Empirical prediction of short-term annual global temperature variability

Patrick Thomas Brown[1] and Ken Caldeira[2]

[1]San Jose State University
[2]Carnegie Institution for Science

November 23, 2022

Abstract

Global mean surface air temperature (T) variability on subdecadal timescales can be of substantial magnitude relative to the long-term global warming signal and such variability has been associated with considerable environmental and societal impacts. Therefore, probabilistic foreknowledge of short-term T evolution may be of value for anticipating and mitigating some course-resolution climate-related risks. Here we present an empirically-based methodology that utilizes global spatial patterns of annual surface air temperature to predict subsequent annual T anomalies via Partial Least Squares Regression. The method's skill is achieved via information on the state of long-term global warming as well as the state and recent evolution of the El Niño-Southern Oscillation and the Interdecadal Pacific Oscillation. We test the out-of-sample skill of the methodology using a "forecast mode" where statistical predictions are made precisely as they would have been if the procedure had been operationalized starting in the year 2000. The forecast errors for lead times of 1 to 4 years are smaller than naïve benchmarks using persistence and perform favorably relative to most dynamical Global Climate Models retrospectively initialized to the observed state of the climate system. Thus, this method can used as a computationally-efficient benchmark for dynamical model forecast systems.

Empirical prediction of short-term annual global temperature variability

Patrick T. Brown[1*], Ken Caldeira[2]

[1] *Department of Meteorology and Climate Science, San Jose State University, San Jose, California*

[2] *Department of Global Ecology, Carnegie Institution for Science, Stanford, California*

*Corresponding author: Patrick T. Brown, E-mail: patrick.brown@sjsu.edu

Abstract

Global mean surface air temperature (T_{global}) variability on subdecadal timescales can be of substantial magnitude relative to the long-term global warming signal and such variability has been associated with considerable environmental and societal impacts. Therefore, probabilistic foreknowledge of short-term T_{global} evolution may be of value for anticipating and mitigating some course-resolution climate-related risks. Here we present an empirically-based methodology that utilizes global spatial patterns of annual surface air temperature to predict subsequent annual T_{global} anomalies via Partial Least Squares Regression. The method's skill is achieved via information on the state of long-term global warming as well as the state and recent evolution of the El Niño-Southern Oscillation and the Interdecadal Pacific Oscillation. We test the out-of-sample skill of the methodology using a "forecast mode" where statistical predictions are made precisely as they would have been if the procedure had been operationalized starting in the year 2000. The forecast errors for lead times of 1 to 4 years are smaller than naïve benchmarks using persistence and perform favorably relative to most dynamical Global Climate Models retrospectively initialized to the observed state of the climate system. Thus, this method can used as a computationally-efficient benchmark for dynamical model forecast systems.

Plain Language Summary

Year-to-year global temperature variability can be large compared to the long-term progression of global warming and such year-to-year variability has been shown to have considerable environmental and societal effects. Thus, approximate foreknowledge of yearly global temperature deviations should be of value for anticipating some climate impacts. This study presents an application a statistical technique, Partial Least Squares Regression, to the problem of year-to-year global temperature prediction. For the task of predicting global temperature one to four years ahead of time, we find that the method is skillful relative to simple benchmarks and it is competitive with predictions produced from much more computationally-expensive Global Climate Models.

Key points: (140 character limit including spaces)

- Global patterns of annual local surface air temperature can be used to predict subsequent annual global temperature deviations
- The state of Global Warming, El Niño and the IPO constrain subsequent annual global temperature to within ~0.45°C ($\pm 2\sigma$)
- The method is skillful relative to a persistence benchmark and it is competitive with hindcasts from initialized Global Climate Models

1. Introduction

Human-caused increases in the atmospheric concentration of well-mixed greenhouse gases are causing pronounced global climate change on decadal to centennial timescales [*Bindoff et al., 2013*]. Perhaps the most-recognized measure of this change is the long-term effect on global mean surface air temperature (T_{global}), which has increased at a rate of ~0.14°C per decade since the middle of the 20th century [*Hansen et al.*, 2010]. Superimposed on top of this externally-forced decadal to centennial-scale warming, is variability which is mostly unforced and spontaneously generated from interactions internal to the ocean-atmosphere-land system [*Bindoff et al., 2013*]. Such T_{global} variability is not as persistent as the contemporary externally-forced signal but it can be substantial in magnitude over subdecadal timescales.

For example, T_{global} increased by ~0.42°C between 2011 and 2016 which is equivalent to approximately three decades worth of long-term warming at the aforementioned historically-observed rate. This suggests that, on subdecadal timescales, deviations in T_{global} can approach 0.5°C in magnitude which has recently been associated with appreciable impacts on climate-related risk [*Guldberg et al.*, 2018]. For instance, the aggregate probability of exceeding the preindustrial-defined 99.9th percentile of daily heat extremes over land increases by 50% to 100% with such a magnitude shift in T_{global} [*Fischer and Knutti*, 2015]. In addition to being of large relative magnitude, subdecadal T_{global} variability also has an extensive spatial footprint in the sense that approximately 87% of the global surface and 99% of the global land surface exemplifies a positive linear relationship between local annual unforced Surface Air Temperature (SAT) deviations and unforced T_{global} deviations (Figure 1).

Given the extensive spatial footprint and large magnitude of subdecadal T_{global} variability, it is perhaps unsurprising that such variability has been linked to considerable environmental and societal impacts. Unforced subdecadal global temperature variability, typically associated with the state of the El Niño-Southern Oscillation [ENSO [*Trenberth et al.*, 2002]] and its related effects [*McPhaden et al.*, 2006], has been connected to global patterns of primary production [*Behrenfeld et al.*, 2001], the distribution of sea bird, marine mammal, and fish populations [*Stenseth et al.*, 2002], as well as coral bleaching [*Walther et al.*, 2002]. Such variability has also been linked to variation in societal phenomena including agricultural output [*David and Christopher*, 2007], gross domestic product growth [*Burke et al.*, 2015], monetary inflation [*Cashin et al.*, 2017], energy demand [*Deschênes and Greenstone*, 2011], human mortality [*Deschênes and Greenstone*, 2011], and civil conflicts [*Hsiang et al.*, 2011]. In addition to these substantive impacts, subdecadal T_{global} variability tends to attract significant attention in popular media [*Gillis*, 2017], which influences the public perception of the urgency/necessity of implementing climate-change mitigation policy.

Given the above concerns, there is potentially substantial utility in the approximate foreknowledge of annual T_{global} values (here, anomalies with respect to the 1951-1980 mean). However, forecasting the particular state of the climate system from months to years ahead of time is notoriously difficult because it is a timescale long enough such that chaos significantly degrades forecasts made based on initial conditions but a timescale too short for typical externally-forced signals to strongly emerge from the 'noise' of unforced variability [*Kirtman et al.*, 2013; *Meehl et al.*, 2009].

Despite these challenges, there has been an emphasis on annual to decadal climate prediction using both statistical [*Krueger and Storch*, 2011; *Newman*, 2013; *Suckling et al.*, 2017; *Sévellec and Drijfhout*, 2018; *Thomas et al.*, 2008] and dynamical models [*Keenlyside et al.*, 2008; *Kirtman et al.*, 2013; *Smith et al.*, 2007]. A particular focus has been placed on the use of dynamical Global Climate Models (GCMs) initialized to the observed state of the climate system and run forward in time, in a manner similar to the procedure used in numerical weather prediction [*Keenlyside et al.*, 2008; *Kirtman et al.*, 2013; *Smith et al.*, 2007]. This method has exhibited potential, especially in certain locations, but these GCM-based predictions come with noteworthy challenges such as large computational expense, incomplete observations for initialization, the necessity to correct for mean biases and the necessity to correct for model drift due to 'coupling shock' [*Meehl et al.*, 2013a]. Furthermore, accurate GCM simulation of internal modes such as ENSO and its teleconnections are of utmost importance for this application but many GCMs still struggle in their simulation of physical processes key to ENSO dynamics [*Bellenger et al.*, 2013].

In this study, we introduce a complement to GCM-based decadal prediction that makes use of historically-observable empirical relationships between the state of the climate system at any given time and subsequent annual T_{global} anomalies. Subdecadal T_{global} anomalies are primarily related to modes of variability in the climate system such as ENSO [*Brown et al.*, 2014; *Trenberth et al.*, 2002], and thus some previous efforts to statistically forecast subdecadal T_{global} have relied primarily on the use of ENSO indices as predictor variables (e.g., [*Smith et al.*, 2007; *Suckling et al.*, 2017]). However, these methods generally require an ad-hoc calculation of an ENSO index which may not fully capture the influence of ENSO on T_{global}. Furthermore, other modes of variability like the Interdecadal Pacific Oscillation (IPO; [*Meehl et al.*, 2013b]], the Atlantic Multidecadal Oscillation (AMO; [*Chylek et al.*, 2014]), the North Atlantic Oscillation [NAO; [*Li et al.*, 2013]), and variability over the Southern Ocean [*Brown et al.*, 2017], have all been suggested to influence global temperature, but they are often neglected as potential predictors of T_{global}. Part of the challenge is that many of the aforementioned modes are correlated with each other (either positively or negatively) which precludes their use in statistical frameworks that assume linear independence of predictors.

With the above considerations in mind, we seek an empirical methodology for predicting unforced T_{global} deviations that (a) allows predictors of T_{global} to be globally comprehensive and thus does not arbitrarily exclude modes of variability originating from, e.g., high latitudes; (b) creates predictors based on the relationship of the data to predictands (as opposed to using, e.g., *a priori* and/or *ad hoc* climate indices); and (c) creates predictors that are uncorrelated with each other. We achieve these goals via the use of Partial Least Squares Regression (PLSR [*Brown and Caldeira*, 2017; *Wold*, 1966]).

2. Methods

We apply Partial Least Squares Regression (PLSR) to the problem of forecasting T_{global} anomalies in the following way. First, observed gridded local SAT (latitude, longitude, annual time) anomalies were obtained from the four primary observational datasets: HadCRUT4 [*Morice et al.*, 2012], NASA GISTEMP [*Hansen et al.*, 2010], NOAA [*Vose et al.*, 2012], and Berkeley Earth Surface Temperature (BEST) [*Rohde et al.*, 2013].

