

THESIS

TOWARDS UNDERSTANDING THE ROLE OF NATURAL VARIABILITY IN CLIMATE
CHANGE

Submitted by

Jingyuan Li

Department of Atmospheric Science

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2017

Master's Committee:

Advisor: David W. J. Thompson

Elizabeth A. Barnes

Daniel Cooley

Copyright by Jingyuan Li 2017

All Rights Reserved

ABSTRACT

TOWARDS UNDERSTANDING THE ROLE OF NATURAL VARIABILITY IN CLIMATE CHANGE

Natural variability plays a large role in determining surface climate on local and regional scales. Understanding the role of natural variability is crucial for accurately assessing and attributing climate trends, both past and future. One successful way to examine the role of natural variability in climate change has been through large ensembles of climate models. This thesis uses one such large ensemble (the NCAR CESM-LE) to test various methods used to quantify natural variability in the context of climate change.

We first introduce a simple analytic expression for calculating the lead time required for a linear trend to emerge in a Gaussian first order autoregressive process. The expression is derived from the standard error of the regression and is tested using the CESM-LE. It is shown to provide a robust estimate of the point in time when the forced signal of climate change has emerged from the natural variability of the climate system with a predetermined level of statistical confidence. The expression provides a novel analytic tool for estimating the time of emergence of anthropogenic climate change and its associated regional climate impacts from either observed or modeled estimates of natural variability and trends.

We next compare and analyze various methods for calculating the effects of internal circulation dynamics on surface temperature. Dynamical adjustment seeks to separate out dynamical contribution to temperature trends, thus reducing the amplitude of natural variability that obscures the signal of anthropogenic forcing. Three specific methods used in the climate literature are examined: principal component analysis (PCR), maximum covariance analysis (MCA), and constructed circulation analogs. An assessment of these methods are given with their respective results from the CESM control run and large ensemble.

ACKNOWLEDGEMENTS

Thanks to my advisor, Dr. David W. J. Thompson, for his guidance and support. Thanks also to Dr. Elizabeth Barnes and Dr. Daniel Cooley for serving on my committee and providing insightful discussions and comments. Thanks to the Thompson research group (Samantha Wills, Brian Crow, Casey, Patrizio, Luke Davis) for helpful input on the analysis.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	v
Chapter 1	1
1.1	1
1.2	3
1.3	5
1.4	6
Chapter 2	9
2.1	9
2.1.1	9
2.1.2	9
2.2	10
2.2.1	10
2.2.2	10
2.2.3	11
Chapter 3	12
3.1	12
3.2	15
3.3	22
Chapter 4	32
4.1	33
4.2	37
Chapter 5	52
5.1	52
5.2	53
5.3	54
Bibliography	56

LIST OF FIGURES

1.1	(left) The standard deviations of 50 year trends in October-March mean surface temperature (top) and precipitation (bottom) from the NCAR CCSM4 40-member large ensemble. Trends are expressed as the total change over the period 2011-2061, and the standard deviations are over the 40 ensemble members. The right panels show wintertime mean temperature and precipitation for two specific grid boxes: Los Angeles and Berlin. The grey lines show results for all 40 ensemble members; the red and blue lines indicate ensemble members with the largest and smallest trends, respectively. Tick marks at every 1°C and 1 mm day^{-1} . Figure taken from [1].	8
3.1	Schematic showing the TOE as calculated from Equation 3.2. The black solid line shows the estimated trend (\hat{b}), and the dashed lines indicate the predicted ranges of trends calculated from Equation 3.1. The blue vertical line indicates the lead time when the forced trend is statistically significant, which is the point when the lower dashed line (lower 95% bound) intersects the zero line.	24
3.2	Comparisons of the standard deviations calculated from the CESM model over the period 1970-2015 using four different methods to remove the long-term forced signal. Panels (a) and (b) show the pooled standard deviations after removing the linear trend from all grid point time series in all ensemble members (as used for Figure 3.4c, d); panels (c) and (d) show pooled standard deviations after removing a 2nd order polynomial fit from all grid point time series in all ensemble members; panels (e) and (f) show pooled standard deviations after removing the ensemble mean time series from all grid point time series in all ensemble members; and panels (g) and (h) show standard deviations from the last 1380 years of the control run. Grey denotes ocean regions.	25
3.3	Trend amplitudes for modeled surface temperature at the grid box collocated with (a) London, UK, (b) central Siberia, (c) Jakarta, Indonesia, using CESM-LE output. The red dots indicate individual trends from all 30 ensemble members and the dashed lines indicate the predicted ranges of trends found by applying Equation 3.1 to the statistics of the model internal variability. The blue vertical line indicates the lead time when the forced trend is statistically significant as per Equation 3.2. See text for details.	26
3.4	Using CESM-LE output to test Equation 3.2. The top panels show the “empirically derived” lead times when the trends emerge from natural variability, calculated as the time step when 29 out of 30 ensemble members exhibit positive trends in the current and all subsequent time steps. The bottom panels show the “analytically derived” lead times (T_{SIG}) derived by applying Equation 3.2 to the model natural variability. Winter corresponds to the October-March means; summer to the April-September means. Note that all lead times beyond the limit of the analysis period (45 years or 2015) are white. Grey denotes ocean regions.	27
3.5	Scatterplots showing the relationship between the “empirically-derived” values of T_{SIG} and the “analytically-derived” values of T_{SIG} for (a) the winter season, and (b) the summer season. The good agreement about the 1:1 ratio (represented as the red line) shows that both methods give similar distributions of T_{SIG}	28

3.6	As in the bottom panels of Figure 3.4, but for lead times calculated by applying Equation 3.2 to the (top) ensemble mean trends from CESM-LE and the observed natural variability and (bottom) linear trends from the observations and the observed natural variability. The observed natural variability is estimated in both panels as the standard deviation of the detrended data. The observations are used over the period 1970-2015. White denotes lead times beyond the limit of the analysis period (larger than 45 years), while grey denotes oceans and any regions of missing data. See text for details.	29
3.7	Top panels show the ratio of amplitudes of natural variability from the CESM-LE to that from observations for (a) winter (October-March) and (b) summer (April-September). Warm colors denote regions where the amplitude of natural variability is larger in the model for the 1970-2015 period, while cool colors denote regions where the amplitude of natural variability is larger in observations for the same period. Bottom panels show differences between the 1970-2015 ensemble-averaged trends from CESM-LE and trends from HadCRUT4 observations for (c) winter and (d) summer. The CESM-LE trends were used in calculating the lead times in the top panels of Figure 3.6, while the observed trends were used in calculating the lead times in the bottom panels of Figure 3.6. The predominance of cool colors for both seasons indicate that observed trends from 1970-2015 were larger than the simulated ensemble mean trends over the same period. Grey denotes oceans and any regions of missing data.	30
3.8	Grid boxes where TOE calculated from HadCRUT4 surface temperature observations fall outside the 95% bounds on TOE calculated for individual ensemble member trends and standard deviations. Only 5% of the observed TOE estimates lie outside the bounds given by the individual ensemble members.	31
3.9	Comparison between lead times calculated using Equation 3.2 (top) and Equation 1.1 where $k = 2$ (bottom). In both cases, the forced signal is given as the ensemble mean temperature trends over 1970-2015, and the natural variability as the detrended observed interannual standard deviation. Grey denotes oceans and any regions of missing data. Note the top panel is reproduced from Figure 3.6a.	31
4.1	SLP (contours) and surface air temperature (shading) regressed onto the first 10 standardized principal components of monthly mean SLP data from the CESM control run. Contours are every 1.5 hPa; negative contours are shown in the dashed lines. DJF values only.	43
4.2	Time series for a single grid box in central Siberia of temperature (top, blue), T_{fit} from the first 10 PCs of SLP (middle, red), and T_{res} (bottom, orange), obtained as the difference between the first two time series.	44
4.3	Regression maps for (a) SLP, and (b) temperature on the temperature timeseries at a single grid box in central Siberia. Data is for DJF monthly anomalies from the CESM-LE control run.	44
4.4	Same as Figure 4.2, but with T_{fit} and T_{res} from the MCA method.	45
4.5	Variance in temperature for DJF months explained by (a) the first 10 PCs of SLP, and (b) the expansion coefficient timeseries, both calculated from DJF monthly anomalies of SLP from the CESM control run. PCR does a better job overall, explaining an average of 40% of the variance in the winter temperature field, while MCA explains an average of 32% of the variance over the Northern Hemisphere extratropics.	45

