Quantifying the role of internal climate variability in future climate trends

David W. J. Thompson[1, 2], Elizabeth A. Barnes[2], Clara Deser[3],

William E. Foust[2], Adam S. Phillips[3]

Submitted to Journal of Climate

December 2014

Revised

March 2015

[1] Corresponding author. Email: davet@atmos.colostate.edu.
[2] Department of Atmospheric Science, Colorado State University, Fort Collins, CO. USA
[3] National Center for Atmospheric Research, Boulder CO. USA

Abstract

Internal variability in the climate system gives rise to large uncertainty in projections of future climate. The uncertainty in future climate due to internal climate variability can be estimated from large ensembles of climate change simulations in which the experiment set-up is the same from one ensemble member to the next but for small perturbations in the initial atmospheric state. However, large ensembles are invariably computationally expensive and susceptible to model bias.

Here we outline an alternative approach for assessing the role of internal variability in future climate based on a simple analytic model and the statistics of the unforced climate variability. The analytic model is derived from the standard error of the regression and assumes that the statistics of the internal variability are roughly Gaussian and stationary in time. When applied to the statistics of an unforced control simulation, the analytic model provides a remarkably robust estimate of the uncertainty in future climate indicated by a large ensemble of climate change simulations.

It is argued that the uncertainty in future climate trends due to internal variability can be robustly estimated from the statistics of the observed variability.

Introduction

The signature of anthropogenic forcing in climate change has and will be superposed on internal climate variability due to a variety of physical processes (e.g., Hawkins and Sutton 2009, 2011; Deser et al. 2012a, 2012b; Wallace et al. 2013; Kirtman and Power et al. 2013; Collins and Knutti et al. 2013; Knutsen et al. 2013; Bindoff and Stott et al. 2013). At most terrestrial locations, a large component of the internal variability in surface climate change arises from variations in the atmospheric circulation (Wallace et al. 1995, 2012, 2013; Deser et al. 2014). On regional spatial scales, the internal variability can overwhelm the signature of anthropogenic forcing not only on year-to-year timescales, but on multidecadal timescales as well (Hawkins and Sutton 2009; Deser et al. 2012a, 2012b; IPCC 2014). The signal of significant warming is thus expected to "emerge" earliest in regions with relatively small natural variability, such as the tropics (Christensen et al. 2007; Mahlstein et al. 2011; Diffenbaugh and Scherer 2012; Hawkins and Sutton 2012). Understanding and predicting the contribution of internal variability to long-term trends in climate is essential for both the adaptation to and mitigation of climate change (IPCC 2014).

What is the most robust way to estimate the role of internal variability in future climate trends? One approach is to generate a large ensemble of climate change simulations in which the individual ensemble members are from the same climate model and subject to the same external forcing, but are initiated with slightly different atmospheric initial conditions. For example, the National Center for Atmospheric Research (NCAR) CCSM3 Large Ensemble Project includes 40 climate change simulations run with the same coupled atmosphere-ocean-sea ice-land model (the NCAR Community Climate System Model 3; CCSM3) and forced with identical

projected changes in greenhouse gases and ozone from 2000-2061 (the SRES A1B Scenario). Since the model and forcing are the same in all ensemble members, the differences in climate trends from one ensemble member to the next derive entirely from the unforced (i.e., internal) variability in the model.

Analyses of the spread in the trends in the NCAR 40-member ensemble make clear the pronounced role of internal climate variability in projections of regional climate change (Deser et al. 2012a, 2012b). For example, the left panels in Figure 1 show the standard deviations of the 50-year (2011-2061) trends in October-March mean near-surface air temperature and precipitation calculated over all ensemble members (i.e., the results indicate the spread in the trends from one ensemble member to the next). The right panels indicate time series of October-March mean surface air temperature and precipitation for all 40 ensemble members at two sample locations. As noted in Deser et al. (2012a), internal variability in the CCSM3 gives rise to temperature trend standard deviations that exceed 1 K/50 years over much of the Northern Hemisphere and precipitation trend standard deviations that exceed 0.5 mm/day/50 years over much of the tropics. Since the spreads in the trends indicated in Fig. 1 arise entirely from stochastic variability in the CCSM3, they may be viewed as the irreducible component of uncertainty in climate change projections in this particular model.

The purpose of this study is to develop a simple analytic model for estimating the uncertainty in projections of future climate trends due to internal climate variability, as exemplified in Fig. 1. The model is derived from the standard error of the regression and is based on two statistics of the unforced climate variability: the standard deviation and autocorrelation. The analytic model is developed in Section 2. It is tested against the NCAR 40-member ensemble in Section 3 and applied to observations in Section 4.

Section 5 includes a discussion of the results. Concluding remarks are given in Section 6.

2. A simple analytic model of the role of internal variability in future climate trends

Consider a time series *x(t)* with mean zero and a linear least-squares trend *b*. The confidence interval (*CI*) on the trend in *x(t)* is expressed as:

$$CI = b \pm e$$

where *e* is the margin of error for the trend. The trend, its confidence interval and its margin of error are all expressed in units $\Delta x / (n_t \Delta t)$, where n_t is the number of time steps and Δt is the time step. For example, if *x(t)* corresponds to 50 years of wintertime mean temperature data, then $n_t = 50$, $\Delta t = 1$ year, and the temperature trend in *x(t)* is expressed in units degrees Celsius/50 years.

If the distribution of the deviations in *x(t)* about its linear trend (i.e., the residuals of the regression) is Gaussian, then the margin of error for the trend in *x(t)* is:

1) $e = t_c s_b$

where the critical *t*-statistic (t_c) is a function of the degrees of freedom and desired confidence interval, and

2) $$s_b = n_t \frac{s_e}{\sqrt{\sum_{i=1}^{n_t} (i - \bar{i})^2}}$$

is the standard error of the trend. In Eq. 2, i denotes time, s_e is the standard error of *x(t)* about its linear trend, and the factor n_t is included so that the standard error is given in units $\Delta x / (n_t \Delta t)$. Equations 1 and 2 are widely used to assess the significance of a trend in climate science (Wilks 1995; von Storch and Zwiers 1999; Santer et al. 2000).

The standard deviation of the time axis (the denominator in Eq. 2) can be expressed as a function of n_t as follows:

3) $$g(n_t) \equiv \frac{1}{\sqrt{\sum_{i=1}^{n_t} (i - \bar{i})^2}} = \sqrt{\frac{12}{n_t^3 - n_t}}$$

where we have used two formulae for consecutive integers to derive the algebraic expression $g(n_t)$ (Appendix). Note that the units on $g(n_t)$ are $1/\Delta t$.

Regarding the standard error of *x(t)* about its linear trend (s_e in Eq. 2): If the residuals of the regression (i.e., the detrended values of the *x(t)* time series) are serially correlated, then s_e must include a scaling factor that accounts for the bias in the sample standard deviation introduced by persistence in the time series (Mitchell et al. 1966; Wilks 1995; von Storch and Zwiers 1999; Santer et al. 2000). As shown in the Appendix, if *x(t)* is well-modeled as Gaussian red noise, then the standard error of *x(t)* about its linear trend can be approximated as:

4) $$s_e = \sigma\gamma(n_t, r_1)$$

where

5) $$\sigma \equiv \sqrt{\frac{1}{n_t - 2}\sum_{i=1}^{n_t}\left[x(i) - bi\right]^2}$$

is the standard deviation of the residuals,

6) $$\gamma(n_t, r_1) \equiv \left(\frac{\left[n_t - 2\right]}{\left[n_t \left(\frac{1 - r_1}{1 + r_1} \right) - 2 \right]} \right)^{1/2}$$

is the scaling factor, and r_1 is the lag-one autocorrelation of the residuals. If the residuals are *not* serially correlated (e.g., the lag-one autocorrelation of the detrended *x(t)* time series is zero), then $s_e = \sigma$. Note that the denominator of Eq. 5 includes the factor $n_t - 2$ (rather than $n_t - 1$, as is the case for estimating the sample standard deviation). The factor $n_t - 2$ arises since estimating the standard deviation of the residuals requires estimating not only the mean of the time series (the y-intercept), but also the regression line (the slope). In the context of climate change, σ can be viewed as the standard deviation of the internal (unforced) variability.