Second, PLSR was performed between predictors of gridded local SAT fields (normalized locally by their standard deviation across time) and predictands of subsequent T_{global} deviations. SAT fields were used as predictor variables of T_{global} because they represent some of the longest and most spatially-comprehensive data available and because much information on the state of various modes of variability in the climate system are contained in the SAT field.

The PLSR framework is similar to that of a Multiple Linear Regression (MLR) problem. In MLR, application to the current problem would entail finding coefficients, $\vec{b}$, such that the mean squared residuals ($\vec{r}$) are minimized in the system,

$$\vec{y}_t = [X]_{t-1}\vec{b} + \vec{r}, \quad (1)$$

where $\vec{y}_t$ would be annual T_{global} anomalies as a function of time (e.g., from 1901-2019) and the matrix $[X]_{t-1}$ would contain the global spatial field of SAT which precedes the unforced T_{global} anomalies in time (e.g., where the rows would correspond to years 1900-2018, and the columns would correspond to each individual gridded location). Because of the high degree of spatial autocorrelation in the SAT predictor fields, the columns in $[X]_{t-1}$ will inevitably be highly collinear, and thus $[X]_{t-1}$ will be well-below full rank. This precludes the application of MLR to this problem. However, PLSR offers a solution to this issue by creating linear combinations of the columns in $[X]_{t-1}$ (PLSR components) that represent a large portion of $[X]_{t-1}$'s variability. The procedure is similar to Principle Component Analysis (PCA) but instead of seeking components that explain the maximum variability in $[X]_{t-1}$ itself, PLSR seeks components in $[X]_{t-1}$ that explain the variability in $\vec{y}_t$. Ultimately, PLSR is akin to the MLR procedure performed on a matrix $[Z]_{t-1}$ containing a relatively low number of PLSR components which represent much of the variability in $[X]_{t-1}$,

$$\vec{y}_t = [Z]_{t-1}\vec{\beta} + \vec{r}. \quad (2)$$

Below we show results using four PLSR components but conclusions are not sensitive to this specific number. We carry out PLSR using the MATLAB™ function 'plsregress' (https://www.mathworks.com/help/stats/plsregress.html). This function performs PLSR regression using the SIMPLS algorithm (see also methods in [*Brown and Caldeira*, 2017] for more details).

Equation (2) is for the specific case of predicting T_{global} anomalies from only the previous year's SAT field. However, in our application, we use the previous two years to predict the next four years of T_{global} deviations. Two lagged years were used because we found that there was some increase in skill by including information on not only the most recent state but also the evolution of predictors.

When using two preceding years, the matrices are horizontally concatenated prior to the application of PLSR. So in our application, equation (1) becomes,

$$\vec{y}_t = ([X]_{t-1} \,|[X]_{t-2})\vec{b} + \vec{r}. \quad (3)$$

The problem was then conducted separately for each lead-time. So, for lead-times of 2, 3 and 4 years, equation (3) becomes,

$$\vec{y}_{t+1} = ([X]_{t-1} \,|[X]_{t-2})\vec{b} + \vec{r}, \quad (4)$$

$$\vec{y}_{t+2} = ([X]_{t-1} \,|[X]_{t-2})\vec{b} + \vec{r}, \quad (5)$$

and

$$\vec{y}_{t+3} = ([X]_{t-1} \,|[X]_{t-2})\vec{\beta} + \vec{r}, \quad (6)$$

respectively.

2.1 LASSO Regularization

In order to prevent overfitting we implement the Least Absolute Shrinkage and Selection Operator LASSO regularization [*Tibshirani*, 1996] (https://www.mathworks.com/help/stats/lasso.html) using only the PLSR forecast and the most recent T_{global} anomaly as the two predictor variables. This serves as a check against overfitting in the sense that it damps the influence of predictors that cause poor predictions on our-of-sample data. The overall effect of this procedure is that under circumstances of little to no out-of-sample skill, the predictions will revert to the mean value of the predictand. Henceforth, we refer to our overall procedure as the BC2020 method.

2.2 Validation

Model validation was achieved via leave-one-out cross validation for years prior to 2000 (which we refer to as "hindcast mode") and through completely out-of-sample predictions made on the post 2000 data (which we refer to a "forecast mode").

Under leave-one-out cross-validation each T_{global} anomaly in the time series took a turn acting as a test year, with the remaining years designated as training years. Each test year was held out of the procedure such that the method was blind to the correct T_{global} anomaly for the test year. PLSR was then performed on the training years and the resulting regression coefficients were used to predict the T_{global} deviation for the test year. If the predictand lead time was one year, and two lagging years were used as predictors, equation (1) could be expanded as,

$$\begin{bmatrix} Tglobal_{t_3} \\ \cancel{Tglobal_{t_4}} \\ Tglobal_{t_5} \\ \cdots \\ Tglobal_{t_n} \end{bmatrix} = \begin{bmatrix} 1 & SAT_{t_2,loc_1} & \cdots & SAT_{t_2,loc_k} & \cdots & SAT_{t_1,loc_1} & \cdots & SAT_{t_1,loc_k} \\ \cancel{1} & \cancel{SAT_{t_3,loc_1}} & \cdots & \cancel{SAT_{t_3,loc_k}} & \cdots & \cancel{SAT_{t_2,loc_1}} & \cdots & \cancel{SAT_{t_2,loc_k}} \\ 1 & SAT_{t_4,loc_1} & \cdots & SAT_{t_4,loc_k} & \cdots & SAT_{t_3,loc_1} & \cdots & SAT_{t_3,loc_k} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & SAT_{t_{n-1},loc_1} & \cdots & SAT_{t_{n-1},loc_k} & \cdots & SAT_{t_{n-2},loc_k} & \cdots & SAT_{t_{n-2},loc_k} \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \cdots \\ b_{n \cdot k} \end{bmatrix} + \begin{bmatrix} r_{t_3} \\ r_{t_4} \\ r_{t_5} \\ \cdots \\ r_{t_n} \end{bmatrix}, \qquad (7)$$

where the subscript t refers to annual time and the subscript loc refers to the global gridded location. If the T_{global} anomaly corresponding to the 4th year in dataset [t_4, 2nd row in equation (7)] was designated as the test year, then the row corresponding to this year would be deleted [represented by being crossed-out in equation (7)] and $T_{global t4}$ would be hindcast with regression coefficients ($\vec{b}$) that were calculated without knowledge of the corresponding predictor-predictand combination. Note, that information form these deleted years (t_4, t_3 and t_2) still appear in the linear system elsewhere (e.g., SAT from t_2 still informs the prediction for $T_{global t3}$).

In hindcast mode, each year took a turn acting as a test year and the difference between the forecast T_{global} anomaly and the observed T_{global} anomaly (Figure 4 and Figure 5) was used to summarize the predictive skill of the method (Figure 3) and inform the confidence intervals of the forecast (Figure 4 and Figure 5).

The predictive skill of the BC2020 model is also quantified using forecast mode where its predictions are tested on completely out-of-sample data in the years following 2000. Forecast mode is conducted just as the model would have been run if it was operationalized staring in the year 2000. That is, no information from future data is used for any part of the training. Specifically, forecasts are made each year (at 1 to 4 year lead times) and model parameters are updated each year prior to the next year's forecast.

2.3 Note on treatment of forced variability.

We do not attempt to partition temperature variability into forced and unforced components. Rather we allow the PLSR procedure itself to partition most of the forced variability into the 1st PLSR component (Figure 6a, 6b and 6c). We choose to do this because, despite much research on the issue [*Frankcombe et al.*, 2015; *Mann et al.*, 2014], the isolation of unforced from total temperature variability remains a major challenge. When trying to do the decomposition, insufficient temporal/spatial variability of historical forcings may cause an insufficient amount of historical variability to be categorized as forced variability, and thus too much variability may be categorized as unforced variability. On the other hand, some studies have suggested that the best estimates of decadally-varying forcing may have been implicitly over-fit to observations to some degree [*Tung and Zhou*, 2013]. Under this view, the observed multidecadal temperature variability would contain a substantial unforced component and attempts to remove to partition variability into a forced and unforced component may be biased in favor of allocating too much variability into the forced designation and necessarily leaving too little variability in the unforced designation. Lacking a clear consensus on how to best decompose forced and unforced variability we choose to allow the PLSR procedure itself to partition most of the forced variability into the 1st PLSR component (Figure 6a, 6b and 6c) which we refer to as the global warming component.

This methodological decision has the added benefit of allowing the procedure to be completely independent of assumptions about the time-evolution of external forcings and/or their representation in dynamical GCM simulations. This is preferable if the method is to serve as a benchmark or point of comparison for dynamical GCMs.

Additionally, a major benefit of not attempting to remove forcing as a preprocessing step is that it eliminates the risk of inadvertently feeding in the model information that it would not have in an operationalized setting. When the model is run in forecast mode, the lack of data preprocessing grantees that there is no information leakage that would cause spuriously low prediction errors. Thus, forecast mode prediction errors incorporate uncertainty in both forced and unforced variability which is desirable when informing confidence intervals going forward in a real-world setting. However, this also means that because the method cannot anticipate time-varying changes in external forcings, it is at an inherent disadvantage compared to GCMs that incorporate retrospective time-varying forcings and statistical models trained on only the unforced component of variability.

2.4 Ability of BC2020 Method to forecast idealized signals

As a demonstration of concept, we employ the BC2020 method on idealized synthetic data. Specifically, we generated synthetic data over the period 1880-2017 which consisted of combinations of sine waves (Figure 2a and Figure 2b) and random noise (not shown). We inserted one oscillation in the Northern Hemisphere grid points and one in the Southern Hemisphere grid points, with each grid point receiving its own random noise time series. We then ran the BC2020 method on the gridded data. Cross-validated hindcasts as well as an out-of-sample forecast are shown in Figure 2c. It can be seen that the methodology is able to learn the relationships between SAT patterns and subsequent T_{global} anomalies. This is particularly apparent in the forecast period (2013-2017 in Figure 2c) where the method predicts an uptick in T_{global} (based on previous patterns) even though T_{global} had been trending down since ~2004.

