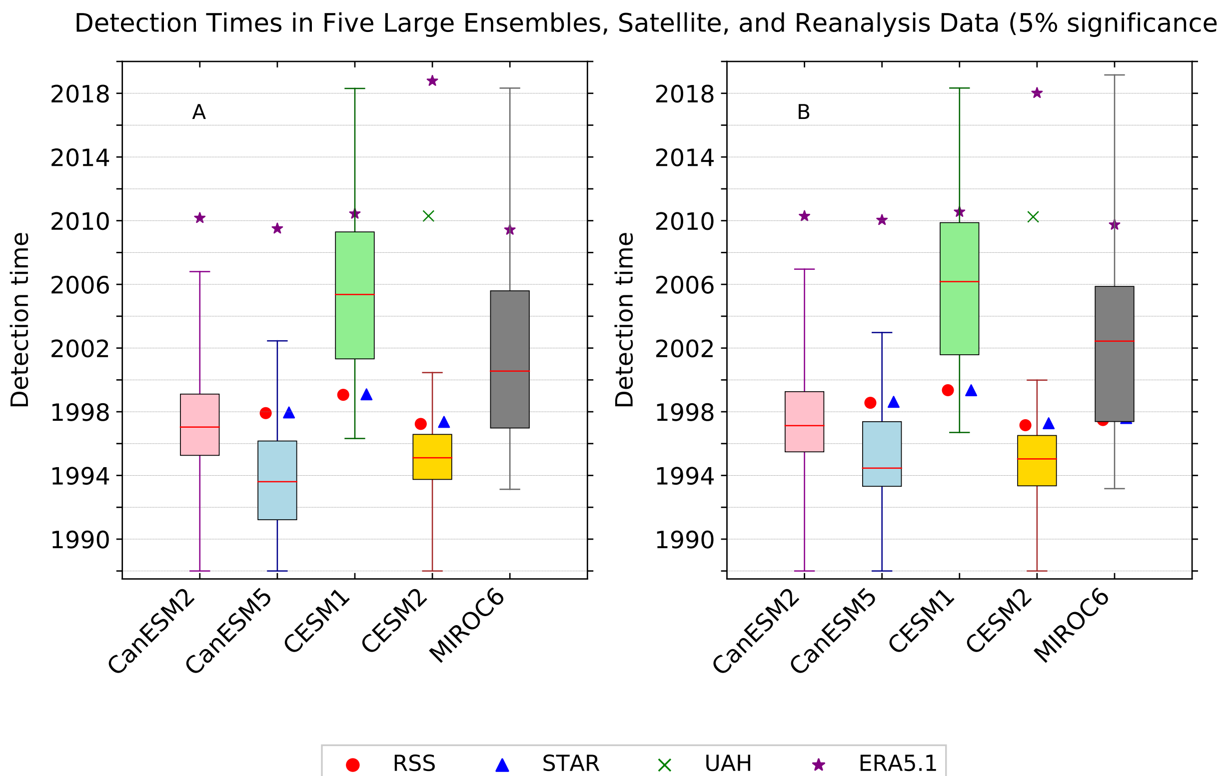

FIG. 6: Stochastic uncertainty in fingerprint detection time in model LEs (box-and-whiskers plots) and actual fingerprint detection time in satellite data (colored symbols). Detection time t_d is defined as the time at which the ratio SN_L first exceeds a stipulated significance threshold (in this case, p = 0.05) and then remains continuously above this threshold as the analysis period L increases. (a) Values of t_d estimated with fingerprints from five different LEs (see first row in Fig. 4) and using the multi-model noise from concatenated pre-industrial control runs performed with 36 CMIP5 models and 30 CMIP6 models. For details of the multi-model noise, refer to Fig. 5 and SM. (b) Fingerprints calculated as in (a), but with noise estimated using the between-realization variability of each LE. In the box-and-whisker plots in both panels, the red horizontal line is the median t_d value in the individual realizations of $T_{\mathrm{AC}}(x,t)$. The box size represents the interquartile t_d range; the whiskers span the full range of detection times in the ensemble.

For each LE, we tested whether the between-realization variability is significantly larger than the multi-model variability. Tests were performed on timescales of 10, 20, 30, and 40 years (see SM for significance test details). There were only two cases in which the between-realization variability was significantly larger at the 5% level: CanESM5 and MIROC6 (for 20- and 40-year timescales, respectively). In these two LEs, the larger single-model noise in Fig. 6b yields slightly later values of $t_{d\{\text{med}\}}$ relative to the corresponding results in Fig. 6a. Single-model noise also exceeds multi-model noise in CESM1, but is not significantly larger at the 5% level on the four timescales we examined. The single-model variability in the CanESM2 and CESM2 LEs is similar in amplitude to the CMIP5 and CMIP6 multi-model variability (respectively). Averaged across the five LEs, the median detection time is 1998.3 for the multi-model noise in Fig. 6a and 1999 for the between-realization variability in Fig. 6b.

There are two key findings from Fig. 6. First, despite model differences in external forcings, equilibrium climate sensitivity (ECS), and the amplitude of MIV (Andrews et al. 2012; Zelinka et al. 2014, 2020; Pallotta and Santer 2020; Fyfe et al. 2021; Po-Chedley et al. 2021), the $F_{\text{AC}}(x)$ patterns in the five LEs are robustly identifiable at the 5% significance level in individual model realizations of satellite-era annual cycle changes. Positive detection occurs in 239 out of 240 cases if multi-model noise is used to calculate the denominator of SN_L and in the same number of cases if single-model noise is employed.[iv]

The second key finding is that the model-predicted $F_{\text{AC}}(x)$ fingerprints are identifiable at the 5% level in 16 out of 20 different combinations of the 5 fingerprints (derived from the 5 LEs) and the 4 observational data sets. This holds for both the multi-model noise in Fig. 6a and the single-model noise in Fig. 6b. The null results in Figs. 6a and b are for the UAH data set. All five fingerprints yield S/N ratios in UAH $T_{\text{AC}}(x,t)$ data that initially exceed the stipulated 5% significance threshold on timescales of $\approx$ 35 years, but then fall below this threshold for UAH S/N ratios calculated over the full satellite era (except in the case of the CESM2 fingerprint; see Fig. 5f).

Finally, we note that removal of all global-mean information from our S/N analysis, as described in Santer et al. (2018), has minimal impact on the detection time results in Fig 6. This illustrates that the identification of model-predicted $F_{\text{AC}}(x)$ patterns in observational data and in individual LE realizations is not solely driven by global-mean changes in annual cycle amplitude – it primarily reflects similarity of large-scale pattern information (see Fig. S1 and Section 5b of SM).

In the following, we refer to the t_d results in Fig. 6b as the "baseline" case. In Section 8, we report on tests which explore the sensitivity of the baseline detection times to use of low- and high-variability subsets of the single-model noise used in Fig. 6b. These subsets of the 240 realizations

[iv]The realization in which the fingerprint cannot be detected is from the MIROC6 LE.

of internal $T_{\mathrm{AC}}(x,t)$ variability are selected based on the power spectral density (PSD) of the model AMO and Niño 3.4 SST time series.

7. Comparison of simulated and observed internal variability spectra

The robust detection of model-predicted $F_{\mathrm{AC}}(x)$ fingerprints in observations and in individual LE realizations has multiple interpretations. Under one interpretation, large-scale forcing by greenhouse gases drives large-scale physical processes that are common to observations and climate models. These processes include summertime drying of mid-latitude continental interiors (Manabe et al. 1981; Wetherald and Manabe 1995; Douville and Plazzotta 2017), expansion of the tropics (Seidel and Randel 2007; Hu and Fu 2007; Quan et al. 2014), and lapse-rate changes (Frierson 2006; Donohoe and Battisti 2013). In contrast, modes of MIV are characterized by smaller-scale patterns of anticorrelated variability that do not project well onto the coherent $F_{\mathrm{AC}}(x)$ patterns (see Fig. 4). This basic difference in the spatial scales of the forced response and MIV favors signal detection (Santer et al. 1994).

A second possible interpretation is that robust detection of model $F_{\mathrm{AC}}(x)$ fingerprints is biased by errors in model representation of MIV (Curry and Webster 2011; O'Reilly et al. 2021). Under this interpretation, models systematically underestimate "observed" MIV, thereby spuriously inflating SN_L and leading to incorrect fingerprint detection claims. This "biased variability" argument is challenging to address because there are large uncertainties in separating externally forced signals from MIV in the single occurrence of signal and noise available in observations (Frankcombe et al. 2015; Kravtsov 2017; Cheung et al. 2017; Kajtar et al. 2019; Pallotta and Santer 2020). This introduces uncertainty in determining the size and significance of model MIV errors.

These two interpretations are not mutually exclusive. We have already shown credible evidence that the first interpretation – dissimilarity of signal and noise patterns – contributes to our high success rate in identifying model $F_{\mathrm{AC}}(x)$ fingerprints in individual LE realizations (see Figs. 4 and 6). In the current section, we consider the plausibility of the second interpretation of our results. In doing so, we make use of the fact that the climate change signals in LEs can be reliably estimated by averaging over many realizations.

We assume that these well-estimated signals, obtained from LEs generated using models with different ECS, MIV, and historical external forcings, encapsulate a significant portion of the true uncertainty in the amplitude and time evolution of forced changes in real-world climate. We apply a regression-based approach (see below) to remove these LE-derived signals from observed time series of three major modes of MIV – the AMO, ENSO, and IPO. Regression-based signal removal is not required in model LEs. The ensemble-mean signal of a given LE is a reasonable estimate of forced changes in that LE, and is simply subtracted from each realization of the LE.

Signal removal in the LEs and observations allows us to isolate the internally generated component of variability in the AMO, ENSO, and IPO time series. We calculate PSD from the "signal removed" residual time series, thus facilitating the direct comparison of simulated and observed MIV. We seek to determine whether there is evidence that the five LEs analyzed here significantly underestimate the observed MIV of the AMO, ENSO, and IPO (Kajtar et al. 2019). Such an error could provide support for the second interpretation of our fingerprint detection results – particularly if the detection time for $F_{\text{AC}}(x)$ fingerprints is sensitive to large inter-model and inter-realization differences in the amplitude of AMO and ENSO variability. Whether such sensitivity exists is explored in Section 8.

Consider results for the AMO first. The amplitude and time evolution of ensemble-mean SST changes in the AMO region varies markedly across the five LEs (Figs. 7a-e). This is unsurprising given model differences in ECS and in direct and indirect anthropogenic aerosol forcings (Zelinka et al. 2014, 2020; Santer et al. 2019).[v] All five ensemble-mean signals show overall SST increases in the AMO region, punctuated by recovery from surface cooling caused by major volcanic eruptions. The SST increases are temporally complex and poorly captured by a linear trend.

Inter-model differences in the median detection time for $F_{\text{AC}}(x)$ fingerprints (Fig. 6) show some correspondence with inter-model differences in the ensemble-mean AMO signal time series in Fig. 7. CanESM5, for example, which has the earliest $t_{d\{\text{med}\}}$ values in Fig. 6, also has the largest and most rapid SST increase in the AMO region (Fig. 7b). Similarly, the smaller and more gradual SST increase in the CESM1 AMO signal appears to be related to the later $t_{d\{\text{med}\}}$ values in CESM1 (compare Figs. 7c and 6).

Removing the ensemble-mean forced SST signals from individual realizations of an LE yields residual AMO variability that is smallest in amplitude in CESM1 and largest in CanESM5 (Figs. 8a-e). Subtracting the unscaled ensemble-mean model signals from observed HadCRUT4 data can produce residuals with large low-frequency variability, primarily because of mismatches between

[v]ECS is 3.7°C and 5.6°C in CanESM2 and CanESM5, 4.0°C and 5.1°C in CESM1 and CESM2, and 2.6°C in MIROC6.

model ECS and the true (but uncertain) real-world ECS (Frankcombe et al. 2015). Model forcing errors also contribute to this large residual variability, thus inflating estimates of "observed" MIV associated with the AMO.