Substituting Eqs. 2, 3, and 4 into Eq. 1 yields the following expression for the margin of error for a trend in *x(t)* in units $\Delta x / \left(n_t \Delta t \right)$:

7) $$e = t_c \cdot n_t \cdot \sigma \cdot \gamma(n_t, r_1) \cdot g(n_t)$$

Equation 7 provides a simple analytic model for the margin of error for a trend in a Gaussian red noise process. It is derived from commonly used estimates of trend significance (e.g., see also Lettenmaier 1976; Santer et al. 2000; and Cassola et al. 2009). And it makes clear that the margin of error on a trend is a function of two statistics of the internal variability, both of which we assume are stationary in time:

1) the standard deviation of the internal (unforced) variability, σ ; and

2) the lag-one autocorrelation of the internal (unforced) variability, r_1 .

Figure 2 shows solutions for Eq. 7 as a function of n_t (trend length) and r_1 (the lag-one autocorrelation of the residuals), where t_c is calculated for the two-tailed 95% confidence level (i.e., 2.5% of the distribution lies in each tail). The results are expressed in units of the margin of error relative to the amplitude of the internal variability, i.e., they show solutions to $\frac{e}{\sigma} = t_c \cdot n_t \cdot \gamma(n_t, r_1) \cdot g(n_t)$. As such, the results indicate the amplitude of the trend required to exceed the 95% confidence level in units of the internal variability. For example, if $\frac{e}{\sigma} = 2$, then the trend must be twice as large as the internal (unforced) variability to exceed its margin of error.

The margins of error on the trends (and thus the trend amplitudes required for significance) increase rapidly as the length of the trend decreases and/or the

autocorrelation increases. If a time series is 40 time steps in length and has autocorrelation $r_1 \approx 0.45$, then the trend in the time series must be twice as large as the standard deviation of the internal variability to exceed the 95% confidence level. If the autocorrelation increases to $r_1 \approx 0.65$, then the trend must be ~three times as large as the internal variability.

As also evidenced in Figure 2, if the autocorrelation of the residuals $r_1 \approx 0$ and the trend length is 50 time steps, then the margin of error on the trend is roughly equal to the standard deviation of the residuals. That is, inserting $r_1 \approx 0$ and $n_t = 50$ into Eq. 7 yields:

8) $e_{95\%} \approx \sigma$ (for $n_t = 50$ and $r_1 \approx 0$)

where $e_{95\%}$ denotes the two-tailed 95% margin of error on the trend in units $\Delta x / (n_t \Delta t)$. Equation 8 provides a rough "rule of thumb" for the uncertainty in 50 year trends in any Gaussian physical process that is not serially correlated from one year to the next: the 95% margin of error on the 50 year trends in a Gaussian process is roughly equal to its interannual standard deviation.

3. Testing the analytic model in a large ensemble of climate change simulations

How well does the analytic model given by Eq. 7 predict the uncertainty in future climate trends? The utility of the analytic model is tested by comparing: 1) the margins of error in trends calculated from a large ensemble of climate change simulations run on

a coupled global climate model (the *actual* margins of error); with 2) the margins of error predicted by applying the analytic model to the statistics of the internal variability of the same coupled global climate model (the *predicted* margins of error). As discussed below, the internal variability of the coupled global climate model is estimated from a long control simulation with fixed anthropogenic forcing.

The *actual* margins of error are derived from 50-year trends in boreal wintertime (October-March) mean near-surface air temperature and precipitation from the NCAR 40-member ensemble of climate change simulations. The NCAR 40-member ensemble is described in detail in Deser et al. 2012a. Briefly, the simulations were run with a fully coupled ocean/land/atmosphere global climate model on a 2.8 x 2.8 degree latitude/longitude grid (the NCAR Community Climate System Model Version 3; CCSM3) and forced with the Special Report on Emissions Scenarios (SRES) A1B scenario. The ensemble members differ only in their initial atmospheric conditions.

The *predicted* margins of error are derived from a 1000 year-long control simulation run on the NCAR CCSM3 in which greenhouse gases are held fixed at 1990 levels. In the analyses shown here, the climate change simulations are examined from 2011-2061 and the control simulation is examined for the last 500 years of the integration. Seasonal-mean surface air temperature and precipitation do not exhibit notable memory from one year to the next at virtually all terrestrial locations in the control run (not shown). So in practice, the predicted 95% margins of error for the 50-year trends derived from Eq. 7 are roughly the same as the interannual standard deviations in the control run (as per Eq. 8).

Figure 3a shows the ensemble-mean 50-year trends in surface air temperature from 2011-2061 averaged over all 40 members in the CCSM3 large ensemble. The

ensemble-mean trends have been discussed in previous work (Deser et al. 2012a) and are shown here to provide context for the amplitude of the internal variability. The warming during the first half of the 21st century is projected to be largest over the Northern Hemisphere, where it exceeds ~3 K/50 years over much of northern North America and Asia (Deser et al. 2012a; Kirtman and Power et al. 2013; Collins and Knutti et al. 2013).

Figure 3b shows the *actual* two-tailed 95% margins of error for the 50-year trends found by: 1) calculating the standard deviations of the trends derived from all 40 ensemble members and 2) multiplying the standard deviations by a factor of two (95% of the normal distribution lies within ~ two standard deviations of the population mean). Note that Fig. 3b is identical to Fig. 1a multiplied by a factor of two. The grey bars in the surrounding panels indicate the histograms of the simulated trends at grid boxes collocated with the indicated cities. The actual margins of error for the trends are due entirely to the internal variability in the NCAR CCSM3, i.e., they are not due to differences in the forcing or the model used in the simulations. As such, they provide a quantitative estimate of the role of internal variability in future climate trends (Deser et al. 2012a, 2012b). By construction, the means of the histograms are equal to the trends in Fig. 3a and the standard deviations of the histograms are equal to *0.5* times the actual margins of errors shown in Fig. 3b. At many terrestrial locations, the margins of error due to internal variability are ~50% as large as the forced signal (compare Figs. 3a and 3b).

Figure 3c shows the *predicted* 95% margins of error for the 50-year trends found by applying the analytic model given by Eq. 7 to the statistics of the control simulation. Stippling indicates regions where the predicted margins are not significantly different

from the actual margins (see Appendix for details). The blue probability density functions in the surrounding panels show the corresponding predicted Gaussian distributions of the trends at grid boxes collocated with the indicated cities, where 95% of the distributions lies between $\pm e_{95\%}$. Comparing Figs. 3b and 3c, it is clear that 1) the margins of error predicted by applying Eq. 7 to the statistics of the control run provide a remarkably accurate prediction for 2) the margin of error on the trends in surface air temperature derived from the large-ensemble of climate change simulations. Over much of the globe, the predicted margins of error are statistically indistinct from the actual margins.

Figure 4 shows analogous results for October-March mean precipitation. The ensemble-mean trends (Fig. 4a) are consistent with increases in precipitation in the deep tropics and high northern latitudes juxtaposed against decreases in precipitation in the subtropics (Kirtman and Power et al. 2013; Collins and Knutti et al. 2013; Held and Soden 2006). As is the case for surface air temperature, the predicted margins of error found by applying Eq. 7 to the statistics of the control run provide a remarkably accurate estimate of both the spatial pattern and amplitude of the actual margins of error throughout much of the globe (compare Figs. 4b and 4c). The predicted margins of error are statistically indistinct from the actual margins over most terrestrial locations. The apparent bimodality in the ensemble member trends at Melbourne (Fig. 4, bottom right) is not reproducible in results for adjacent grid boxes, and is thus likely an artifact of sampling variability.

We reproduced the above analyses for the boreal summer season months April-September (results not shown). To first order, the similarities between the predicted and

actual margins of error in April-September mean surface air temperature and precipitation are comparable to those indicated in Figs. 3 and 4.

4. Application to observations

The results in the previous section indicate that the role of internal variability in a large ensemble of climate change simulations can be quantified to a high degree of accuracy from the statistics of the variability in an unforced control simulation. The results highlight the importance of simulating correctly the internal variability in a control simulation: If the standard deviation and/or autocorrelation of the simulated internal variability are biased relative to the observations, then those biases will project directly onto the uncertainty in simulations of climate change. Since model simulations inevitably contain biases, the internal variability of the *real-world* is arguably best estimated from the *real-world* itself, i.e., from observations.

The analytic model is next applied to estimates of internal variability derived from two observational data sources: 1) precipitation data from the Global Precipitation Climatology Project (GPCP) Version 2.2 Combined Precipitation Data Set (Adler et al. 2003), and 2) surface air temperature data from the HadCRUT4 dataset (Kennedy et al. 2011; Osborn and Jones 2014). The precipitation data are analyzed on a 2.5x2.5 mesh and were obtained from the NOAA Physical Sciences Division; the surface air temperature data are analyzed on a 5x5 mesh and were obtained from the Climatic Research Unit at the University of East Anglia.