2.5 CMIP5 decadal hindcast experiments

An alternative method for predicting T_{global} deviations comes from the observationally-initialized GCMs that participated in the CMIP5 decadal hindcast experiments [*Taylor et al.*, 2011]. These GCMs were initialized to various aspects of the observed state of the climate system (Table 1) and run forward in time, incorporating retrospective estimates of external forcings over the given forecast period (e.g., hindcasts starting in 1990 incorporated forcing in 1991 associated with the Mt. Pinatubo volcanic eruption even though that forcing was not predictable in advance). Therefore, the decadal hindcast experiments incorporate retrospective information on external forcing and thus their hindcast performance is at an advantage relative to the BC2020 method. There are 18 GCMs that participated in this experiment. Several GCMs use multiple initialization methods. These GCMs have some ensemble members which are initialized to the absolute observed anomalies of variables (full field initialization), while some ensemble members are initialized with anomalies from observed climatology (anomaly initialization). We treat GCMs' ensemble sets that use different initialization methods as being effectively different GCMs. This treatment has the effect of expanding the number of GCMs in this experiment from 18 to 24 (Table 1).

When GCMs are initialized to observations they have a tendency to drift towards their own preferred climate state which is often biased relative to observations. In order to correct for both the bias and the drift we apply the standard method of drift correction recommended by the International CLIVAR Project office. This method is described below and illustrated for a single GCM (bcc-csm1-1) in Fig S1.

The raw T_{global} hindcasts are represented as $Y_{j\tau}$ where j is the initial forecast time (j = 1, …n) and τ is the forecast lead time in years (τ = 1,…m; Figure S1a). The initial forecast time depends on the model (Table 1). The corresponding observations for which the hindcasts are compared against are represented as $X_{j\tau}$ (Figure S1c). For the main results, we show RMSEs with respect to GISTEMP but RMSEs with respect to the other three observational datasets are shown in Figure S4. The average anomalies over the entire series of forecasts are given by,

$$\bar{Y}_\tau = \frac{1}{n}\sum_{j=1}^{n} Y_{j\tau},$$

$$\bar{X}_\tau = \frac{1}{n}\sum_{j=1}^{n} X_{j\tau},$$

(Figure S1d and S1e respectively).

Drift is calculated on a model-by-model basis as the difference between the ensemble mean forecasts and observations over all cases,

$$d_\tau = \bar{Y}_\tau - \bar{X}_\tau,$$

(Figure S1f).

The drift (which implicitly contains any mean bias) is then subtracted from the raw hindcasts to obtain bias/drift corrected hindcasts,

$$\hat{Y}_{j\tau} = Y_{j\tau} - d_\tau = \bar{X}_\tau + \left(Y_{j\tau} - \bar{Y}_\tau\right) = \bar{X}_\tau + Y'_{j\tau},$$

(Figure S1g and S2).

where $Y'_{j\tau} = Y_{j\tau} - \bar{Y}_\tau$ is the anomaly of the raw forecast relative to the forecast average over all forecast periods. We perform this procedure in a hold-one-out cross-validated manner such that the anomalies over any given hindcast time period of evaluation are not included in the bias/drift calculation for that time period.

Other studies have suggested yet more involved post-processing of the decadal hindcast experiments in which the time-dependent aspect of the drift is taken into account. Such post-processing requires free-running historical experiments which are not available for all of the CMIP5 GCMs considered here and thus we use time-independent drift correction.

2.6 Comparison to other statistical methods.

Several other studies have published statistical procedures capable of predicting T_{global} on subdecadal timescales [*Krueger and Storch*, 2011; *Newman*, 2013; *Suckling et al.*, 2017; *Sévellec and Drijfhout*, 2018; *Thomas et al.*, 2008]. None of these procedures are directly comparable to the method outlined in this work because different choices are made regarding the treatment of the forced component of variability, the target T_{global} dataset, the timespan of the training data, the timespan of the test data, and the rigor of the cross-validation (Table S1). Nevertheless, we provide comparisons of the BC2020 method to the most analogous results from these previous studies (green lines and magenta dots in Figure S4). However, a rigorous comparison of methods would require a standardization of method protocols, training datasets, evaluation datasets, etc. and is beyond the scope of this study.

3. Results

3.1 Hindcast skill comparison

Figure 3 shows the root-mean-square error (RMSE) of T_{global} hindcasts relative to observations for the BC2020 method (blue), a persistence benchmark (black), and GCM decadal hindcast experiments (red) using the NASA GISTEMP dataset as observations (other three datasets shown in Figure S4). The RMSEs of the BC2020 method for both cross-validated 1900-2000 predictions (hindcast mode) and completely out-of-sample predictions post 2000 (forecast mode) are shown. The persistence benchmark uses the average of the previous five years to forecast the year in question (using five years minimized the error from persistence).

Averaged across the prediction lead times of 1 to 4 years, the BC2020 method produced lower RMSEs than the persistence benchmark and lower RMSEs than the mean RMSE of the initialized GCMs. Perhaps surprisingly, the mean GCM had a larger RMSE than the persistence benchmark which highlights the challenges associated with dynamical climate prediction at this timescale and suggests that the average GCM has difficulty simulating downstream teleconnections between various modes of variability and subsequent T_{global} deviations.

3.2 Hindcasts and forecasts

Figure 4 and Figure 5 shows the BC2020 method's hindcast and forecast predictions of T_{global} at lead-times of 1 to 4 years applied to the NASA GISTEMP dataset (the other 3 datasets are shown in Figure S3).

Scientific discussion of decadal climate variability and a potential short-term hiatus in global warming began emerging around the latter portion of the 2000s decade [*Easterling and Wehner*, 2009]. In 2004, the BC2020 method would have predicted a slight cooling through 2008 (the 4-year lead-time forecast valid in 2008, magenta dot, was close but below the actual value), and thus might have hinted at the emergence of what would come to be known as the hiatus [*Medhaug et al.*, 2017]. By 2012, there was much scientific discussion of the global warming hiatus [*Kaufmann et al.*, 2011; *Meehl et al.*, 2011; *Solomon et al.*, 2011; *Solomon et al.*, 2010] and its potential to persist. In that year, the BC2020 method predicted the T_{global} was primed to experience an uptick over the subsequent 4 years, and it predicted a new historical record in 2016 (magenta dot for 2016). This surge in T_{global} did come to pass, and 2016 did set the historical record, although it surpassed the magnitude of the BC2020 forecasted anomaly. The BC2020 method has produced some forecasts that were off by a large margin as well. For example, the 1-year lead time errors for 2006 and 2014 were both about -0.2°C. Note, however, that these errors help inform the confidence intervals for both the hindcasts and forecasts over the period 2020-2023.

The spatial patterns of and time evolution of SAT variability, that are the most relevant to the prediction of subsequent T_{global} anomalies, are illustrated in Figure 6. Figure 6a, 6d and 6g show the PLSR scores (analogous to the principle component time series in Principle Component Analysis, PCA) and the loadings (analogous to the Empirical Orthogonal Functions (EOFs)) of the first three PLSR components which explain 85%, 5% and 3% of the variance in subsequent T_{global} variability at the 1-year lead-time. Positive PLSR loadings displayed on the maps denote where local SAT anomalies are associated with subsequent warm unforced T_{global} anomalies and negative PLSR loadings denote where local cool SAT anomalies are associated with subsequent warm T_{global} anomalies.

The first PLSR component (Figure 6a-6c) largely corresponds to the externally forced global warming signal and is the dominant explainer of variability since 1900. The externally forced nature of the component is apparent from the spatially coherent nature of the pattern (Figure 6b and 6c).

The second PLSR component (Figure 6d-6f) shows that warm T_{global} anomalies (apart from the global warming signal) are preceded by cool local SAT anomalies in the tropical Pacific two years prior (Figure 6e). However, one year prior to a warm T_{global} anomaly (again relative to the long-term global warming signal), there tends to be a warm anomaly over the equatorial Pacific (Figure 6f). This pattern indicates that unforced warm T_{global} deviations are associated with a transition from La Niña-like conditions two years prior, to El Niño-like conditions one year prior to the year in consideration.

The 3rd PLSR component shows that warm T_{global} anomalies are associated with a positive IPO. The sharpening of the IPO pattern from two years to one year prior to the prediction year indicates that positive T_{global} anomalies are associated with an antecedent strengthening of an already positive IPO state. These findings are consistent with independent methodologies that have highlighted the Pacific's prominence in the modulation of unforced T_{global} variability [*Brown et al.*, 2014; *England et al.*, 2014].

Overall, the first three PLSR components agree with the broader literature that the long-term T_{global} evolution can be predicted by the state of global warming, and subsequent refinements of the T_{global} anomaly for any given year can be made by incorporating information on the state of ENSO and of the IPO.

3.3 True forecast.

Training on data from 1900-2019 and using SAT deviations from 2018-2019 as predictor fields, the BC2020 method forecasts T_{global} anomalies, with 2σ uncertainty ranges of +0.98°C (± 0.2), +0.98°C (±0.22), +0.91°C (±0.24) and +0.95°C (±0.24) above the 1951-1980 mean for the NASA GISTEMP dataset (Figure 4 and 5). The BC2020 method's forecast for 2020 and 2021 suggests values nearly equal to that of 2019 before slightly lower values in 2022 and 2023 (Figure 4 and Figure 5). Thus the BC2020 method does not necessarily foresee the 2016 global temperature record being broken in the forecast period.

4. Conclusion

The empirical method laid out in this study shows skill in hindcasting global mean surface temperature (T_{global}) anomalies and even preforms better than most dynamical Global Climate Models (GCMs) at this task. This indicates that a large fraction of the information necessary to constrain short-term T_{global} evolution is contained in the antecedent global surface air temperature field (the only predictor variable used here). Given that short-term T_{global} variability is of substantial magnitude relative to the long-term trend and it has an extensive global spatial footprint with broad ecological and societal impacts, this tool represents a computationally inexpensive means of anticipating and possibly mitigating some short-term climate effects.

Nevertheless, the physical mechanisms that can be surmised from statistical relationships are necessarily limited. Insight on the mechanistic underpinnings of the BC2020 method's skill can be informed from PLSR loading patterns (Figure 6) but the most compressive physical understanding of interannual T_{global} variability will ultimately involve the use GCMs. Furthermore, GCMs provide geographic-specific predictions for many variables which are necessary for developing process understanding and for anticipating many impacts. Thus, the method presented here should not be considered a replacement for GCM decadal prediction but rather it should be viewed as a complement and/or a benchmark to which GCM predictions can be compared.

Acknowledgments.