4.6	Plots showing (left) the standard error on 50 year DJF averaged trends, calculated as (left) two times the standard deviation across all 34 chunks of 50 year trends, and (right) the interannual standard deviation of DJF averaged anomaly data. The calculations are shown for (top) temperature, (middle) dynamically induced temperature (T_{fit} via PCR), and (bottom) the dynamically adjusted temperature (T_{res}). From the CESM control run.	46
4.7	50-year DJF surface temperature trends for each ensemble member of the CESM-LE (labeled 1-30) for the period 1955-2004.	47
4.8	50-year dynamically-adjusted DJF temperature trends for each ensemble member of the CESM-LE (labeled 1-30) for the period 1955-2004. The dynamically-adjusted temperature was calculated as the residual obtained by subtracting the temperature fitted to first 10 PCs for each ensemble member from the original temperature data. . .	48
4.9	95% standard error on DJF averaged trends across all ensemble members for 1955-2004 of (a) surface air temperature (b) T_{fit} from internal circulation dynamics found by applying the PCR method, and (c) T_{res} , the residual temperature not associated with internal dynamics.	49
4.10	The correlation between varying numbers of ensemble members used and the ensemble-mean DJF temperature trend. The left panel shows the spread given by the ensemble members for the actual DJF temperature trends for the period 1955-2004, while the right panel shows the same spread for the temperature with circulation dynamics removed via the PC regression method. For the ensemble mean, the residual temperature trends and actual temperature trends have a correlation of 0.94.	49
4.11	Probability density distributions of the spatial correlations shown in Figure 4.10. The left panels show the correlations for surface air temperature data, while the right panels show the correlations for the dynamically-adjusted temperature. Comparisons are for 100 runs of random averaged values for 1, 2, and 5 ensemble members.	50
4.12	PDFs of (left) RMSE between the ensemble mean temperature trend and temperature trends from 30 ensemble members; and (right) RMSE between the ensemble mean temperature trend and temperature trends from 30 dynamically adjusted ensemble members.	51

Chapter 1

Introduction

One of the most important issues for the 21st century is predicting how local and regional climate will change due to anthropogenic forcing. Increasing greenhouse gas concentrations have already contributed to at least half of the observed warming seen since 1951 ([2]). While the radiative forcing due to increasing greenhouse gas concentrations will be the key driver of future climate, it is difficult to distinguish the signal from the noise intrinsic to the climate system. This thesis seeks to distinguish between these two influences (natural and anthropogenic) to help elucidate future effects on climate.

Section 1.1 reviews the concept of natural variability and examines previous studies on its role in climate, Section 1.2 introduces the concept of “time of emergence” (TOE), a key variable in climate prediction, and reviews previous studies on TOE. Section 1.3 examines past studies on dynamical adjustment techniques, which attempt to remove circulation-induced variability in surface temperature. Section 1.4 outlines the motivation and goals of this thesis.

1.1 Natural and Internal Variability

Uncertainty in future climate projections arise from three main sources: 1) uncertainty in anthropogenic emissions, 2) uncertainty in the model response (differences amongst climate models), and 3) uncertainty from the internal variability of the climate system ([3]). Internal variability is defined as the natural (innate) variability of the climate system arising from processes and coupled interactions between the atmosphere, oceans, land, and cryosphere. While we can better specify emissions scenarios and improve our climate models, the random processes in the unforced climate system makes it unlikely that the uncertainty from internal variability can be reduced. Thus, internal variability poses an inherent limit to climate predictability. Although human-induced climate change is likely to dominate over internal variability for time periods longer than a decade on the global scale, internal variability can overwhelm anthropogenic forcing on regional and local

scales, even on timescales of up to 50 years at middle and high latitudes ([4], [3]). At the same time, opposing signs of internal variability may mask or hide the effects of anthropogenic forcing. Thus, in order to more accurately detect the impacts of anthropogenic climate change, the role of internal variability must be accurately assessed.

It is important to note the difference between variability due to innate processes within the climate system and variability due to processes that are natural but not part of the internal oscillatory processes of the climate system. Such “external natural forcings” include processes such volcanic eruptions and solar irradiance changes, which has been called “total” natural variability in previous studies ([5]), but in this thesis will be referred to simply as natural variability. Meanwhile, the innate processes or coupled interactions between the atmosphere, oceans, land, and cryosphere will be referred to as internal variability, although many previous studies have also called this natural variability.

Previous studies have looked at the climate system in the context of internal or natural variability both for the past century and for future projections of climate change. For instance, [6] found differences up to 6°C in projections of temperature over North America for the period 2005-2060 due solely to internal variability. Precipitation showed an even wider spread than temperature. This makes it even harder to forecast future climate projections, as different model results could be due to internal or natural variability along with differences between model physics and parameterizations. Indeed, [3] suggested that at least half of the inter-model spread amongst CMIP3 models for the period 2005-2060 is due to internal variability.

The uncertainty in future climate due to internal variability is often assessed by running large ensembles of climate change simulations. In a large ensemble, every ensemble member is run using the same climate model and the same external forcing but slightly different initial atmospheric conditions. Thus, any differences between the ensemble members are solely due to internal variability. This is because after initial condition memory is lost (on the order of weeks in the atmosphere), each ensemble develops from internal fluctuations characteristic of a random, stochastic process ([7]). Figure 1.1 shows an example from one such large ensemble, the NCAR CCSM 40-member

ensemble. The right panels shows two examples of the spread of temperature and precipitation, one at the grid box collocated with Los Angeles and one collocated with Berlin. The grey lines show the output from each of the 40 ensemble members, whose differences are solely due to internal variability. Since the real world is only a single realization (one of the grey lines), Figure 1.1 shows graphically the obstacle presented by internal variability to climate prediction. The left panels extend the example grid box results to the whole globe, showing the standard deviations of 50-year wintertime trends of temperature and precipitation given by the ensemble members. We can also see geographical differences; temperature variance due to internal variability is greatest in the extratropics, while precipitation variance is greatest in the tropics. While this has been a successful way to estimate the uncertainty due to internal variability, large ensembles are not only computationally expensive but also include their own model biases. [1] demonstrated an alternative approach to calculating the uncertainty in trends due to internal variability. If the internal variability is roughly Gaussian, then the margin of error on a trend can be calculated analytically from two statistics of the unforced climate variability: its standard deviation and its autocorrelation. The analytical model has been tested and found robust by comparing the range of trends found in all 40 members of the NCAR CCSM3 large ensemble of climate change simulations with those predicted by the statistics of its associated pre-industrial control simulation.

1.2 Time of Emergence

A key variable in climate detection is the time of emergence (TOE), defined as the point in time when the signal of climate change “emerges” from the underlying noise of background natural variability (e.g., [3,8–15]). This emergence is said to occur when the new climate state has deviated significantly from a reference climate state for a period of time. It is helpful for anticipating when the impacts of climate change are likely to have significant effects across societies and ecosystems, and can inform risk assessments, mitigation policies, and climate adaptation planning. The IPCC AR5 concluded that it is “virtually certain that internal variability alone cannot account for the

observed global warming since 1951" ([2]). But when are we confident that such a change can be detected? And what metric is used to reach that conclusion?

The IPCC AR5 ([16]) has also noted that there is "no single metric" for estimating TOE, and as such, it is difficult to reach a consensus on the changing state of the climate. But for the most part, TOE is estimated as the first lead time when the anthropogenic signal in climate change exceeds a predetermined factor of the amplitude of the natural variability, often presented as a signal-to-noise problem. In this case, the TOE for a time series $x(t)$ is expressed as:

$$n_{TOE} = \frac{k s_e}{b} \quad (1.1)$$

where n_{TOE} is the time of emergence (the number of time steps when the anthropogenic signal in climate change has "emerged"), k is the required signal to noise ratio of the forced signal to the natural variability (generally between 1 and 3), s_e is the amplitude of the internal (unforced) variability, and b is the linear trend per time step.

Most previous studies of TOE are based on empirical estimates of the first lead time when Equation 1.1 (or a closely related variant) is satisfied. The differences lie in the methodologies used to determine b , s_e , and k . For example, [12], [13], [17], and [15] all define b as the change in the climate state averaged over an ensemble of climate change simulations, where the forced signal is smoothed with different averaging periods. [10] estimates b using a generalized least squares regression model. [14] define b as the linear projection of regional temperatures onto smoothed values of simulated global-mean temperatures. [12], [13], and [14] estimate b from ordinary least squares (OLS) linear regression and prescribe a signal to noise ratio (k) that is an integer factor of the natural variability. [11], [3] and [15] consider various "epoch differences" and a value of k derived from the t-statistic for the difference of means. [18] also considers differences between epochs when assessing the time of emergence, and apply a Kolmogorov-Smirnov test to assess whether two sample epochs are significantly different (i.e., rather than a t-statistic).

The existing literature on TOE provides valuable insight into the point in time when anthropogenic climate change will emerge from natural climate variability on regional spatial scales. But

it also has several shortcomings. The methodologies used to estimate the trend (b in Equation 1.1), the amplitude of the natural variability (s_e in Equation 1.1), and the predetermined signal to noise ratio (k in Equation 1.1) vary widely from one study to the next, which makes it difficult to reproduce and compare different estimates of the TOE. Times of emergence defined on the basis of a fixed signal to noise ratio [14, e.g.,] do not correspond to a particular level of statistical significance. Several existing methods require smoothing the data using a wide range of methodologies. Furthermore, many of the methods are based on empirical rather than analytical techniques. Chapter 3 of this thesis will introduce an alternative model for estimating the lead time required for a linear trend to emerge from natural variability at a predetermined level of statistical confidence.