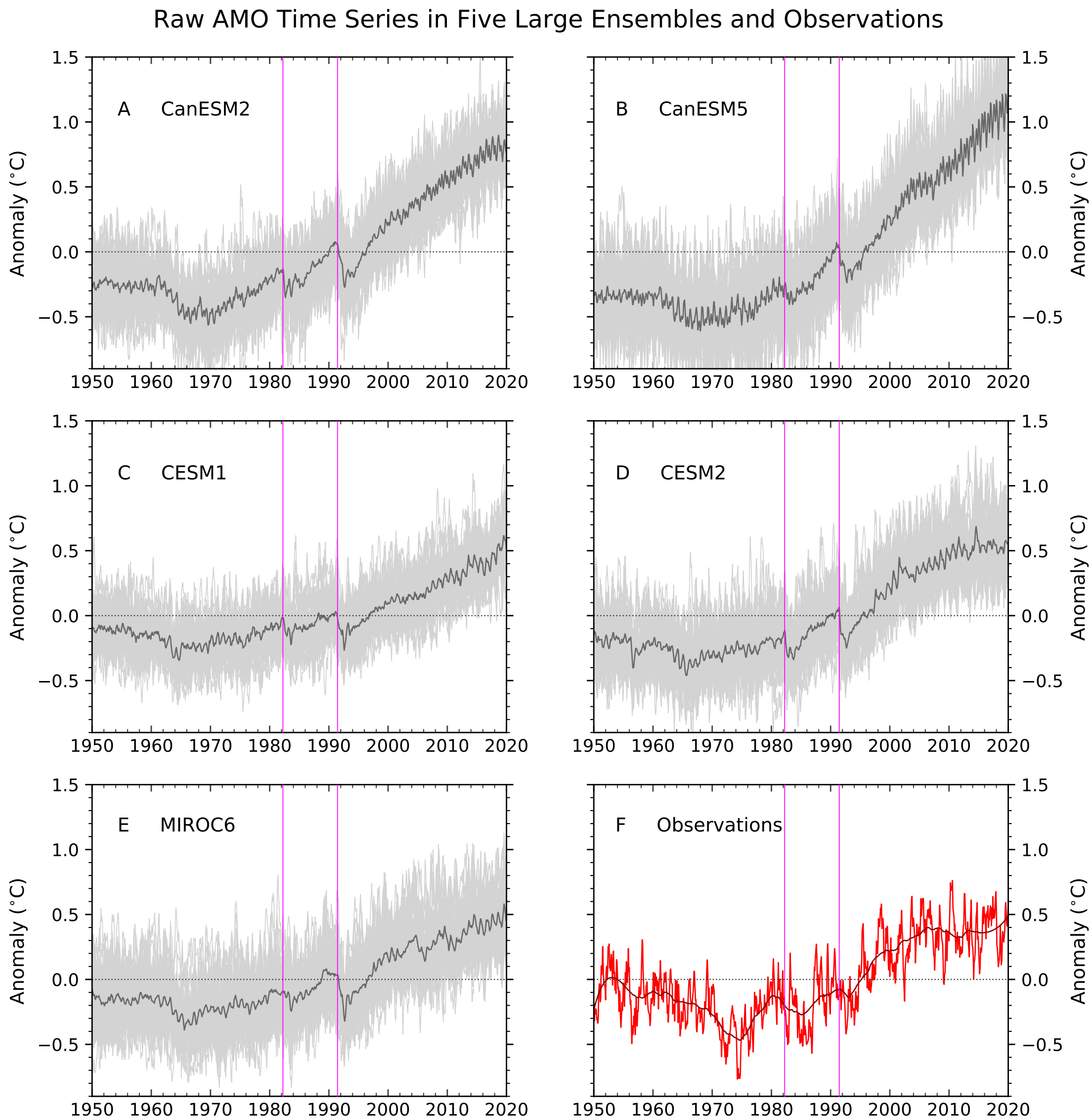

FIG. 7: Simulated and observed time series of the Atlantic Multidecadal Oscillation (AMO). Results are for SST changes spatially averaged over 0°–60°N and 80°W–0° (see Enfield et al. 2001, and SM). (a-e) AMO time series calculated from individual realizations (light grey) and multi-model averages (dark grey) of five LEs. (f) Raw (red) and filtered (dark red) AMO time series calculated from HadCRUT4 SST data. A Savitzky-Golay filter was applied to smooth the observations. The filter used a window width of 141 months and a third-order polynomial. The vertical magenta lines denote the eruption dates of El Chichón in March 1982 and Pinatubo in June 1991.

We therefore subtract scaled model AMO signals from observations (Frankcombe et al. 2015; Steinman et al. 2015). Scaling involves $Y(t) = a + b\overline{X}(t) + \epsilon(t)$, the regression between the observed AMO time series, $Y(t)$, and $\overline{X}(t)$, the ensemble-mean AMO time series for an individual LE. The residual $\epsilon(t)$ is the "signal removed" AMO time series. Subtraction of $b\overline{X}(t)$ from the HadCRUT4 AMO time series markedly damps the residual low-frequency variability (Figure 8f). For example, at 284 months (23.7 years), regression-based removal of scaled AMO signals decreases the observed PSD range by 92% relative to the range obtained with unscaled signal subtraction (not shown).

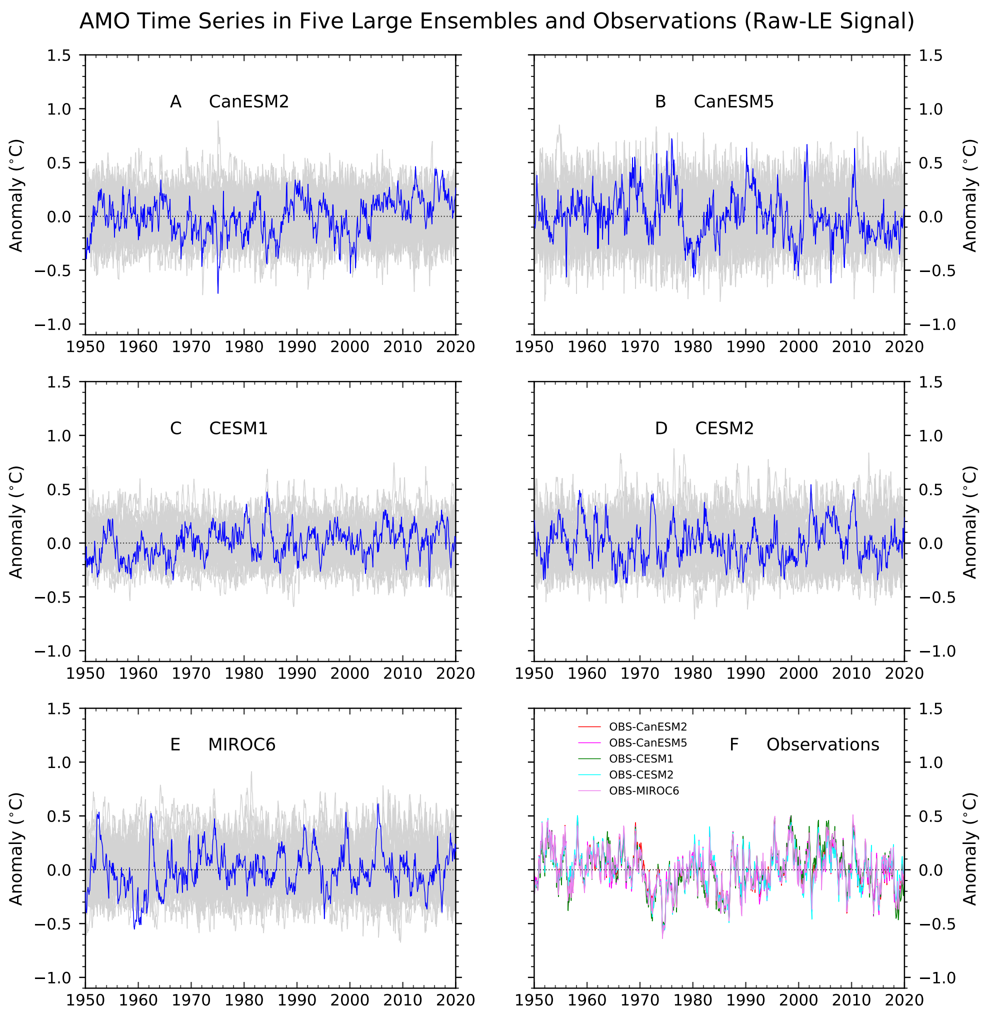

Fig. 8: Simulated and observed time series of the Atlantic Multidecadal Oscillation (AMO) after removing externally forced SST signals. (a-e) "Signal removed" AMO time series (thin grey lines) after subtracting ensemble-mean AMO SST changes in a given LE from each realization of the LE. The blue line is the "signal removed" time series for the last realization in the LE. (f) Observed "signal removed" time series. The five ensemble-mean AMO signal time series in Figs. 7a-e were each subtracted from the HadCRUT4 AMO time series using regression-based scaling.

Simulated and observed "signal removed" spectra for AMO SSTs are shown in Fig. 9. While the observed spectrum and the spectra for both CanESM models are well-described by simple power law fits, the CESM models and MIROC6 exhibit more complex spectral shape, with noticeable flattening of PSD at periods greater than 100 months. Of greatest interest here is the comparison of P_{LOW}, the PSD at 284 months. This is the longest period that can be usefully resolved from the 852 months (71 years) of the observed AMO and Niño 3.4 SST time series. Systematic model underestimation of observed P_{LOW} has the potential to spuriously inflate the signal-to-noise ratio SN_L, thereby biasing fingerprint detection times towards earlier and more ubiquitous detection.

We compare simulated and observed P_{LOW} in two ways. First, we determine the total number of model realizations in the five LEs with P_{LOW} values exceeding the smallest of the observed P_{LOW} values in Fig. 9f (see bottom edge of red bands). Second, for each LE, we determine the number of realizations in that LE with P_{LOW} values exceeding the corresponding observed P_{LOW} value.[vi] We refer to these two comparisons subsequently as Method 1 and Method 2 (respectively). They are simple measures of the consistency between simulated and observed low-frequency PSD.[vii]

For the AMO, Method 1 and Method 2 yield 56 and 50 realizations exceeding observed P_{LOW} (23% and 21% of the total number of realizations).[viii] We conclude from this that the five model LEs analyzed here show evidence of underestimating the amplitude of observed low-frequency AMO variability (Kajtar et al. 2019), but that this underestimate is not statistically significant at the 5% level. If it were, we would expect a smaller fraction of model exceedances of observed P_{LOW} (5% or less).

Qualitatively and quantitatively different results are obtained for SST variability in the Niño 3.4 region of the tropical Pacific (Fig. 10). SST changes in this region are a common proxy for ENSO variability. Fluctuations in ENSO have substantial impact on global surface temperature (Kosaka and Xie 2013), tropospheric temperature (Po-Chedley et al. 2021), and many other climatic variables (Bonfils et al. 2015).

SST variability in the Niño 3.4 region is markedly larger than in the AMO region (c.f. Figs. 10 and 7), so that even with ensemble sizes of 40 to 50 realizations, there is still substantial residual noise in the ensemble-mean Niño 3.4 SST time series (Figs. 10a-e). This noise displays power at

[vi] For example, if P_{LOW} in CanESM2 is being evaluated, we compare P_{LOW} values in individual CanESM2 realizations with the observed P_{LOW} estimated by subtraction of the ensemble-mean CanESM2 AMO signal from the HadCRUT4 AMO time series.

[vii] See Pallotta and Santer (2020) for more sophisticated PSD comparisons.

[viii] For both methods, most of the model realizations exceeding P_{LOW} are from CanESM5, CESM2, and MIROC6 (see Figs. 9b,d,e).

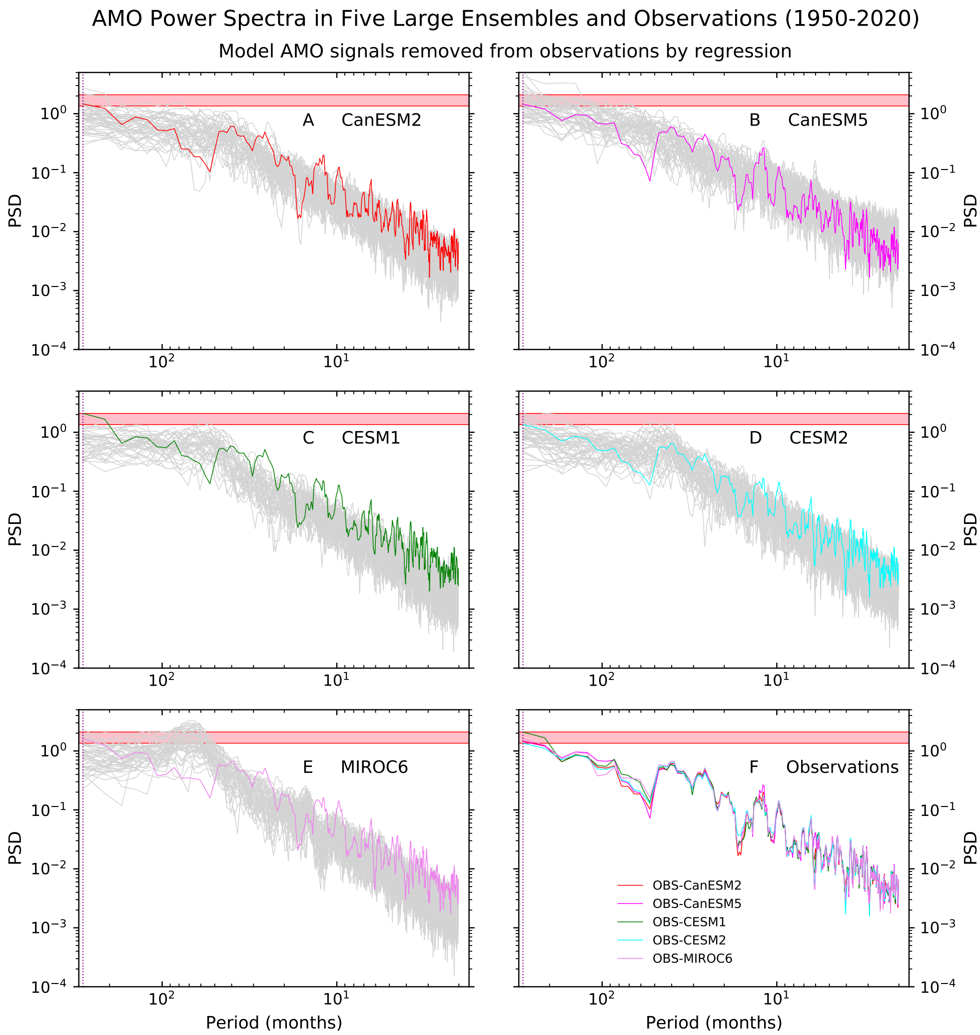

FIG. 9: Power spectral density (PSD) in simulated and observed AMO time series. (a-e) PSD in individual realizations (grey lines) of "signal removed" AMO time series shown in Figs. 8a-e. (f) PSD in five "signal removed" observed AMO time series. The (scaled) forced component of AMO SST changes for each LE was subtracted from the HadCRUT4 AMO time series. Individual observed "signal removed" AMO time series in panel f are also plotted in panels a-e for their corresponding LE (i.e., for the LE used to estimate and subtract an AMO signal from observations). The red horizontal band delimits the lowest and highest values of PSD at a period of 284 months in the five "signal removed" observational spectra. The vertical dotted purple line at the left of each panel corresponds to this 284-month period (see SM for further technical details).

a period of 12 months, most clearly in MIROC6 (Figs. 11a-e). This residual power is consistent with a change over the satellite era in the seasonal cycle of Niño 3.4 SSTs.