This study was supported by the Fund for Innovative Climate and Energy Research and the Carnegie Institution for Science endowment. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for the Coupled Modelled Intercomparison Project (CMIP), and we thank the climate modelling groups for producing and making available their model output. For CMIP the US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

Author contributions: PTB conceived of the study, performed the initial analysis and wrote a first draft of the manuscript. PTB and KC both contributed to interpretation of the results and refinement of the manuscript.

Competing interests: The authors declare no competing interests.

Data and materials availability: The MATLAB code used for prediction can be accessed at www.github.com/____. The CMIP5 data used for this study can be accessed at http://pcmdi9.llnl.gov/. The forcing data can be accessed at https://data.giss.nasa.gov/modelforce/Miller_et_2014/Fi_Miller_et_al14_upd.txt. The GISTEMP observations can be accessed at https://data.giss.nasa.gov/gistemp/. The HadCRUT4 observations can be accessed at https://crudata.uea.ac.uk/cru/data/temperature/. The NOAA observations can be accessed at https://www.ncdc.noaa.gov/data-access/marineocean-data/noaa-global-surface-temperature-noaaglobaltemp. The Berkeley Earth observations can be accessed at http://berkeleyearth.org/data/. Other data and material requests are available from the corresponding author.

Figures

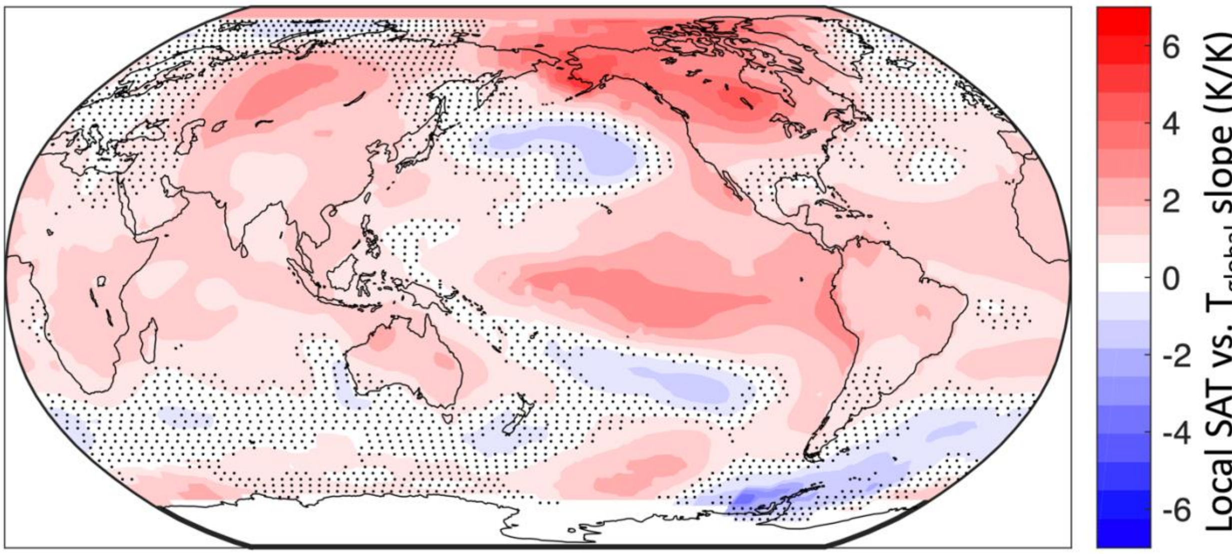

Figure 1. Spatial footprint of annual unforced global mean surface air temperature variability showing that the majority of the surface displays a positive relationship with the global mean. Colors represent the magnitude of the linear regression coefficient (slope) between local unforced (subdecadal timescale) annual surface air temperature deviations and global mean unforced (subdecadal timescale) annual surface air temperature deviations. Stippling indicates that the linear regression coefficient is not statistically different from zero at the 90% confidence level or above. Data is from the GISTEMP dataset and spans 1950-2019. This timespan was selected due to it being the longest time period with near global spatial coverage. Unforced variability was isolated from forced variability so that subdecadal variations could be highlighted rather than the long-term trend. This decomposition between forced and unforced variability was achieved via multiple linear regression against historical radiative forcings.

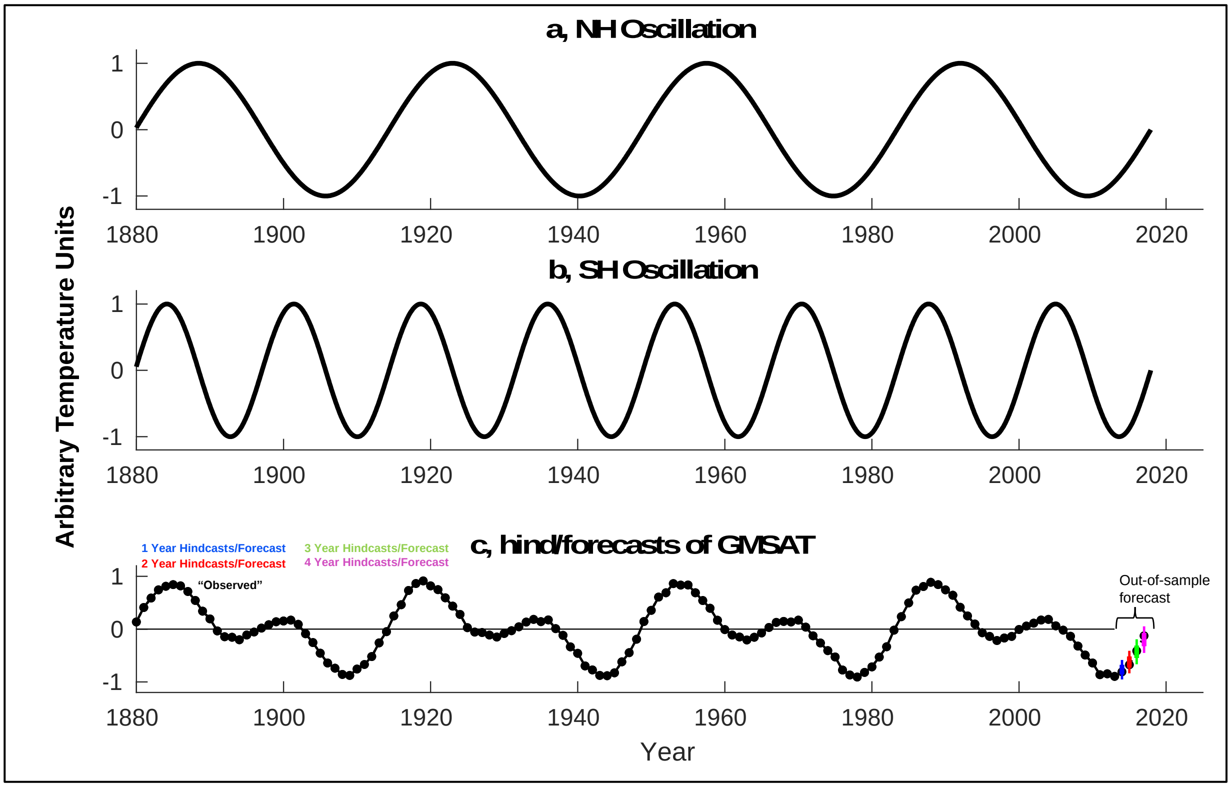

Figure 2. BC2020 method applied to idealized synthetic data. (a-b) Sine curves placed in the Northern Hemisphere grid points (a) and Southern Hemisphere grid points (b) in addition to random noise (not shown). c) Analogous to Figure 5 but applied to this idealized case.

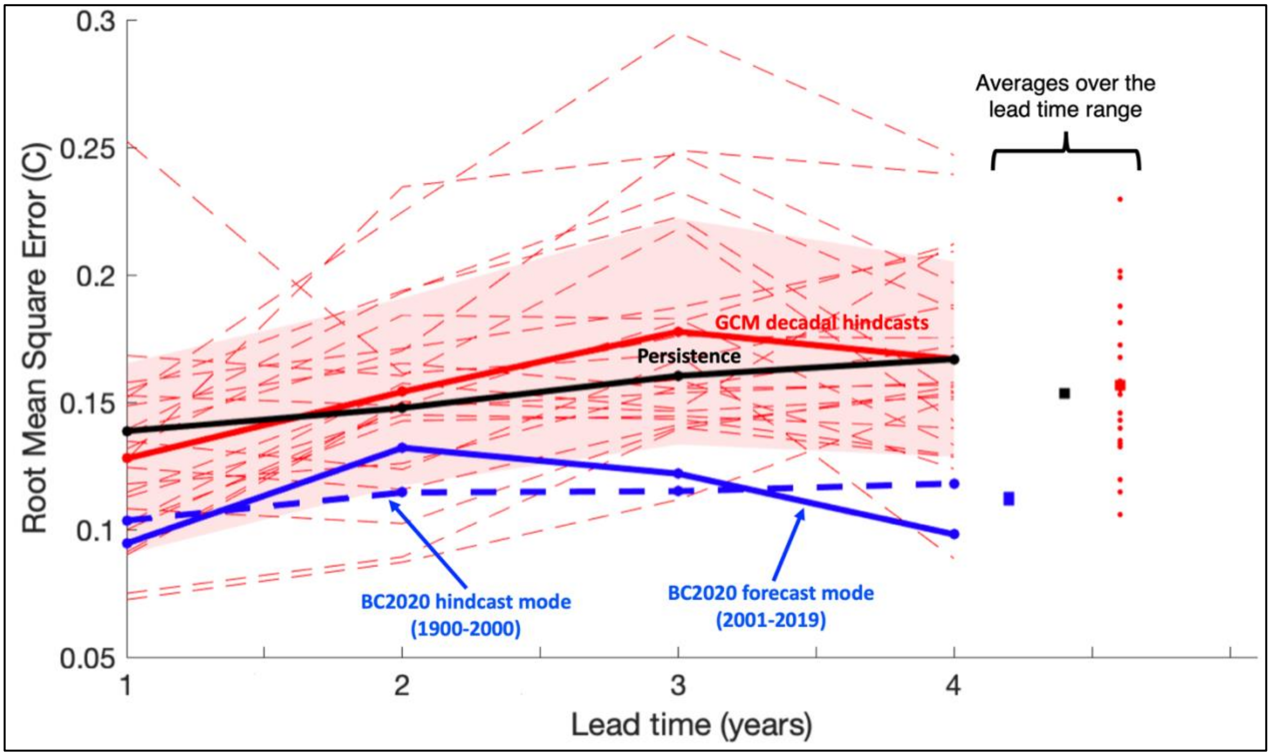

Figure 3. Prediction errors for various methods of anticipating global mean surface air temperature anomalies. The root mean square error (RMSE) is shown between predictions of global mean surface air temperature anomalies and observed global mean surface air temperature anomalies as a function of prediction lead-time. The RMSEs for the GCM decadal hindcasts are calculated over time periods that vary by model as described in Table S1. Note that the GCM decadal hindcast experiments incorporate retrospective information on external forcing and thus their hindcast performance is at an advantage relative to the BC2020 method.

