AGU ADVANCING EARTH AND SPACE SCIENCES

Earth and Space Science

RESEARCH ARTICLE
10.1029/2022EA002468

Comparing CMIP6 Climate Model Simulations of Annual Global Mean Temperatures to a New Combined Data Product

Peter F. Craigmile[1] and **Peter Guttorp**[2]

[1]Department of Mathematics and Statistics, Hunter College, CUNY, New York, NY, USA, [2]Norwegian Computing Center, Oslo, Norway

Key Points:

- The trends in global annual mean temperatures in sixth Climate Model Intercomparison Project (CMIP6) historical runs often disagree with that of a new combined observation product
- Warming in global mean temperatures from 1880–1899 to 1995–2014 is more often less across CMIP6 historical runs than observations suggest
- Very few CMIP6 historical runs have a time series structure that is compatible with the data product on the annual global scale

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

P. F. Craigmile,
peter.craigmile@hunter.cuny.edu

Citation:

Craigmile, P. F., & Guttorp, P. (2023). Comparing CMIP6 climate model simulations of annual global mean temperatures to a new combined data product. *Earth and Space Science*, *10*, e2022EA002468. https://doi.org/10.1029/2022EA002468

Received 31 MAY 2023
Accepted 18 SEP 2023

Author Contributions:

Conceptualization: Peter F. Craigmile, Peter Guttorp
Data curation: Peter F. Craigmile, Peter Guttorp
Formal analysis: Peter F. Craigmile, Peter Guttorp
Investigation: Peter F. Craigmile, Peter Guttorp
Methodology: Peter F. Craigmile, Peter Guttorp

Abstract A new statistical approach to validating climate models is introduced. First, five observational estimates of global mean surface temperature with estimated standard errors are combined into one data product, latent observed annual global temperature anomalies for the years 1880–2014, using a Bayesian hierarchical statistical approach. Summarizing these observed anomalies, estimates of smooth trend, levels of warming, and residual dependence as summarized by the spectral density function, are provided with simultaneous 95% credible bands. Then, corresponding estimates of smooth trend, levels of warming, and residual dependence are produced for sixth Climate Model Intercomparison Project (CMIP6) historical simulations analyzed at the annual global temperature anomaly scale, and compared to these bands. Among our results, we find that 93 out of the 318 CMIP6 historical model runs contain trends fitting inside the simultaneous bands for the smooth trend constructed from the data products, and for residual temporal dependence 69 out of 318 model runs contain spectral density functions that are within the corresponding data-product-based-bands. We estimate the mean global temperature increase from 1995–2014 relative to 1880–1899, from the data product, to be 0.896°C with a 95% credible interval of between 0.877 and 0.915. We find that 14 CMIP6 model runs agree with this interval, 197 model runs lead to a smaller temperature increase globally, and 107 model runs lead to a larger temperature increase.

1. Introduction

Global mean temperature is a common indicator of climate change (Williams & Eggleston, 2017). Recently several of the global temperature series (Morice et al., 2020; Lenssen et al., 2019; Huang et al., 2019; Rohde & Hausfather, 2020) have been updated with estimates of uncertainty. Craigmile and Guttorp (2022) used statistical methods to combine such estimates into a single one with attendant uncertainty. Together with the recent historical simulations that are part of the sixth Climate Model Intercomparison Project (CMIP6; Eyring et al. (2016)) this allows a new approach to assessing how well the climate model runs compare to the data series of historical development of global mean temperature.

Fan et al. (2020) compares a smaller group of the CMIP6 models than ours to two data sets: an earlier HadCRU set than ours, and a data set due to Willmott and Robeson (1995), finding relatively high pattern correlation between models and data but with substantial regional variability between models. They also find that the ensemble mean follows the data trend in global mean temperature over different time periods. There has long been a discussion about whether all climate model runs should be weighted equally in ensemble analysis, or one should weight them with regard to how well they reproduce historical data when forced by historical greenhouse gas emissions. It appears that the CMIP6 runs have forced this issue, in that models with very high climate sensitivity tend to do a poor job describing historical data (Hausfather et al., 2022; Zelinka et al., 2020). Several researchers (e.g., Ribes et al. (2021); Liang et al. (2020)) now argue for various kinds of weighting of climate model ensembles relative to each model's historical performance.

In this work we produce a joint data product from five major observational products using a Bayesian hierarchical modeling framework. This modeling approach, first introduced in Craigmile and Guttorp (2022), accounts for inherent uncertainties in each observational product but also allows us to identify discrepancies between the different products. The posterior samples produced provide possible reconstructions of observed global temperature anomalies for the years 1880–2014, which we use to compare to global temperature anomalies for the same years calculated for different CMIP6 model runs in a number of ways. First, we compare changes in the global mean temperature anomalies over the period 1880–2014 by calculating nonlinear trends for both the joint data

Software: Peter F. Craigmile, Peter Guttorp
Validation: Peter F. Craigmile
Visualization: Peter F. Craigmile, Peter Guttorp
Writing – original draft: Peter F. Craigmile, Peter Guttorp
Writing – review & editing: Peter F. Craigmile, Peter Guttorp

product and the CMIP6 model runs. Second, we look for possible warming by calculating the difference in the mean temperatures from 1995–2014 relative to 1880–1899. Third, we compare the dependence structure of the residuals from the trend for the joint data product with the CMIP6 runs using plots of the spectrum density function. Finally, we compare the distributions of the time series residuals (the so-called innovations) after detrending using a shift plot.

It is important to note that our statistical modeling approach is not intended to emulate all features present in global temperature anomalies, such as decomposing the effect of the naturally variability from the various forcing effects (e.g., Sundberg et al. (2012); Hind et al. (2012); Moberg et al. (2015); Tingley et al. (2015)). Instead, our intention is to compare useful statistical summaries of trend, warming, and dependence and to point to CMIP6 models and runs that agree or disagree with the distributions of these summaries inferred from the observational data products. This is a necessary first step to learning about the inherent time series structure of global temperature anomalies expressed by different CMIP6 models.

In Section 2, we describe the data used: CMIP6 model runs and five observational global annual mean temperature series. We then describe the hierarchical statistical methods used in Craigmile and Guttorp (2022) that produces a reconstruction of the observed annual global temperature anomaly. In Section 3 we use this reconstruction to build various simultaneous credible bands for the observed trend, the spectral density function (a summary of the dependence in the observed series), and the warming that we observe. We use these bands to compare the CMIP6 model runs to our observation product on an annual global scale. We close with a discussion and point to further work in Section 4.

2. Material and Methods

2.1. Climate Model Simulations

We use 318 runs from 58 models using the CMIP6 historical forcings (see Text S1 in Supporting Information S1 for a complete list, with references). For each of the models we use all runs available with standard initialization, physics and forcing (typically called i1p1f1). The gray curves in Figure 1 shows the ensemble of global annual mean temperature, anomalized relative to the entire period. The anomalization is done by month and grid square. For each month/grid square combination we subtract off the mean temperature for that grid square and month over all years. We then calculate a mean global anomaly by area weighted averaging over grid squares. Finally a yearly anomaly is obtained by averaging monthly anomalies for each month in the year. The main reason for our approach to computing anomalies is that we are trying to mimic as much as possible the calculation of anomalies from land stations series. Among the advantages with our approach is that we take out seasonal variability and that we can deal with varying grid sizes in a consistent matter.

2.2. Temperature Data Products

Data products are obtained from Berkeley Earth (Rohde & Hausfather, 2020), GISS (Lenssen et al., 2019), Hadley Center (Morice et al., 2020), NOAA (Huang et al., 2019) and the Cowtan and Way (2014) global annual mean temperature series. We use data from 1880 through 2014, anomalizing each series with respect to the entire time period. We omit the Tokyo Climate Center series (Ishihara, 2007), as it only goes back to 1890, and appears to be biased downwards due to lack of Arctic information (McKinnon, 2022), and are unaware of any other data series that have reliable uncertainty estimates.

The five global annual mean series used are shown in black in Figure 1. In order to combine the five series we use a hierarchical Bayesian model. There is a fairly substantial literature that uses hierarchical statistical models in climate science. An early example is the prediction of ENSO-events (Berliner et al., 2000). Some other examples include paleoclimatology analysis (Tingley et al., 2015), the assessment of climate model bias and dependence (Jun et al., 2008), and climate projection uncertainty (Carvalho & Rickershauser, 2013). In this work, we extend the ideas of Craigmile and Guttorp (2022), who summarized global temperature anomalies.