All five LEs have small positive warming trends in their ensemble-mean Niño 3.4 time series. Observed warming in this region is more muted (Fig. 10f), partly due to the phasing of ENSO and

IPO variability over 1950 to 2020 (Kosaka and Xie 2013; Trenberth 2015; Meehl et al. 2011, 2016; England et al. 2014; Fyfe et al. 2016; Po-Chedley et al. 2021).

Because of the relatively small externally forced component in simulated Niño 3.4 SST changes and the large residual noise in this component, model ensemble-mean Niño 3.4 SST time series are only weakly correlated with the raw observed Niño 3.4 SST time series, with r ranging from 0.02 in MIROC6 to 0.17 in CESM1.[ix] Scaling and subtraction of these Niño 3.4 SST signals from observations has only small impact on the original observed Niño SST time series, yielding the spectra shown in Fig. 11f.

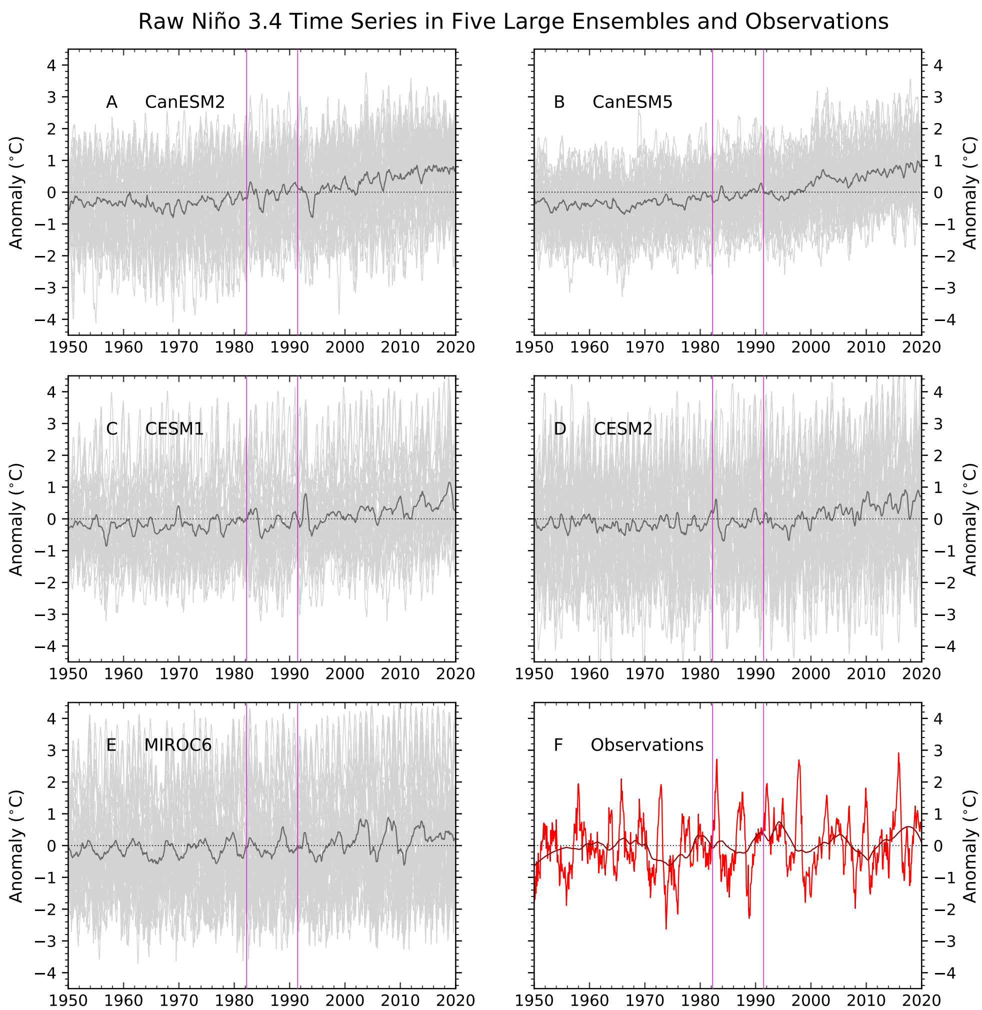

FIG. 10: As for Fig. 7 but for simulated and observed time series of SST spatially averaged over the Niño 3.4 region (5°N-5°S, 120°W-170°W).

[ix]For the corresponding calculation with AMO SST time series, r ranges from 0.78 for CESM1 to 0.81 for CESM2.

All simulated and observed Niño 3.4 SST spectra in Fig. 11 have a discrete peak within the canonical 3- to 7-year range of ENSO variability (AchutaRao and Sperber 2002). This peak is more narrowly defined in MIROC6 than in the other LEs or observations. Simulated Niño 3.4 spectra show a noticeable decrease in PSD for periods longer than approximately 7 years. This PSD decrease is less pronounced in observations. In contrast to the AMO results, Methods 1 and 2 yield 185 and 178 exceedances of observed P_{LOW} – i.e., 77% and 74% of the LE realizations have power at 284 months that is higher than in observations. There is no evidence from our analysis, therefore, that the LEs examined here systematically underestimate the observed low-frequency variability of ENSO. This is consistent with other findings (Lienert et al. 2011).

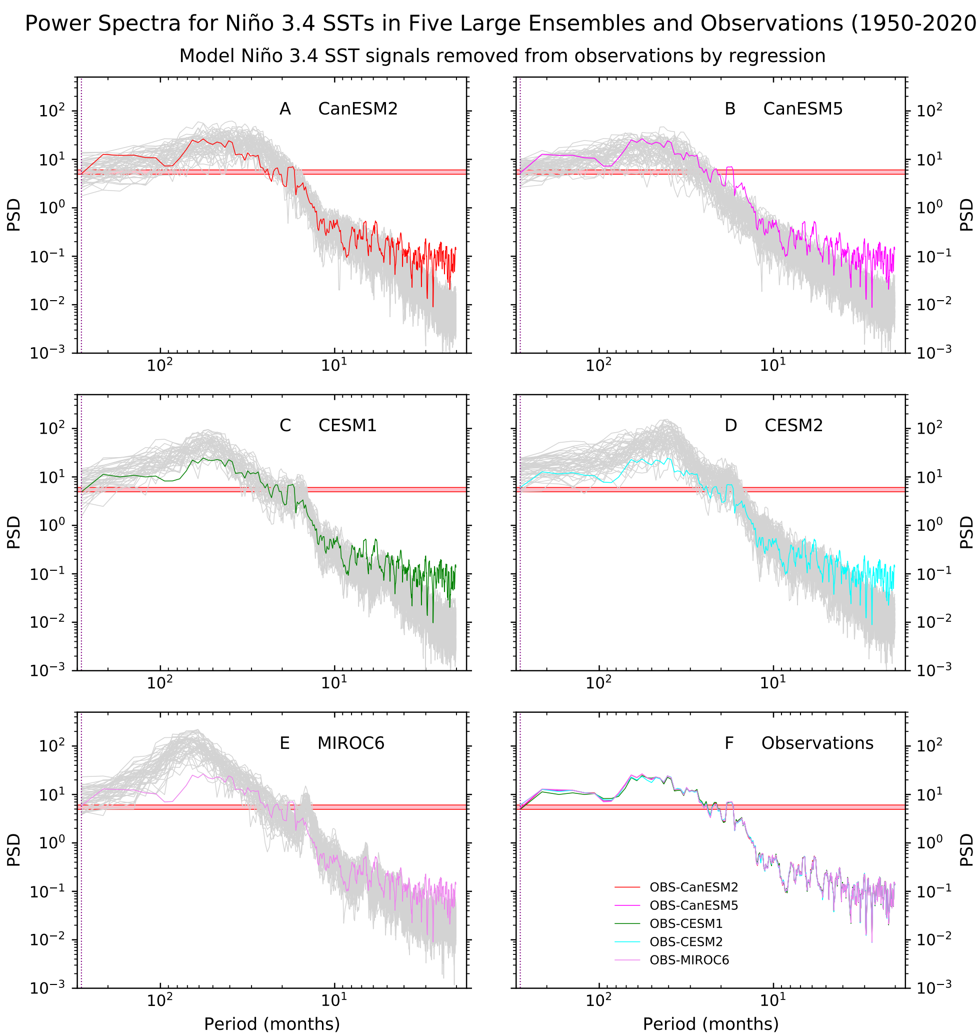

FIG. 11: As for Fig. 9 but for spectra of simulated and observed "signal removed" Niño 3.4 SST time series.

An analysis of the IPO (not shown) leads to a similar conclusion. Unlike Niño 3.4 SSTs, the IPO is influenced by both the tropical and extratropical variability of Pacific SSTs (Meehl et al. 2016; Trenberth 2015; Henley et al. 2015, 2017). For the IPO, we find 116 and 101 exceedances of observed P_{LOW} for Methods 1 and 2, corresponding to 48% and 42% of LE realizations with low-frequency PSD that is larger than in the "signal removed" observations (Kajtar et al. 2019). Possible implications of such simulated and observed P_{LOW} differences for fingerprint detection time are explored in the next section.

8. Detection time sensitivity tests

Other previously published studies considered the links between fingerprint detection and model performance in simulating observed global-scale variability (Hegerl et al. 1996; Allen and Tett 1999) or investigated the sensitivity of D&A results to large inter-model differences in variability (Santer et al. 2009; Sippel et al. 2021). There have, however, been few studies of links between detection time results and the behavior of individual modes of MIV.

We explore these links here using sensitivity tests (Fig. 12). We repeat the "baseline" S/N analysis shown in Fig. 6b with two 50-member subsets of the 240 individual samples of between-realization $T_{\mathrm{AC}}(x,t)$ variability. These two 50-member subsets[x] correspond to low- and high-amplitude variability of a specific mode of MIV at a specific timescale. The mode amplitude is estimated from the spectra of "signal removed" time series (see Figs. 9a-e and Figs. 11a-e). There are four separate sensitivity tests, one for each mode (the AMO and ENSO) and each timescale of interest (284 months and 70 months). The procedure for conducting these sensitivity tests is described in detail in the SM.

Recall that the internal variability of $T_{\mathrm{AC}}(x,t)$ is used to calculate the denominator of our S/N ratios, which in turn are used to estimate fingerprint detection times (Section 6). Comparing detection times obtained for $T_{\mathrm{AC}}(x,t)$ subsets – with subsetting based on the low and high PSD values of key modes of MIV – allows us to explore possible links between the simulated mode amplitude and our D&A results.

Our analysis timescales of 284 months and 70 months (23.7 and 5.8 years, respectively) were selected for the following reasons. Detection of a slowly-evolving externally forced fingerprint requires information on the background noise of MIV. Given 852-month (71-year) record lengths

[x]For each subset, there are 10 members from each LE. This reduces the impact on the sensitivity test of MIV biases in a single LE.