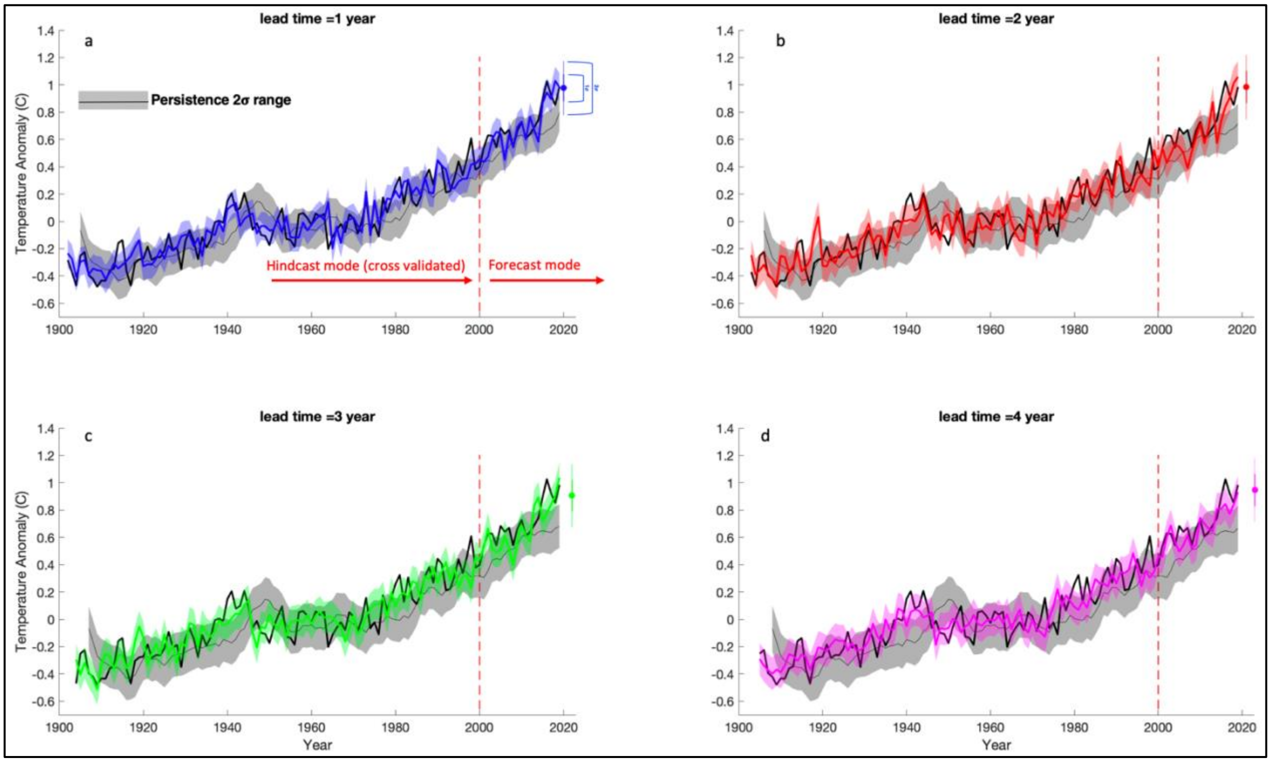

Figure 4. Out-of-sample predictions for each lead-time as well as the forecasts for the years 2020-2023. Predictions for years through 2000 utilize leave-one-out cross validation and predictions for the years following 2000 are made precisely as they would have been if the method was operationalized starting in the year 2000 (model parameters are tuned only on past data and progressively updated each year to make the forecasts). Forecasts for 2020-2023 show ±1σ (thick lines) and ±2σ (thin lines) confidence intervals which are derived from the RMSE of the forecast mode errors. The grey shading is the 2σ naïve persistence forecast which projects the next year's global temperature anomaly as being the average of the previous 5 years' global temperature anomalies (averaging over 5 years minimized the persistence error). The GISTEMP dataset is used for observations here but results are similar for the other three global temperature datasets considered (Figure S3).

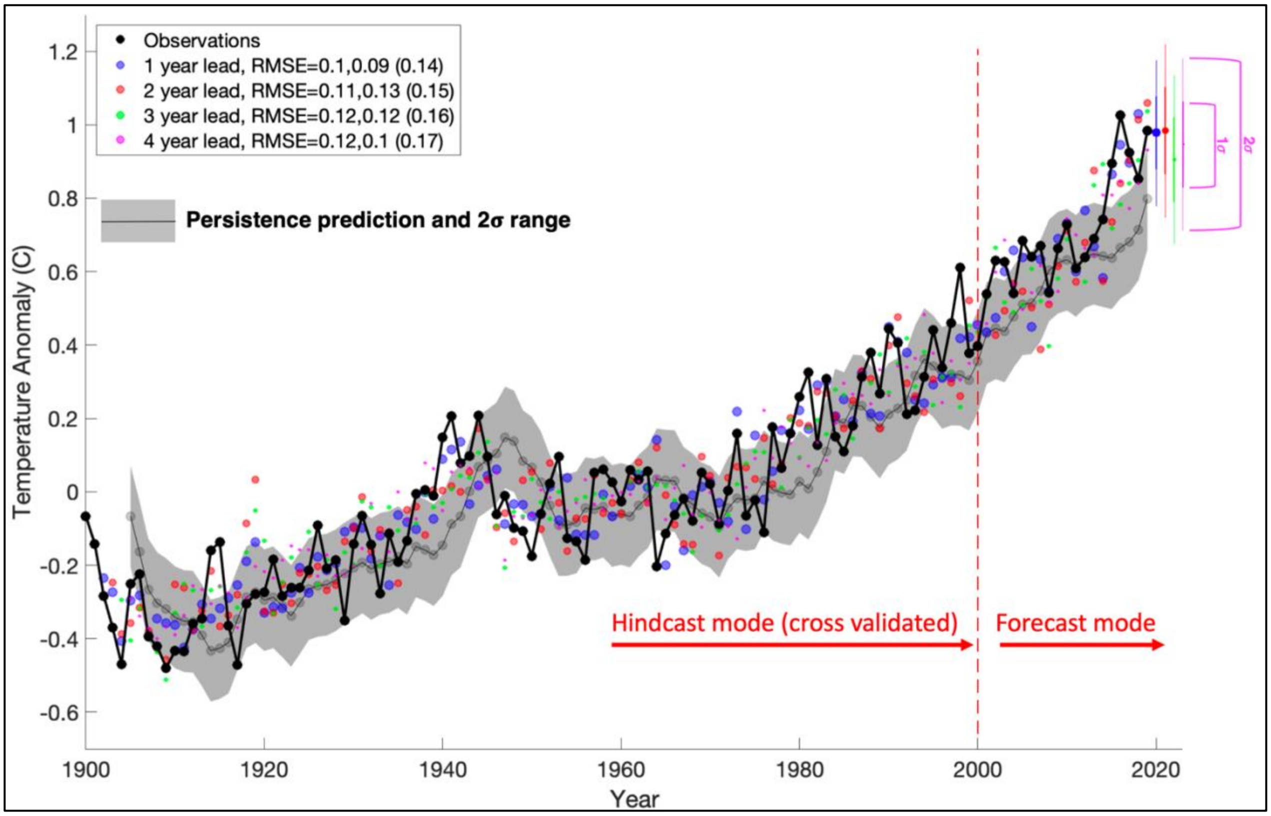

Figure 5. Same information as Figure 4 but displayed on a single figure. The legend displays the prediction Root Mean Square Errors (RMSEs) for hindcast mode, forecast mode and for the persistence forecast respectively for each lead time.