The observed internal climate variability is assumed to be closely approximated by the statistics of the detrended, seasonal-mean grid point values over the period 1979-2013. In principle: 1) the anthropogenic forcing of the past few decades is not perfectly

linear; and 2) the amplitude of the internal variability on decadal timescales may be underestimated in the relatively short 1979-2013 period (the GPCP precipitation data are only available after 1979). However, in practice: 1) the statistics of the grid point surface air temperature and precipitation observations are effectively identical whether the anthropogenic signal is modeled as a first order (linear trend) or second order polynomial fit; and 2) variations on decadal timescales account for a relatively small fraction of the total variance in surface air temperature and precipitation on regional scales (not shown).

Figures 5 and 6 compare the 95% margins of error for the 50-year October-March mean surface air temperature and precipitation trends derived from a) the observations (top panels; solid distributions in surrounding panels) and b) the CCSM3 control simulation (bottom panels; dashed distributions). As is the case for the control simulation output, observed October-March mean surface air temperature and precipitation do not exhibit statistically significant memory from one year to the next at virtually all terrestrial locations (not shown). Hence, in practice: 1) the 95% margins of error on the 50-year trends predicted by applying Eq. 7 to the observations are effectively equal to 2) the standard deviations of the (detrended) October-March mean observations, as per Eq. 8. Note that the results in Figs. 5b and 6b are identical to those shown in Figs. 3c and 4c, respectively, except that: 1) the stippling in Figs. 5b and 6b indicates regions where the modeled and observed interannual variances are significantly *different* from each other at the 95% confidence level (see Fig. 5 caption); and 2) the control simulation output used in Figs. 5 and 6 has been interpolated to the same mesh as the observations before calculating the interannual standard deviations

(i.e., the temporal variance of area-mean surface air temperature and precipitation generally decreases when averaged over successively larger spatial regions).

The margins of error predicted by the CCSM3 control simulation and observations exhibit similar spatial patterns but have significantly different amplitudes over large regions of the globe (stippling). For example, the control simulation exhibits significantly different margins of error in surface air temperature over much of western North America, southern Asia, and tropical South America and Africa (Fig. 5). It also exhibits significantly different margins of error in precipitation over much of North America, South America and eastern Asia (Fig. 6). The differences between the margins of error predicted by the observed and control interannual standard deviations are visually apparent at several of the indicated cities (probability distribution functions). Comparable differences are found during the April-September season in both surface air temperature and precipitation (results not shown).

5. Discussion

The analytic model developed in Section 2 is based on two primary assumptions. One, it assumes that the internal variability is roughly Gaussian and is not dominated by, say, bimodal or oscillatory behavior. The climate system exhibits various forms of quasi-periodic variability other than the seasonal cycle, e.g., the Madden-Julian Oscillation (Zhang 2005) and El-Nino/Southern Oscillation. But a substantial fraction of climate variability is well-modeled as a Gaussian process, particularly at extratropical locations (Hartmann and Lo 1998; Feldstein 2000; Newman et al. 2003) and on interannual timescales. The strong similarities between the actual and predicted margins of errors in Fig. 4 suggest that even precipitation is sufficiently Gaussian on

seasonal-mean timescales to justify the assumptions that underlie the analytic model.

Two, it assumes that the standard deviation and autocorrelation of the internal variability are stationary in time. There is evidence that the standard deviation of surface air temperature will change over select locations in response to climate change, with decreases in temperature variance over the high latitudes of the Northern Hemisphere during winter (e.g., Gregory and Mitchell 1995; Screen 2014; Schneider et al. 2015) and increases over various terrestrial regions in summer (Fischer and Schär 2009). The most noticeable changes in interannual variance in the NCAR CCSM3 large-ensemble of climate change simulations are found over eastern Europe/western Russia, where the standard deviations of surface air temperature decrease during the forcing period (Appendix Fig. A1). But even in this region, the differences between the predicted and actual margins of error are not statistically significant (Fig. 3c). As demonstrated in Figs. 3-4, the analytic model provides a remarkably robust estimate of the uncertainty in simulated climate change due to internal variability over the vast majority of the globe.

In fact, the analytic model given by Eq. 7 may be viewed as a "null hypothesis" for the role of internal variability in future climate on *any* timescale in *any* physical field that is well-modeled as a Gaussian process. The results in Figures 7a and 7b show trends in temperature and precipitation as a function of trend length at the model grid point collocated with Los Angeles (results for the same grid point are highlighted in Figs. 1, 3 and 4). Red dots indicate trends derived from all 40 ensemble members for the period starting 2011 and ending on the year indicated on the abscissa. For example, the red dots at 2061 indicate trends derived from all 40 ensemble members for the period 2011-2061 (in units 50 years^{-1}), and are identical to the results used to generate the grey histograms in the top left panels of Figs. 3 and 4. The dashed lines indicate the ranges of

the trends given by $bn_t \pm e$, where b is the ensemble mean trend and e is found by applying Eq. 7 to the statistics of the control run. Figure 7c indicates analogous results for October-March mean precipitation averaged over grid points that lie within the Colorado River watershed. Figure 7d indicates results for the model North Atlantic Oscillation (i.e., northern annular mode), calculated here as the standardized difference between October-March mean sea-level pressure anomalies collocated with Iceland and the Azores.

Figure 7 highlights two main points: 1) It makes clear that the analytic model given by Eq. 7 provides a remarkably accurate estimate of the uncertainty in climate change projections on an array of timescales, not just the 50 year timescale exemplified in Figs. 3 and 4; and 2) It highlights the utility of the analytic model for estimating the uncertainty in surface climate trends averaged a relatively large spatial region, and in the large-scale atmospheric circulation. The uncertainties in projected changes in precipitation averaged over the Colorado River watershed (Fig. 5c) and in the North Atlantic Oscillation (Fig. 5d) indicated by the large-ensemble are closely approximated by applying Eq. 7 to the statistics of the control simulation.

6. Concluding remarks

Uncertainty in projections of future climate change can arise from three different factors (e.g., Hawkins and Sutton 2009): 1) "model uncertainty" - the uncertainty that arises from variations in the forced response from one model to the next; 2) "scenario uncertainty" - the uncertainty that arises from differences in forcing scenarios; and 3) "uncertainty due to internal variability", the uncertainty that arises from unpredictable

internal variability in the climate system. Model and scenario uncertainty can be reduced as models and our understanding of future forcing scenarios improve. The uncertainty due to internal climate variability arises from the chaotic nature of the climate system and - to the extent that it is unpredictable - is presumably irreducible.

The analytic model given by Eq. 7 provides a simple basis for estimating the uncertainty in climate change projections due to internal variability using the statistics of the unforced climate. It provides a zeroth-order estimate of the uncertainty in future trends due to internal variability in a range of physical fields including, for example, precipitation averaged over a watershed, surface air temperature averaged over an agricultural region, and the atmospheric circulation at middle latitudes.

The analytic model is based on two assumptions: 1) the internal variability is well-modeled as Gaussian; and 2) the standard deviation and/or autocorrelation of the internal climate variability do not change in response to anthropogenic forcing. The robustness of the model to both assumptions is strongly supported by the close similarities between: 1) the uncertainties in climate trends estimated by the statistics of an unforced control simulation and 2) the uncertainties found in a large-ensemble of climate change simulations. To the extent that the assumptions of the analytic model hold, the results suggest that large-ensembles provide little information on the role of internal variability in future climate that can not be inferred from the statistics of an unforced control simulation.

The analytic model also makes clear the direct relationship between 1) biases in a model control simulation and 2) the uncertainty in projections of future climate change. If the statistics of a control simulation in a given climate model are biased relative to the observations, then so is the apparent role of internal variability in future climate change

simulated by the same model. To the extent that albeit imperfect observational records provide a more realistic representation of the real-world than a climate model, it follows that the role of internal variability in future climate trends is arguably best estimated not from a long control simulation or a large-ensemble of climate change simulations, but from observational estimates of internal climate variability.

Acknowledgments

We thank John M. Wallace, Susan Solomon and an anonymous reviewer for insightful comments on the study. NCAR is sponsored by the National Science Foundation. DWJT, EAB and WEF are supported by the National Science Foundation Climate Dynamics program.