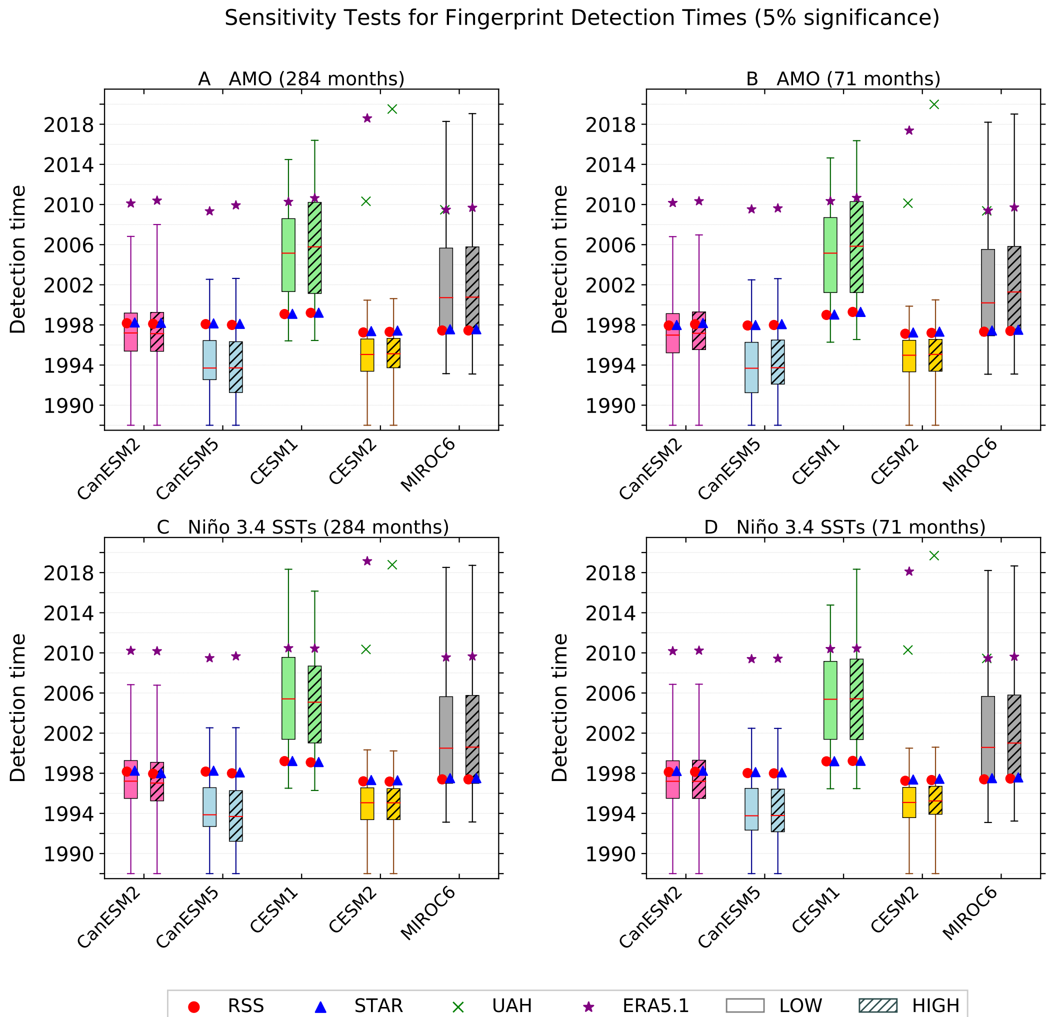

FIG. 12: Stochastic uncertainty in fingerprint detection time t_d in model LEs (box-and-whiskers plots) and actual fingerprint detection time in satellite data (colored symbols). Results are for sensitivity tests involving the selection of 50-member subsets from the 240 realizations of unforced $T_{\mathrm{AC}}(x,t)$ variability. (a-b) Partitioning of internal $T_{\mathrm{AC}}(x,t)$ variability into low- and high-variability subsets is based on the PSD values at 284 and 70 months in spectra calculated from "signal removed" AMO time series (panels a and b, respectively). (c-d) As for panels a and b but for the use of spectra from simulated "signal removed" Niño 3.4 SST time series. See Section 8 and SM for further information on sensitivity tests. The caption of Fig. 6 provides details of box-and-whiskers plots. The shaded bars in each panel display t_d results for high-variability subsets of $T_{\mathrm{AC}}(x,t)$. Unshaded bars show t_d for low-variability $T_{\mathrm{AC}}(x,t)$ subsets.

for the AMO and Niño 3.4 SST time series, the longest noise timescale we can usefully resolve is 284 months. The choice of the shorter 70-month timescale was driven by the presence of a spectral peak close to this period in the "signal removed" MIROC6 AMO and Niño 3.4 SST time series (see Figs. 9e and 11e).

On both of timescales considered here, and for both the AMO and Niño 3.4 SSTs, the average PSD is typically a factor of 3-4 larger in the high-variability subset of spectra than in the low-variability subset. This indicates that for each mode and timescale, the amplitude differences between the high- and low-variability subsets are sufficiently large to justify investigating the implications of these differences for unforced $T_{\mathrm{AC}}(x,t)$ variability and fingerprint detection time.

Our sensitivity tests yield three main results (Fig. 12). First, in each sensitivity test and for each LE, the "low PSD" and "high PSD" subsets of unforced $T_{\mathrm{AC}}(x,t)$ variability yield similar values of the median detection time $t_{d\{\mathrm{med}\}}$, with $t_{d\{\mathrm{med}\}}$ differences < 1 year. Second, the "baseline" $t_{d\{\mathrm{med}\}}$ results in Fig. 6b are relatively unaffected by repeating the D&A analysis with "low PSD" and "high PSD" subsets of the original 240 realizations of unforced $T_{\mathrm{AC}}(x,t)$ variability. All sensitivity tests preserve the relative differences in $t_{d\{\mathrm{med}\}}$ found in the "baseline" case – e.g., the earliest fingerprint detection is still in CanESM5 and the latest detection is still in CESM1. Third, the model-predicted $F_{\mathrm{AC}}(x)$ fingerprints are statistically identifiable in 75% of the 160 sensitivity tests in Fig. 12 that involve satellite and reanalysis data.[xi]

Figure S2 in the SM shows SN_L for one of the four sensitivity tests: selecting subsets of unforced $T_{\mathrm{AC}}(x,t)$ variability based on PSD at 284 months in the simulated AMO spectra (Figs. 9a-e). In all five LEs, the "low PSD" subset yields larger S/N ratios (relative to the "high PSD" subset) for analysis periods longer than $\approx$ 25-30 years (Figs. S2a-e). This means that low-amplitude AMO variability at 284 months tends to correspond to lower-amplitude multidecadal $T_{\mathrm{AC}}(x,t)$ variability, which damps the denominator of S/N and increases S/N ratios. Conversely, high-amplitude AMO variability at 284 months tends to correspond to higher-amplitude multidecadal $T_{\mathrm{AC}}(x,t)$ variability, thereby decreasing S/N ratios. Qualitatively similar "low PSD-versus-high PSD" differences in SN_L are also found for the other three sensitivity tests (not shown).

The results in Fig 12 and in Fig. S2 raise several questions. The first question is why the "low PSD-versus-high PSD" S/N differences in Figs. S2a-e have relatively small impact on $t_{d\{\mathrm{med}\}}$. The answer is that these S/N differences are small for $L <\approx$ 25-30 years. This explains why the median detection times in Fig. 12a are so similar in the "low PSD" and "high PSD" cases, particularly for CanESM2, CanESM5, and CESM2. In these three models, the S/N ratios for almost all individual realizations exceed the 5% significance threshold in less than 30 years, well before the "low PSD-versus-high PSD" S/N differences become pronounced.

[xi] 160 = 4 satellite data sets × 5 different fingerprints × 2 variability subsets (low PSD and high PSD) × 4 sensitivity tests.

The second question is why are our "baseline" fingerprint detection times are robust to partitioning the original 240 realizations of unforced $T_{\mathrm{AC}}(x,t)$ variability into "low PSD" and "high PSD" subsets. Recall that the annual cycle fingerprints in the five LEs are spatially uncorrelated with the dominant $T_{\mathrm{AC}}(x,t)$ noise modes (Fig. 4). This was true for both the multi-model CMIP5 and CMIP6 noise and for the single-model between-realization variability in each LE. Quasi-orthogonality of fingerprint and noise patterns also applies to the noise subsets in all of our "low PSD" and "high PSD" sensitivity tests. Because fingerprint and leading noise patterns are so dissimilar, differences in the amplitude of unforced $T_{\mathrm{AC}}(x,t)$ variability associated with low- and high-amplitude behavior of the AMO and ENSO have relatively small impact on $t_{d\{\mathrm{med}\}}$.

Put differently, our fingerprint analysis reveals coherent, global-scale externally forced responses common to all five LEs. Examples include decreases in $T_{\mathrm{AC}}(x,t)$ over the Arctic and mid-latitude $T_{\mathrm{AC}}(x,t)$ increases in NH continental interiors (Figs. 4a-e). These distinctive features are absent in patterns of unforced $T_{\mathrm{AC}}(x,t)$ fluctuations associated with the AMO, ENSO, and other modes, which are characterized by variability at smaller spatial scales (Figs. 4f-o). This mismatch between the spatial scales of fingerprint and noise helps to explain why inter-model and inter-realization differences in the amplitude of key modes of MIV have limited impact on $t_{d\{\mathrm{med}\}}$.

9. Annual cycle changes in aquaplanet simulations

Santer et al. (2018) discussed some of the possible physical mechanisms involved in producing the distinctive patterns of observed and simulated $T_{\mathrm{AC}}(x,t)$ changes shown in Fig. 1. They noted that there are pronounced hemispheric asymmetries in both the climatological mean state of $T_{\mathrm{AC}}(x,t)$ and in its satellite-era trends. Climatological asymmetries in $T_{\mathrm{AC}}(x,t)$ are related to NH-versus-SH differences in land fraction, heat capacity (through the differences in land fraction), and sea ice coverage. Hemispheric asymmetries in the $T_{\mathrm{AC}}(x,t)$ trends over 1979 to 2020 are influenced not only by these factors, but also by hemispherically asymmetric external forcings. Examples of the latter include anthropogenic aerosol forcing (Bonfils et al. 2020; Kang et al. 2021) and the forcing and circulation response associated with stratospheric ozone depletion (see Fig. S3 and Gillett et al. 2004; Thompson et al. 2011; Bandoro et al. 2014; Randel et al. 2017; Solomon et al. 2017). Low-frequency changes in modes of internal variability may also contribute to variations in the Hadley circulation (Mantsis and Clement 2009) and are another possible influence on $T_{\mathrm{AC}}(x,t)$.

One of the most prominent aspects of the patterns in Fig. 1 is the increase in annual cycle amplitude at mid-latitudes in both hemispheres, with larger increases in the NH than the SH. These "ridges" in $T_{\mathrm{AC}}(x,t)$ trends arise from larger tropospheric warming in the summer hemisphere (Santer et al. 2018). Possible causes of these features include changes in the meridional temperature gradient or in atmospheric SW absorption that result in seasonally-dependent changes in stability (Frierson 2006; Donohoe and Battisti 2013; Santer et al. 2018), poleward expansion of the Hadley circulation and the tropics (Held 2000; Fu et al. 2006; Seidel and Randel 2007; Frierson et al. 2007; Kang and Liu 2012; Quan et al. 2014), lapse-rate changes unrelated to tropical expansion (Brogli et al. 2019), and summertime drying of the land surface (Manabe et al. 1981; Wetherald and Manabe 1995; Douville and Plazzotta 2017). Other factors may also be relevant, such as the response to land-sea warming contrast, the direct radiative effects of CO_2, and SST trend patterns (He and Soden 2017). These explanations are not mutually exclusive.

To explore the influence of land and ice albedo on $T_{\mathrm{AC}}(x,t)$ changes, we analyzed existing aquaplanet simulations performed with GFDL-AM2.1 (Feldl et al. 2017) and new simulations with CESM2-CAM6. These numerical experiments involve running an atmospheric model in aquaplanet configuration with a realistic seasonal cycle of insolation, a 30-meter fixed-depth slab ocean, and quadrupled CO_2. A key difference is that CESM2-CAM6 includes sea-ice thermodynamics; GFDL-AM2.1 does not. In both sets of simulations, parameters influencing ice albedo were systematically varied in order to evaluate the effect of sea-ice changes on atmospheric heat transport and feedback strength. We show results for one selected value of these parameters. Results for other values are qualitatively similar (see SM and Figs. S4 and S5).

In GFDL-AM2.1, annual-mean TMT changes between the $4\times CO_2$ and control simulations are largest in the tropics (Fig. 13a), where the net feedback in the simulations is positive and large (Feldl et al. 2017). The largest annual-mean TMT changes in CESM2-CAM6 occur in high-latitude regions of pronounced sea ice extent decrease (Fig. 13c). In terms of annual cycle changes, the most salient feature of Figs. 13b,d is that even without land and land-ocean warming contrasts, the aquaplanet simulations capture the mid-latitude increases in $T_{\mathrm{AC}}(x,t)$ evident in satellite data and in ESMs with realistic geography (Fig. 1). Unlike the observations and ESMs, however, these mid-latitude "ridges" are more hemispherically symmetric in the aquaplanet runs. Relative to GFDL-AM2.1, mid-latitude $T_{\mathrm{AC}}(x,t)$ increases are larger and further poleward in CESM2-CAM6.