2.3. Hierarchical Modeling

The basic idea behind the hierarchical approach to statistical modeling in this context is that the measurements correspond to a latent quantity, one interpretation of which may be the actual annual global temperature anomaly,

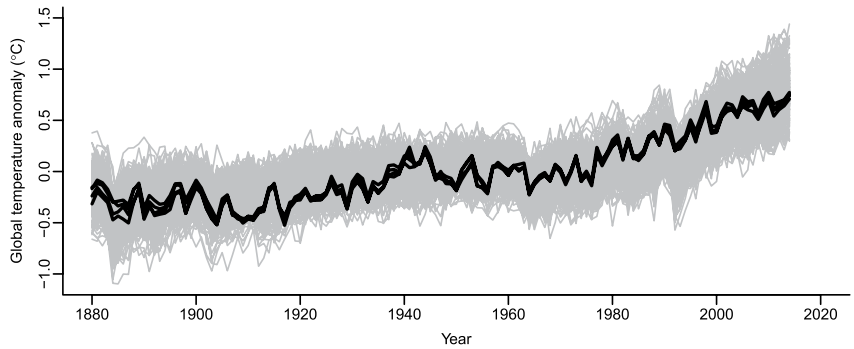

Figure 1. Time series plots of the annual global temperature anomalies for 318 CMIP6 historical model runs (in gray) as well as annual global temperature anomalies for the five data products (in black).

observed with error. The latent temperature anomaly is represented as a smooth trend with (residual) time series structure. In climate science, the term trend is commonly used to denote a linear trend. As we and others have pointed out elsewhere (e.g., Craigmile and Guttorp (2011); Stocker et al. (2013) [Box 2.2]), when looking at long-term changes in global mean temperature, the trend is clearly not linear. We will therefore, in this paper, use the term trend to denote a smooth, nonlinear trend function. Letting $\{Y_t : t = 1, \ldots, N\}$ denote the latent annual global mean temperature anomaly that we wish to infer upon, we assume that

$$Y_t = \mu_t + \nu_t, \quad t = 1, \ldots, N. \tag{1}$$

The trend component $\{\mu_t\}$ is modeled by a linear combination of b cubic b-spline basis functions $\{\boldsymbol{x}^T(t)\}$ (de Boor, 1978):

$$\mu_t = \boldsymbol{x}^T(t)\,\boldsymbol{\beta}, \quad t = 1, \ldots, N. \tag{2}$$

We choose $b = 8$ basis functions which corresponds to using a varying bandwidth kernel on the order of 15–30 years to estimate a smooth temperature trend. The serial dependence about the trend, $\{\nu_t\}$, is modeled using a stationary Gaussian autoregressive process of order $p = 4$ (AR(4)). (We discuss the robustness to other choices of basis functions, b, and autoregressive orders, p in Text S3 of Supporting Information S1).

We assume a model for the annual global temperature anomalies for the five data products $\{D_{j,t}\}$, $j = 1, \ldots, 5$ that depends on the latent annual global mean temperature anomaly $\{Y_t\}$ but also contains extra processes. Mathematically,

$$D_{j,t} = Y_t + \delta_{j,t} + \epsilon_{j,t}, \quad t = 1, \ldots, N. \tag{3}$$

The process $\{\epsilon_{j,t}\}$ captures the variability of each data product. A priori we assume that this process is independent over time and centered around zero, but with a variance that is equal to $v_{j,t}$, a measure of uncertainty that is given by each group that produce the data product j at each time t. In addition since there is overlap in the measurements used to compute the global mean temperature we assume correlation between different data products; $\text{corr}(\epsilon_{j,t}, \epsilon_{j',t}) = \rho_{j,j'}$. The process $\{\delta_{j,t}\}$ captures an additional discrepancy for each data product j, that apriori is assumed to be independent over data products and time t, with a normal distribution centered around zero with a constant variance that must be estimated.

For further details of the model and a full discussion of the priors and computational details see Craigmile and Guttorp (2022). All computations were carried out in the R software package (R Core Team, 2023). We generate samples from the posterior distribution of the latent processes $\{Y_t\}$, $\{\delta_{j,t}\}$, $\{\epsilon_{j,t}\}$ and other model parameters given the data products using Markov chain Monte Carlo (MCMC). The samples of latent processes can also be considered an ensemble from the distribution; that is, the possible observed annual global temperature anomaly paths for the years 1800–2014 in the case of the $\{Y_t\}$ process. Figure 2 shows summaries of these latent observed annual global anomalies $\{Y_t\}$ for these years. In panel (a), the black line denotes the posterior mean latent observed annual global temperature anomalies and the gray band is a simultaneous 95% credible band, whereas in (b) the uncertainties are summarized by calculating posterior standard deviations of the anomalies by year. As expected, we see evidence of substantial warming on a global annual scales, especially since 1980. Except for during the

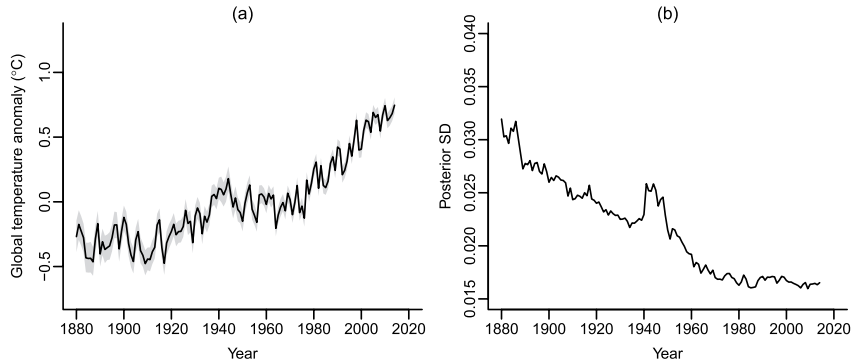

Figure 2. (a) Time series plots of the posterior mean global temperature anomalies for 1880–2014 (black lines), along with simultaneous 95% credible intervals (gray regions). (b) A plot of the posterior standard deviation of the anomalies by year.

two world wars, the posterior standard deviation decreases from 1880 to 1980. The posterior standard deviation is more stable after 1980. In what follows we may also refer to the posterior latent observed annual global temperature anomalies $\{Y_t\}$ as our data reconstruction.

The methodology designed for computing simultaneous credible bands for latent Gaussian processes can be found in Bolin and Lindgren (2015), with a description of the R software package excursions given in Bolin and Lindgren (2018). A simultaneous $(1 - \alpha)$-credible band for a process $X(t)$ is a region $\{(t, x)$: $t = 1, \ldots, T$, $q_\rho(t) \le x \le q_{1-\rho}(t)\}$, where $q_\rho(t)$ is obtained from the distribution of $X(s)$ by choosing ρ so that the joint probability $P(q_\rho(t) \le x \le q_{1-\rho}(t), 1 \le t \le T) \ge 1 - \alpha$. To determine the quantiles $q_\rho(t)$ and $q_{1-\rho}(t)$ we must integrate the posterior distribution of $X(t)$. The papers cited, as well as the excursions package, contain algorithms for fast computation of such integrals.

Figure 3a displays the trend estimate based on fitting the hierarchical statistical model to the five data product series (we summarize the posterior distribution of the nonlinear trend component $\{\mu_t\}$.) The gray area is a 95% simultaneous credible band for the trend. The band being simultaneous means that it is valid for all years 1880–2014 at the same time. One can for example, fit horizontal lines to see when the trend went above previous highs. We see that the global mean temperature anomaly trend since around 1990 is higher than at any previous time in the record.

A spectral analysis is a insightful summary of the dependence of a stationary time series in terms of a decomposition of sinusoids at different frequencies. The spectral density function expresses the variability of these sinusoids as a function of frequency (see, e.g., Brillinger (1981) and Percival and Walden (1993)). For an autoregressive process of order p the spectral density function $S(f)$ at frequency f is (Percival & Walden, 1993, Equation 488b)

$$S(f) = \frac{\sigma^2}{|1 - \sum_{j=1}^{p} \phi_j e^{-i2\pi f j}|^2}, \quad f \in \left[-\frac{1}{2}, \frac{1}{2}\right],$$

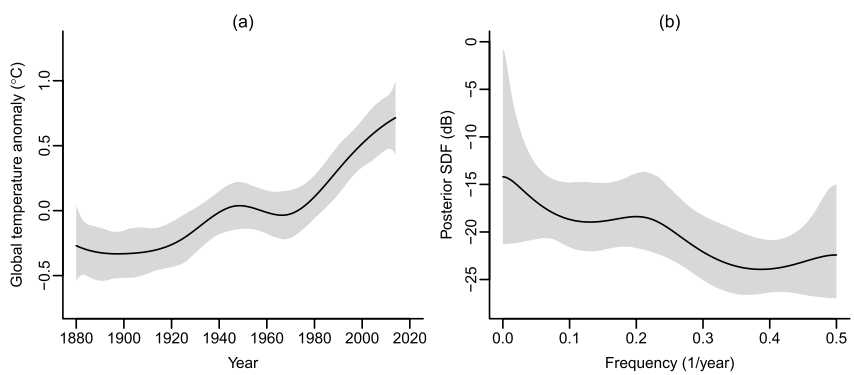

Figure 3. (a) Time series plots of the posterior mean trend for 1880–2014 (black lines), along with simultaneous 95% credible intervals (gray regions); (b) Posterior mean spectral density function (SDF) by frequency, of the AR(4) time series component $\{\nu_t\}$ (black lines), along with simultaneous 95% credible intervals (gray regions).

where $\phi_1, \ldots, \phi_p$ are the coefficients of the autoregressive process and the innovation variance is σ^2. By evaluating the spectral density function over posterior draws of the model parameters characterizing the time series component $\{\nu_t\}$, we obtain the posterior distribution of the spectral density function. For each draw, we transform the spectral density function to the decibel ($10 \log_{10}$) scale. After verifying at each frequency that a Gaussian approximation for the log spectral density function draws was good using normal quantile-quantile plots, we calculate simultaneous 95% bands for the spectral density function on the decibel scale using the method of Bolin and Lindgren (2015). Figure 3b summarizes the posterior distribution of the spectral density function on the decibel scale for the latent observed annual global anomaly process. In this figure, a straight horizontal line would correspond to a white noise process, and we see that no such line fits inside the credible band, indicating that the time series is correlated over time. The peak at low frequency is confounded with the trend, and the spectral peak between frequencies 0.15 and 0.25 indicates increased time series dependence in the global observed temperature anomaly on a scale of 4–7 years.