1.3 Dynamical Adjustment

Given the large uncertainty in climate projections due to natural and internal variability, dynamical adjustment arises as a way to remove circulation-induced temperature variability that obscures the signal of anthropogenic forcing. Although the circulation itself may respond to anthropogenic forcing (e.g., a poleward shift in the annular modes), this response is generally much weaker than the surface temperature response ([3]). Many previous studies have used a variety of methods to examine the role of internal circulation dynamics on surface climate over different regions and time periods. The earliest studies done by [19] and [20] were of teleconnection patterns, in which linear combinations of surface temperature and sea level pressure were used to define teleconnection indices. These studies also correlated temperature and precipitation at various locations around the world and attempted to separate out large-scale variability patterns, defining indices for the Southern Oscillation, the North Atlantic Oscillation, and the South Pacific Oscillation. Later studies attempted to explicitly isolate the contribution to surface temperature from circulation dynamics. [21] estimated the dynamically induced contribution to global-mean temperatures from a cold-ocean-warm-land (COWL) pattern. The COWL pattern, while explaining a substantial amount of the month-to-month variance in hemispheric and global-mean temperatures, cannot be considered to be solely dynamically-induced since greenhouse gases warms the continents more

rapidly than the oceans ([22]). [23] used multivariate linear regression to show that much of the temperature trends in the Northern Hemisphere from the mid 1970s to early 1990s were due to natural variability patterns (rather than anthropogenic forcing). For instance, changes in the North Atlantic Oscillation (NAO) resulted in cooling in the northwest Atlantic and warming across Europe and Eurasia, while teleconnections from ENSO and its feedbacks in the midlatitudes were linked to temperature changes in the Pacific basin and North America. [24] applied principal component (PC) based indices from the sea level pressure field (the NAM and the PNA patterns) to surface air temperature trends over the time period 1958-1999. [25] found that dynamical adjustment in observed Northern Hemisphere winter temperatures over the period 1965-2000 decreased the respective temperature trend by 41% and brought the observed trends much closer to those simulated by climate models used in the IPCC AR4. [26] used maximum covariance analysis between the sea level pressure and temperature fields to estimate the dynamical contribution to global-mean surface temperature over the period 1900-2009, along with the signatures of temperature due to ENSO and volcanic eruptions. The study also showed that the dynamical contribution to temperature is uncorrelated with the global mean SST. [27] applied three variants of partial least squares (PLS) regression on sea level pressure to derive dynamically-adjusted temperature and compared such methods to a PC regression method. [28] applied dynamical adjustment to a large ensemble of model simulations (the CESM-LE) to examine temperature trends over North America for the past 50 years (1963-2012) using a variation of the constructed analog method developed in [29]. The authors concluded that the atmospheric circulation accounted for one-third of the surface temperature trend over all of North America during this time period, and that removing the effects of internal dynamics narrows the ensemble distribution and substantially advances the time of emergence for the forced SAT anomalies due to the smaller amplitude of the noise.

1.4 Thesis Goal

As summarized in this chapter, previous studies have clearly demonstrated the importance of natural and internal variability, as well as the usefulness of a large ensemble to more accurately

quantify their effects. The goal of this thesis is twofold. First, we introduce a simple and novel expression for estimating the lead time required for a linear trend to emerge from natural variability at a predetermined level of statistical significance. The expression is developed from the standard error of the regression, which is widely used in climate research, but as far as we know has not been exploited for the explicit purpose of calculating TOE. We then test the resulting expression in a large ensemble of climate change simulations and extend the analysis to compare with observations. The results demonstrate the robustness of the assumptions that underlie the expression, and make clear its utility for assessing the emergence of linear trends in climate data. Second, we specifically compare three methods prevalent in the climate literature for calculating the contribution to surface air temperature from internal circulation dynamics, using both a prehistorical control run and a large ensemble of climate simulations. The dynamically-induced temperature is subtracted from the original temperature time series, and the ensemble distribution of the resulting residual temperature is compared to that of the original temperature data. The application to a large ensemble allows us to examine both the uncertainty of the dynamically-adjusted temperature results and its comparisons to a “true” forced signal (the ensemble mean), something that would not be possible from only observational data.

Chapter 2 describes the data and methods used in this study. Chapter 3 derives a novel method for time of emergence and discusses its application to climate trends is explored. Chapter 4 explores the results of dynamically-adjusted temperature methods in the context of the large ensemble. Conclusions and an outline for future research are provided in chapter 5.

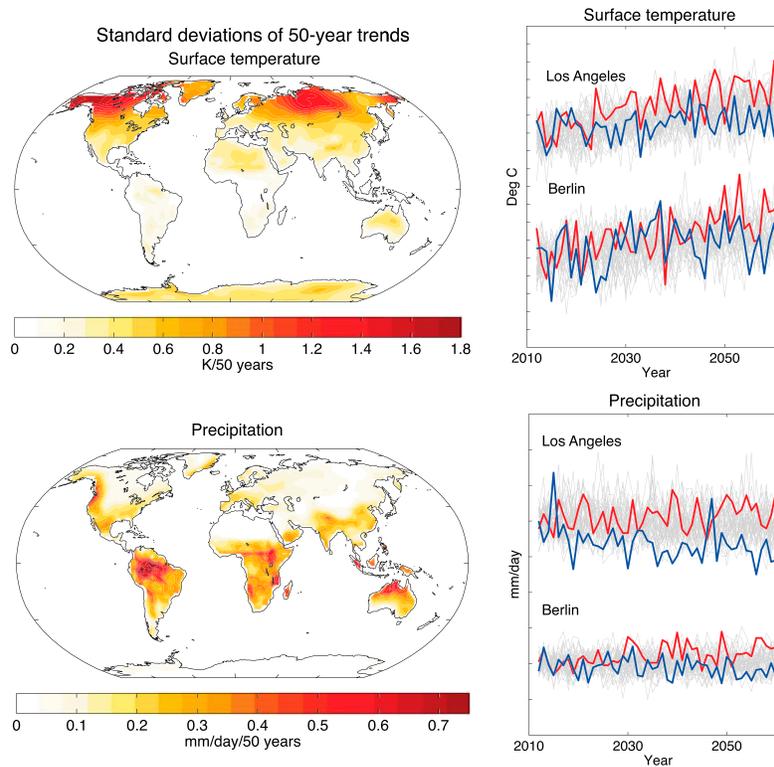


Figure 1.1: (left) The standard deviations of 50 year trends in October-March mean surface temperature (top) and precipitation (bottom) from the NCAR CCSM4 40-member large ensemble. Trends are expressed as the total change over the period 2011-2061, and the standard deviations are over the 40 ensemble members. The right panels show wintertime mean temperature and precipitation for two specific grid boxes: Los Angeles and Berlin. The grey lines show results for all 40 ensemble members; the red and blue lines indicate ensemble members with the largest and smallest trends, respectively. Tick marks at every 1°C and 1 mm day^{-1} . Figure taken from [1].

Chapter 2

Methods

2.1 Data

2.1.1 CESM1 Model

In this thesis, the main way of quantifying natural and internal variability is through the NCAR Community Earth System Model Large Ensemble (CESM-LE). The CESM-LE is a set of 40 climate change simulations at approximately 1° resolution for both latitude and longitude run using a single CMIP5 coupled climate model: the Community Earth System Model version 1 (CESM1), with the Community Atmosphere Model, Version 5 [CESM1(CAM5)]. The CESM1(CAM5) consists of coupled atmosphere, ocean, land, and sea ice prognostic component models. All simulations (consisting of the period 1920-2100) have the same model configuration and external forcings (historical until 2005 and RCP8.5 thereafter) but begin from slightly different initial conditions. The only differences in initial conditions are in the surface air temperature field (on the order of 10^{-14}K), with all other initial conditions held constant across ensemble members ([7]). In this thesis, we use the original 30 simulations released in 2014, along with a 1800-year preindustrial control run of the same model, forced with 1850 carbon dioxide levels. Further details of the simulations can be found in [30] and [7].

2.1.2 Observational Datasets

To examine where the observations fall in the CESM-LE distribution, we use observed temperature data from CRUTEM4, a gridded dataset of monthly near-surface air temperature anomalies over land at a resolution of $5^\circ \times 5^\circ$ ([31]). It is one of the datasets from the Met Office Hadley Centre and was obtained from the Climatic Research Unit at the University of East Anglia. The

data is available from 1950 to the present day, and we use the time period April 1970 to March 2016.