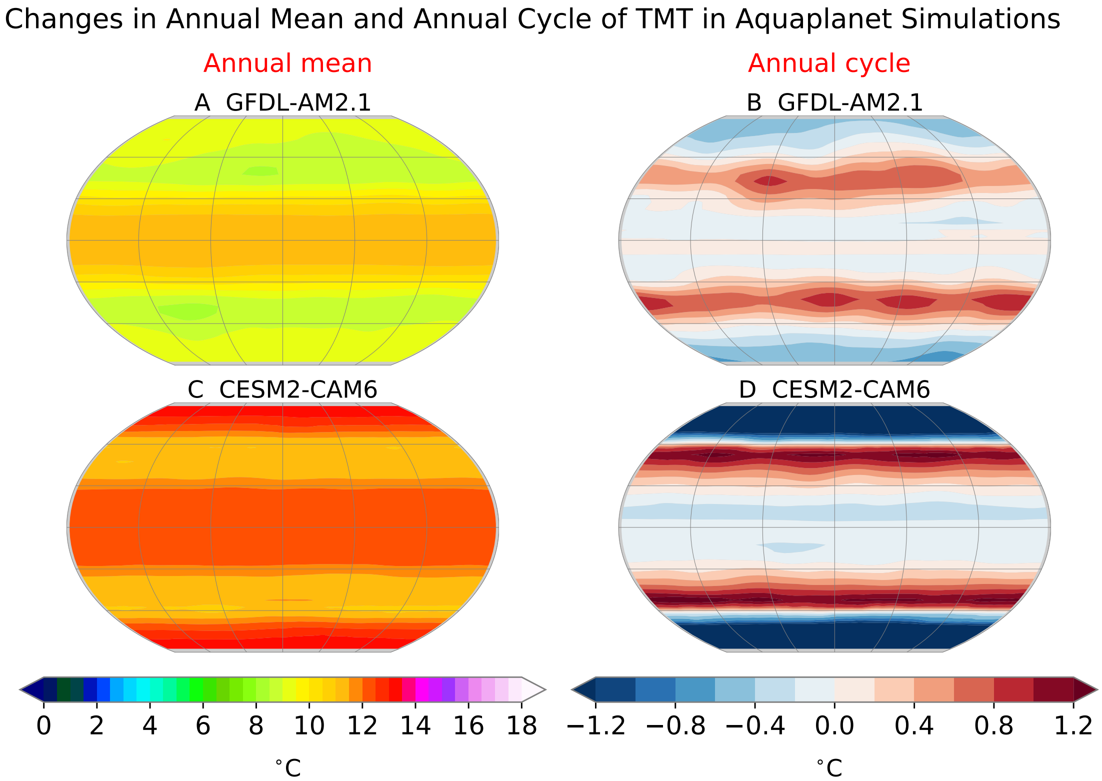

FIG. 13: Changes in uncorrected TMT (°C) in aquaplanet simulations. (a-b) Simulations with GFDL-AM2.1 (Feldl et al. 2017). (c-d) Simulations with CESM2-CAM6. (left column) Annual-mean TMT changes. (right column) Changes in the amplitude of the annual cycle of TMT. In GFDL-AM2.1, ocean albedo was set to values of 0.45 at grid-points where the surface temperature was less than 270K. In CESM2-CAM6, the parameter used for tuning snow albedo, r_snw, was set to 0.7. Changes in the annual mean and annual cycle of TMT were calculated by differencing climatologies computed from averages of a 4×CO_2 experiment and a control run with pre-industrial atmospheric CO_2. The climatologies are of length 30 years for GFDL-AM2.1 and 100 years for CESM2-CAM6 (see SM).

We draw three inferences from these results. First, they suggest that in observations and ESMs, the zonal structure of mid-latitude increases in $T_{AC}(x,t)$ is partly driven by GHG-induced changes in static stability and Hadley circulation that are superimposed on the climatological seasonal cycle of the thermal equator, ITCZ location, and Hadley cell poleward edge (Frierson 2006; Kang and Liu 2012; Donohoe and Battisti 2013). Second, Fig. 13 (right column) implies that the presence of realistic geography contributes to the observed and ESM-simulated hemispheric asymmetry in mid-latitude $T_{AC}(x,t)$ trends, likely through the combined effect of summertime drying over land (Manabe et al. 1981; Wetherald and Manabe 1995; Douville and Plazzotta 2017) and hemispheric differences in land fraction and heat capacity. Third, relative to the observations and ESMs, larger changes in annual mean TMT in the aquaplanet simulations yield proportionately smaller mid-latitude increases in annual cycle amplitude. The reasons for this are unclear.

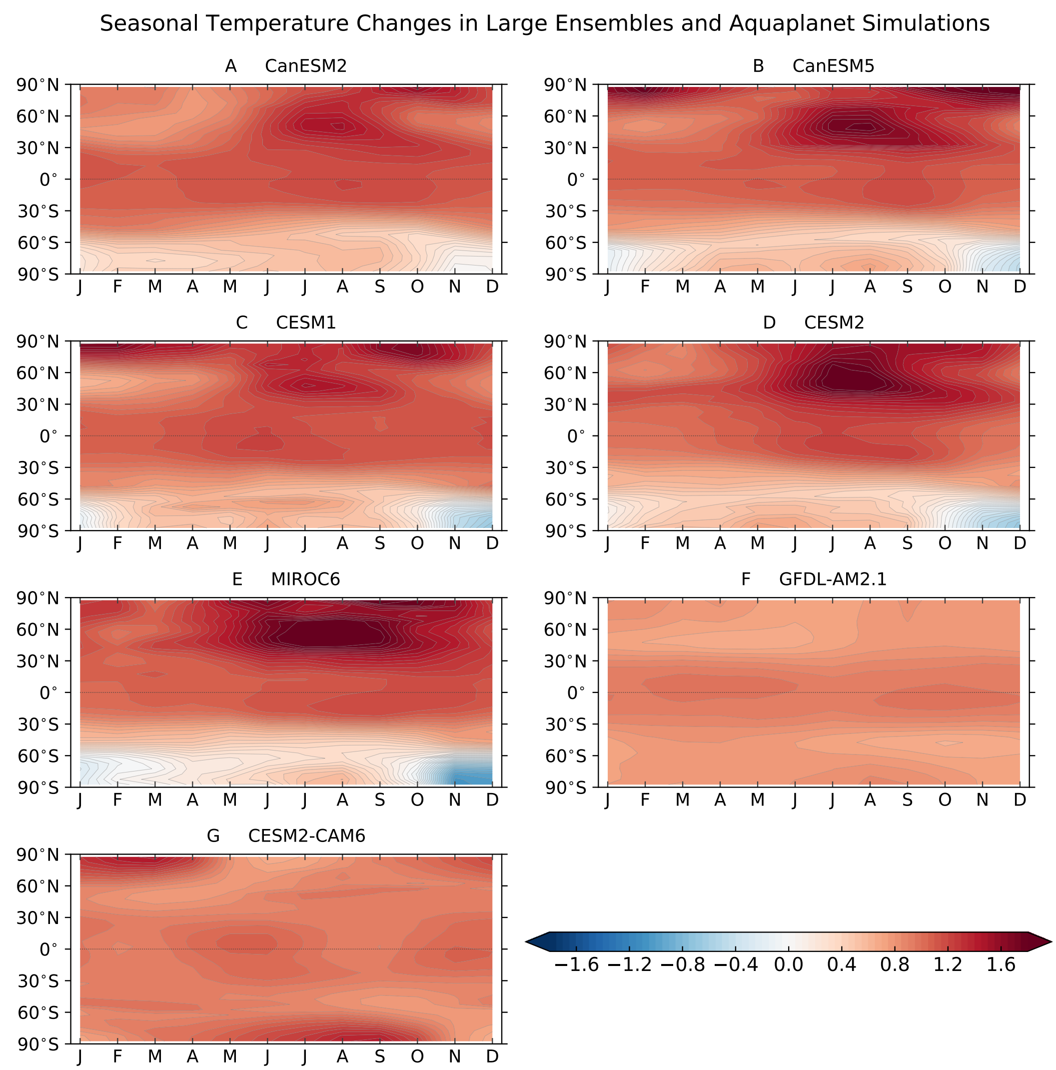

Fig. 14: Scaled seasonal changes (dimensionless) in uncorrected TMT in LEs and two aquaplanet simulations. (a-e) Scaled total linear changes in zonal-mean TMT over 1979 to 2020 in five LEs. Total linear changes are calculated separately for each month from the ensemble-average monthly-mean temperatures of the LE. Scaling is with the global-mean annual-mean total linear change in each LE. (f) Scaled zonally-averaged differences in uncorrected TMT between the 30-year averages of an aquaplanet perturbation experiment and control run performed with the GFDL-AM2.1 model. Ocean albedo α is set to 0.45 in the simulation shown here. (g) As for panel f but for $4\times CO_2$ and CTL simulations performed with the CESM2-CAM6 model and for differences between 100-year climatologies. The parameter used for tuning snow albedo, r_snw, has been set to 0.7. The scaling in panels f and g is with the global-mean annual-mean temperature change between the time averages of the perturbation experiment and CTL.

We consider next the seasonality that drives the $T_{AC}(x,t)$ changes in the ESMs and the aquaplanet simulations. To compare the magnitude of seasonal changes in different types of simulation (transient and quadrupled CO_2) with very different radiative forcing, we scale results by global-mean annual-mean TMT changes. In the ESMs, zonal-mean warming over 1979 to 2020 occurs in every month and at every latitude, except poleward of 60°S during austral summer (Figs. 14a-e). This high-latitude SH seasonal cooling signal arises from the temperature and circulation changes caused by Antarctic stratospheric ozone depletion (Solomon et al. 2012; Eyring et al. 2013). In the aquaplanet simulations, the equilibrium response to CO_2 quadrupling is also characterized by warming at all latitudes and in all months, with largest warming in the tropics in GFDL-AM2.1 and poleward of 70° in CESM2-CAM6 (Figs. 14f and g, respectively).

To more easily discern the seasonality of TMT changes, we express the monthly changes as departures from annual-mean changes at each latitude band (Fig. 15). This reveals that the five ESMs have seasonal trends in zonal-mean TMT similar to those found in the CMIP5 multi-model average (Santer et al. 2018), with maximum mid-latitude warming in NH summertime (Fig. 15a-e). This summertime warming signal is the primary driver of the increase in the amplitude of the annual cycle of NH mid-latitude tropospheric temperature.

As in the ESMs, the aquaplanet simulations display strong seasonality in mid-latitude TMT changes, with larger warming in late summer and fall in both summer hemispheres (Figs. 15f,g). Maximum mid-latitude warming is delayed by several months relative to the ESMs. This phase lag is likely due to the absence of land and to the fact that the 30-meter slab oceans have larger heat capacity compared to land.

The amplitude of annual cycle changes over the Arctic and Antarctic is markedly smaller in GFDL-AM2.1 than in CESM2-CAM6 (compare Figs. 15f,g). This difference is due to the absence of sea-ice thermodynamics in the GFDL-AM2.1 aquaplanet simulation and to the substantial impact of sea-ice thermodynamics on the high-latitude seasonal cycle.[xii] With thermodynamic sea ice in CESM2-CAM6, there is greater seasonality in ice extent in the $1\times CO_2$ climate compared to the GFDL-AM2.1 "albedo-only" representation of sea ice effects. The larger amplitude of the control run seasonal cycle yields larger high-latitude $T_{AC}(x,t)$ changes in CESM2-CAM6 in

[xii] In the GFDL-AM2.1 aquaplanet integrations, sea ice is represented as a temperature-dependent surface albedo following classic energy-balance model theory. In the CESM2-CAM6 simulations, sea ice is modeled as a slab of ice that conducts heat vertically through the ice assuming a linear temperature profile. Snow can accumulate on top of the ice. The treatment of sea ice in the CESM2-CAM6 aquaplanet integrations is still somewhat simplified compared to the most comprehensive ice models, which include multiple layers of ice and brine pockets within the ice.

response to the quadrupling of CO_2 and the complete loss of sea ice. Greater initial sea-ice coverage in CESM2-CAM6 (and a more equatorward ice edge) may also explain some of the differences between the mid-latitude $T_{AC}(x,t)$ changes in the CESM2-CAM6 and GFDL-AM2.1 aquaplanet runs (Figs. 15f,g).