(In Supporting Information S1 see Table S1 for a discussion of the correlations in the time series errors over data products, and Figure S1 in Supporting Information S1 for a summary of the discrepancies attributable to each of the five original data products.)

3. Results

On the basis of the posterior summaries of the latent observed annual global temperature anomaly obtained by fitting the hierarchical model to the five data products, we now investigate how well the CMIP6 historical model runs, calculating annual global temperature anomalies, mimic these data products over the time period 1880–2014. We will look at four different comparisons: comparing the nonlinear trends over 1880–2014; comparing the difference in the mean temperatures from 1995 to 2014 relative to 1880–1899; comparing the temporal dependence as measured by the spectral density function of the time series error component; and comparing the overall distribution of the time series innovations (i.e., the estimated independent errors making up the time series process $\{\nu_t\}$).

3.1. Comparing Trends

For each annual global anomaly calculated from each CMIP6 historical model run, we fit the time series model given by Equations 1 and 2. We choose $b = 8$ basis functions and an autoregressive order of $p = 4$ to match the model assumed for the latent observed annual global temperature anomalies. Figures 4 and 5 display the estimated CMIP6-based-trends for each path plotted on top of the 95% simultaneous credible bands for the data-based trends calculated from the latent observed annual global temperature anomaly process $\{Y_t\}$. Different lines colors denote different CMIP6 modeling groups—the same colors are used in subsequent figures. The percentages of CMIP6 trends inside the credible bands are tabulated in the top right hand corner of each panel. Since the data-based trend band is a simultaneous band, if the model trends agree with the data-based trends they should rarely be outside the credible band. More formally for each CMIP6 model, if the CMIP6 trends calculated for each global temperature anomaly run were compatible with the trends generated from the latent observed annual global temperature anomaly process, then 95% of runs should lie within the simultaneous band. We find that 10 CMIP6 models each have 100% of runs that have estimated annual global trends that are within the simultaneous band (CMCC-CM2-HR4, CNRM-CM6-1-HR, CNRM-CM6-1, FGOALS-f3-L, FGOALS-g3, GISS-E2-2-H, INM-CM4-8, MPI-ESM1-2-HR, MPI-ESM1-2-LR, NorCPM1). In total, 225 out of the total number of 318 runs contain trends that are outside the simultaneous band. Of the 48 CMIP6 models with some disagreement, 33 models do not have any runs for which the trends are compatible with the latent observed annual global temperature anomaly process trends whereas 15 models have at least one run that is compatible with the latent observed trend.

3.2. Assessing Warming

To summarize the temperature shifts that we see in the annual global temperature anomalies for the CMIP6 model runs as compared to the latent observed annual global temperature anomaly, we calculate the difference in the mean temperatures from 1995–2014 relative to 1880–1899. This is similar to the comparison in IPCCs AR6 (Eyring et al., 2021), where the comparison is to 1850–1899. Figure 6 shows this difference for each CMIP6

10.1029/2022EA002468

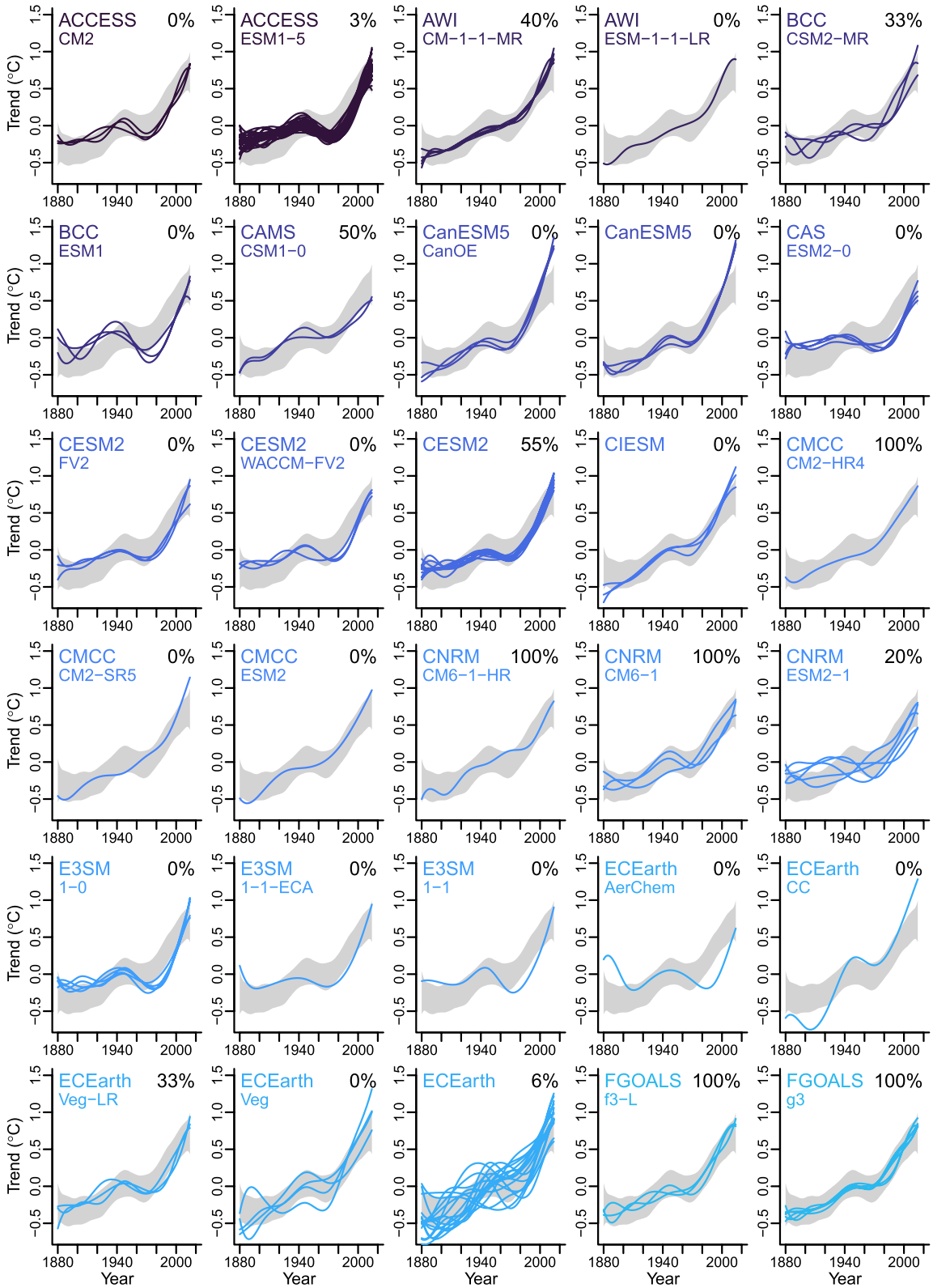

Figure 4. Trends for runs from the first 30 models, plotted on top of gray simultaneous 95% credible band for the trends calculated using the latent observation annual global temperature anomaly process. Different colors denote different CMIP6 modeling groups. The percentage of estimated trends for each CMIP6 model run that are included in the credible band are tabulated in the top right hand corner.