2.2 Statistical Methodology

2.2.1 Linear Regression

The most widely used statistical technique in this thesis is simple linear regression analysis. Given time series x and y , we can write an estimate of y as:

$$\hat{y} = \hat{a}_1 x + \hat{a}_0 \quad (2.1)$$

Equation 2.1 gives the estimate of y from linear regression, modeled as a straight line. \hat{a}_1 is the estimated regression coefficient and \hat{a}_0 is the estimated y-intercept. Both are calculated by minimizing the sum of the squared errors of y , and are given by the following equations:

$$\hat{a}_1 = \frac{\overline{x'y'}}{\overline{x'^2}} \quad (2.2)$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x} \quad (2.3)$$

where overbars denote the mean and primes denote the deviation from the mean.

To derive an analytical method for the time of emergence we apply this linear regression to a first order autoregressive process (Section 3.1). In this case the x time series is time, while the y time series is surface air temperature.

2.2.2 EOF analysis

To look at the dynamically-induced temperature, we also make use of Empirical Orthogonal Function (EOF) analysis, also called Principal Component Analysis (PCA). This was performed

on monthly anomalies of SLP for the winter season (December through February) for the area 20N to 90N. The time mean was subtracted and the data cosine weighted to account for the decrease in area towards the pole. The anomaly data is organized into a matrix X , and the cosine weighted data into a matrix X_w , both with dimensions $[M \times N]$, where M is the time dimension and N is the spatial dimension. The spatial dispersion matrix is calculated as

$$C = \frac{1}{N} X_w X_w^T \quad (2.4)$$

We then perform eigenanalysis on the spatial dispersion matrix, and the resulting eigenvectors are the PC time series. The PC time series are each standardized, and the resulting EOF pattern is calculated by

$$d_i = \frac{1}{M} z_i^T X \quad (2.5)$$

where z_i is the i th standardized PC time series, and d_i is the EOF pattern associated with one standard deviation of the corresponding PC time series.

2.2.3 Correlation and Variance Explained

To assess how well each of our three dynamical adjustment methods explain the variance in surface air temperature, we make use of the correlation coefficient, defined as:

$$r = \frac{\overline{x'y'}}{\sigma_x \sigma_y} \quad (2.6)$$

The linear correlation coefficient r is a value between -1 and 1 and is a measure of the strength and direction of the linear relationship between the variables x and y . The square of the correlation coefficient (r^2) is the percent of the variance in y that is explained by the linear least-squares fit with x .

Chapter 3

Results: Part I

The goal of this chapter is to use the standard error of the regression to develop a simple analytic expression for the lead time required for a linear trend to emerge from natural variability at a predetermined level of statistical significance. The expression is derived from an AR(1) process and then explicitly applied to surface air temperature data as a metric of calculating the time of emergence. We test the usefulness of this metric for assessing the emergence of significant linear trends in the CESM-LE, taking advantage of the ability to calculate natural and internal variability from a large ensemble. Lastly, the expression is applied to observed temperature data over the past 45 years to compare any differences between simulated and observed natural variability as well as where the observations as a single realization fall within the spread of the large ensemble.

3.1 An Analytic Expression for the Lead Time Required for a Linear Trend to Emerge from Natural Variability

Consider the case of a first order autoregressive (AR(1)) time series of length N with a linear trend b imposed upon it such that

$$x(n_t) = bn_t + \alpha x(n_t - 1) + \beta \epsilon(n_t)$$

where $n_t = 1, 2, \dots, N$ is the number of time steps, $x(0) = 0$ by assumption, and ϵ is white noise (independent Gaussian noise with mean of zero and variance of σ_ϵ^2). The parameter α is between 0 and 1 and represents the memory in the time series $x(n_t)$ from one time step to the next.

Here we estimate the trend b using simple linear regression, where \hat{b} is the regression estimator of the trend. The parameter α can be estimated as the lag-one autocorrelation of the time series (r_1). The confidence interval in the total change in $x(n_t)$ over time n_t can thus be expressed as:

$$CI = \hat{b}n_t \pm e$$

where n_t is again the number of time steps and e is the uncertainty in the change in $x(n_t)$ over time n_t given by $\hat{b}n_t$. The margin of error (e) is given in units $\Delta x/(n_t\Delta t)$, where Δt is the time step. The trend \hat{b} is given in units $\Delta x/\Delta t$ so that $\hat{b}n_t$ is the change over the length of the record and has the same units as e .

Following [1], if detrended values of $x(n_t)$ are well-modeled as an AR(1) process, then the margins of error on the linear trend in $x(n_t)$ can be expressed as:

$$e = t_c \cdot n_t \cdot s \cdot \gamma(n_t, r_1) \cdot g(n_t) \quad (3.1)$$

where

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

and

$$g(n_t) \equiv \sqrt{\frac{12}{n_t^3 - n_t}}$$

In Equation 3.1, t_c is the t-statistic for the desired confidence level and s is the standard deviation of the residuals of the regression (i.e., of detrended values of $x(t)$). The expressions for $\gamma(n_t, r_1)$ and $g(n_t)$ account for 1) the effects of persistence on the estimate of s , where r_1 is the lag-one autocorrelation of the residuals, and 2) the standard deviation of the time axis, respectively. Note that Equation 3.1 is simply the standard error of the regression for the case where 1) the predictor is time, and 2) detrended values of the predictand are well-modeled as an AR(1) process (e.g., [1, 32]).

The trend in $x(t)$ is statistically significant when it exceeds its margins of error. Setting $e = \hat{b}n_t$ in Equation 3.1 yields:

$$\frac{(T_{SIG}^3 - T_{SIG})}{12} \cdot \frac{T_{SIG} \left(\frac{1 - r_1}{1 + r_1} \right) - 2}{[T_{SIG} - 2]} = \left(\frac{t_c s}{\hat{b}} \right)^2 \quad (3.2)$$

where T_{SIG} denotes the lead time when the trend in $x(t)$ is statistically significant (in units of time steps). That is: *Given our parameter estimates of \hat{b} , s , and r_1 , T_{SIG} is the estimated number of time steps required for the trend to be statistically significant at the desired confidence level.* The value of T_{SIG} can be trivially calculated given the estimated amplitude of the forced signal (\hat{b}), the amplitude of the natural climate variability (s), the lag-one autocorrelation of the natural climate variability (r_1), and the desired significance level (t_c). It requires no subjective analysis choices (such as the length of the periods used in epoch differences) and no smoothing of the data.

The parameters \hat{b} , s , and r_1 are calculated using the entire length of the time series (n_t), while t_c is a function of n_t . Thus, to solve for T_{SIG} , Equation 3.2 is calculated iteratively at each time step. For instance, a time series with a T_{SIG} of 10 time steps is calculated using the two-tailed 95% value of $t_c = \pm 2.26$, while a time series with a T_{SIG} of 50 time steps is calculating using the respective value of $t_c = \pm 2.01$. We use the entire length of the time series to determine the values of \hat{b} , s , and r_1 . This provides the most accurate estimates given that the linear trend and the Gaussian AR(1) distribution are consistent throughout the entire time period, as calculations using short time series can produce erroneous values. A graphical representation of the calculation of T_{SIG} can be seen in Figure 3.1, denoted as the time step when the lower 95% error bound intersects zero.

Equation 3.2 can be simplified greatly given two conditions: 1) detrended values of $x(n_t)$ are not serially correlated ($r_1 \approx 0$). This condition holds for climate variability at most terrestrial locations on interannual timescales, since there is very little memory in the internal variability of land surface climate from one year to the next (see discussion in [1]). 2) The trend length is at least ~ 10 time steps. In this case, $T_{SIG}^3 \gg T_{SIG}$ and the two tailed t-statistic for the 95% confidence level is $t_c \sim 2$. Applying both conditions yields a simplified version of Equation 3.2 that is suitable for cases where the internal variability is not serially correlated from one time step to the next:

$$T_{95\%} \approx 3.6 \left(\frac{s}{\hat{b}} \right)^{2/3} \quad (3.3)$$

where $T_{95\%}$ is the lead time when the trend in $x(t)$ is statistically significant at the 95% confidence level. Equation 3.3 places Equation 3.2 in a “signal to noise” format that is similar but not identical to that used in many previous studies and provides a useful back-of-the-envelope estimate for T_{SIG} . All analyses in this thesis use the general expression of Equation 3.2 for accuracy.

3.2 Application to Climate Trends

In this section, we test the robustness of Equation 3.2 (T_{SIG}) for assessing the point in time when the signature of anthropogenic warming emerges from the background noise of natural climate variability (i.e., achieves statistical significance) on regional scales. We perform the assessment for land surface temperature changes at individual grid boxes. To do so, we exploit a large ensemble of climate change simulations.