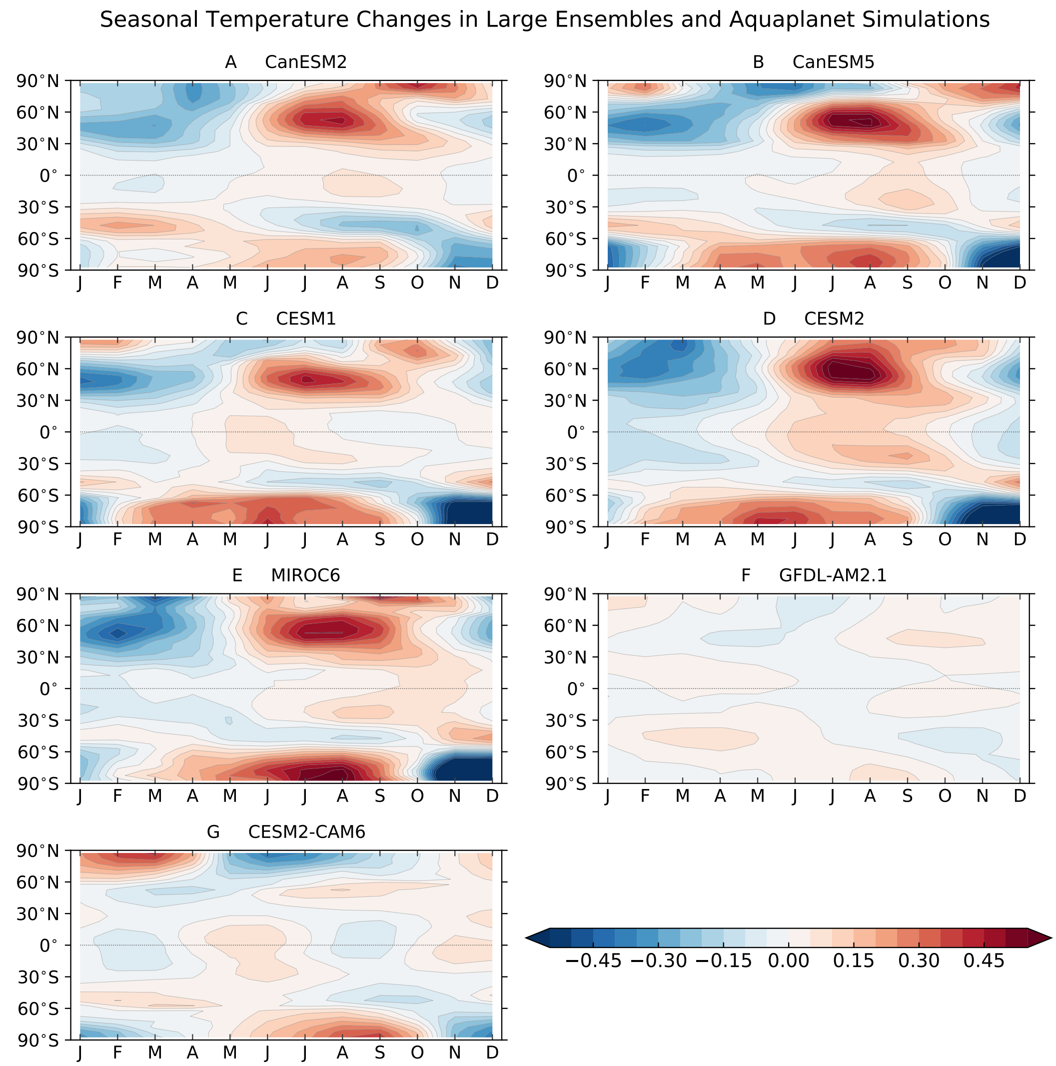

FIG. 15: As for Fig. 14 but with results expressed as monthly-mean departures from annual-mean changes. The scalings in each panel are identical to those used in Fig. 14.

10. Conclusions

There will always be some irreducible uncertainty in partitioning observed climate records into multidecadal internal variability (MIV) and forced responses (Frankcombe et al. 2015; Kravtsov

2017; Cheung et al. 2017; Kajtar et al. 2019; Pallotta and Santer 2020). Partitioning difficulties arise from multiple sources. One source is in the model-predicted signals that are removed from observations to isolate MIV (Section 7). These signals are affected by uncertainties and errors in estimates of historical external forcings (Solomon et al. 2011, 2012; Fyfe et al. 2021) and in the simulated responses to external forcings (Fyfe et al. 2021). Forced modulation of internal variability is an additional complication in separating signal and noise (Maher et al. 2015), along with structural uncertainties and residual errors in the observations (Mears et al. 2011; Mears and Wentz 2017; Zou and Wang 2011; Po-Chedley et al. 2015).

In the real world, for example, there are uncertainties in the amplitude and patterns of MIV and in our quantitative understanding of low-frequency changes in net anthropogenic aerosol forcing (Mann and Emanuel 2006; Carslaw et al. 2013). This complicates separation of MIV from the response to aerosol forcing. In consequence, there is uncertainty in estimating the contribution of MIV to the positive detection of an anthropogenic fingerprint in observed $T_{\mathrm{AC}}(x,t)$ changes.

It is conceivable, therefore, that fortuitous phasing of modes of Pacific and Atlantic MIV may have favored the positive detection of a seasonal cycle fingerprint in satellite $T_{\mathrm{AC}}(x,t)$ data (Santer et al. 2018). Large initial condition ensembles (LEs) are a valuable virtual laboratory for exploring this possibility. The five LEs analyzed here span a wide range of phase space in equilibrium climate sensitivity (from 2.6°C to 5.6°C), the amplitude of MIV, and the size of net anthropogenic aerosol forcing (Zelinka et al. 2014, 2020). Despite this wide phase space, and despite differences in the phasing of MIV in the 240 realizations of historical $T_{\mathrm{AC}}(x,t)$ changes examined here, our D&A results are remarkably robust. We obtain positive detection of model seasonal cycle fingerprints in 239 of the 240 realizations (Fig. 6).

We also used LEs to investigate concerns regarding the reliability of model MIV estimates (Curry and Webster 2011; O'Reilly et al. 2021). For the AMO, between 21% and 23% of the model LE realizations have values of P_{LOW} (the spectral density at ≈ 24 years) that exceed P_{LOW} in the "signal removed" HadCRUT4 SST data. Model-data agreement in low-frequency variability is closer for the IPO and Niño 3.4 SSTs. We find no evidence that the LEs analyzed here significantly underestimate the observed low-frequency power of major modes of internal variability.

More importantly, our sensitivity studies (Section 8) explicitly show that even in the presence of large (factor of 3-4) inter-model and inter-realization differences in the amplitude of AMO and

ENSO variability, the seasonal cycle fingerprints shown in Figs. 4a-e are robustly identifiable in models and satellite data. This is primarily due to the fact that the fingerprint patterns are spatially dissimilar to the patterns of internal $T_{\mathrm{AC}}(x,t)$ variability associated with the AMO and ENSO.

The robustness of the seasonal cycle D&A results shown here, taken together with the evidence from the aquaplanet simulations (Section 9), suggests that basic physical processes are dictating a common pattern of forced $T_{\mathrm{AC}}(x,t)$ response in observations and in the five LEs. The key processes involved are likely to include GHG-induced expansion of the tropics, lapse-rate changes, land surface drying, and sea ice decrease (Manabe et al. 1981; Wetherald and Manabe 1995; Held 2000; Fu et al. 2006; Frierson 2006; Frierson et al. 2007; Seidel and Randel 2007; Kang and Liu 2012; Donohoe and Battisti 2013; Quan et al. 2014; Douville and Plazzotta 2017; Feldl et al. 2017).

Our study clearly illustrates that the analysis of multiple LEs provides diagnostic benefits for D&A research, enabling analysts to explore the robustness of fingerprint detection results in novel ways. Additional diagnostic benefit arises through comparisons of idealized aquaplanet simulations with results from full Earth System Models – and through comparing aquaplanet simulations with very different representation of climate processes associated with sea ice (Feldl et al. 2017, 2020; Feldl and Merlis 2021). Such comparisons may help to improve understanding of the physical mechanisms influencing seasonal cycle fingerprints and of the expected seasonal cycle changes over the 21st century.

Acknowledgments. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison (PCMDI) provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. This work was performed under the auspices of the U.S. Department of Energy (DOE) by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. At LLNL, S.P.-C., C.B., G.P., and M.D.Z. were supported by the Regional and Global Model Analysis Program (RGMA) of the Office of Science at the DOE. S.P.-C. was also supported under LDRD 18-ERD-054. B.D.S. was funded under LDRD task number 21-FS-035. N.F. was supported by NSF award AGS-1753034. M.F.S. was supported by NOAA's Climate Program Office Modeling, Analysis, Predictions, and Projections (MAPP) program grant

NA20OAR4310445. The work of K.B.R. was supported by the Institute for Basic Sciences (IBS), Republic of Korea, under IBS-R028-D1. Q.F. was in part supported by NSF Grant AGS-1821437. S.S. was funded in part by NSF Grant AGS-1848863. N.R. was supported by the RGMA component of the Earth and Environmental System Modeling Program of the U.S. Department of Energy's Office of Biological & Environmental Research (BER) via National Science Foundation IA 1844590. We thank Jeff Painter (LLNL) for calculating synthetic satellite temperatures from CMIP5 simulation output and Frank Wentz (RSS) for providing observational TMT and TLS data. Dáithí Stone and two anonymous reviewers provided constructive and helpful comments.

Data availability statement. Synthetic satellite temperatures calculated from model simulations and the ERA 5.1 reanalysis are provided at: https://pcmdi.llnl.gov/research/DandA/.

References

AchutaRao, K., and K. R. Sperber, 2002: Simulation of the El Niño Southern Oscillation: Results from the Coupled Model Intercomparison Project. *Cli. Dyn.*, **19**, 191–209, https://doi.org/10.1007/s00382-001-0221-9.

Allen, M. R., and S. F. B. Tett, 1999: Checking for model consistency in optimal fingerprinting. *Cli. Dyn.*, **15**, 419–434.

Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor, 2012: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophys. Res. Lett.*, **39**, L09712, https://doi.org/10.1029/2012GL051607, http://10.1029/2012GL051607.

Bandoro, J., S. Solomon, A. Donohoe, D. W. J. Thompson, and B. D. Santer, 2014: Influences of the Antarctic ozone hole on Southern Hemisphere summer climate change. *J. Clim.*, **27**, 6245–6264.

Barnett, T. P., D. Pierce, K. AchutaRao, P. Gleckler, B. D. Santer, J. Gregory, and W. Washington, 2005: Penetration of human-induced warming signal into the world's oceans. *Science*, **309**, 284–287.

Bindoff, N. L., and Coauthors, 2013: Detection and Attribution of Climate Change: from Global to Regional. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I* *to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, Eds., Cambridge University Press, 867–952.

Bintanja, R., and E. C. van der Linden, 2013: The changing seasonal climate in the Arctic. *Sci. Rep.*, **3**, https://doi.org/10.1038/srep01556.

Bonfils, C., B. D. Santer, J. C. Fyfe, K. Marvel, T. J. Phillips, and S. R. H. Zimmerman, 2020: Human influence on joint changes in temperature, rainfall and continental aridity. *Nat. Clim. Change*, **10**, 726–731, https://doi.org/10.1038/s41558-020-0821-1.

Bonfils, C., B. D. Santer, T. J. Phillips, K. Marvel, L. R. Leung, C. Doutriaux, and A. Capotondi, 2015: Relative contributions of mean-state shifts and ENSO-driven variability to precipitation changes in a warming climate. *J. Clim.*, **28**, 9997–10 013.

Brogli, R., N. Kröner, S. L. Sørland, D. Lüthi, and C. Schär, 2019: The role of Hadley circulation and lapse-rate changes for the future European summer climate. *J. Climate*, **32**, 385–404, https://doi.org/10.1175/JCLI-D-18-0431.1.

Carslaw, K. S., and Coauthors, 2013: Large contribution of natural aerosols to uncertainty in indirect forcing. *Nature*, **503**, 67–71, https://doi.org/10.1038/nature12674.

Cheung, A. H., M. E. Mann, B. A. Steinman, L. M. Frankcombe, M. H. England, and S. K. Miller, 2017: Reply to "Comment on comparison of low-frequency internal climate variability in CMIP5 models and observations". *J. Climate*, **30**, 9773–9782, https://doi.org/10.1175/JCLI-D-17-0531.1.

Cohen, J. M., M. J. Lajeunesse, and J. R. Rohr, 2018: A global synthesis of animal phenological responses to climate change. *Nat. Clim. Change*, **8**, 224–228, https://doi.org/10.1038/s41558-018-0067-3.

Curry, J., and P. Webster, 2011: Climate science and the uncertainty monster. *Bull. Amer. Met. Soc.*, **92**, 1667–1682, https://doi.org/10.1175/2011BAMS3139.1.

Deser, C., A. Phillips, V. Bourdette, and H. Teng, 2012: Uncertainty in climate projections: The role of internal variability. *Cli. Dyn.*, **38**, 527–546.

Deser, C., A. S. Phillips, M. A. Alexander, and B. V. Smoliak, 2014: Projecting North American climate over the next 50 years: Uncertainty due to internal variability. *J. Clim.*, **27**, 2271–2296.