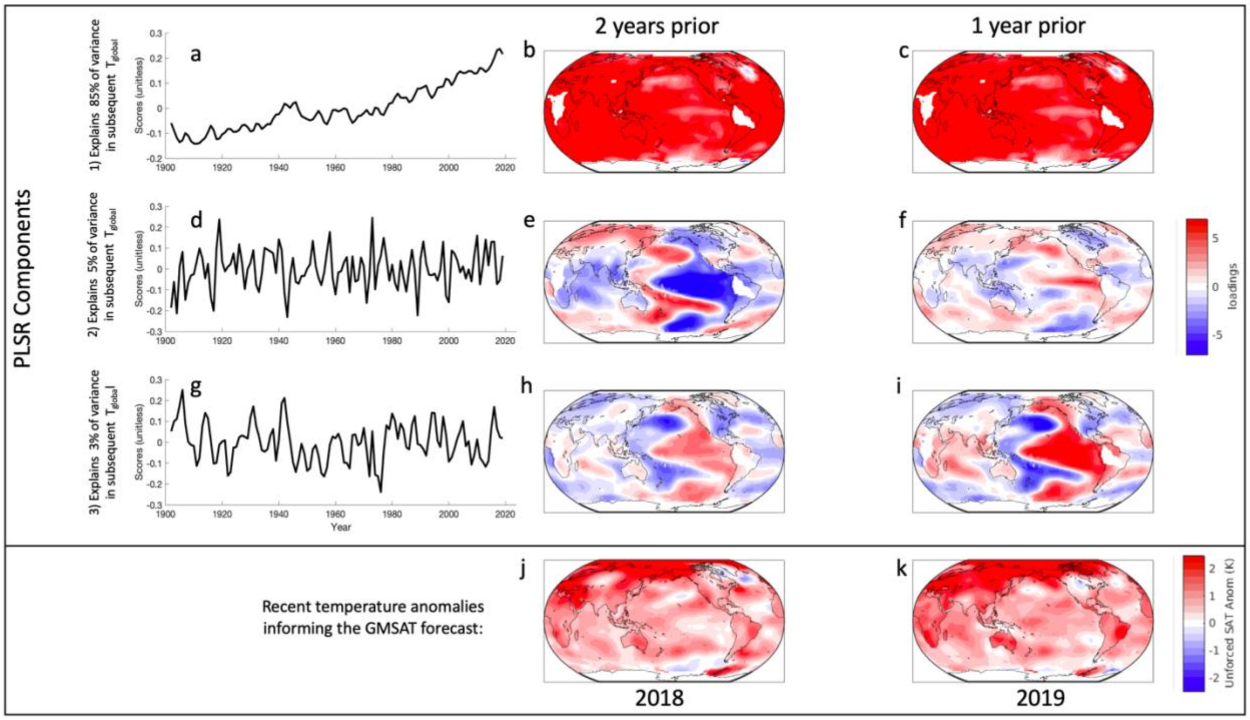

Figure 6. (a,d,g) Partial Least Squares Regression scores (analogous to Principle Component time series) illustrating the temporal variation of the modes of variability shown in the maps of the corresponding row. (b, c, e, f, h, i) Partial Least Squares Regression loading maps (analogous to Empirical Orthogonal Function maps) for the first three components associated with a subsequent year's global temperature anomaly (i.e., 1-year lead time). Positive loadings indicate that a warm unforced local surface air temperature anomaly at that location is associated with a warm subsequent global temperature anomaly and negative loadings indicate that a cool unforced local surface air temperature anomaly at that location is associated with a warm subsequent global temperature anomaly. (j,k) Unforced local surface air temperature anomalies for two years (2018-2019) informing the BC2020 method's forecast for 2020-2023.

Table 1. Information of the CMIP5 GCMs that participated in the decadal hindcast experiments. RMSEs in Figure 3 are calculated over all available ensemble members and over all start years for each effective GCM.

Effective Model Number	Model Name	Number of ensemble members	Initialization Method	Start Years
1	bcc-csm1-1	4	i1	1961-2007
2	CanCM4_i2	10	i1	1962-1980, 1982-2005, 2007-2012
3	CCSM4	10	i1	1976, 1986, 1991, 1996, 2001-2007
4	CFSv2-2011	4	i1	1981, 1982, 1984, 1986, 1991, 1994, 1996, 1997, 1999, 2001, 2004, 2006, 2007, 2010, 2011
5	CMCC-CM	1	i1	1961, 1966, 1971, 1976, 1981, 1986, 1991, 1996, 2001, 2006
6	CNRM-CM5	10	i1	1960, 1961, 1965, 1966, 1970, 1971, 1975, 1976, 1980, 1981, 1985, 1986, 1990, 1991, 1995, 1996, 2000, 2001, 2005, 2006
7	EC-EARTH	10	i1	1961, 1966, 1971, 1976, 1981, 1986, 1991, 1996, 2001, 2006
8	FGOALS-g2	3	i1	1961, 1966, 1971, 1976, 1981, 1986, 1991, 1996, 2001, 2006
9	FGOALS-s2	3	i1	1966, 1971, 1976, 1981, 1986, 1996, 2001
10	GEOS-5	3	i1	1961-1981, 1986-2010
11	GFDL-CM2p1	10	i1	1962-2013
12	HadCM3	10	i2	1961-2010
13	IPSL-CM5A-LR	6	i1	1961, 1966, 1971, 1976, 1981, 1986, 1991, 1996, 2001, 2006
14	MIROC4h	6	i1	1966, 1971, 1976, 1986, 1991, 1996, 2001
15	MIROC5	6	i1	1960, 1962-1980, 1982-2005, 2007-2011
16	MPI-ESM-LR	10	i1	1961-2011
17	MPI-ESM-MR	3	i1	1961, 1966, 1971, 1976, 1981, 1986, 1991, 1996, 2001-2011
18	MRI-CGCM3	9	i1	1961, 1966, 1971, 1976, 1981, 1986, 1991, 1996, 2001, 2006, 2011, 2012
19	CanCM4_i2	10	i2	1966, 1971, 1976, 1986, 1991, 1996, 2001-2005, 2007-2009
20	CFSv2-2011_i2	3	i2	1961, 1966, 1971, 1976, 1981, 1986, 1991
21	CMCC-CM_i2	1	i2	1961, 1966, 1971, 1976, 1981, 1986, 1991, 1996, 2001, 2006
22	CMCC-CM_i3	1	i3	1961, 1966, 1971, 1976, 1981, 1986, 1991, 1996, 2001, 2006
23	EC-EARTH_i3	10	i3	1961-2006
24	HadCM3_i3	10	i3	1961-2010

Supplementary Tables and Figures

Table S1. Other notable studies concerned with statistical prediction of global mean surface air temperature and how their methods and evaluation procedures differ from BC2020. Comparison of hindcast errors between these studies and BC2020 are shown in Figure S4 (all are green in Figure S4 except for Sevellec and Drijfhout (2018) which is represented with a magenta square corresponding to their reported RMSE over lead times of 1-5 years). A rigorous comparison of methods with BC2020 would require a standardization of method protocols, training datasets, evaluation datasets, etc. and is beyond the scope of this study.

Reference	Method name	Reference Figure	Target T_{global} dataset	Hindcast evaluation time period	Other major distinctions with BC2020
Laepple et al., 2008	IENS	Fig. 3a, black solid line	GISTEMP	1930-2006	Forced T_{global} variability not removed - CMIP3 20C3M experiment used to represent forced T_{global}; no implicit foreknowledge of volcanic forcing
Krueger and Von Storch, 2011	Prediction model (10)	Fig. 2b, black solid line	HadCRUT3	1930-2006	Represents forced T_{global} variability with atmospheric CO_2 concentration; 10-year block cross-validation
Newman, 2013	LIM	Fig. 3f, blue line	HadCRUT3	1901-2009	Forced T_{global} variability not removed; 10-year block cross-validation
Suckling et al., 2017	Real-time	Fig. 7b, red line	Cowtan & Way	1960-2014	Forced T_{global} variability not removed - predictors include external forcing time series; no implicit foreknowledge of forcing included; several model parameters (GHG forcing lag, NINO3.4 lag, predictor screening) are set outside of and prior to any cross-validation; RMSE calculated with respect to an ensemble mean
Suckling et al., 2017	Prescribed natural forcing	Fig. 7b, green line	Cowtan & Way	1960-2014	Forced T_{global} variability not removed - predictors include external forcing time series; foreknowledge of volcanic eruption forcing included; several model parameters (GHG forcing lag, NINO3.4 lag, predictor screening) are set outside of and prior to any cross-validation; RMSE calculated with respect to an ensemble mean
Suckling et al., 2017	Exploiting the trend	Fig. 7b, blue line	Cowtan & Way	1960-2014	Forced T_{global} variability not removed - predictors include external forcing time series; several model parameters (GHG forcing lag, NINO3.4 lag, predictor screening) are set outside of and prior to any cross-validation; RMSE calculated with respect to an ensemble mean
Sevellec and Drijfhout, 2018	PROCAST	N/A	GISTEMP	1880-2017	Forced T_{global} variability removed with Multiple Linear Regression but different forcings used

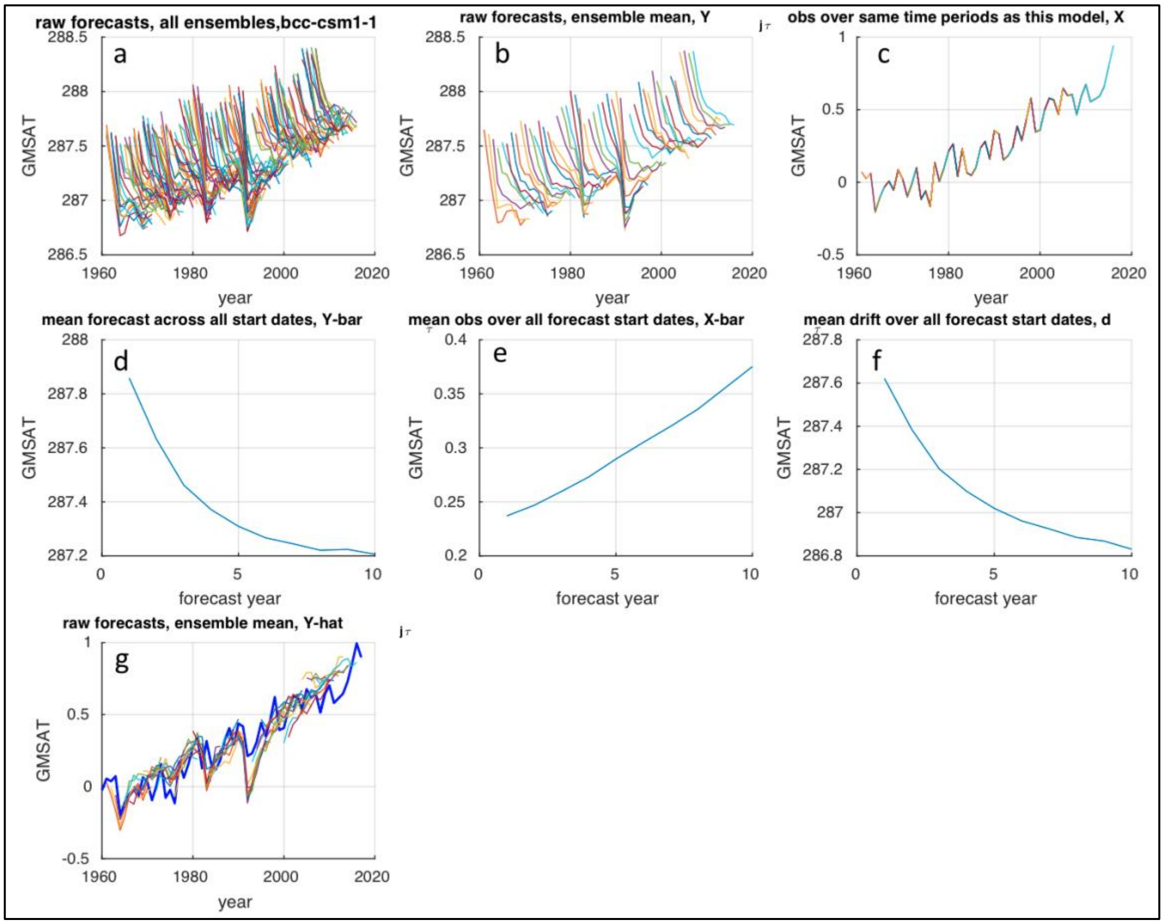

Figure S1. Illustration of bias and drift correction performed on the CMIP5 decadal hindcast experiments. See text for details.