In a large ensemble of climate change simulations, each individual ensemble member may be viewed as a unique realization of “model reality.” Here we test the expression for T_{SIG} using output from the NCAR Community Earth System Model Large Ensemble (CESM-LE). The analyses are based on seasonal-mean values of surface temperature for the Northern Hemisphere (NH) cold (October-March) and warm (April-September) season months over the 1970-2015 period. There are three reasons for choosing this time period: 1) We wish to focus on the period with the largest global warming observed to date [2]; 2) We wish to compare results derived from the CESM-LE with results derived from observations, which are relatively sparse before 1970; and 3) Our analytic expression is based on a linear least-squares fit to the forced signal, which is approximately linear over the selected period (the linear assumption is discussed in more detail in the final section). The simulated trends in global-mean surface temperature from the CESM-LE are not linear over the full simulation period 1920-2100, i.e., the trends increase from roughly zero in the mid-20th century to roughly 0.5 K/decade in the latter part of the 21st century [[7], c.f., Fig. 2]. However, they are approximately linear over the comparatively short 1970-2015 period examined here. We tested the linearity assumption by comparing residual temperature time series derived by subtracting a linear fit to the data with those derived by subtracting second and third order fits to the data. The

higher-order fits do not significantly change the variance explained by the residual time series (see Figure 3.2).

The expression for T_{SIG} is tested as follows. First, we calculate the “empirically-derived T_{SIG} ” as the first time step when 29 out of 30 ensemble members exhibit trends of the same sign as that of the model forced signal in the current and all subsequent time steps. In the context of large ensembles, for a confidence level of 95%, the expression for T_{SIG} (Equation 3.2) should thus correspond to the lead time when 97.5% of all possible realizations of “model reality” exhibit trends of the same sign as the forced signal (for a two-tailed confidence interval). Given that the CESM-LE only consists of 30 members, 29 is the closest approximation to our 95% confidence level. Note that the empirically-derived T_{SIG} does not correspond to a strict statistical quantity and is calculated primarily to explore the robustness of Equation 3.2 in the context of a large ensemble of climate simulations. The additional requirement that 29 out of 30 ensemble members must also exhibit trends of the same sign as that of the model forced signal in *all subsequent* time steps is to control for any false positives in the T_{SIG} results (i.e., a T_{SIG} that has “emerged” but then falls below the 29/30 threshold at a future time step).

Second, we calculate the “analytically-derived T_{SIG} ” at all grid boxes by solving Equation 3.2 for T_{SIG} using: 1) the ensemble mean trends in temperature calculated over the period 1970-2015 (\hat{b}); 2) the standard deviations of the residuals of the regression (i.e., the variability about the long-term trends; s); and 3) the lag-1 autocorrelations of the residuals of the regression (r_1). The ensemble mean trends are assumed to reflect the forced signal in surface temperature. The standard deviation and lag-1 autocorrelation of the residuals are found by 1) detrending the seasonal-mean temperature time series in each of the ensemble members and at each grid box and 2) calculating the pooled standard deviations and ensemble-averaged lag-1 autocorrelations of the residual time series. The resulting values of s and r_1 are assumed to reflect the amplitude and persistence of the model’s natural variability. We note that for short time series (less than 20 time steps) or time series with large memory, using detrended residuals from ordinary least squares can result in erroneous

lag-1 autocorrelation values. In such cases, it is advisable to use a generalized least squares (GLS) estimator of the trend to calculate r_1 .

In principle, the model's natural variability can be isolated using a variety of different methodologies. Figure 3.2 explores four different approaches for estimating the parameter s in Equations 3.1-3.3 using the CESM output: 1) removing a linear fit to the temperature time series in each of the ensemble members (as done in our analyses), 2) removing a second order polynomial (rather than linear) fit to the grid box temperature time series in all ensemble members, thus retaining natural external forcings due to, say, volcanic eruptions and allowing for exponential changes in temperature; 3) removing the grid box *ensemble-mean* temperature time series from the grid box temperature time series in all ensemble members, thus explicitly removing the signals of all forms of forced variability from the ensemble members and allowing for forced variability on a range of timescales; and 4) taking the last 1380 years of the CESM preindustrial control run, in which there are no natural external forcings (e.g., volcanoes or solar irradiance changes). In practice, all four methods yield very similar estimates of internal/natural climate variability and thus very similar estimates of the time of emergence (Figure 3.2).

The four methods have various advantages and disadvantages. The advantages of Method 1 are that a) the residuals of the linear fit correspond directly to the residuals of the regression that form the basis for s in Equation 3.2; and b) a similar method can be applied to observations in the absence of climate model output. The disadvantages are that a) the anthropogenic signal is not necessarily best modeled as a linear trend; b) the linear fits include a component of the internal variability, since stochastic variability includes a trend component; and c) the linear fit does not account for externally forced variability due to, say, volcanic eruptions. Method 2 is similar to Method 1, but has the additional advantage that it allows for exponential changes in temperature. However, the residuals of the second order polynomial fit do not - strictly speaking - correspond to the residuals of the regression that form the basis for s in Equation 3.2. The residuals of Method 3 also do not form the basis for σ in Equation 3.2, but removing the ensemble mean time series arguably reflects the most robust method for removing the variability due to all forms of external

forcing in the CESM-LE, including anthropogenic forcings (e.g., due to increasing greenhouse gases) and external natural forcings (e.g., due to volcanoes). Method 4 is the simplest and most accurate for estimating pure internal variability, as the preindustrial control run does not include either anthropogenic forcings or natural external forcings such as volcanic eruptions (which are included in the forced simulations). Note that the first two methods include both internal climate variability and natural variability due to volcanic forcings and solar irradiance changes, whereas the latter two methods only include internal variability.

Figure 3.3 illustrates the analytically and empirically-derived values of T_{SIG} in NH wintertime surface temperatures at three grid boxes: one from Northern Hemisphere midlatitudes (at a grid box whose node is close to London, UK); one from a region of relatively high temperature variance (located in central Siberia); and one from a region of relatively low temperature variance (close to Jakarta, Indonesia). The sloping black lines in all three panels indicate the ensemble mean trend over the 1970-2015 period at each location. As noted above, the ensemble mean trend is interpreted as the “forced signal” of climate change. The small red dots indicate the trends in surface temperature from all 30 individual ensemble members, where the trends start in 1970 and end on the date indicated on the ordinate axis. The units on all trends are K/length of the record. The dashed lines in all three panels indicate the 95% margins of error on the “forced signal,” where the margins of error are derived from Equation 3.1. Note that the close agreement of the 95% margins of error given by Equation 3.1 (dashed lines) and the large ensemble (red dots) attest to the robustness of Equation 3.1 for estimating the role of natural variability in climate trends (see [1]).

The analytically-derived values of T_{SIG} are calculated at each terrestrial location by inserting the estimated forced signal and natural variability for each grid point into Equation 3.2. For example, in the case of London, the estimated forced signal is $\hat{b} = 0.02$ K/year, the amplitude of the natural variability is $s = 0.6$ K, and the winter-to-winter autocorrelation is not significantly different from zero ($r_1 \sim 0$). Inserting the above values into Equation 3.2 yields $T_{SIG} = 41$ years, or 2011, which by definition is the lead time when the lower bound of the 95% confidence lev-

els intersects zero (the intersection is marked by the vertical blue line in Figure 3.3). The forced signal and natural variability both vary from one location to the next in Figure 3.3, but in general the latter dominates the variations in T_{SIG} . For example, T_{SIG} is longer over Siberia where the interannual temperature variance is much larger ($s = 2$ K), but shorter over Indonesia where the interannual temperature variance is relatively small ($s = 0.2$ K). The inverse relationship between regional temperature variance and the emergence of the forced signal has been noted extensively in previous works (e.g., [11, 13]). The key point in Figure 3.3 is that the expression given in Equation 3.2 for T_{SIG} clearly provides a simple and robust estimate of the first lead time at which 95% of the realizations of model “reality” (as given by individual ensemble members) exhibit warming.

Figure 3.4 shows the results for a similar test at all terrestrial grid boxes during the NH winter and summer seasons. The top panels indicate the “empirically-derived” values of T_{SIG} found by empirically calculating the lead time when 29 of the 30 ensemble members exhibit warming in the current and subsequent time steps. The bottom panels in Figure 3.4 indicate the analytically-derived values of T_{SIG} from Equation 3.2 (very similar results are derived for Equation 3.3, since the lag-one autocorrelation of seasonal-mean surface temperature is not significantly different from zero at most terrestrial grid boxes). Warm colors indicate lead times of 45 years since 1970 (e.g., times of emergence less than 2015). White denotes lead times that exceed the analysis period (TOE beyond 2015), while grey denotes oceans and any missing data. The top and bottom panels of Figure 3.4 are very similar, as further shown in Figure 3.5. Here, we compare TOE results from the empirical and analytical methods using a scatterplot. The 1:1 ratio line, given in red, represents all locations where the TOE calculated from the empirical method is identical to the TOE calculated from the analytical method. Each blue dot represents the TOE calculated from both methods at a single grid box. Overall, the two methods have a good agreement about the 1:1 ratio line, although there is a large spread in the TOE results.

The strong similarities between the top and bottom panels in Figure 3.4 are important. They suggest that the lead time given by Equation 3.2 provides a reliable estimate of the geographical pattern of detection time - the time at which virtually all possible realizations of model “reality”

indicate trends of the same sign as the forced signal. They also support the assumptions that underlie Equation 3.2, e.g., that the natural variability is sufficiently Gaussian and the forced signal sufficiently linear to warrant use of the standard error of the regression. As noted in numerous previous studies (e.g., [11, 13, 14]), the forced signal in surface temperature emerges earliest in regions where the variance is smallest, i.e., the tropics during all seasons and the extratropics during the warm season months.