10.1029/2022EA002468

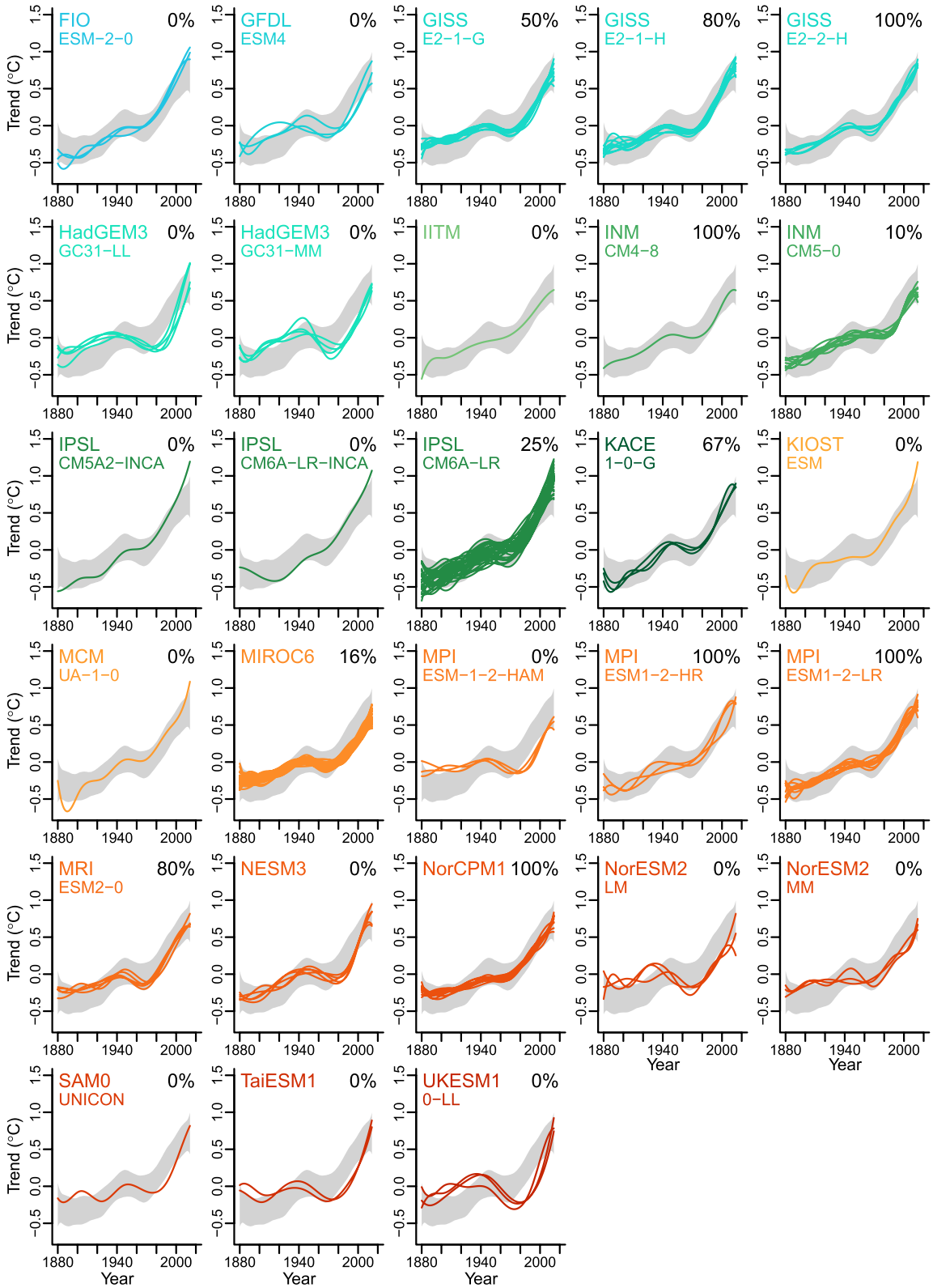

Figure 5. Trends for runs from the remaining models, plotted on top of gray simultaneous 95% credible band for the trends calculated using the latent observation annual global temperature anomaly process. Different colors denote different CMIP6 modeling groups. The percentage of estimated trends for each CMIP6 model run that are included in the credible band are tabulated in the top right hand corner.

10.1029/2022EA002468

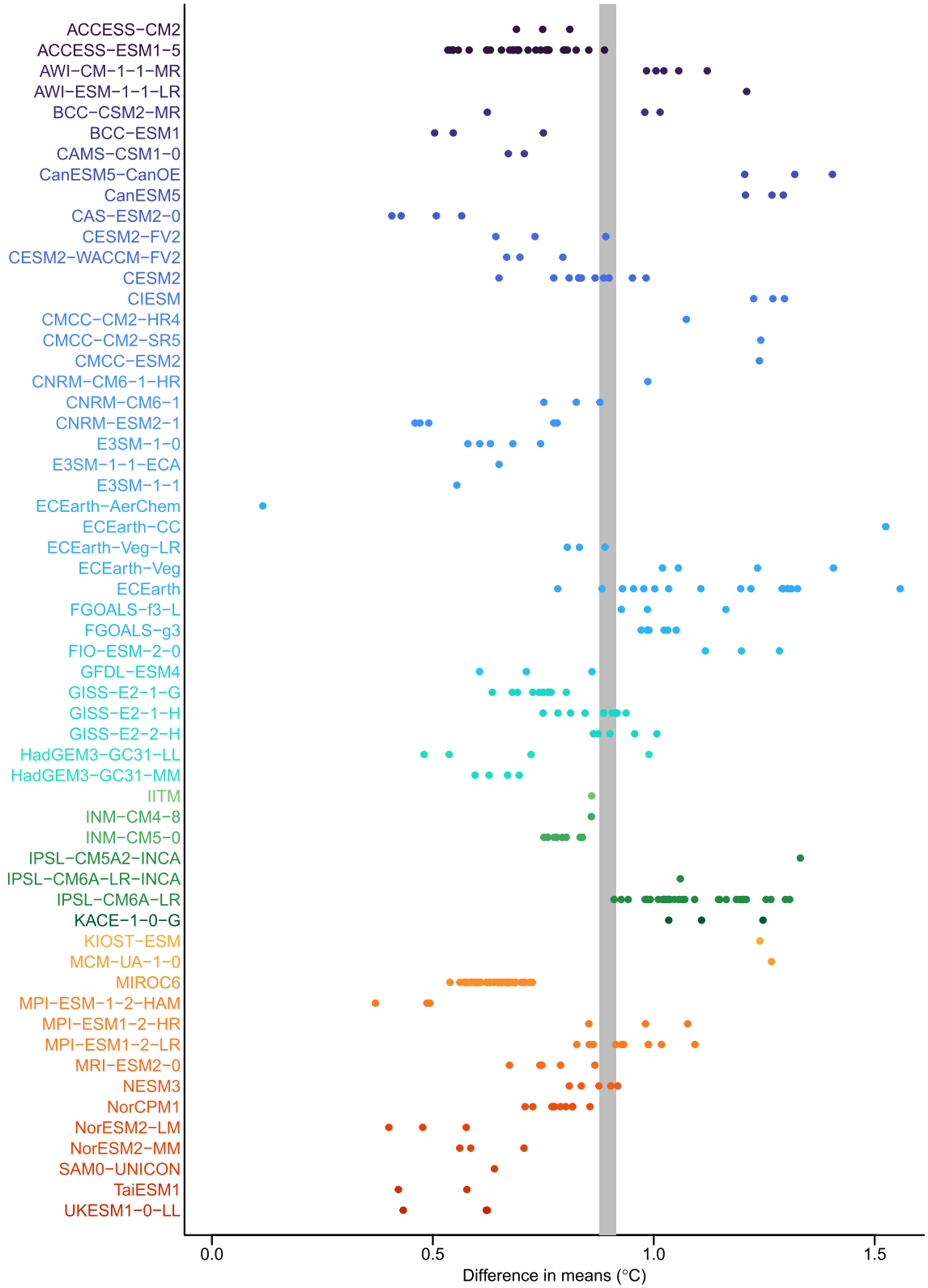

Figure 6. The difference in the mean temperature from 1995–2014 relative to 1880–1899 for each CMIP6 model run. Different colors denote different CMIP6 modeling groups. The gray region is a 95% credible interval calculating the same difference in mean temperature between the same time periods using the latent observed annual global temperature anomaly.

Table 1
With Respect to the Difference in the Mean Temperature From 1995–2014 Relative to 1880–1899, the Lists of the CMIP6 Models for Which all Model Runs Are Cooler, Similar, or Warmer as Compared to 95% Credible Intervals Calculated for Latent Observed Annual Global Temperature Anomalies

25 CMIP6 models where all runs are cooler than the observed credible band
ACCESS-CM2 BCC-ESM1 CAMS-CSM1-0 CAS-ESM2-0 CESM2
CNRM-ESM2-1 E3SM-1-0 E3SM-1-1-ECA E3SM-1-1 ECEarth-CC
GFDL-ESM4 GISS-E2-1-G HadGEM3-GC31-MM IITM INM-CM4-8
INM-CM5-0 MIROC6 MPI-ESM-1-2-HAM MRI-ESM2-0 NorCPM1
NorESM2-LM NorESM2-MM SAM0-UNICON TaiESM1 UKESM1-0-LL
0 CMIP6 models where all runs are within the observed credible band
19 CMIP6 models where all runs are warmer than the observed credible band
AWI-CM-1-1-MR AWI-ESM-1-1-LR CanESM5-CanOE CanESM5 CIESM
CMCC-CM2-HR4 CMCC-CM2-SR5 CMCC-ESM2 CNRM-CM6-1 ECEarth-Veg-LR
ECEarth-Veg FGOALS-f3-L FGOALS-g3 FIO-ESM-2-0 IPSL-CM5A2-INCA
IPSL-CM6A-LR KACE-1-0-G KIOST-ESM MCM-UA-1-0
9 CMIP6 models where any runs cover the observed credible band
BCC-CSM2-MR CESM2 ECEarth GISS-E2-1-H GISS-E2-2-H
HadGEM3-GC31-LL MPI-ESM1-2-HR MPI-ESM1-2-LR NESM3

Note. The final list indicates which CMIP6 models have at least one run with values that cover the observed credible band.

model run, as compared to a 95% credible interval from the posterior distribution estimated from the posterior distribution of the latent observed annual global temperature anomalies, $\{Y_t\}$, given the data products (the gray region). We estimate the mean temperature increase according to the data to be 0.896°C with a 95% credible interval of between 0.877 and 0.915. The corresponding result in AR6 is 0.85°C (0.69–0.95). As expected, our result which accounts for and combines information from multiple data products lead to a 95% interval that is narrower. For comparison, the CMIP6 ensemble mean of the amount of warming is 0.835°C.

In terms of this measure, 14 CMIP6 model runs agree with the latent observed annual global temperature anomaly. The mean temperature increase is smaller for 197 model runs and larger for 107 model runs. To further investigate warming by model, the first three lists of Table 1 indicate the collection of CMIP6 models for which all model runs are cooler, similar, or warmer than the latent observed annual global temperature anomalies (i.e., have levels of warming that are all less, inside, or higher than the observed simultaneous credible band). In addition, to guard against the possibility that the variability implied by our analysis of the data products may be different than the variability implied by the CMIP6 historical model runs, we list the collection of CMIP6 models which have any runs that cover the observed credible interval in the final list of Table 1.

Examining Figure 6 and Table 1 we find substantial homogeneity in the mean temperature increases among the runs for each model, and across different versions of models from the same modeling group. For example, both NorESM2 models tend to have smaller temperature increases than implied by the latent observed anomalies, whereas except for the single ECEarth3-AerChem run, the ECEarth models are more likely to have larger increases.