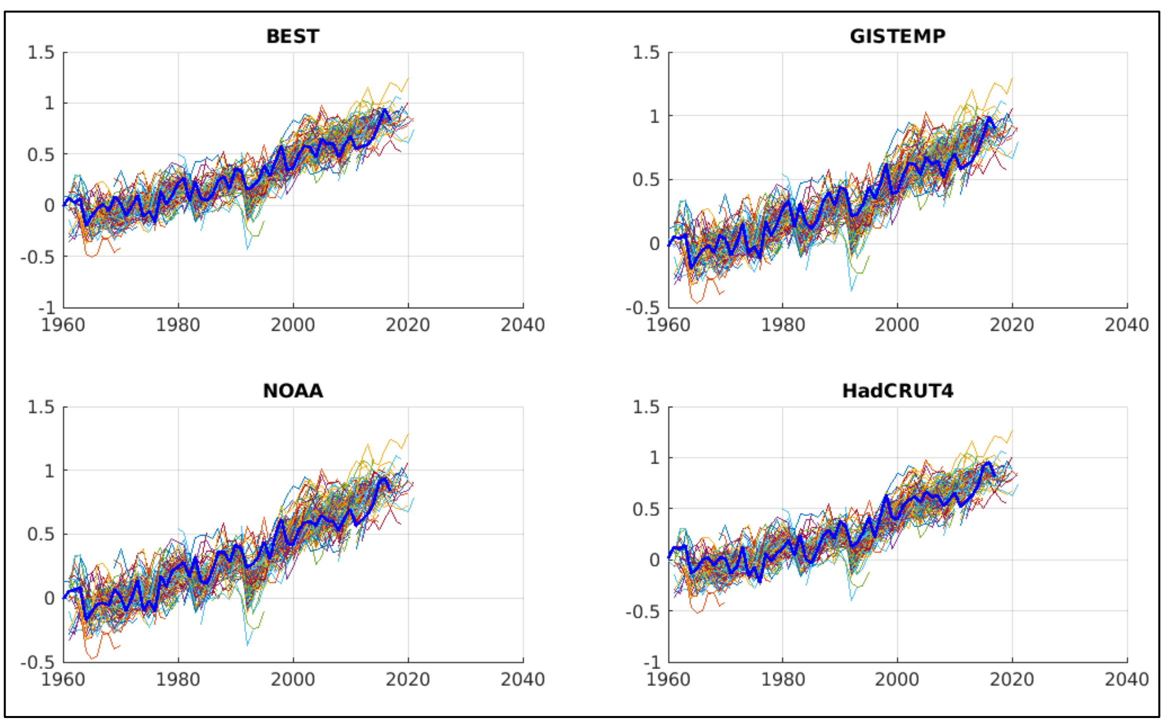

Figure S2. All CMIP5 decadal hindcast experiment results after bias and drift correction.

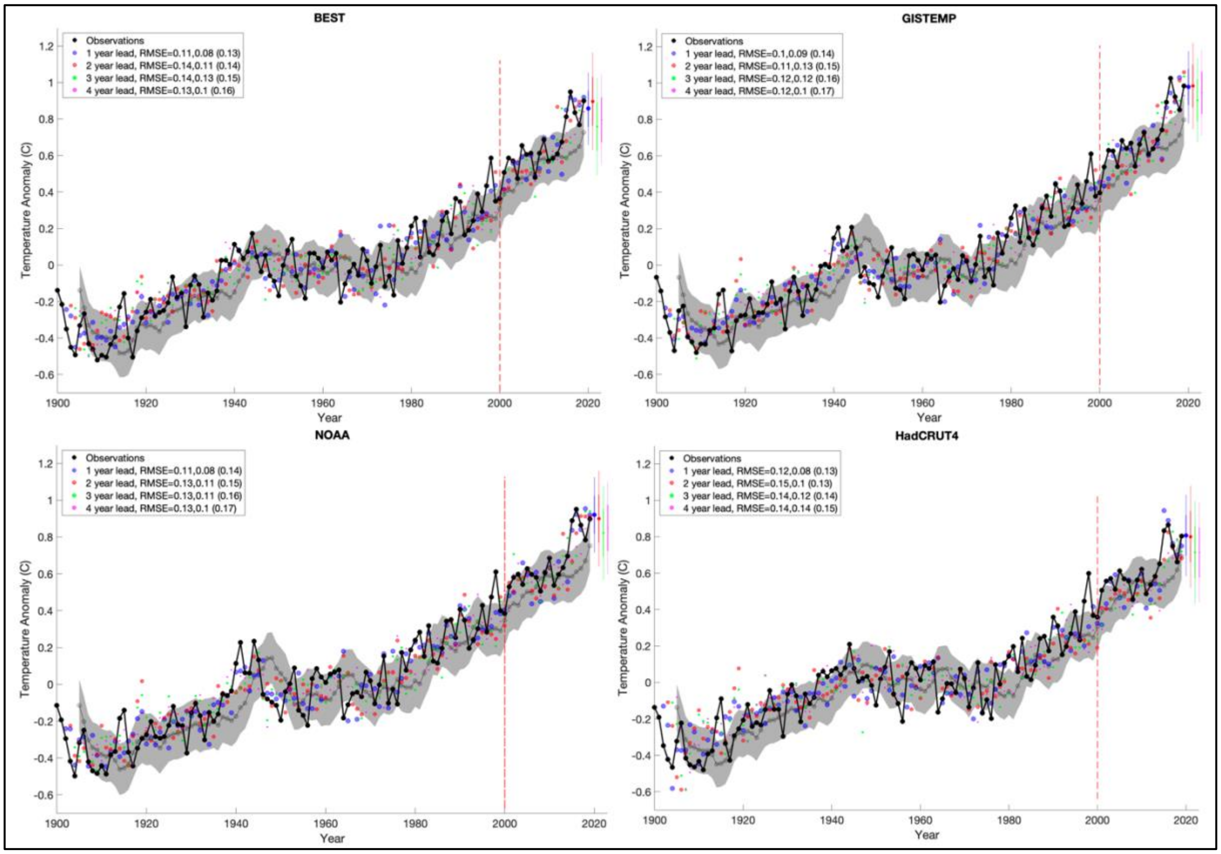

Figure S3. Same as Figure 5 but showing all four temperature datasets.

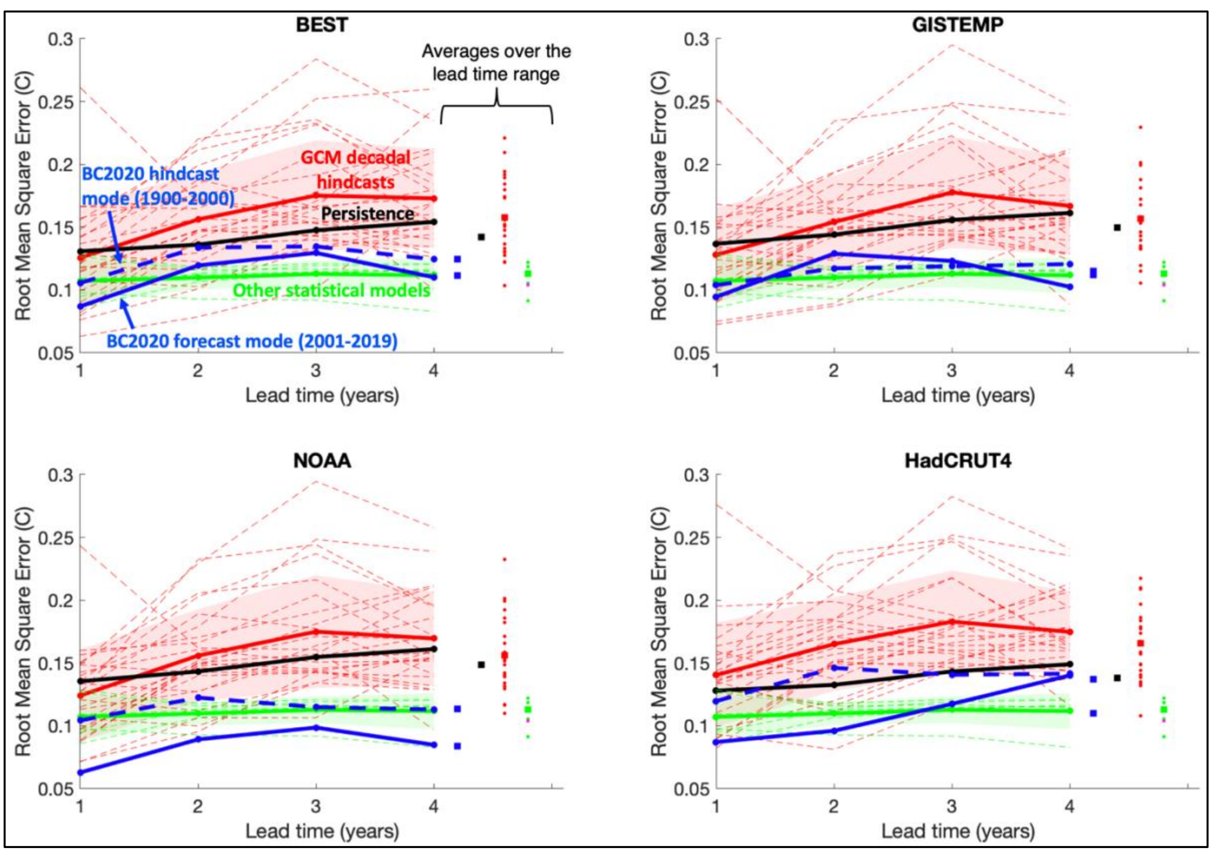

Figure S4. Same as Figure 3 but showing all four temperature datasets.

References

Behrenfeld, M. J., et al. (2001), Biospheric Primary Production During an ENSO Transition, *Science*, *291*(5513), 2594-2597, doi:10.1126/science.1055071.