The top panels in Figure 3.6 examine analogous results, but for the case where 1) the estimated forced signal (\hat{b}) is again derived from the CESM-LE but 2) the natural variability (s and r_1) is derived from observations of surface temperature from the HadCRUT4 dataset. The HadCRUT4 data are obtained from the Climatic Research Unit at the University of East Anglia and are analyzed on a 5x5 degree grid for the time period January 1970 to September 2015. The advantage of using observations to estimate the natural variability is that - by definition - they best reflect the variance of the "real-world." The disadvantages are that: 1) the observed record may be too short to fully sample variability on decadal timescales; and 2) the observed record includes missing data and may include residual errors that influence estimates of the observed variability. As in the case of Figure 3.4, \hat{b} is defined as the ensemble-mean trend from CESM over 1970-2015. In contrast to Figure 3.4, s is found by 1) detrending the *observed* wintertime-mean surface temperature data at each grid box, and 2) calculating the standard deviation of the resulting time series. Note that the detrending methodology is identical to that applied to individual ensemble members (except for the pooling) in Figure 3.4.

Results based on the amplitude of observed natural variability are similar but not identical to those based on the natural variability displayed by the CESM-LE. Regions of strong agreement between the top panels in Figure 3.6 and bottom panels of Figure 3.4 correspond to areas where the variability in the CESM-LE closely corresponds to that in the observations. Regions where the top panels in Figure 3.6 and bottom panels in Figure 3.4 are notably different point to areas where differences between the simulated and observed natural variability lead to differences in the lead time when surface warming emerges in a statistically significant sense. These differences can

be seen more clearly in the top panels of Figure 3.7, which show the ratios of the amplitudes of natural variability derived from the CESM-LE to those derived from observations. For the most part, the CESM-LE overestimates the variance of surface temperature and thus underestimates the times of emergence over much of the Northern Hemisphere midlatitudes.

The bottom panels in Figure 3.6 show analogous results to those in the top panels, but in this case both the natural variability and the forced signal (the linear trends) are estimated from observations. That is: the forced signal (\hat{b}) is defined as the linear trend calculated from observations over the period 1970-2015, and the natural variability (s) is found in an identical manner to the top panel. The observed trends reflect only one realization of reality and are therefore noisier than the model ensemble-mean trends, particularly over regions of large temperature variance such as the Northern Hemisphere midlatitudes during winter (e.g., [6]). Nevertheless, the resulting lead times are interesting in that they provide a purely observational estimate of the lead time when the observed warming emerges from the observed natural climate variability in a statistically significant sense.

The differences between the upper and lower panels in Figure 3.6 arise solely from differences between trends from the CESM-LE ensemble mean and the observations. The CESM-LE ensemble mean trends from 1970-2015 are weaker than those derived from the observations over much of the tropical land areas, Europe and East Asia during summer (see bottom panels of Figure 3.7). Hence, the purely observational lead times in these regions are shorter than those derived from the ensemble-mean trends.

Figure 3.8 explores whether the TOE estimates obtained solely from observations lie outside the range of TOE estimates derived from all individual ensemble members. To address this question, we calculated the TOE at all grid boxes and for all ensemble members using the individual ensemble member trends and detrended standard deviations as estimates of the forced signal and natural variability (i.e., we treated output from individual ensemble members as we treated the observations in the lower panel of Figure 3.6). Interestingly, the observed TOE estimates given in the

bottom panel of Figure 3.6 lie within the 95% bounds on TOE estimates derived from individual ensemble members over 95% of all land areas (Figure 3.8).

3.3 Discussion

The standard error of the regression is widely used in climate research. But to the best of our knowledge, it has not been explicitly used to develop an expression for the time of emergence of anthropogenic climate change. The resulting expression for T_{SIG} provides a novel and general “rule of thumb” for assessing the lead time when anthropogenic climate change will emerge from natural climate variability. The methodology has some disadvantages relative to existing methods, e.g., it assumes that the natural variability is Gaussian, which is not required in existing metrics based on the Kolmogorov-Smirnov test (e.g., [18]). However, it also has several key advantages:

1) The expression for T_{SIG} given by Equation 3.2 (and Equation 3.3 for the case where the data are not serially correlated) indicates the lead time when the forced signal of the trend has emerged in a statistically significant sense. Some previous studies explicitly consider TOE in the context of statistical significance (e.g., [6, 11, 15]). But others consider it in the context of specific values of the natural variability. For example, consider the case of TOE defined as the first lead time when the forced signal exceeds two times the amplitude of the natural variability (e.g., one of the criteria outlined in [14]). At the grid box close to London, the TOE for $k = 2$ in Equation 1.1 occurs at a lead time of 74 years, which is more than three decades longer than the point in time when the trend is significant (Figure 3.3). Similarly large differences are found throughout much of the extratropics (Figure 3.9).

2) The expression for T_{SIG} exploits linear regression instead of epoch differences to estimate the linear trend. For example, [11], [3], and [15] all consider statistical significance when assessing the time of emergence, but consider the differences in means between epochs of various lengths rather than linear trends. The distinction is important. Linear regression uses all of the data in a time series, while epoch differences only take data from the beginning and end of the time series. Additionally, the variance of the epoch difference estimator varies greatly depending on the length

of the epoch used, and is always larger than the variance of the linear trend estimator for AR(1) time series with lag-1 autocorrelations less than about 0.85 ([33]). Thus, for all time series with a lag-1 autocorrelation less than 0.85, we believe the linear regression estimator to be preferable to epoch differences.

3) The expression for T_{SIG} is not subject to a multiple testing problem. Many previous methods of calculating TOE have relied on stepping through continuous time steps and defining the TOE as the first time step when the criteria is met (e.g., the first time step when the forced signal exceeds two times the amplitude of the natural variability). This sequential testing increases the rate of Type I errors (false positives in the results).

4) The expression for T_{SIG} can be solved analytically and requires no additional modifications to the data. Hence the resulting estimate of TOE can be easily reproduced from one study to the next, and readily compared across different model configurations and forcing scenarios.

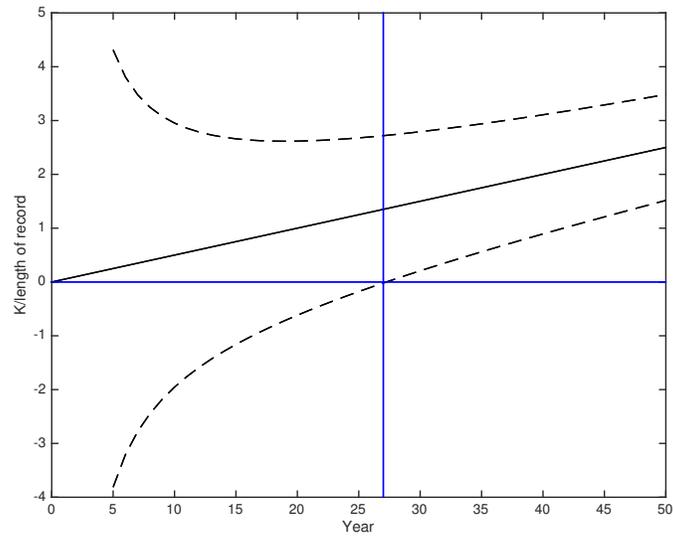


Figure 3.1: Schematic showing the TOE as calculated from Equation 3.2. The black solid line shows the estimated trend (\hat{b}), and the dashed lines indicate the predicted ranges of trends calculated from Equation 3.1. The blue vertical line indicates the lead time when the forced trend is statistically significant, which is the point when the lower dashed line (lower 95% bound) intersects the zero line.