Appendix

Expanding the standard deviation of the time axis

The standard deviation of the time axis (the denominator in Eq. 2) can be expanded as a function of n_t, since the time axis corresponds to a series of consecutive integers. Using two formulae for consecutive integers:

$$\sum_{i=1}^{n_t} i = \frac{n_t(n_t+1)}{2} \qquad \text{and} \qquad \sum_{i=1}^{n_t} i^2 = \frac{n_t(n_t+1)(2n_t+1)}{6}$$

it follows that:

$$\bar{i} = \frac{1}{n_t}\sum_{i=1}^{n_t} i = \frac{(n_t+1)}{2}$$

and thus:

$$\begin{aligned}
&\sum_{i=1}^{n_t}(i-\bar{i})^2 = \sum_{i=1}^{n_t} i^2 - \sum_{i=1}^{n_t} 2i\cdot\bar{i} + \sum_{i=1}^{n_t}\bar{i}^2 \\
&= \sum_{i=1}^{n_t} i^2 - (n_t+1)\sum_{i=1}^{n_t} i + \sum_{i=1}^{n_t}\left(\frac{n_t+1}{2}\right)^2 \\
&= \frac{n_t(n_t+1)(2n_t+1)}{6} - \frac{n_t(n_t+1)^2}{2} + \frac{n_t(n_t+1)^2}{4} \\
&= \frac{{n_t}^3 - n_t}{12}
\end{aligned}$$

Accounting for autocorrelation in the standard error of the residuals

A simple and commonly used method for accounting for the bias in the sample standard deviation is to substitute an effective sample size n_{eff} for the sample size n_t in the calculation of the standard deviation (i.e., Eq. 5). If *x(t)* is well-modeled as Gaussian red noise (and thus its autocorrelation function decays exponentially with lag), then n_{eff} can be approximated as (Mitchell et al. 1966; Santer et al. 2000):

$$n_{eff} \sim n_t \left(\frac{1 - r_1}{1 + r_1} \right)$$

Note that the time between independent samples in a red-noise time series is equal to ~twice the e-folding time of the autocorrelation function (Leith 1973).

Substituting n_{eff} for n_t in the denominator of Eq. 5 yields the following relationship between the standard error (s_e) and the standard deviation (σ) of the residuals:

$$s_e = \sigma \gamma(n_t, r_1)$$

where

$$\gamma(n_t, r_1) \equiv \left(\frac{[n_t - 2]}{\left[n_t \left(\frac{1 - r_1}{1 + r_1} \right) - 2 \right]} \right)^{1/2}.$$

Estimating statistical significance in Figures 3c and 4c

The stippling in Figure 3c denotes regions where the predicted margins shown in Figure 3c fall within the 95% confidence intervals of the actual margins shown in Figure 3b. The 95% confidence intervals of the actual margins are found by: 1) calculating the 95% confidence intervals of the interannual standard deviations from the control simulation using the Chi-squared distribution assuming 40 degrees of freedom; 2) multiplying the resulting confidence intervals by a factor of 2 to convert them to confidence intervals of the predicted margins of error. The stippling in Figure 4c is found in an analogous manner.

Changes in interannual variance over the 2011-2060 period

Figure A1 indicates the ratio of variances between the periods 2051-2060 and 2011-2020 from the NCAR 40-member ensemble of climate change simulations. See caption for details.

References

Adler, R.F., et al, 2003: The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present). J. Hydrometeor., 4, 1147-1167.

Bindoff, N.L., P.A. Stott, et al., 2013: Detection and Attribution of Climate Change: from Global to Regional. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the IPCC [Stocker, T.F., et al. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Casola, J. H. et al. 2009: Assessing the Impacts of Global Warming on Snowpack in the Washington Cascades. J. Climate, 22, 2758-2772.

Christensen, J. H., B. Hewitson, et al., 2007: Regional Climate Projections. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Collins, M., R. Knutti, et al., 2013: Long-term Climate Change: Projections, Commitments and Irreversibility. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the IPCC [Stocker, T.F., et al. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Deser, C., Phillips, A., Bourdette, V. & Teng, H., 2012a: Uncertainty in climate change projections: the role of internal variability. Clim. Dynam. 38, 527–547.

Deser, C., R. Knutti, S. Solomon, & A. S. Phillips, 2012b: Communication of the role of natural variability in future North American climate. Nat. Clim. Change, 2, 775-779, doi: 10.1038/nclimate1562.

Deser, C., A. S. Phillips, M. A. Alexander & B. V. Smoliak, 2014:. Projecting North American Climate over the next 50 years: Uncertainty due to internal variability. J. Climate, 27, 2271-2296.

Diffenbaugh, N. S., & Scherer, M., 2011: Observational and model evidence of global emergence of permanent, unprecedented heat in the 20th and 21st centuries. Clim. Change, 107, 615-624.

Feldstein, S. B., 2000: Teleconnections and ENSO: The timescale, power spectra, and climate noise properties. J. Climate, 13, 4430.

Fischer, E. & Schär, C, 2009: Future changes in daily summer temperature variability: driving processes and role for temperature extremes. Clim. Dynam. 33, 917–935.

Gregory, J. M. & Mitchell, J. F. B., 1995: Simulation of daily variability of surface temperature and precipitation over Europe in the current and 2xCO2 climates using the UKMO high-resolution climate model. Q. J. R. Meteorol. Soc., 121, 1451-1476.

Hartmann, D. L. & Lo, F., 1998: Wave-driven zonal flow vacillation in the Southern Hemisphere. J. Atmos. Sci., 55, 1303–1315.

Hawkins, E. & Sutton, R., 2009: The potential to narrow uncertainty in regional climate predictions. Bull. Amer. Meteor. Soc., 90, 1095–1107.

Hawkins, E. & Sutton, R, 2011: The potential to narrow uncertainty in projections of regional precipitation change. Clim. Dynam. 37, 407–418.

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

Held, I. M. & Soden, B, J., 2006: Robust Responses of the Hydrological Cycle to Global Warming. J. Climate, 19, 5686–5699.

IPCC, 2014: Summary for policymakers. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., et al. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1-32.

Kay, J. E., et al. The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bull. Amer. Met. Soc.*, in press.

Kennedy J.J., Rayner, N.A., Smith, R.O., Saunby, M. and Parker, D.E., 2011: Reassessing biases and other uncertainties in sea-surface temperature observations measured in situ since 1850 part 2: biases and homogenisation. Journal of Geophysical Research 116, D14104, doi:10.1029/2010JD015220.

Kirtman, B., S.B. Power, et al., 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 IPCC [Stocker, T.F., et al. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Knutson, T. R., F. Zend, & A, T. Wittenberg, 2013: Multimodel Assessment of Regional Surface Temperature Trends: CMIP3 and CMIP5 Twentieth-Century Simulations, 26, 8709-8743.

Leith, C. E., 1973: The standard error of time-averaged estimates of climatic means. J. Appl. Meteor., 12, 1066-1069.

Lettenmaier, D. P., 1976: Detection of trends in water quality data from records with dependent observations. Water Resources Research. 12, 1037-1046.

Mahlstein, I., R. Knutti, S. Solomon, and R. W. Portmann, 2011: Early onset of significant local warming in low latitude countries, Environ. Res. Lett., 6(3), 034009, doi:10.1088/1748-9326/6/3/034009.

Mitchell, J. M., Jr., B. Dzerdzeevskii, H. Flohn, W. L. Hofmeyr, H. H. Lamb, K. N. Rao & C. C. Wallen, 1966: Climatic Change, Techn. Note 79, 79 pp., World Meteorol. Org., Geneva.

Newman, M., Compo G. P., & Alexander M. A., 2003: ENSO-forced variability of the Pacific decadal oscillation. J. Climate, 23, 3853-3857.

Osborn, T.J. and Jones, P.D., 2014: The CRUTEM4 land-surface air temperature data set: construction, previous versions and dissemination via Google Earth. Earth System Science Data 6, 61-68, doi:10.5194/essd-6-61-2014.

Santer, B. D., T. M. L. Wigley, J. S. Boyle, D. J. Gaffen, J. J. Hnilo, D. Nychka, D. E. Parker & K. E. Taylor, 2000: Statistical significance of trends and trend differences in layer average atmospheric temperature time series, J. Geophys. Res., 105, 7337–7356.

Schneider, T., T. Bischoff, and H. Plotka, 2015: Physics of changes in synoptic midlatitude temperature variability. Journal of Climate, submitted.