3.3. Comparing Temporal Dependence

Again using the time series model given by Equation 1 and 2 fit to each CMIP6 model run, we evaluate the spectral density function for the autoregressive time series process. Figures 7 and 8 display the spectral density functions on the decibel scale estimated for each CMIP6 model run plotted on top of the 95% simultaneous credible bands for the spectral density function on decibel scale for the latent observed annual temperature anomaly process. (The method to calculate the simultaneous bands for spectral density functions is described at the end of Section 2.3.) The percentages of CMIP6 spectral density functions inside the credible bands are tabulated in the top right hand corner of each panel. In terms of dependence, all the historical runs from three CMIP6

10.1029/2022EA002468

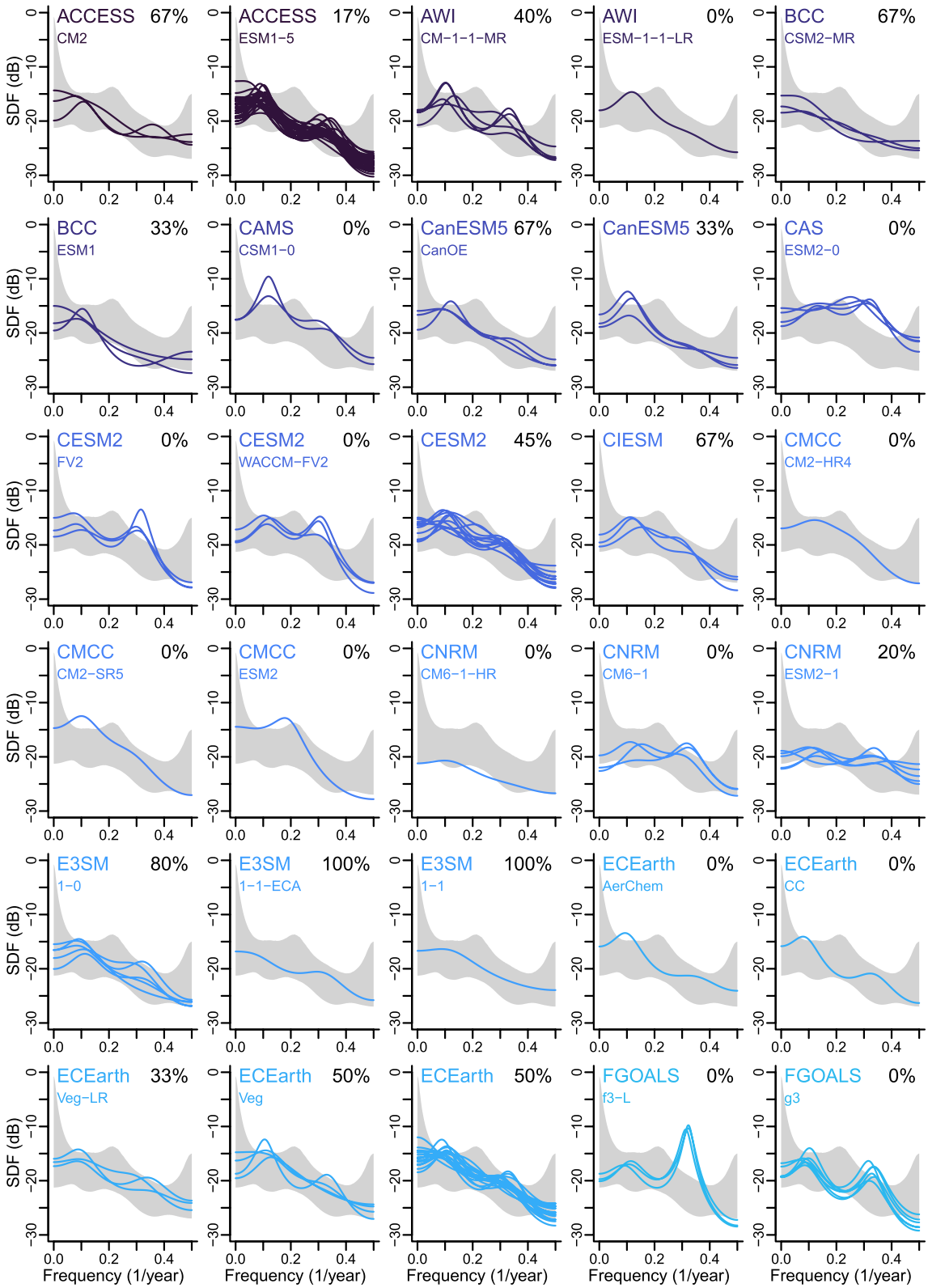

Figure 7. Estimated spectral density functions on the decibel scale for runs from the first 30 models, plotted on top of the gray simultaneous 95% confidence band for the spectral density function on the decibel scale calculated using the latent observation annual global temperature anomaly process. Different colors denote different CMIP6 modeling groups. The percentage of estimated spectral density functions for each CMIP6 model run that are included in the credible band are tabulated in the top right hand corner.

10.1029/2022EA002468

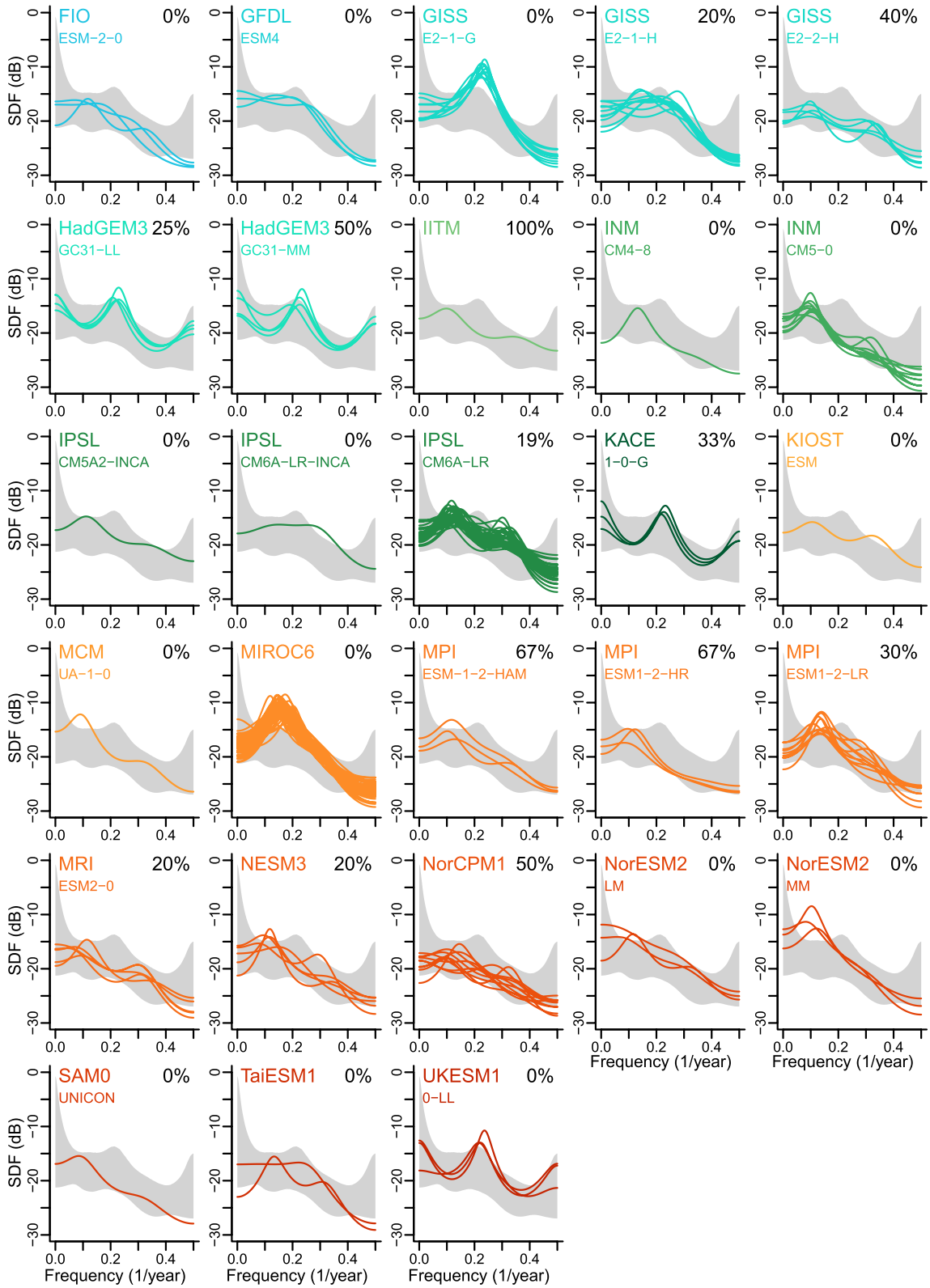

Figure 8. Estimated spectral density functions on the decibel scale for runs from the remaining models, plotted on top of the gray simultaneous 95% confidence band for the spectral density function on the decibel scale calculated using the latent observation annual global temperature anomaly process. Different colors denote different CMIP6 modeling groups. The percentage of estimated spectral density functions for each CMIP6 model run that are included in the credible band are tabulated in the top right hand corner.

models (E3SM-1-1-ECA, E3SM-1-1, and IITM) agree with the data-based bands, calculated from the latent observed annual global temperature anomalies. In total, 249 out of the total number of 318 runs contain spectral density functions that are outside the data-based simultaneous band. Of the 55 CMIP6 models with some disagreement, 29 models do not have any runs for which the dependence are compatible with the data-based bands whereas 26 models have at least one run that is compatible with the data-based bands. Among the models that do not agree, we often see spectral peaks at different frequencies from those given by the data. An AR(4) time series model permits estimated spectral density functions with one or two peaks. Some models have single peaks (e.g., E3SM-1-1, INM-CM4-8 and both MPI models) whereas other models have double peaks (e.g., two of the CESM2 models and both FGOALS models). While the spectral peaks lie outside the bands calculated for the latent observed annual global temperature anomalies, the location of the single spectral peak matches the data reconstruction quite well for a number of different models such as GISS-E2-1-G, MIROC6, both HadGEM3 models, and UKESM1-0-LL.