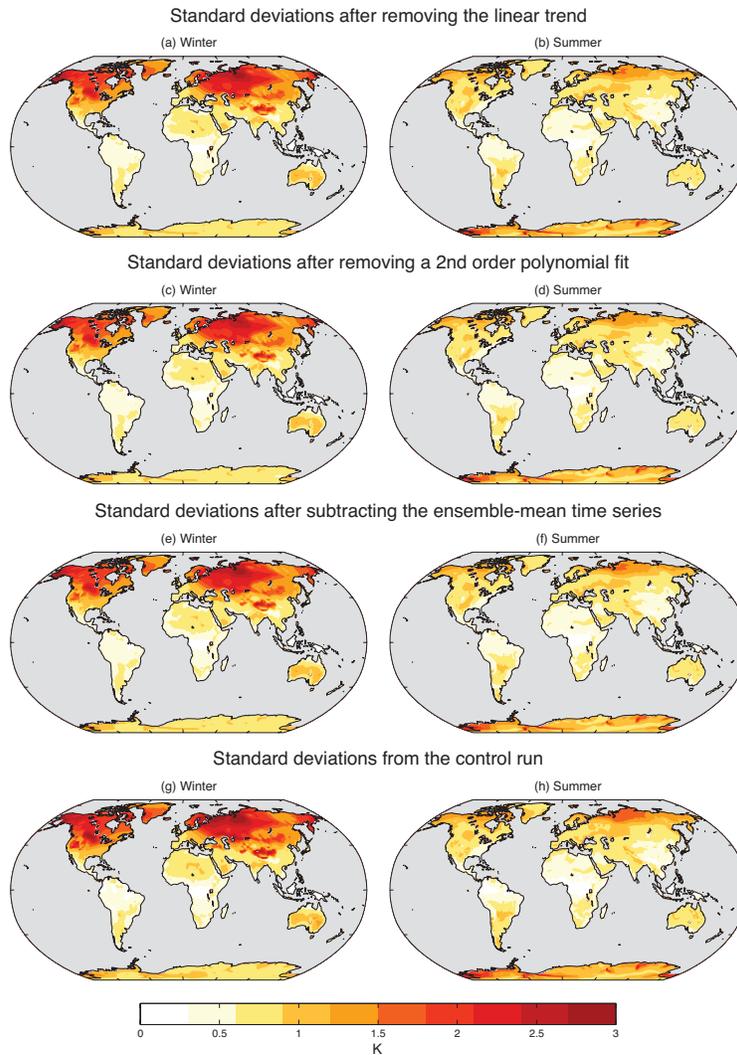


Figure 3.2: Comparisons of the standard deviations calculated from the CESM model over the period 1970-2015 using four different methods to remove the long-term forced signal. Panels (a) and (b) show the pooled standard deviations after removing the linear trend from all grid point time series in all ensemble members (as used for Figure 3.4c, d); panels (c) and (d) show pooled standard deviations after removing a 2nd order polynomial fit from all grid point time series in all ensemble members; panels (e) and (f) show pooled standard deviations after removing the ensemble mean time series from all grid point time series in all ensemble members; and panels (g) and (h) show standard deviations from the last 1380 years of the control run. Grey denotes ocean regions.

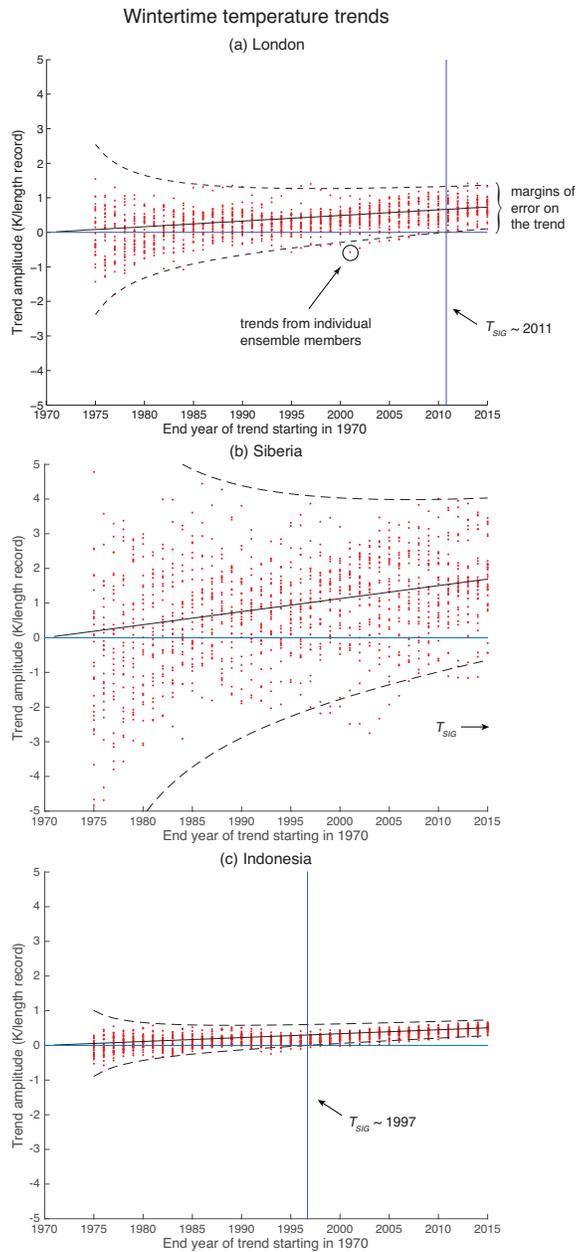


Figure 3.3: Trend amplitudes for modeled surface temperature at the grid box collocated with (a) London, UK, (b) central Siberia, (c) Jakarta, Indonesia, using CESM-LE output. The red dots indicate individual trends from all 30 ensemble members and the dashed lines indicate the predicted ranges of trends found by applying Equation 3.1 to the statistics of the model internal variability. The blue vertical line indicates the lead time when the forced trend is statistically significant as per Equation 3.2. See text for details.

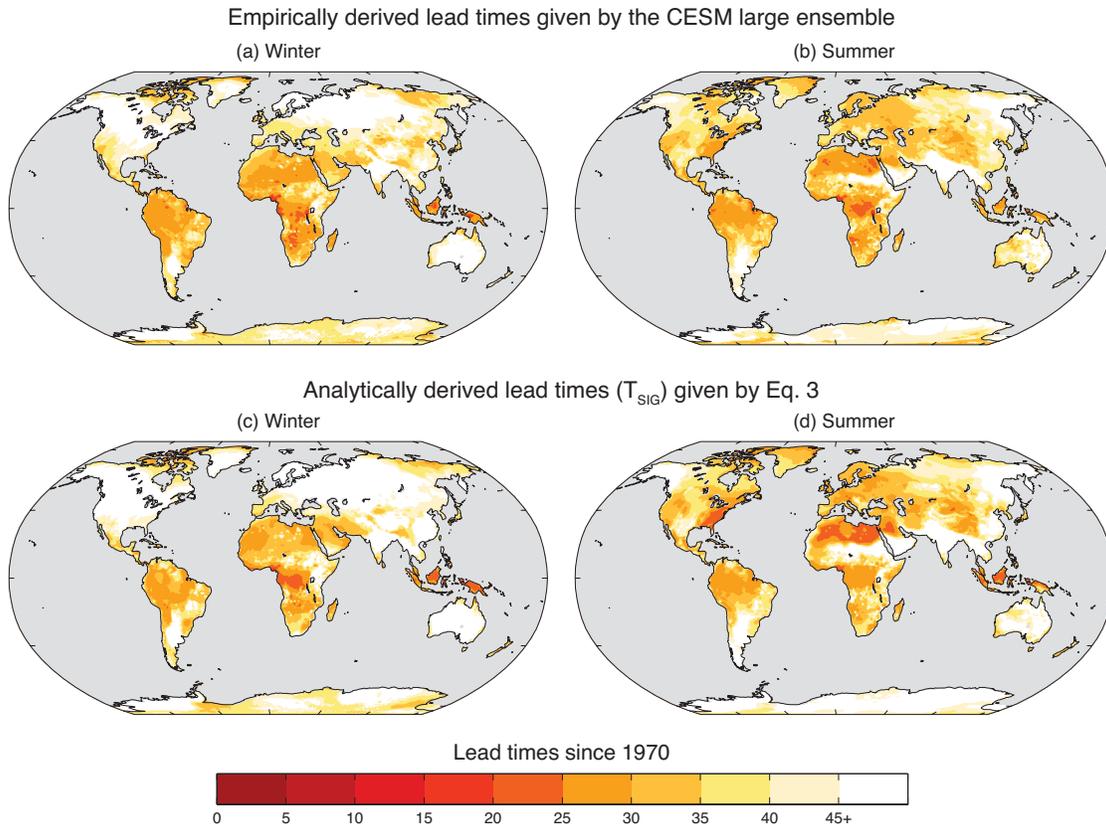


Figure 3.4: Using CESM-LE output to test Equation 3.2. The top panels show the “empirically derived” lead times when the trends emerge from natural variability, calculated as the time step when 29 out of 30 ensemble members exhibit positive trends in the current and all subsequent time steps. The bottom panels show the “analytically derived” lead times (T_{SIG}) derived by applying Equation 3.2 to the model natural variability. Winter corresponds to the October-March means; summer to the April-September means. Note that all lead times beyond the limit of the analysis period (45 years or 2015) are white. Grey denotes ocean regions.