Screen J. A., 2014: Arctic amplification decreases temperature variance in Northern mid- to high-latitudes, Nature Climate Change, DOI:10.1038/nclimate2268.

von Storch, H. & Zwiers, F. W., 1999: Statistical Analysis in Climate Research Cambridge University Press, 484 pp.

Wallace, J. M., Y. Zhang &J. A. Renwick, 1995: Dynamic contribution to hemispheric

mean temperature trends, Science, 270, 780-783.

Wallace, J. M., Q. Fu, B. V. Smoliak, P. Lin & C. M. Johanson, 2012:. Simulated versus observed patterns of warming over the extratropical Northern Hemisphere continents during the cold season. Proc. Natl. Acad. Sci., 109, 14332-14337.

Wallace, J.M. C. Deser, B.V. Smoliak, & A.S. Phillips, 2013: Attribution of climate change in the presence of internal variability, In Climate Change: Multidecadal and Beyond (Eds: C.P. Chang, M. Ghil, M. Latif, and J. M. Wallace). World Scientific Series on Asia-Pacific Weather and Climate, Vol. 6.

Zhang, C., 2005: Madden-Julian Oscillation. Rev. Geophys, 43, RG2003, Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences: An Introduction. Academic Press, 467 pp.

Figure Captions

Figure 1. Left. The standard deviations of the 50 year trends in October-March mean surface temperature (top) and precipitation (bottom) based on output from the NCAR 40-member ensemble of climate change simulations. Trends are expressed as the total change over the 50 year period 2011-2061. The trend standard deviations indicate the spread in the trends derived from all 40 ensemble members. Right. Wintertime mean time series of surface temperature (top) and precipitation (bottom) for grid boxes collocated with the indicated locations. The grey lines show results for all 40 ensemble members; the red and blue lines indicate the ensemble members with the largest and smallest trends over the 2011-2061 period, respectively. Tickmarks at 1 deg C and 1 mm/day.

Figure 2. Analytic solutions for the uncertainty in future climate due to internal variability. Trend amplitude required to exceed the 95% margin of error relative to the standard deviation of the internal variability. For example, a trend of "2" indicates that the trend must be twice as large as the internal (unforced) variability to exceed the 95% margin of error. Results are derived from Equation 7 and are shown as a function of the trend length (in timesteps) and the lag-one autocorrelation (r1). Contours are spaced at trend amplitudes of 0.5.

Figure 3. Using the control run to estimate the 95% margins of error on 50-year trends in October-March mean surface temperature. a) The "forced response" defined as the linear trends in October-March mean surface temperature averaged over all 40 ensemble members in K/50 years. b) The "actual" 95% margins of error on the 50 year

trends derived from all 40 ensemble members. c) The "predicted" 95% margins of error on the 50 year trends derived by applying Eq. 7 to the statistics of the control run. Stippling indicates regions where the predicted margins are *not* significantly different from the actual margins (see Appendix for details). (Surrounding panels) The probability distribution functions of the 50 year trends at grid boxes collocated with the indicated cities. The grey bars denote the histograms derived from all 40 ensemble members. The standard deviations of the blue curves correspond to the predicted margins of error on the trends found by applying Eq. 7 to the statistics of the control run; the means of the blue curves correspond to the ensemble mean trends. The areas under the blue curves are normalized so that they match the areas under the attendant gray bars.

Figure 4. As in Fig. 3, except for October-March mean precipitation. Trends are expressed in mm/day/50 years.

Figure 5. Using observations to estimate the 95% margins of error on 50-year trends in October-March mean surface temperature. (top) The "predicted" 95% margins of error on the 50 year trends derived by applying Eq. 7 to the statistics of the observations. (bottom) The "predicted" 95% margins of error derived by applying Eq. 7 to the statistics of the control run. The results in the bottom panel are the same as those shown in Fig. 3c except that 1) the stippling indicates regions where the modeled and observed margins of error *are* significantly different from each other at the 95% confidence level (ratios $>1.5:1$ or $<1:1.5$ exceed the 95% confidence level based on a test of the F-statistic assuming one degree of freedom per year); and 2) the model output has been

interpolated to the same mesh as the observations. (Surrounding panels) The probability distribution functions of the 50 year trends for grid boxes collocated with the indicated cities. Solid and dashed curves denote the distribution functions predicted by applying Eq. 7 to the statistics of the observations and the interpolated control simulation output, respectively. Distributions are normalized so that they have the same area.

Figure 6. As in Fig. 5, but for October-March mean precipitation. The observations have been interpolated to the same mesh as the model output.

Figure 7. Actual trends (red) and the predicted ranges of the trends (black dashed lines) in fields indicated. Trend periods start in 2011 and end on the year indicated on the abscissa. Units are change over the length of the trend period. Red dots indicate the "actual" trends derived from all 40 ensemble members. Black dashed lines indicate the "predicted" ranges of the trends found by applying Eq. 7 to the statistics of the control run as a function of trend length. The margins of error on the trends indicated by the black dashed lines (e in Eq. 1) are given by the statistics of the control run; the amplitudes of the forced signal (b in Eq. 1) are given by the ensemble mean trends. The units on the North Atlantic Oscillation index are standard deviations of the interannual variability.

Appendix Figure A1. Ratio of variances between the periods 2051-2060 and 2011-2020 from the NCAR 40-member ensemble of climate change simulations. The variances are calculated as the pooled detrended seasonal-mean data from all ensemble members.

Ratios greater than ~1.4 and less than ~0.71 (indicated by stippling) are significant at the 95% level based on the F-statistic.

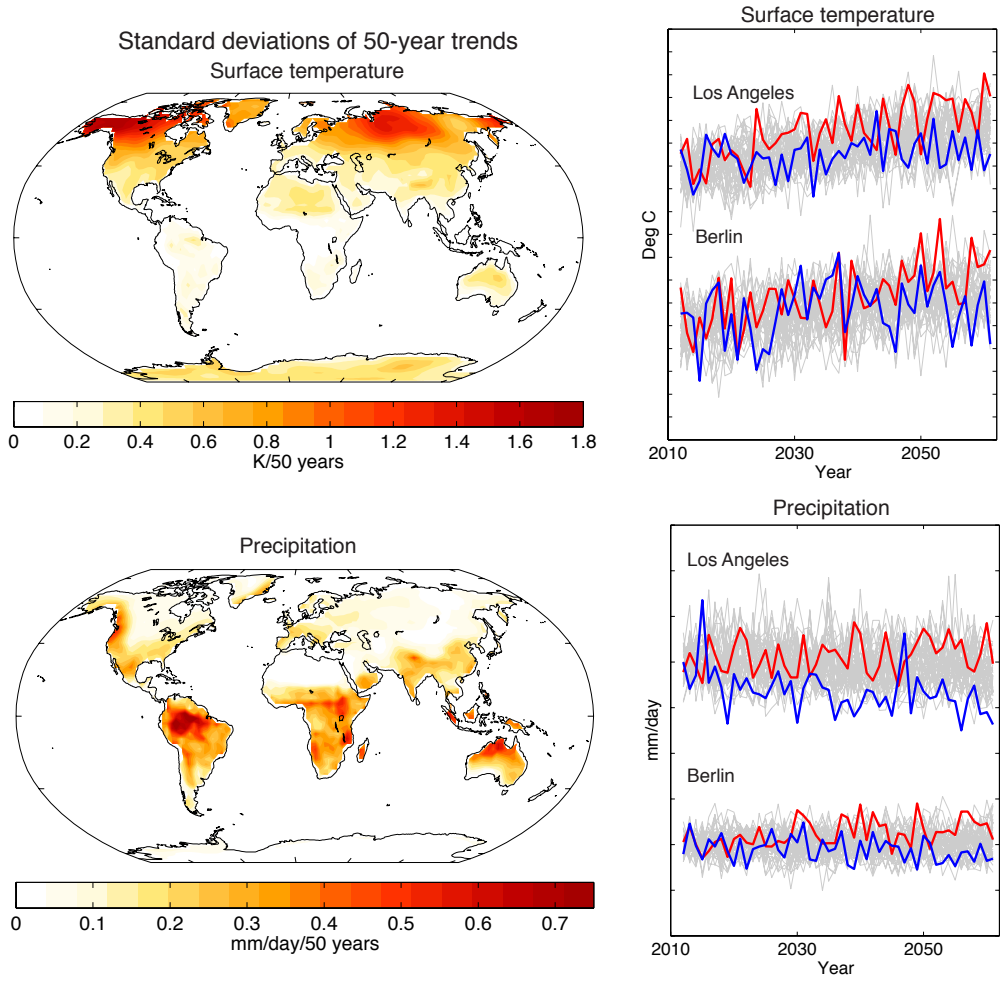


Figure 1. Left. The standard deviations of the 50 year trends in October-March mean surface temperature (top) and precipitation (bottom) based on output from the NCAR 40-member ensemble of climate change simulations. Trends are expressed as the total change over the 50 year period 2011-2061. The trend standard deviations indicate the spread in the trends derived from all 40 ensemble members. **Right.** Wintertime mean time series of surface temperature (top) and precipitation (bottom) for grid boxes collocated with the indicated locations. The grey lines show results for all 40 ensemble members; the red and blue lines indicate the ensemble members with the largest and smallest trends over the 2011-2061 period, respectively. Tickmarks at 1 deg C and 1 mm/day.