3.4. Comparing Residuals From Trend

If we think of the trend as describing the changing climate, the residuals from the trend roughly describe the noise in the temperature system. For both the latent observed annual global temperature anomalies and the CMIP6 model runs we assume an autoregressive process to capture this noise. If the model fits well, the residuals of the Gaussian autoregressive process, also known as the time series innovations, should be centered around zero with a constant spread, and independent over time.

In this section we will look at these innovations with a very broad brush, in that we do not look at climate model innovations separately for each model. Rather, we compare the innovations calculated after fitting the time series model given by Equation 1 and 2 to the five original global data products with the innovations calculated to all the CMIP6 models runs using the same structure of time series model. For this comparison, we use what is called a shiftplot (Doksum & Sievers, 1976). Consider two distribution functions F and G, and define the shift function

$$D(x) = F^{-1}(G(x)) - x.$$

The shift function is identically zero when the continuous distributions F and G are equal. We estimate $D(x)$ from a sample $x_1, \ldots x_n, y_1, \ldots, y_m$ from the distributions F and G, respectively, using the empirical distribution function $F_n(X) = \sum_{i=1}^{n} 1(X_i \leq x)/n$, and the corresponding estimate $G_m(x)$, and plugging these into the definition of $D(x)$, using the right inverse of $F_n(x)$, also called the empirical quantile function. We then plot the estimated shift function against x. Following Doksum and Sievers (1976) we get a simultaneous confidence band for the function from the distribution of the Kolmogorov-Smirnov statistic. The idea of this shiftplot (sometimes called residual quantile-quantile plot) is to assess whether the distributions of yearly innovations (over all years and over all runs or products) are the same. If so, the plot should look like a horizontal line at height 0. A horizontal line at a different height indicates a location shift, while a line with nonzero slope indicates a change in variability. Figure 9 shows that the red horizontal line at zero falls well inside the 95% simultaneous confidence band, indicating that the two innovation distributions (one for the latent observed annual global temperature anomalies and one for all the CMIP6 model runs) cannot be distinguished statistically.

4. Conclusions and Discussion

The question we are trying to answer in this paper is whether the historical climate simulations in CMIP6 are a reasonable description of the observed temperatures. To first learn about observed temperatures we use modern hierarchical statistical tools to produce latent posterior reconstructions of observed annual global temperature anomalies from 1880 to 2014 using five major observational products. With these reconstructions, we are able to calculate simultaneous credible bands for both the trend, smooth nonlinear changes in the mean global temperature anomaly, and the spectral density function that characterizes the residual time series dependence. We then determine whether estimates of the trend and spectral density function from CMIP6 model runs are compatible with the 95% simultaneous bands calculated on the basis of the observed data products. In addition, our approach allows us to compare how well the models capture observed warming relative to the early years. While there are homogeneities across CMIP6 modeling groups and model runs with respect to the trends and temporal dependencies inherent in global mean temperature anomalies, there are substantial differences between these CMIP6 model runs and the posterior distribution of observed global mean temperature anomalies as constructed

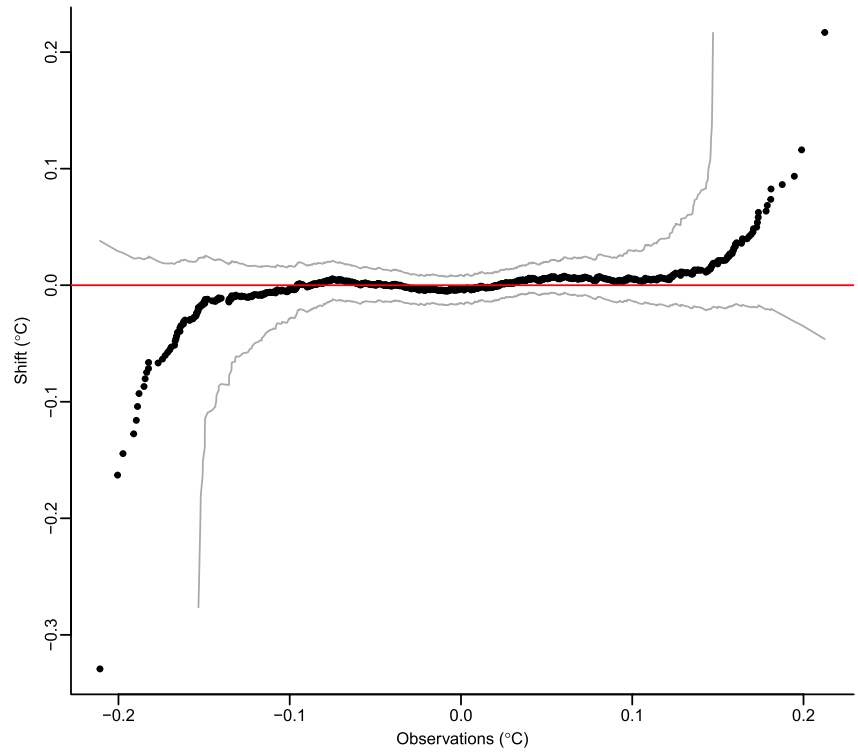

Figure 9. Shiftplot of the time series innovations calculated using the latent observed annual global temperature anomaly relative to the innovations calculated from all CMIP6 model runs. The thin black curves form a 95% simultaneous confidence band, and the red horizontal line corresponds to the two distributions being equal.

from the five data products. Both the data reconstruction and CMIP6 model runs show warming in recent years, but the amount of warming varies substantially. In addition, the CMIP6-based model runs often indicate different temporal structure as compared to the data reconstruction.

In Craigmile and Guttorp (2022), we performed a more stringent assessment of two other data products, the ERA5 reanalysis and the UAH satellite global temperature series. Instead of looking at the trend behavior, we required the two series to fall within the simultaneous credibility band for the posterior latent observed annual global anomaly. The reanalysis satisfied this criterion, while the satellite series did not.

Our analysis is based on estimating a smooth trend using $b = 8$ cubic b-spline basis functions, and estimating the temporal dependence using a stationary Gaussian autoregressive process of order $p = 4$. If we choose to estimate the trend or temporal dependence in some other fashion, we may obtain different results for the trend and spectral density function. (Again, the Supporting Information S1 contains a robustness study, indicating that our reconstruction of the latent observed annual global temperature anomaly is robust to both changes the number of basis functions used, b, in the nonlinear trend component, as well as the autoregressive model order, p.) Also, throughout our analyses we have compared the estimated trend or spectral density function for each CMIP6 model run to simultaneous credible bands calculated from the data. As a consequence we summarize whether each quantity is wholly inside the simultaneous credible bands. Inspecting the trends in Figures 4 and 5 further, we find that there are some CMIP6 model trends with good agreement over most years, but poor agreement in the early and later years.

We have chosen an annual time scale in order to avoid having to fit seasonal effects that would be necessary on a monthly time scale. As shown by Thomson (1995), the dominant cycle in global monthly temperature is the anomalistic year (time between perihelia) rather than the slightly shorter tropical year (time between vernal equinoxes), and therefore deseasonalization must be done very carefully. It may also be of interest to split the data into southern and northern hemispheres. A similar model for data products to that used for the global data could be used, however, this was not attempted as it would require us to model the bivariate time series dependence that is present between and within the two hemispheres. A comparison of the CMIP6 model simulations with a

reconstructed latent observed global temperature anomaly on a seasonal and hemispheric scale is left for further research. Furthermore, one could compare gridded products. The main difficulty here is that the data products are produced on different grids, which is also true for the CMIP6 models. In order to align all these products to a common grid, it is necessary to model the spatio-temporal correlation structure of each product. We have not attempted this, as it would stretch our computational assets beyond their capacity.

An issue with computing global mean temperature from observational data is that the ocean air temperature is replaced by sea surface temperature (SST). In the CERA-20C reanalysis the difference between the two measurements is fairly constant on a global scale at 1.5°C, with the SST warmer (Feng et al., 2018). However, when the measurements are replaced by anomalies, this difference disappears, so the anomalies from these two quantities can be considered exchangeable on an annual-global spatio-temporal scale. Recent work, using relatively sparse data from the TAO buoys, indicates that the difference may recently have been increasing in the equatorial Pacific (Rubino et al., 2020). We are currently investigating this issue in other research.

A future aim of our research is to model the climate simulations, which would be substantially more complicated than the analysis presented in this paper, and necessitate some kind of sampling framework, perhaps in the style of Ribes et al. (2017). Rather than comparing model simulations to reconstructed data products as we do in this paper, modeling the climate simulations directly could allow us to learn about time-varying commonalities and differences between climate models and between modeling groups. Such an approach requires a different hierarchical model, which must include an assessment of the dependence between paths, climate models, and climate modeling groups, using tools from, for example, Knutti et al. (2013); Brunner et al. (2020). (Also see, e.g., Smith et al. (2009); Tebaldi and Sansó (2009); Guinness and Hammerling (2017); Chang and Guillas (2018); Zumwald et al. (2020); Sansom et al. (2021) for the hierarchical statistical modeling of climate model simulations.) We envision some variant of functional analysis of variance (e.g., Meiring (2007); Kaufman and Sain (2010)), and we leave this for another paper.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The computations use version 4.3.0 of the R software package (R Core Team, 2023). All code, and R objects with data sets and processed climate model runs are freely available at Craigmile and Guttorp (2023).