Bellenger, H., E. Guilyardi, J. Leloup, M. Lengaigne, and J. Vialard (2013), ENSO representation in climate models: from CMIP3 to CMIP5, *Climate Dynamics*, 1-20, doi:10.1007/s00382-013-1783-z.

Bindoff, N. L., P.A. Stott, K.M. AchutaRao, M.R. Allen, N. Gillett, D. Gutzler, K. Hansingo, G. Hegerl, Y. Hu, S. Jain, I.I. Mokhov, J. Overland, J. Perlwitz, R. Sebbari and X. Zhang In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*Rep.*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Brown, P. T., and K. Caldeira (2017), Greater future global warming inferred from Earth's recent energy budget, *Nature*, *552*, 45, doi:10.1038/nature24672.

Brown, P. T., W. Li, and S.-P. Xie (2014), Regions of significant influence on unforced global mean surface air temperature variability in climate models, *Journal of Geophysical Research: Atmospheres*, 2014JD022576, doi:10.1002/2014JD022576.

Brown, P. T., Y. Ming, W. Li, and S. A. Hill (2017), Change in the magnitude and mechanisms of global temperature variability with warming, *Nature Climate Change*, *7*, 743, doi:10.1038/nclimate338.

Burke, M., S. M. Hsiang, and E. Miguel (2015), Global non-linear effect of temperature on economic production, *Nature*, *527*(7577), 235-239, doi:10.1038/nature15725.

Cashin, P., K. Mohaddes, and M. Raissi (2017), Fair weather or foul? The macroeconomic effects of El Niño, *Journal of International Economics*, *106*, 37-54, doi:https://doi.org/10.1016/j.jinteco.2017.01.010.

Chylek, P., J. D. Klett, G. Lesins, M. K. Dubey, and N. Hengartner (2014), The Atlantic Multidecadal Oscillation as a dominant factor of oceanic influence on climate, *Geophysical Research Letters*, 2014GL059274, doi:10.1002/2014GL059274.

David, B. L., and B. F. Christopher (2007), Global scale climate–crop yield relationships and the impacts of recent warming, *Environmental Research Letters*, *2*(1), 014002. doi:10.1088/1748-9326/2/1/014002

Deschênes, O., and M. Greenstone (2011), Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US, *American Economic Journal: Applied Economics*, *3*(4), 152-185, doi: 10.1257/app.3.4.152.

Easterling, D. R., and M. F. Wehner (2009), Is the climate warming or cooling?, *Geophysical Research Letters*, *36*(8), L08706, doi:10.1029/2009GL037810.

England, M. H., S. McGregor, P. Spence, G. A. Meehl, A. Timmermann, W. Cai, A. S. Gupta, M. J. McPhaden, A. Purich, and A. Santoso (2014), Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus, *Nature Clim.*, doi:10.1038/nclimate2106.

Fischer, E. M., and R. Knutti (2015), Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes, *Nature Climate Change*, *5*, 560, doi:10.1038/nclimate2617.

Frankcombe, L. M., M. H. England, M. E. Mann, and B. A. Steinman (2015), Separating Internal Variability from the Externally Forced Climate Response, *Journal of Climate*, *28*(20), 8184-8202, doi:10.1175/jcli-d-15-0069.1.

Gillis, J. (2017), Earth Sets a Temperature Record for the Third Straight Year., in *New York Times*.

Guldberg, O. H., D. Jacob., M. Taylor, et. al. (2018), Chapter 3: Impacts of 1.5ºC global warming on natural and human systems. In IPCC Special Report on 1.5ºC global warming.

Hansen, J., R. Ruedy, M. Sato, and K. Lo (2010), GLOBAL SURFACE TEMPERATURE CHANGE, *Reviews of Geophysics*, *48*(4), RG4004, doi:10.1029/2010RG000345.

Hsiang, S. M., K. C. Meng, and M. A. Cane (2011), Civil conflicts are associated with the global climate, *Nature*, *476*, 438, doi:10.1038/nature10311.

Kaufmann, R. K., H. Kauppi, M. L. Mann, and J. H. Stock (2011), Reconciling anthropogenic climate change with observed temperature 1998–2008, *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.1102467108.

Keenlyside, N. S., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner (2008), Advancing decadal-scale climate prediction in the North Atlantic sector, *Nature*, *453*(7191), 84-88, doi: 10.1038/nature06921

Kirtman, B., S.B. Power, J.A. Adedoyin, G.J. Boer, R. Bojariu, I. Camilloni, F.J. Doblas-Reyes, A.M. Fiore, M. Kimoto, G.A. Meehl, M. Prather, A. Sarr, C. Schär, R. Sutton, G.J. van Oldenborgh, G. Vecchi and H.J. Wang (2013), Near-term Climate Change: Projections and Predictability. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*Rep.*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Krueger, O., and J.-S. V. Storch (2011), A Simple Empirical Model for Decadal Climate Prediction, *Journal of Climate*, *24*(4), 1276-1283, doi:10.1175/2010jcli3726.1.

Li, J., C. Sun, and F.-F. Jin (2013), NAO implicated as a predictor of Northern Hemisphere mean temperature multidecadal variability, *Geophysical Research Letters*, *40*(20), doi:10.1002/2013GL057877.

Mann, M. E., B. A. Steinman, and S. K. Miller (2014), On Forced Temperature Changes, Internal Variability and the AMO, *Geophysical Research Letters*, 2014GL059233, doi:10.1002/2014GL059233.

McPhaden, M. J., S. E. Zebiak, and M. H. Glantz (2006), ENSO as an Integrating Concept in Earth Science, *Science*, *314*(5806), 1740-1745, doi:10.1126/science.1132588.

Medhaug, I., M. B. Stolpe, E. M. Fischer, and R. Knutti (2017), Reconciling controversies about the ‘global warming hiatus’, *Nature*, *545*, 41, doi:10.1038/nature22315.

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, *Nature Clim. Change*, *1*(7), 360-364, doi: 10.1038/nclimate1229

Meehl, G. A., et al. (2013a), Decadal Climate Prediction: An Update from the Trenches, *Bulletin of the American Meteorological Society*, *95*(2), 243-267, doi:10.1175/BAMS-D-12-00241.1.

Meehl, G. A., et al. (2009), Decadal Prediction, *Bulletin of the American Meteorological Society*, *90*(10), 1467-1485, doi:10.1175/2009BAMS2778.1.

Meehl, G. A., A. Hu, J. M. Arblaster, J. Fasullo, and K. E. Trenberth (2013b), Externally Forced and Internally Generated Decadal Climate Variability Associated with the Interdecadal Pacific Oscillation, *Journal of Climate*, *26*(18), 7298-7310, doi:10.1175/JCLI-D-12-00548.1.

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, *Journal of Geophysical Research: Atmospheres*, *117*(D8), doi:10.1029/2011JD017187.

Newman, M. (2013), An Empirical Benchmark for Decadal Forecasts of Global Surface Temperature Anomalies, *Journal of Climate*, *26*(14), 5260-5269, doi:10.1175/jcli-d-12-00590.1.

Rohde R, M. R., Jacobsen R, Muller E, Perlmutter S, et al. (2013), A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011, *Geoinfor Geostat: An Overview 1:1*, doi:10.4172/2327-4581.1000101.

Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy (2007), Improved Surface Temperature Prediction for the Coming Decade from a Global Climate Model, *Science*, *317*(5839), 796-799, doi:10.1126/science.1139540.

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*(6044), 866-870, doi:10.1126/science.1206027.

Solomon, S., K. H. Rosenlof, R. W. Portmann, J. S. Daniel, S. M. Davis, T. J. Sanford, and G.-K. Plattner (2010), Contributions of Stratospheric Water Vapor to Decadal Changes in the Rate of Global Warming, *Science*, *327*(5970), 1219-1223, doi:10.1126/science.1182488.

Stenseth, N. C., A. Mysterud, G. Ottersen, J. W. Hurrell, K.-S. Chan, and M. Lima (2002), Ecological Effects of Climate Fluctuations, *Science*, *297*(5585), 1292-1296, doi:10.1126/science.1071281.

Suckling, E. B., G. J. van Oldenborgh, J. M. Eden, and E. Hawkins (2017), An empirical model for probabilistic decadal prediction: global attribution and regional hindcasts, *Climate Dynamics*, *48*(9), 3115-3138, doi:10.1007/s00382-016-3255-8.

Sévellec, F., and S. S. Drijfhout (2018), A novel probabilistic forecast system predicting anomalously warm 2018-2022 reinforcing the long-term global warming trend, *Nature Communications*, *9*(1), 3024, doi:10.1038/s41467-018-05442-8.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2011), An Overview of CMIP5 and the Experiment Design, *Bulletin of the American Meteorological Society*, *93*(4), 485-498, doi:10.1175/BAMS-D-11-00094.1.

Thomas, L., J. Stephen, and C. Katie (2008), Interannual temperature predictions using the CMIP3 multi-model ensemble mean, *Geophysical Research Letters*, *35*(10), doi:10.1029/2008GL033576.

Tibshirani, R. (1996), Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267-288.

Trenberth, K. E., J. M. Caron, D. P. Stepaniak, and S. Worley (2002), Evolution of El Niño–Southern Oscillation and global atmospheric surface temperatures, *Journal of Geophysical Research: Atmospheres*, *107*(D8), AAC 5-1-AAC 5-17, doi:10.1029/2000JD000298.

Tung, K.-K., and J. Zhou (2013), Using data to attribute episodes of warming and cooling in instrumental records, *Proceedings of the National Academy of Sciences*, *110*(6), 2058-2063, doi:10.1073/pnas.1212471110.

Vose, R. S., et al. (2012), NOAA's Merged Land–Ocean Surface Temperature Analysis, *Bulletin of the American Meteorological Society*, *93*(11), 1677-1685, doi:10.1175/BAMS-D-11-00241.1.

Walther, G.-R., E. Post, P. Convey, A. Menzel, C. Parmesan, T. J. C. Beebee, J.-M. Fromentin, O. Hoegh-Guldberg, and F. Bairlein (2002), Ecological responses to recent climate change, *Nature*, *416*, 389, doi:10.1038/416389a.

Wold, H. (1966), Estimation of Principal Components and Related Models by Iterative Least squares, in *Multivariate Analysis.*, edited, pp. 391-420, Academic Press.