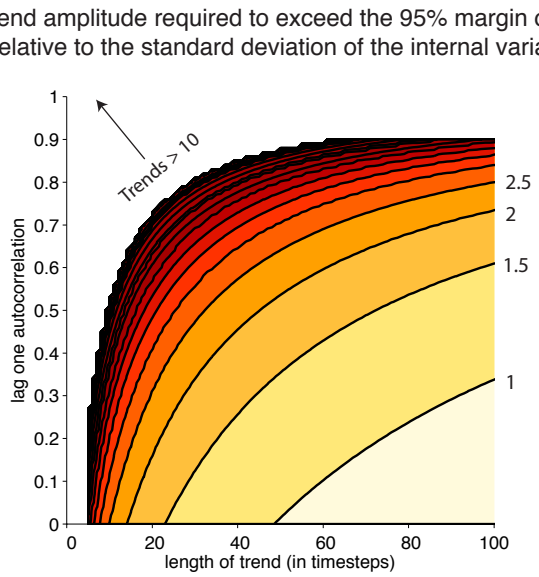


Figure 2. Analytic solutions for the uncertainty in future climate due to internal variability. Trend amplitude required to exceed the 95% margin of error relative to the standard deviation of the internal variability. For example, a trend of "2" indicates that the trend must be twice as large as the internal (unforced) variability to exceed the 95% margin of error. Results are derived from Equation 7 and are shown as a function of the trend length (in timesteps) and the lag-one autocorrelation (r_1). Contours are spaced at trend amplitudes of 0.5.

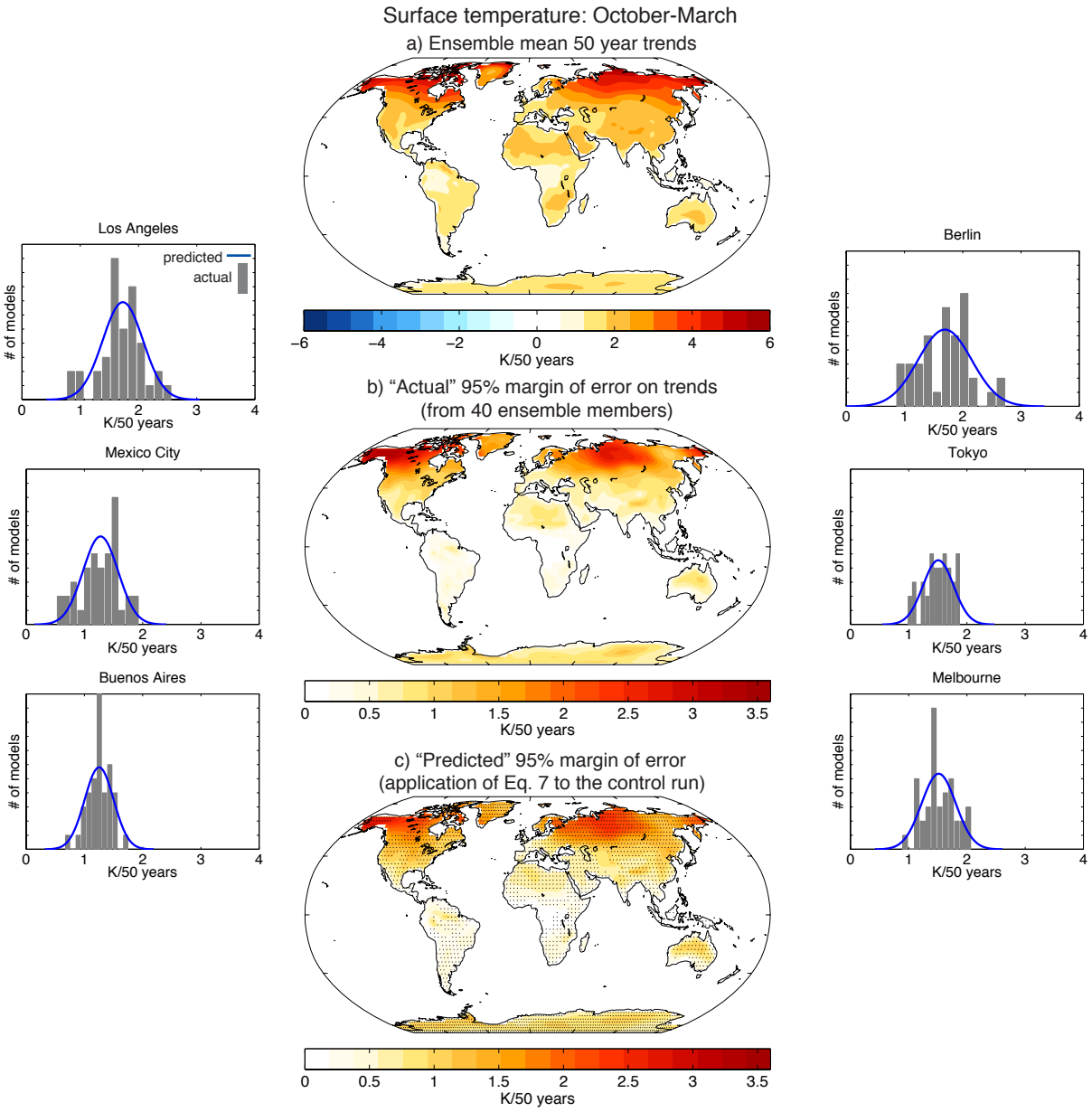


Figure 3. Using the control run to estimate the 95% margins of error on 50-year trends in October-March mean surface temperature. a) The “forced response” defined as the linear trends in October-March mean surface temperature averaged over all 40 ensemble members in K/50 years. b) The “actual” 95% margins of error on the 50 year trends derived from all 40 ensemble members. c) The “predicted” 95% margins of error on the 50 year trends derived by applying Eq. 7 to the statistics of the control run. Stippling indicates regions where the predicted margins are not significantly different from the actual margins (see Appendix for details). (Surrounding panels) The probability distribution functions of the 50 year trends at grid boxes collocated with the indicated cities. The grey bars denote the histograms derived from all 40 ensemble members. The standard deviations of the blue curves correspond to the predicted margins of error on the trends found by applying Eq. 7 to the statistics of the control run; the means of the blue curves correspond to the ensemble mean trends. The areas under the blue curves are normalized so that they match the areas under the attendant gray bars.