Acknowledgments
None.

References

Berliner, L., Wikle, C., & Cressie, N. (2000). Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. *Journal of Climate*, *13*(22), 3953–3968. https://doi.org/10.1175/1520-0442(2001)013<3953:llpops>2.0.co;2

Bolin, D., & Lindgren, F. (2015). Excursion and contour uncertainty regions for latent Gaussian models. *Journal of the Royal Statistical Society: Series B*, *77*(1), 85–106. https://doi.org/10.1111/rssb.12055

Bolin, D., & Lindgren, F. (2018). Calculating probabilistic excursion sets and related quantities using excursions. *Journal of Statistical Software*, *86*(5), 1–20. https://doi.org/10.18637/jss.v086.i05

Brillinger, D. (1981). *Time series: Data analysis and theory*. Holt.

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., & Knutti, R. (2020). Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth System Dynamics*, *11*(4), 995–1012. https://doi.org/10.5194/esd-11-995-2020

Carvalho, C., & Rickershauser, J. (2013). Characterizing the uncertainty of climate change projections using hierarchical models. In A. O'Hagan & M. West (Eds.), *The Oxford handbook of applied Bayesian analysis (chap. 20)*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198703174.013.20

Chang, K.-L., & Guillas, S. (2018). Computer model calibration with large non-stationary spatial outputs: Application to the calibration of a climate model. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *68*(1), 51–78. https://doi.org/10.1111/rssc.12309

Cowtan, K., & Way, R. (2014). Global temperature reconstructions version 2. https://doi.org/10.15124/20ee85c3-f53c-4ab6-8e50-270b0ddd3686

Craigmile, P. F., & Guttorp, P. (2011). Space-time modelling of trends in temperature series. *Journal of Time Series Analysis*, *32*(4), 378–395. https://doi.org/10.1111/j.1467-9892.2011.00733.x

Craigmile, P. F., & Guttorp, P. (2022). A combined estimate of global temperature. *Environmetrics*, *33*(3), e2706. https://doi.org/10.1002/env.2706

Craigmile, P. F., & Guttorp, P. (2023). Comparing CMIP6 climate model simulations of annual global mean temperatures to a new combined data product, v1.0.0 [software][data]. Zenodo. https://doi.org/10.5281/zenodo.8347096

de Boor, C. (1978). *A practical guide to splines*. Springer-Verlag.

Doksum, K. A., & Sievers, G. L. (1976). Plotting with confidence: Graphical comparisons of two populations. *Biometrika*, *63*(3), 421–434. https://doi.org/10.1093/biomet/63.3.421

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Eyring, V., Gillett, N., Achuta Rao, N., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., et al. (2021). Human influence on the climate system. In V. Masson-Delmotte, et al. (Eds.), *Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change* (pp. 423–552). Cambridge University Press.

Fan, X., Duan, Q., Shen, C., Wu, Y., & Xing, C. (2020). Global surface air temperatures in CMIP6: Historical performance and future changes. *Environmental Research Letters*, *15*(10), 104056. https://doi.org/10.1088/1748-9326/abb051

Feng, X., Haines, K., & de Boisson, E. (2018). Coupling of surface air and sea surface temperatures in the CERA-20C reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *144*(710), 195–207. https://doi.org/10.1002/qj.3194

Guinness, J., & Hammerling, D. (2017). Compression and conditional emulation of climate model output. *Journal of the American Statistical Association*, *113*(521), 56–67. https://doi.org/10.1080/01621459.2017.1395339

Hausfather, Z., Marvel, K., Schmidt, G. A., Nielsen-Gammon, J. W., & Zelinka, M. (2022). Climate simulations: Recognize the "hot model" problem. *Nature*, *605*(7908), 26–29. https://doi.org/10.1038/d41586-022-01192-2

Hind, A., Moberg, A., & Sundberg, R. (2012). Statistical framework for evaluation of climate model simulations by use of climate proxy data from the last millennium – Part 2: A pseudo-proxy study addressing the amplitude of solar forcing. *Climate of the Past*, *8*(4), 1355–1365. https://doi.org/10.5194/cp-8-1355-2012

Huang, B., Menne, M. J., Boyer, T., Freeman, E., Gleason, B. E., Lawrimore, J. H., et al. (2019). Uncertainty estimates for sea surface temperature and land surface air temperature in NOAAGlobalTemp version 5. *Journal of Climate*, *33*(4), 1351–1379. https://doi.org/10.1175/JCLI-D-19-0395.1

Ishihara, K. (2007). Estimation of standard errors in global average surface temperatures (in Japanese). *Weather Service Bulletin*, *74*, 19–26.

Kaufman, C. G., & Sain, S. R. (2010). Bayesian functional ANOVA modeling using Gaussian process prior distributions. *Bayesian Analysis*, *5*(1), 123–149. https://doi.org/10.1214/10-ba505

Knutti, R., Masson, D., & Gettelman, A. (2013). Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters*, *40*(6), 1194–1199. https://doi.org/10.1002/grl.50256

Lenssen, N., Schmidt, G., Hansen, J., Menne, M., Persin, A., Ruedy, R., & Zyss, D. (2019). Improvements in the GISTEMP uncertainty model. *Journal of Geophysical Research: Atmospheres*, *124*(12), 6307–6326. https://doi.org/10.1029/2018jd029522

Liang, Y., Gillett, N. P., & Monahan, A. (2020). Climate model projections of 21st century global warming constrained using the observed warming trend. *Geophysical Research Letters*, *47*(12), e2019GL086757. https://doi.org/10.1029/2019gl086757

McKinnon, K. A. (2022). Discussion on A combined estimate of global temperature. *Environmetrics*, *33*(3). https://doi.org/10.1002/env.2721

Meiring, W. (2007). Oscillations and time trends in stratospheric ozone levels: A functional data analysis approach. *Journal of the American Statistical Association*, *102*(479), 788–802. https://doi.org/10.1198/016214506000000825

Moberg, A., Sundberg, R., Grudd, H., & Hind, A. (2015). Statistical framework for evaluation of climate model simulations by use of climate proxy data from the last millennium – Part 3: Practical considerations, relaxed assumptions, and using tree-ring data to address the amplitude of solar forcing. *Climate of the Past*, *11*(3), 425–448. https://doi.org/10.5194/cp-11-425-2015

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., et al. (2020). An updated assessment of near-surface temperature change from 1850: The HadCRUT5 dataset. *Journal of Geophysical Research: Atmospheres*, *126*(3), e2019JD032361. https://doi.org/10.1029/2019JD032361

Percival, D. B., & Walden, A. T. (1993). *Spectral analysis for univariate time series*. Cambridge University Press.

R Core Team. (2023). R: A language and environment for statistical computing [computer software manual]. R-project. Retrieved from https://www.R-project.org/

Ribes, A., Qasmi, S., & Gillett, N. P. (2021). Making climate projections conditional on historical observations. *Science Advances*, *7*(4), eabc0671. https://doi.org/10.1126/sciadv.abc0671

Ribes, A., Zwiers, F. W., Azaïs, J.-M., & Naveau, P. (2017). A new statistical approach to climate change detection and attribution. *Climate Dynamics*, *48*(1–2), 367–386. https://doi.org/10.1007/s00382-016-3079-6

Rohde, R. A., & Hausfather, Z. (2020). The Berkeley Earth land/ocean temperature record. *Earth System Science Data Discussions*, *12*(4), 3469–3479. https://doi.org/10.5194/essd-12-3469-2020

Rubino, A., Zanchettin, D., De Rovere, F., & McPhaden, M. (2020). On the interchangeability of sea-surface and near-surface air temperature anomalies in climatologies. *Scientific Reports*, *10*(1), 7433. https://doi.org/10.1038/s41598-020-64167-1

Sansom, P. G., Stephenson, D. B., & Bracegirdle, T. J. (2021). On constraining projections of future climate using observations and simulations from multiple climate models. *Journal of the American Statistical Association*, *116*(534), 1–12. https://doi.org/10.1080/01621459.2020.1851696

Smith, R. L., Tebaldi, C., Nychka, D., & Mearns, L. O. (2009). Bayesian modeling of uncertainty in ensembles of climate models. *Journal of the American Statistical Association*, *104*(485), 97–116. https://doi.org/10.1198/jasa.2009.0007

Stocker, T. F., Dahe, Q., Plattner, G.-K., Tignor, M. M., Allen, S. K., Boschung, J., et al. (2013). *Climate change 2013: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press.