Deser, C., and Coauthors, 2020: Insights from Earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Change*, **10**, 277–286.

Donohoe, A., and D. S. Battisti, 2013: The seasonal cycle of atmospheric heating and temperature. *J. Clim.*, **25**, 4962–4980, https://doi.org/10.1175/JCLI-D-12-00713.1.

Douville, H., and M. Plazzotta, 2017: Midlatitude summer drying: An underestimated threat in CMIP5 models? *Geophys. Res. Lett.*, **44**, 9967–9975, https://doi.org/10.1002/2017GL075353.

Duan, J., and Coauthors, 2019: Detection of human influences on temperature seasonality from the 19th century. *Nat. Sustainability*, **2**, 484–490, https://doi.org/10.1038/s41893-019-0276-4.

Dwyer, J. G., M. Biasutti, and A. H. Sobel, 2012: Projected changes in the seasonal cycle of surface temperature. *J. Clim.*, **25**, 6359–6374, https://doi.org/10.1175/JCLI-D-11-00741.1.

Enfield, D. B., A. M. M.-N. nez, and P. J. Trimble, 2001: The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S. *Geophys. Res. Lett.*, **28**, 2077–2080.

England, M. H., and Coauthors, 2014: Recent intensification of wind–driven circulation in the Pacific and the ongoing warming hiatus. *Nat. Clim. Change*, **4**, 222–227, https://doi.org/10.1038/nclimate2106.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Mod. Dev.*, **9(5)**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Eyring, V., and Coauthors, 2013: Long-term ozone changes ozone and associated climate impacts in CMIP5 simulations. *J. Geophys. Res.*, **118**, 5029–5060.

Fasullo, J. T., J. F. Lamarque, C. Hannay, N. Rosenblum, S. Tilmes, P. DeRepentigny, A. Jahn, and C. Deser, 2021: Spurious late historical-era warming in CESM2 and other CMIP6 climate simulations driven by prescribed biomass burning emissions. *Nat. Clim. Change*, submitted.

Feldl, N., S. Bordoni, and T. M. Merlis, 2017: Coupled high-latitude climate feedbacks and their impact on atmospheric heat transport. *J. Climate*, **30**, 189–201, https://doi.org/10.1175/JCLI-D-16-0324.1.

Feldl, N., and T. M. Merlis, 2021: Polar amplification in idealized climates: The role of ice, moisture, and seasons. *Geophys. Res. Lett.*, **48**, https://doi.org/10.1029/2021GL094130.

Feldl, N., S. Po-Chedley, H. K. A. Singh, S. Hay, and P. J. Kushner, 2020: Sea ice and atmospheric circulation shape the high-latitude lapse rate feedback. *npj. Clim. Atmos. Sci.*, **41**, https://doi.org/10.1038/s41612-020-00146-7.

Frankcombe, L. M., M. H. England, M. E. Mann, and B. A. Steinman, 2015: Separating internal variability from the externally forced climate response. *J. Clim.*, **28**, 8184–8202, https://doi.org/10.1175/JCLI-D-15-0069.1.

Frierson, D. M. W., 2006: Robust increases in midlatitude static stability in simulations of global warming. *Geophys. Res. Lett.*, **33**, L24816, https://doi.org/10.1029/2006GL027504.

Frierson, D. M. W., J. Lu, and G. Chen, 2007: Width of the Hadley cell in simple and comprehensive general circulation models. *Geophys. Res. Lett.*, **34**, L18804, https://doi.org/10.1029/2007GL031115.

Fu, Q., and C. M. Johanson, 2004: Stratospheric influences on MSU-derived tropospheric temperature trends: A direct error analysis. *J. Clim.*, **17**, 4636–4640.

Fu, Q., C. M. Johanson, J. M. Wallace, and T. Reichler, 2006: Enhanced mid-latitude tropospheric warming in satellite measurements. *Science*, **312**, 1179.

Fu, Q., C. M. Johanson, S. G. Warren, and D. J. Seidel, 2004: Contribution of stratospheric cooling to satellite-inferred tropospheric temperature trends. *Nature*, **429**, 55–58.

Fyfe, J. C., V. Kharin, B. D. Santer, R. N. S. Cole, and N. P. Gillett, 2021: Significant impact of forcing uncertainty in a large ensemble of climate model simulations. *Proc. Nat. Acad. Sci.*, (in review).

Fyfe, J. C., and Coauthors, 2016: Making sense of the early-2000s warming slowdown. *Nat. Clim. Change*, **6**, 224–228.

Fyfe, J. C., and Coauthors, 2017: Large near-term projected snowpack loss over the western United States. *Nature Communications*, **8**, https://doi.org/10.1038/ncomms14996.

Gillett, N. P., B. D. Santer, and A. J. Weaver, 2004: Quantifying the influence of stratospheric cooling on satellite-derived tropospheric temperature trends. *Nature*, **432**, https://doi.org/10.1038/nature03209.

Gillett, N. P., F. W. Zwiers, A. J. Weaver, and P. A. Stott, 2003: Detection of human influence on sea-level pressure. *Nature*, **422**, 292–294.

Hasselmann, K., 1979: *On the signal-to-noise problem in atmospheric response studies*, 251–259. Roy. Met. Soc., London.

Hawkins, E., and R. Sutton, 2012: Time of emergence of climate signals. *Geophys. Res. Lett.*, **39**, L01702, https://doi.org/10.1029/2011GL050087.

He, J., and B. J. Soden, 2017: A re-examination of the projected subtropical precipitation decline. *Nat. Clim. Change*, **7**, 53–57, https://doi.org/10.1038/NCLIMATE3157.

Hegerl, G. C., H. v. Storch, K. Hasselmann, B. D. Santer, U. Cubasch, and P. D. Jones, 1996: Detecting anthropogenic climate change with an optimal fingerprint method. *J. Clim.*, **9**, 2281–2306.

Hegerl, G. C., and Coauthors, 2007: Understanding and Attributing Climate Change. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, Eds., Cambridge University Press, 663–745.

Held, I. M., 2000: The general circulation of the atmosphere. Available at http://gfd.whoi.edu/proceedings/2000/PDFvol2000.html.

Henley, B. J., J. Gergis, D. J. Karoly, S. Power, J. Kennedy, and C. K. Folland, 2015: A tripole index for the Interdecadal Pacific Oscillation. *Cli. Dyn.*, **45**, 3077–3090, https://doi.org/10.1007/s00382-015-2525-1.

Henley, B. J., and Coauthors, 2017: Spatial and temporal agreement in climate model simulations of the Interdecadal Pacific Oscillation. *Env. Res. Lett.*, **12**, https://doi.org/10.1088/1748-9326/aa5cc8.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Q. J. Roy. Met. Soc.*, **146**, 1999–2049.

Hu, Y. Y., and Q. Fu, 2007: Observed poleward expansion of the Hadley circulation since 1979. *Atmos. Chem. Phys.*, **7**, 5229–5236.

IPCC, 2021: Summary for Policymakers. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou, Eds., Cambridge University Press, 41 pages.

Kajtar, J. B., M. Collins, L. M. Frankcombe, M. H. England, T. J. Osborn, and M. Juniper, 2019: Global mean surface temperature response to large-scale patterns of variability in observations and CMIP5. *Geophys. Res. Lett.*, **46**, 2232–2241, https://doi.org/10.1029/2018GL081462.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.*, **77**, 437–471.

Kamae, Y., H. Shiogama, M. Watanabe, M. Ishii, H. Ueda, and M. Kimoto, 2015: Recent slowdown of tropical upper tropospheric warming associated with Pacific climate variability. *Geophys. Res. Lett.*, **42**, 2995–3003, https://doi.org/10.1002/2015GL063608.

Kang, S. M., and J. Liu, 2012: Expansion of the Hadley Cell under global warming: Winter versus summer. *J. Clim.*, **25**, 8387–8393.

Kang, S. M., S.-P. Xie, C. Deser, and B. Xiang, 2021: Zonal mean and shift modes of historical climate response to evolving aerosol distribution. *Sci. Bulletin*, in press.

Kay, J. E., and Coauthors, 2015: The Community Earth System Model: Large ensemble project. *Bull. Amer. Met. Soc.*, **96**, 1333–1349.

Kirchmeier-Young, M. C., F. W. Zwiers, and N. P. Gillett, 2017: Attribution of extreme events in Arctic sea ice extent. *J. Clim.*, **30**, 553–571.

Kosaka, Y., and S.-P. Xie, 2013: Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature*, **501**, 403–407.

Kravtsov, S., 2017: Comment on "Comparison of low-frequency internal climate variability in CMIP5 models and observations". *J. Climate*, **30**, 9763–9772, https://doi.org/10.1175/JCLI-D-17-0438.1.

Lienert, F., J. C. Fyfe, and W. J. Merryfield, 2011: Do climate models capture the tropical influences on North Pacific sea surface temperature variability? *J. Clim.*, **24**, 6203–6209, https://doi.org/10.1175/JCLI-D-11-00205.1.

Maher, N., S. McGregor, M. H. England, and A. S. Gupta, 2015: Effects of volcanism on tropical variability. *Geophys. Res. Lett.*, **42**, 6024–6033, https://doi.org/10.1002/2015GL064751.

Mahlstein, I., G. Hegerl, and S. Solomon, 2012: Emerging local warming signals in observational data. *Geophys. Res. Lett.*, **39**, L21711, https://doi.org/10.1029/2012GL053952.

Manabe, S., R. T. Wetherald, and R. J. Stouffer, 1981: Summer dryness due to an increase of atmospheric CO_2 concentration. *Climatic Change*, **3**, 347–386.

Mann, M. E., and K. A. Emanuel, 2006: Atlantic hurricane trends linked to climate change. *EOS*, **87**, 233, 238, 241.

Mantsis, D. F., and A. C. Clement, 2009: Simulated variability in the mean atmospheric meridional circulation over the 20th century. *Geophys. Res. Lett.*, **36**, L06704, https://doi.org/10.1029/2008GL036741.

Marvel, K., M. Biasutti, C. Bonfils, K. E. Taylor, Y. Kushnir, and B. I. Cook, 2017: Observed and projected changes to the precipitation annual cycle. *J. Clim.*, **30**, 4983–4995, https://doi.org/10.1175/JCLI-D-16-0572.1.

Marvel, K., and C. Bonfils, 2013: Identifying external influences on global precipitation. *Proc. Nat. Acad. Sci*, **110**, 19 301–19 306.

Mears, C., and F. J. Wentz, 2017: A satellite-derived lower-tropospheric atmospheric temperature dataset using an optimized adjustment for diurnal effects. *J. Clim.*, **30**, 7695–7718.

Mears, C., F. J. Wentz, P. Thorne, and D. Bernie, 2011: Assessing uncertainty in estimates of atmospheric temperature changes from MSU and AMSU using a Monte-Carlo technique. *J. Geophys. Res.*, **116**, D08112, https://doi.org/10.1029/2010JD014954.

Meehl, G. A., J. M. Arblaster, J. T. Fasullo, A. Hu, and K. E. Trenberth, 2011: Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods. *Nat. Clim. Change*, **1**, 360–364, https://doi.org/10.1038/nclimate1229.

Meehl, G. A., A. Hu, B. D. Santer, and S.-P. Xie, 2016: Interdecadal Pacific Oscillation contributions to multi-decadal variability of 20th century globally averaged surface temperatures. *Nat. Clim. Change*, **6**, 1005–1008.

Meinshausen, M., and Coauthors, 2011: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, **109**, 213–241.

Min, S. K., X. Zhang, F. W. Zwiers, and T. Agnew, 2008: Human influence on Arctic sea ice detectable from early 1990s onwards. *Geophys. Res. Lett.*, **35**, L21701.

Min, S. K., X. Zhang, F. W. Zwiers, P. Friederichs, and A. Hense, 2009: Signal detectability in extreme precipitation changes assessed from twentieth century climate simulations. *Cli. Dyn.*, **32**, 95–111, https://doi.org/10.1007/s00382-008-0376-8.

Mitchell, J. F. B., and D. J. Karoly, 2001: Detection of climate change and attribution of causes. *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson, Eds., Cambridge University Press, 695–738.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res.*, **117**, D08101, https://doi.org/10.1029/2011JD017187.

North, G. R., K. Y. Kim, S. S. P. Shen, and J. W. Hardin, 1995: Detection of forced climate signals. Part 1: Filter theory. *J. Clim.*, **8**, 401–408.

O'Reilly, C. H., D. J. Befort, A. Weisheimer, T. Woollings, A. Ballinger, and G. Hegerl, 2021: Projections of northern hemisphere extratropical climate underestimate internal variability and associated uncertainty. *Comm. Earth & Env.*, **2**, https://doi.org/10.1038/s43247-021-00268-7.

Pallotta, J., and B. D. Santer, 2020: Multi-frequency analysis of simulated versus observed variability in tropospheric temperature. *J. Clim.*, **33**, 10 383–10 402, https://doi.org/10.1175/JCLI-D-20-0023.1.