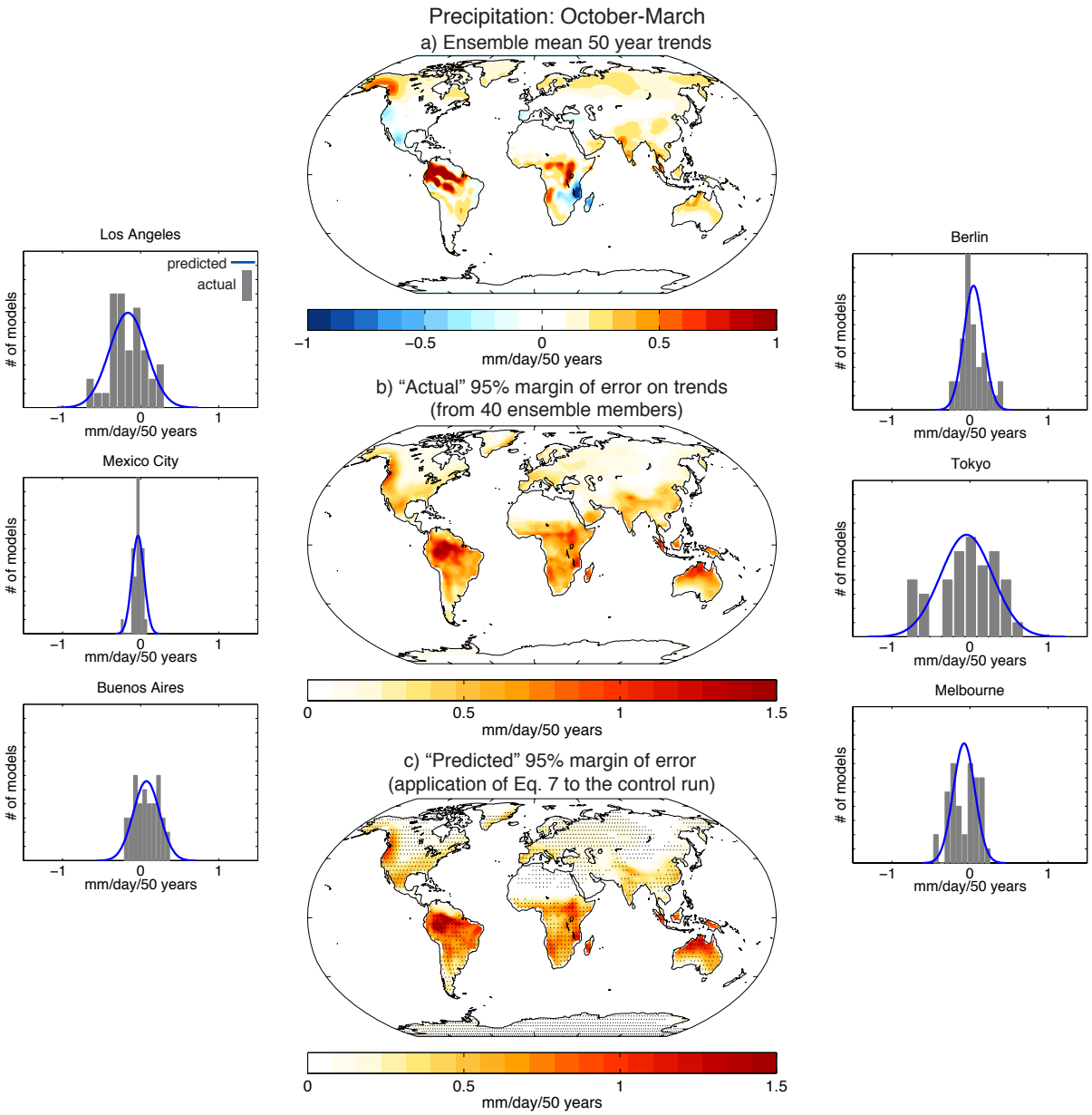


Figure 4. As in Fig. 3, except for October-March mean precipitation. Trends are expressed in mm/day/50 years.

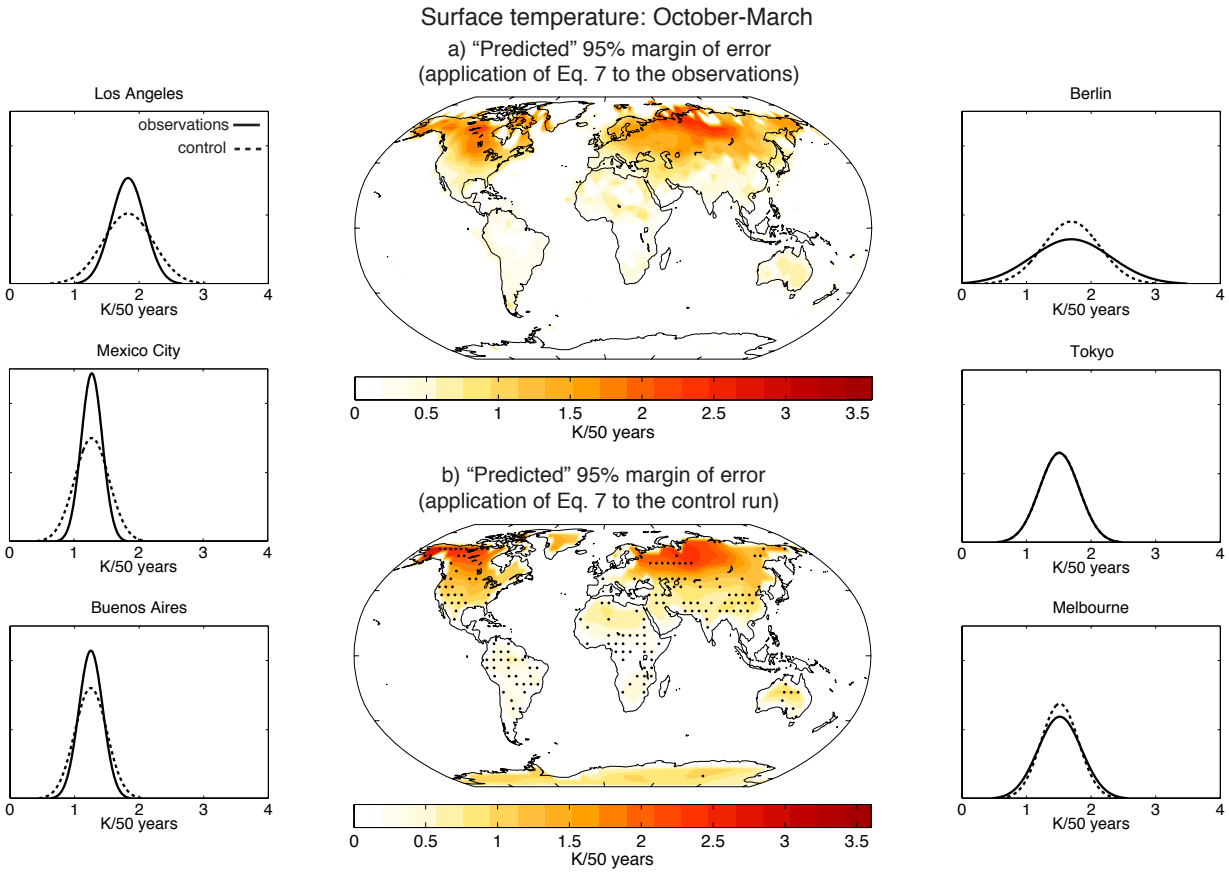


Figure 5. Using observations to estimate the 95% margins of error on 50-year trends in October-March mean surface temperature. (top) The "predicted" 95% margins of error on the 50 year trends derived by applying Eq. 7 to the statistics of the observations. (bottom) The "predicted" 95% margins of error derived by applying Eq. 7 to the statistics of the control run. The results in the bottom panel are the same as those shown in Fig. 3c except that 1) the stippling indicates regions where the modeled and observed margins of error are significantly different from each other at the 95% confidence level (ratios or exceed the 95% confidence level based on a test of the F-statistic assuming one degree of freedom per year); and 2) the model output has been interpolated to the same mesh as the observations. (Surrounding panels) The probability distribution functions of the 50 year trends for grid boxes collocated with the indicated cities. Solid and dashed curves denote the distribution functions predicted by applying Eq. 7 to the statistics of the observations and the interpolated control simulation output, respectively. Distributions are normalized so that they have the same area.