Sundberg, R., Moberg, A., & Hind, A. (2012). Statistical framework for evaluation of climate model simulations by use of climate proxy data from the last millennium – Part 1: Theory. *Climate of the Past*, *8*(4), 1339–1353. https://doi.org/10.5194/cp-8-1339-2012

Tebaldi, C., & Sansó, B. (2009). Joint projections of temperature and precipitation change from multiple climate models: A hierarchical Bayesian approach. *Journal of the Royal Statistical Society: Series A*, *172*(1), 83–106. https://doi.org/10.1111/j.1467-985x.2008.00545.x

Thomson, D. J. (1995). The seasons, global temperature, and precession. *Science, New Series*, *268*(5207), 59–68. https://doi.org/10.1126/science.268.5207.59

Tingley, M., Craigmile, P. F., Haran, M., Li, B., Mannshardt, E., & Rajaratnam, B. (2015). On discriminating between GCM forcing configurations using Bayesian reconstructions of Late-Holocene temperatures. *Journal of Climate*, *28*(20), 8264–8281. https://doi.org/10.1175/jcli-d-15-0208.1

Williams, M., & Eggleston, S. (2017). *Using indicators to explain our changing climate to policymakers and the public* (Vol. 66). WMO Bulletin. Retrieved from https://public.wmo.int/en/resources/bulletin/using-indicators-explain-our-changing-climate-policymakers-and-public

Willmott, C., & Robeson, S. (1995). Climatologically aided interpolation (CAI) of terrestrial air temperature. *International Journal of Climatology*, *15*(2), 221–229. https://doi.org/10.1002/joc.3370150207

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., et al. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, *47*(1), e2019GL085782. https://doi.org/10.1029/2019gl085782

Zumwald, M., Knüsel, B., Baumberger, C., Hirsch Hadorn, G., Bresch, D. N., & Knutti, R. (2020). Understanding and assessing uncertainty of observational climate datasets for model evaluation using ensembles. *WIREs Climate Change*, *11*(5). https://doi.org/10.1002/wcc.654

References From the Supporting Information

Bader, D. C., Leung, R., Taylor, M., & McCoy, R. B. (2019). *E3SM-Project E3SM1.1 model output prepared for CMIP6 C4MIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.11441

Bader, D. C., Leung, R., Taylor, M., & McCoy, R. B. (2020). *E3SM-Project E3SM1.1ECA model output prepared for CMIP6 C4MIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.11443

Bentsen, M., Oliviè, D. J. L., Seland, y., Toniazzo, T., Gjermundsen, A., Graff, L. S., et al. (2019). *NCC NorESM2-MM model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.506

Bethke, I., Wang, Y., Counillon, F., Kimmritz, M., Fransner, F., Samuelsen, A., et al. (2019). *NCC NorCPM1 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.10843

Boucher, O., Denvil, S., Levavasseur, G., Cozic, A., Caubel, A., Foujols, M.-A., et al. (2018). *IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1534

Boucher, O., Denvil, S., Levavasseur, G., Cozic, A., Caubel, A., Foujols, M.-A., et al. (2020). *IPSL IPSL-CM6A-LR-INCA model output prepared for CMIP6 AerChemMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.13581

Boucher, O., Denvil, S., Levavasseur, G., Cozic, A., Caubel, A., Foujols, M.-A., et al. (2021). *IPSL IPSL-CM5A2-INCA model output prepared for CMIP6 AerChemMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.13641

Brovkin, V., Wieners, K.-H., Giorgetta, M., Jungclaus, J., Reick, C., Esch, M., et al. (2019). *MPI-M MPIESM1.2-LR model output prepared for CMIP6 C4MIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.748

Byun, Y.-H., Lim, Y.-J., Sung, H. M., Kim, J., Sun, M., & Kim, B.-H. (2019). *NIMS-KMA KACE1.0-G model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.2241

Cao, J., & Wang, B. (2019). *NUIST NESMv3 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.2021

Chai, Z. (2020). *CAS CAS-ESM2.0 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1944

Danabasoglu, G. (2019a). *NCAR CESM2-FV2 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.11281

Danabasoglu, G. (2019b). *NCAR CESM2 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.2185

Danabasoglu, G. (2019c). *NCAR CESM2-WACCM model output prepared for CMIP6 AerChemMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.10023

Danek, C., Shi, X., Stepanek, C., Yang, H., Barbi, D., Hegewald, J., & Lohmann, G. (2020). *AWI AWI-ESM1.1LR model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.9301

Dix, M., Bi, D., Dobrohotoff, P., Fiedler, R., Harman, I., Law, R., et al. (2019). *CSIRO-ARCCSS ACCESS-CM2 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.2281

EC-Earth Consortium. (2019). *EC-Earth-Consortium EC-Earth3 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.181

EC-Earth Consortium. (2020a). *EC-Earth-Consortium EC-Earth3-AerChem model output prepared for CMIP6 AerChemMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.699

EC-Earth Consortium. (2020b). *EC-Earth-Consortium EC-Earth-3-CC model output prepared for CMIP6 C4MIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.650

EC-Earth Consortium. (2020c). *EC-Earth-Consortium EC-Earth3-Veg-LR model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.643

EC-Earth Consortium. (2020d). *EC-Earth-Consortium EC-Earth3-Veg-LR model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.4707

Fogli, P. G., Iovino, D., & Lovato, T. (2020). *CMCC CMCC-CM2-SR5 model output prepared for CMIP6 OMIP OMIP1*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.13230

GISS. (2018). *NASA-GISS GISS-E2.1H model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1421

GISS. (2019). *NASA-GISS GISS-E2-1-G-CC model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.11657

GISS. (2019). *NASA-GISS GISS-E2.2H model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.15861

Huang, W. (2019). *THU CIESM model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1352

Jun, M., Knutti, R., & Nychka, D. W. (2008). Spatial analysis to quantify numerical model bias and dependence. *Journal of the American Statistical Association*, *103*(483), 934–947. https://doi.org/10.1198/016214507000001265

Jungclaus, J., Bittner, M., Wieners, K.-H., Wachsmann, F., Schupfner, M., Legutke, S., et al. (2019). *MPI-M MPIESM1.2-HR model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.741

Kim, Y., Noh, Y., Kim, D., Lee, M.-I., Lee, H. J., Kim, S. Y., & Kim, D. (2019). *KIOST KIOST-ESM model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1922

Krasting, J. P., John, J. G., Blanton, C., McHugh, C., Nikonov, S., Radhakrishnan, A., et al. (2018). *NOAA-GFDL GFDL-ESM4 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1407

Lee, W.-L., & Liang, H.-C. (2019). *AS-RCEC TaiESM1.0 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.9684

Li, L. (2019). *CAS FGOALS-g3 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1783

Lin, P. (2019). *CAS FGOALS-f3-L model output prepared for CMIP6 OMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.2044

Lovato, T., Peano, D., & Butenschön, M. (2021). *CMCC CMCC-ESM2 model output prepared for CMIP6 C4MIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.13163

Neubauer, D., Ferrachat, S., Siegenthaler-Le Drian, C., Stoll, J., Folini, D. S., Tegen, I., et al. (2019). *HAMMOZ-Consortium MPI-ESM1.2-HAM model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1622

Park, S., & Shin, J. (2019). *SNU SAM0-UNICON model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1489
Prajesh, A., Ayantika, D., Modi, A., Narayanasetti, S., Raghavan, K., Panickal, S., & Singh, M. (2019). *CCCR-IITM IITM-ESM model output data prepared for CMIP6 CMIP/DECK*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.44
Qin, Y., Klein, S. A., Zelinka, M. D., & Golaz, C. (2021). *LLNL E3SM1.0 model output prepared for CMIP6 CFMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.15093
Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., & Andrews, T. (2019). *MOHC HadGEM3-GC31-MM model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.420
Rong, X. (2019). *CAMS CAMS_CSM1.0 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1399
Scoccimarro, E., Bellucci, A., & Peano, D. (2020). *CMCC CMCC-CM2-HR4 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1358
Seferian, R. (2018). *CNRM-CERFACS CNRM-ESM2-1 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1391
Seland, O., Bentsen, M., Oliviè, D. J. L., Toniazzo, T., Gjermundsen, A., Graff, L. S., et al. (2019). *NCC NorESM2-LM model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.502
Semmler, T., Danilov, S., Rackow, T., Sidorenko, D., Barbi, D., Hegewald, J., et al. (2018). *AWI AWI-CM1.1MR model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.2686
Song, Z., Qiao, F., Bao, Y., Shu, Q., Song, Y., & Yang, X. (2019). *FIO-QLNM FIO-ESM2.0 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.9047
Stouffer, R. (2019). *U of Arizona MCM-UA-1-0 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.2421
Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019a). *CCCma CanESM5-CanOE model output prepared for CMIP6 C4MIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.10203
Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019b). *CCCma CanESM5-CanOE model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.10205
Tang, Y., Rumbold, S., Ellis, R., Kelley, D., Mulcahy, J., Sellar, A., et al. (2019). *MOHC UKESM1.0-LL model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1569
Tatebe, H., & Watanabe, M. (2018). *MIROC MIROC6 model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.5603
Voldoire, A. (2018). *CNRM-CERFACS CNRM-CM6-1 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1375
Voldoire, A. (2019). *CNRM-CERFACS CNRM-CM6-1-HR model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1385
Volodin, E., Mortikov, E., Gritsun, A., Lykossov, V., Galin, V., Diansky, N., et al. (2019a). *INM INM-CM4-8 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1422
Volodin, E., Mortikov, E., Gritsun, A., Lykossov, V., Galin, V., Diansky, N., et al. (2019b). *INM INM-CM5-0 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1423
Webb, M. (2019). *MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CFMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.435
Xin, X., Zhang, J., Zhang, F., Wu, T., Shi, X., Li, J., et al. (2018). *BCC BCC-CSM2MR model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1725
Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., et al. (2019). *MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.621
Zhang, J., Wu, T., Shi, X., Zhang, F., Li, J., Chu, M., et al. (2018). *BCC BCC-ESM1 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1734
Ziehn, T., Chamberlain, M., Lenton, A., Law, R., Bodman, R., Dix, M., et al. (2019). *CSIRO ACCESS-ESM1.5 model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.2288