Parmesan, C., and G. Yohe, 2003: A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.

Pierce, D., P. J. Gleckler, T. P. Barnett, B. D. Santer, and P. J. Durack, 2012: The fingerprint of human-induced changes in the ocean's salinity and temperature fields. *Geophys. Res. Lett.*, **39**, L21704, https://doi.org/10.1029/2012GL053389.

Po-Chedley, S., B. D. Santer, S. Fueglistaler, M. D. Zelinka, P. Cameron-Smith, J. F. Painter, and Q. Fu, 2021: Natural variability drives model-observational differences in tropical tropospheric warming. *Proc. Nat. Acad. Sci.*, **118**, e2020962118, https://doi.org/10.1073/pnas.2020962118.

Po-Chedley, S., T. J. Thorsen, and Q. Fu, 2015: Removing diurnal cycle contamination in satellite-derived tropospheric temperatures: Understanding tropical tropospheric trend discrepancies. *J. Clim.*, **28**, 2274–2290.

Qian, C., and X. Zhang, 2015: Human influences on changes in the temperature seasonality in mid- to high-latitude land areas. *J. Clim.*, **28**, 5908–5921, https://doi.org/10.1175/JCLI-D-14-00821.1.

Quan, X.-W., M. P. Hoerling, J. Perlwitz, H. F. Diaz, and T. Xu, 2014: How fast are the tropics expanding? *J. Clim.*, **27**, 1999–2013, https://doi.org/10.1175/JCLI-D-13-00287.1.

Randel, W. J., L. Polvani, F. Wu, D. E. Kinnison, C.-Z. Zou, and C. Mears, 2017: Troposphere-stratosphere temperature trends derived from satellite data compared with ensemble simulations from WACCM. *J. Geophys. Res.*, **122**, 9651–9667, https://doi.org/10.1002/2017JD027158.

Riahi, K., and Coauthors, 2017: The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Glob. Env. Change*, **42**, 153–168, https://doi.org/10.1016/j.gloenvcha.2016.05.009.

Risser, M. D., and M. F. Wehner, 2017: Attributable human-induced changes in the likelihood and magnitude of the observed extreme precipitation during Hurricane Harvey. *Geophys. Res. Lett.*, **44**, 12 457–12 464, https://doi.org/10.1002/2017GL075888.

Rodgers, K. B., J. Lin, and T. L. Frölicher, 2015: Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an earth system model. *Biogeosci.*, **12**, 3301–3320.

Rodgers, K. B., and Coauthors, 2021: Ubiquity of human-induced changes in climate variability. *Earth Syst. Dynam.*, **12**, 1393–1411, https://doi.org/10.5194/esd-12-1393-2021.

Root, T. L., D. P. MacMynowski, M. D. Mastrandrea, and S. H. Schneider, 2005: Human-modified temperatures induce species changes: Joint attribution. *Proc. Nat. Acad. Sci.*, **102**, 7465–7469, https://doi.org/10.1073/pnas.0502286102.

Santer, B. D., W. Brüggemann, U. Cubasch, K. Hasselmann, H. Höck, E. Maier-Reimer, and U. Mikolajewicz, 1994: Signal-to-noise analysis of time-dependent greenhouse warming experiments. *Cli. Dyn.*, **9**, 267–285.

Santer, B. D., J. Fyfe, S. Solomon, J. Painter, C. Bonfils, G. Pallotta, and M. Zelinka, 2019: Quantifying stochastic uncertainty in detection time of human-caused climate signals. *Proc. Nat. Acad. Sci.*, **116**, 19 821–19 827, https://doi.org/10.1038/s41558-019-0424-x.

Santer, B. D., T. M. L. Wigley, T. P. Barnett, and E. Anyamba, 1995: Detection of climate change and attribution of causes. *Climate Change 1995: The Science of Climate Change. Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change*, J. T. Houghton, L. G. M. Filho, B. A. Callander, N. Harris, A. Kattenberg, and K. Maskell, Eds., Cambridge University Press, 407–443.

Santer, B. D., and Coauthors, 1996: A search for human influences on the thermal structure of the atmosphere. *Nature*, **382**, 39–46.

Santer, B. D., and Coauthors, 2003: Influence of satellite data uncertainties on the detection of externally forced climate change. *Science*, **300**, 1280–1284.

Santer, B. D., and Coauthors, 2009: Incorporating model quality information in climate change detection and attribution studies. *Proc. Nat. Acad. Sci.*, **106**, 14 778–14 783, https://doi.org/10.1073/pnas.090173106.

Santer, B. D., and Coauthors, 2018: Human influence on the seasonal cycle of tropospheric temperature. *Science*, **361**, eaas8806, https://doi.org/10.1126/science.aas8806.

Santer, B. D., and Coauthors, 2021: Using climate model simulations to constrain observations. *J. Climate*, **34**, 6281–6301, https://doi.org/10.1175/JCLI-D-20-0768.1.

Seidel, D. J., and W. J. Randel, 2007: Recent widening of the tropical belt: Evidence from tropopause observations. *J. Geophys. Res.*, **112**, D20113, https://doi.org/10.1029/2007JD008861.

Serreze, M. C., and R. G. Barry, 2011: Processes and impacts of Arctic amplification: a research synthesis. *Glob. Planet. Change*, **77**, 85–96, https://doi.org/10.1016/j.gloplacha.2011.03.004.

Simmons, A., and Coauthors, 2020: Global stratospheric temperature bias and other stratospheric aspects of ERA5 and ERA5.1. Technical Memo 859, European Centre for Medium-Range Weather Forecasts, 40 pp.

Sippel, S., N. Meinshausen, E. M. Fischer, E. Székely, and R. Knutti, 2020: Climate change now detectable from any single day of weather at global scale. *Nat. Clim. Change*, **10**, 35–41, https://doi.org/10.1038/s41558-019-0666-7.

Sippel, S., N. Meinshausen, E. Székely, E. Fischer, A. G. Pendergrass, F. Lehner, and R. Knutti, 2021: Robust detection of forced warming in the presence of potentially large climate variability. *Science Advances*, **7**, eabh4429, https://doi.org/doi/10.1126/sciadv.abh4429.

Smith, R. D., J. K. Dukowicz, and R. C. Malone, 1992: Parallel ocean general circulation modeling. *Physica D*, **60**, 38–61, https://doi.org/10.1016/0167-2789(92)90225-C.

Solomon, S., J. S. Daniel, R. R. Neely, J.-P. Vernier, E. G. Dutton, and L. W. Thomason, 2011: The persistently variable "background" stratospheric aerosol layer and global climate change. *Science*, **333**, 866–870.

Solomon, S., P. J. Young, and B. Hassler, 2012: Uncertainties in the evolution of stratospheric ozone and implications for recent temperature changes in the tropical lower stratosphere. *Geophys. Res. Lett.*, **39**, L17706, https://doi.org/10.1029/2012GL052723.

Solomon, S., and Coauthors, 2017: Mirrored changes in Antarctic ozone and stratospheric temperature in the late 20th versus early 21st centuries. *J. Geophys. Res.*, **122**, 8940–8950, https://doi.org/10.1002/2017JD026719.

Spencer, R. W., J. R. Christy, and W. D. Braswell, 2017: UAH version 6 global satellite temperature products: Methodology and results. *Asia-Pac. J. Atmos. Sci.*, **53**, 121–130, https://doi.org/10.1007/s13143-017-0010-y.

Steinman, B. A., M. E. Mann, and S. K. Miller, 2015: Atlantic and Pacific multidecadal oscillations and Northern Hemisphere temperatures. *Science*, **347**, 988–991.

Stine, A. R., and P. Huybers, 2012: Changes in the seasonal cycle of temperature and atmospheric circulation. *J. Clim.*, **25**, 7362–7380, https://doi.org/10.1175/JCLI-D-11-00470.1.

Stott, P. A., D. A. Stone, and M. R. Allen, 2004: Human contribution to the European heatwave of 2003. *Nature*, **432**, 610–614.

Stott, P. A., S. F. B. Tett, G. S. Jones, M. R. Allen, J. F. B. Mitchell, and G. J. Jenkins, 2000: External control of 20th century temperature by natural and anthropogenic forcings. *Science*, **290**, 2133–2137.

Stott, P. A., and Coauthors, 2016: Attribution of extreme weather and climate-related events. *J. Atmos. Sci.*, **7**, 23–41, https://doi.org/10.1002/wcc.380.

Stouffer, R. J., G. Hegerl, and S. Tett, 2000: A comparison of surface air temperature variability in three 1000-yr coupled ocean–atmosphere model integrations. *J. Clim.*, **13**, 513–537.

Suárez-Gutiérrez, L., C. Li, P. W. Thorne, and J. Marotzke, 2017: Internal variability in simulated and observed tropical tropospheric temperature trends. *Geophys. Res. Lett.*, **44**, 5709–5719, https://doi.org/10.1002/2017GL073798.

Swart, N. C., S. T. Gille, J. C. Fyfe, and N. P. Gillett, 2018: Recent Southern Ocean warming and freshening driven by greenhouse gas emissions and ozone depletion. *Nat. Geosci.*, **11**, 836–841.

Swart, N. C., and Coauthors, 2019: The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, **12**, 4823–4873, https://doi.org/10.5194/gmd-12-4823-2019.

Tatebe, H., and Coauthors, 2019: Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, **12 (7)**, 2727–2765, https://doi.org/10.5194/gmd-12-2727-2019.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Taylor, P. C., M. Cai, A. Hu, J.Meehl, W. Washington, and G. J. Zhang, 2013: A decomposition of feedback contributions to polar warming amplification. *J. Clim.*, **26**, 7023–7043, https://doi.org/10.1175/JCLI-D-12-00696.1.

Tett, S. F. B., T. C. Johns, and J. F. B. Mitchell, 1997: Global and regional variability in a coupled AOGCM. *Cli. Dyn.*, **13**, 303–323.

Tett, S. F. B., J. F. B. Mitchell, D. E. Parker, and M. R. Allen, 1996: Human influence on the atmospheric vertical temperature structure: Detection and observations. *Science*, **274**, 1170–1173.

Thompson, D. W. J., J. J. Kennedy, J. M. Wallace, and P. D. Jones, 2008: A large discontinuity in the mid-twentieth century in observed global-mean surface temperature. *Nature*, **453**, 646–649, https://doi.org/10.1038/nature06982.

Thompson, D. W. J., S. Solomon, P. J. Kushner, M. H. England, K. M. Grise, and D. J. Karoly, 2011: Signatures of the Antarctic ozone hole in southern hemisphere surface climate change. *Nat. Geosci.*, **4**, 741–749, https://doi.org/10.1038/ngeo1296.

Thorne, P. W., and Coauthors, 2002: Assessing the robustness of zonal mean climate change detection. *Geophys. Res. Lett.*, **29**, https://doi.org/10.1029/2002GL015717.

Trenberth, K. E., 2015: Has there been a hiatus? *Science*, **349**, 791–792.

Wetherald, R. T., and S. Manabe, 1995: The mechanisms of summer dryness induced by greenhouse warming. *J. Climate*, **8**, 3096–3108.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, San Diego, 1–467 pp.

Willett, K. M., N. P. Gillett, P. D. Jones, and P. W. Thorne, 2007: Attribution of observed surface humidity changes to human influence. *Nature*, **449**, 710–713, https://doi.org/10.1038/nature06207.

Yettella, V., and M. R. England, 2018: The role of internal variability in Twenty-First-Century projections of the seasonal cycle of Northern Hemisphere surface temperature. *J. Geophys. Res.*, **123**, 13 149–13 167, https://doi.org/10.1029/2018JD029066.

Zelinka, M. D., T. Andrews, P. M. Forster, and K. E. Taylor, 2014: Quantifying components of aerosol-cloud-radiation interactions in climate models. *J. Geophys. Res.*, **119**, 7599–7615, https://doi.org/10.1002/2014JD021710.

Zelinka, M. D., T. A. Myers, D. T. McCoy, S. Po-Chedley, P. M. Caldwell, P. Ceppi, S. A. Klein, and K. E. Taylor, 2020: Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.*, **47**, https://doi.org/10.1029/2019GL085782.

Zhang, X., F. W. Zwiers, G. C. Hegerl, F. H. Lambert, N. P. Gillett, S. Solomon, P. A. Stott, and T. Nozawa, 2007: Detection of human influence on twentieth-century precipitation trends. *Nature*, **448**, 461–466, https://doi.org/10.1038/nature06025.

Zou, C.-Z., M. D. Goldberg, and X. Hao, 2018: New generation of U.S. satellite microwave sounder achieves high radiometric stability performance for reliable climate change detection. *Sci. Adv.*, **4**, eaau0049.

Zou, C.-Z., and W. Wang, 2011: Inter-satellite calibration of AMSU-A observations for weather and climate applications. *J. Geophys. Res.*, **116**, D23113, https://doi.org/10.1029/2011JD016205.