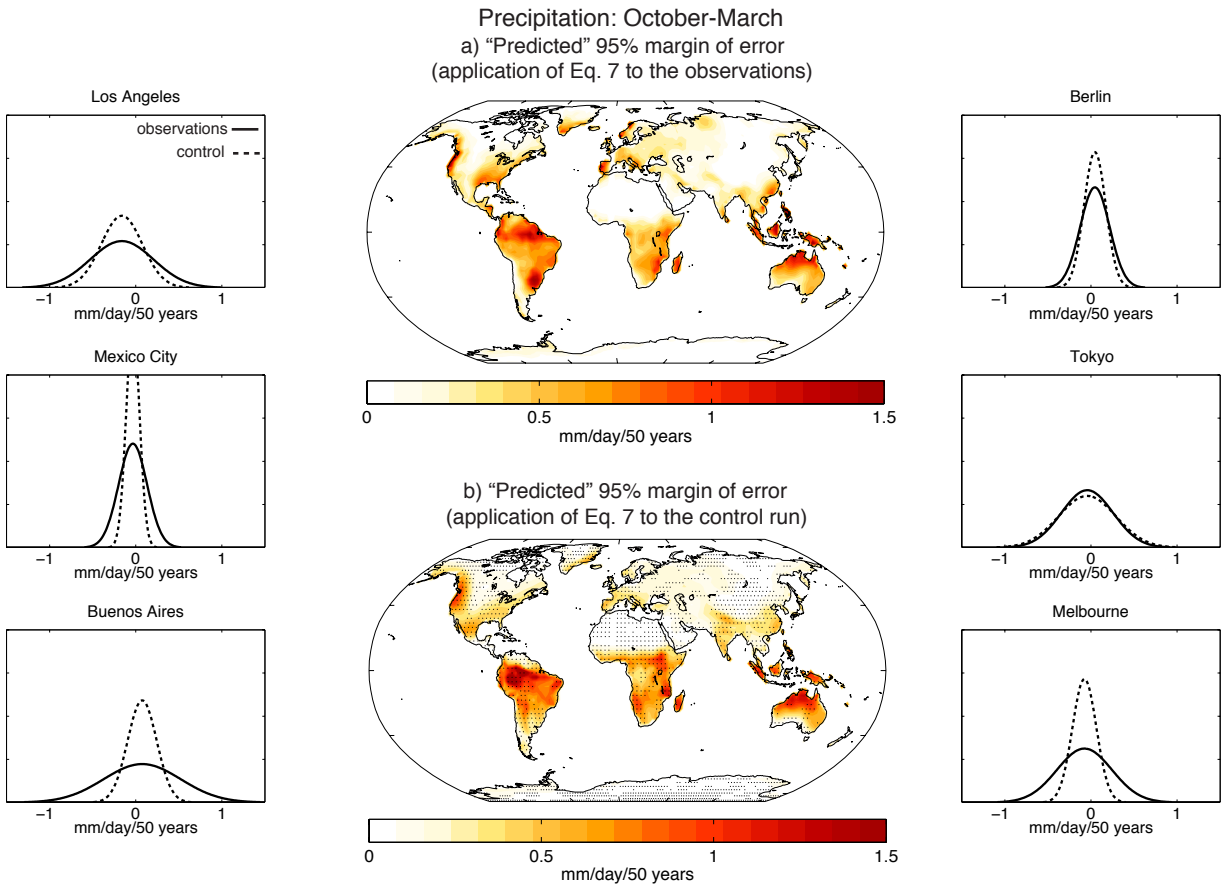


Figure 6. As in Fig. 5, but for October-March mean precipitation. The observations have been interpolated to the same mesh as the model output.

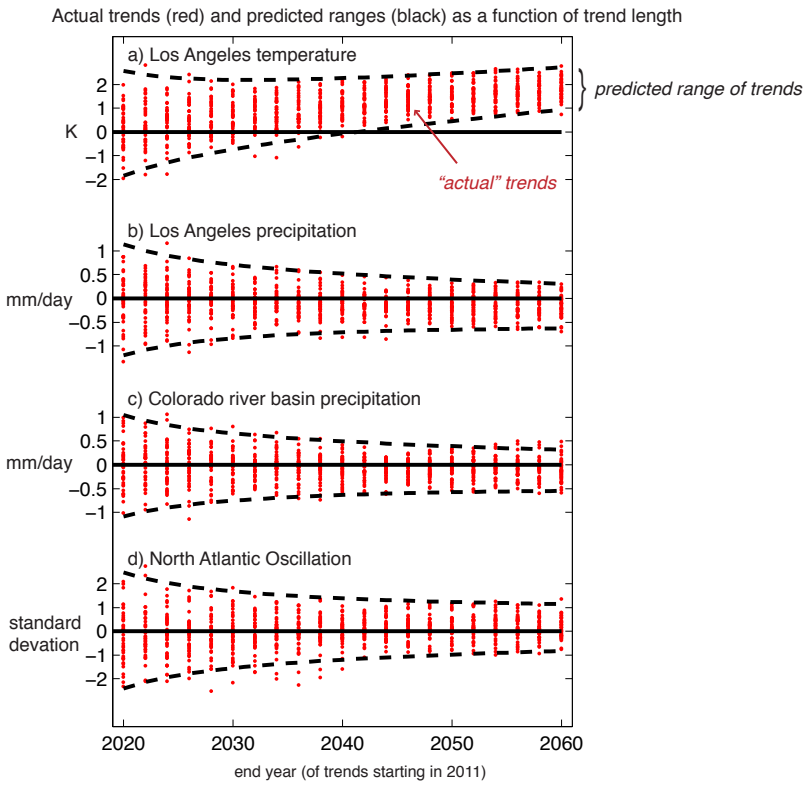
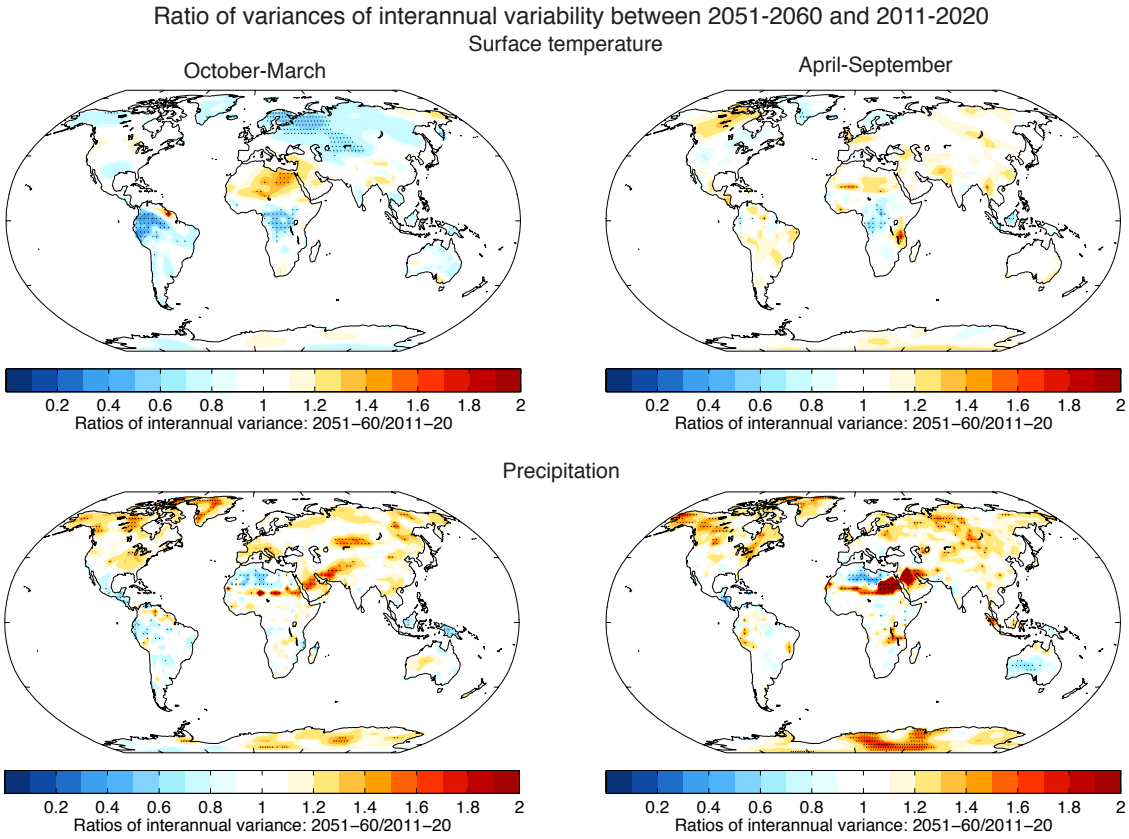


Figure 7. Actual trends (red) and the predicted ranges of the trends (black dashed lines) in fields indicated. Trend periods start in 2011 and end on the year indicated on the abscissa. Units are change over the length of the trend period. Red dots indicate the "actual" trends derived from all 40 ensemble members. Black dashed lines indicate the "predicted" ranges of the trends found by applying Eq. 7 to the statistics of the control run as a function of trend length. The margins of error on the trends indicated by the black dashed lines (*e* in Eq. 1) are given by the statistics of the control run; the amplitudes of the forced signal (*b* in Eq. 1) are given by the ensemble mean trends. The units on the North Atlantic Oscillation index are standard deviations of the interannual variability.



Appendix Figure A1. Ratio of variances between the periods 2051-2060 and 2011-2020 from the NCAR 40-member ensemble of climate change simulations. The variances are calculated as the pooled detrended seasonal-mean data from all ensemble members. Ratios greater than ~1.4 and less than ~0.71 (indicated by stippling) are significant at the 95% level based on the F-statistic.