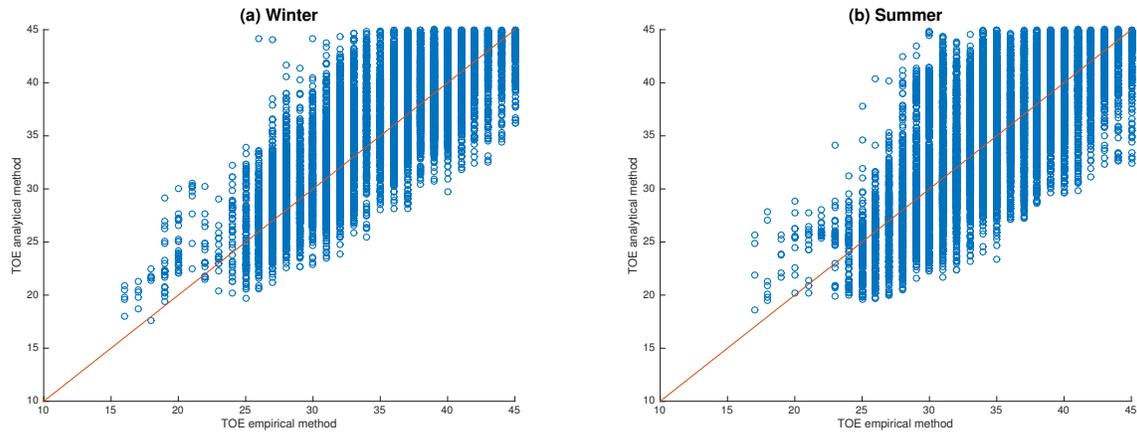


Figure 3.5: Scatterplots showing the relationship between the “empirically-derived” values of T_{SIG} and the “analytically-derived” values of T_{SIG} for (a) the winter season, and (b) the summer season. The good agreement about the 1:1 ratio (represented as the red line) shows that both methods give similar distributions of T_{SIG} .

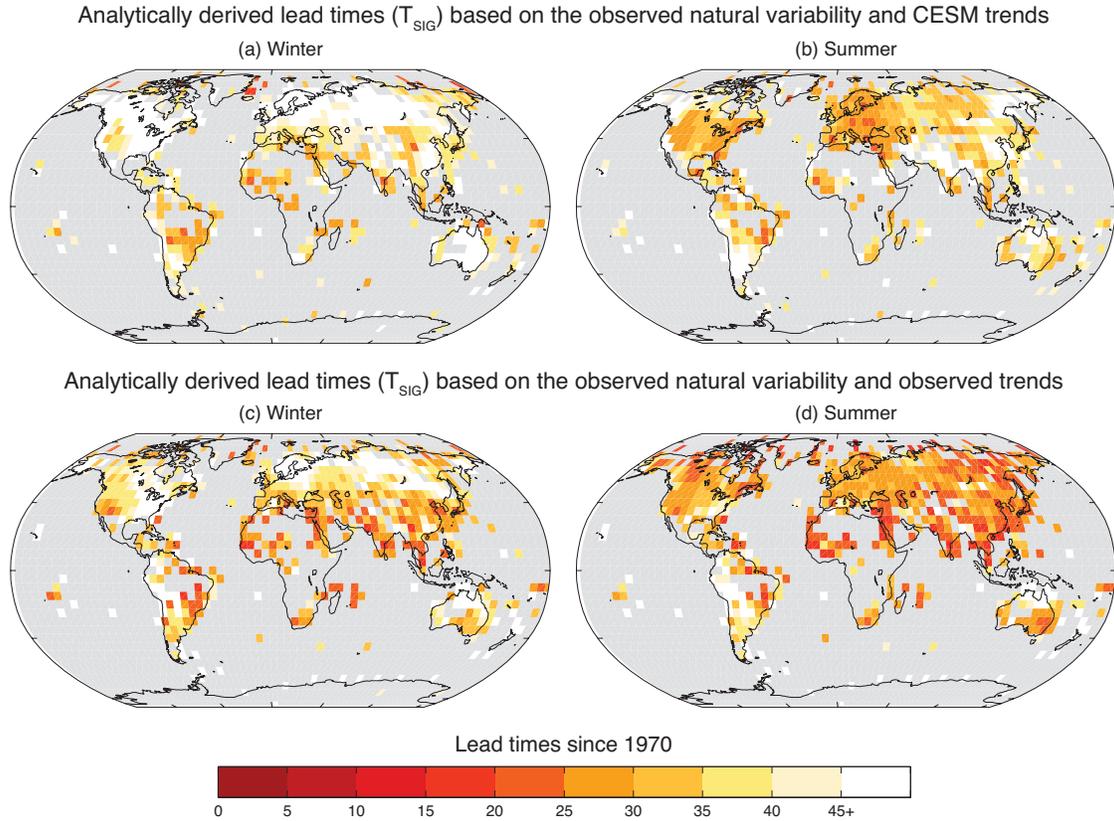


Figure 3.6: As in the bottom panels of Figure 3.4, but for lead times calculated by applying Equation 3.2 to the (top) ensemble mean trends from CESM-LE and the observed natural variability and (bottom) linear trends from the observations and the observed natural variability. The observed natural variability is estimated in both panels as the standard deviation of the detrended data. The observations are used over the period 1970-2015. White denotes lead times beyond the limit of the analysis period (larger than 45 years), while grey denotes oceans and any regions of missing data. See text for details.

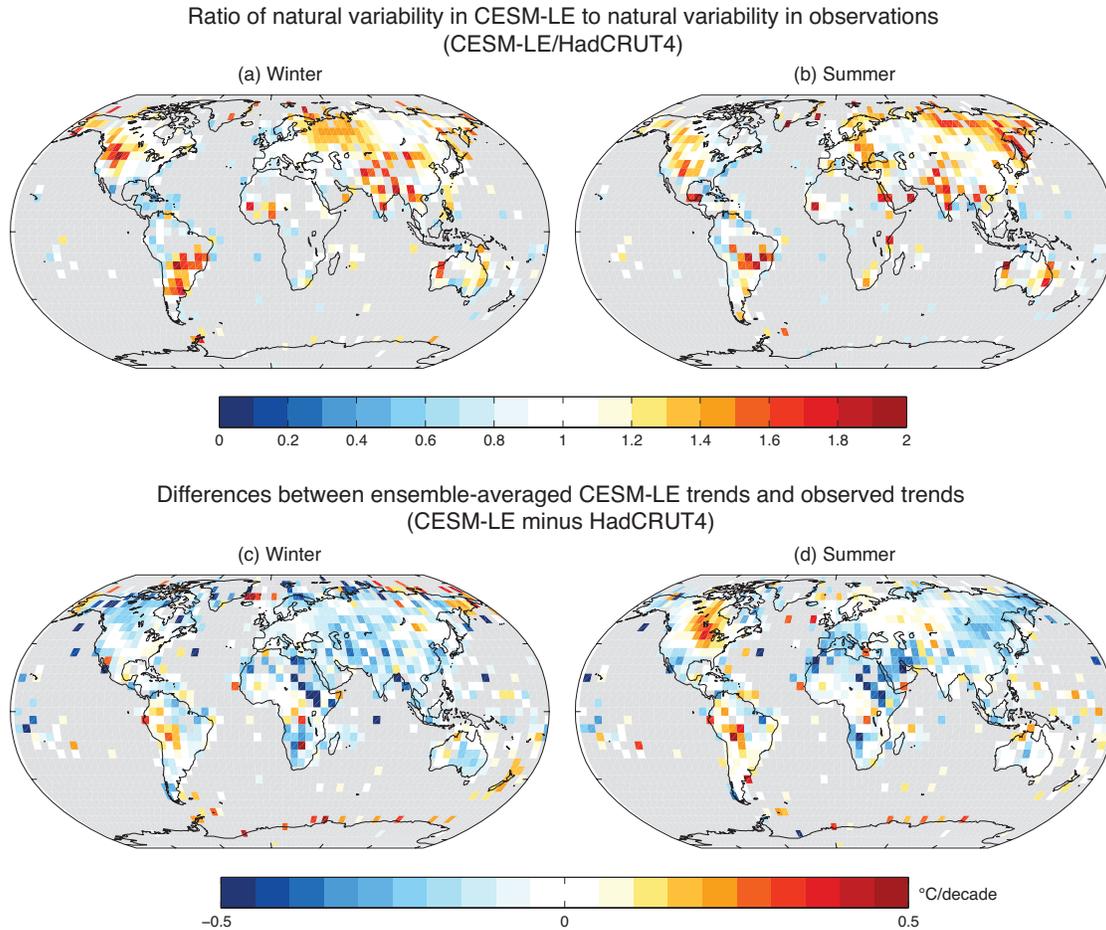


Figure 3.7: Top panels show the ratio of amplitudes of natural variability from the CESM-LE to that from observations for (a) winter (October-March) and (b) summer (April-September). Warm colors denote regions where the amplitude of natural variability is larger in the model for the 1970-2015 period, while cool colors denote regions where the amplitude of natural variability is larger in observations for the same period. Bottom panels show differences between the 1970-2015 ensemble-averaged trends from CESM-LE and trends from HadCRUT4 observations for (c) winter and (d) summer. The CESM-LE trends were used in calculating the lead times in the top panels of Figure 3.6, while the observed trends were used in calculating the lead times in the bottom panels of Figure 3.6. The predominance of cool colors for both seasons indicate that observed trends from 1970-2015 were larger than the simulated ensemble mean trends over the same period. Grey denotes oceans and any regions of missing data.

Comparison of TOE between HadCRUT4 and CESM large ensemble

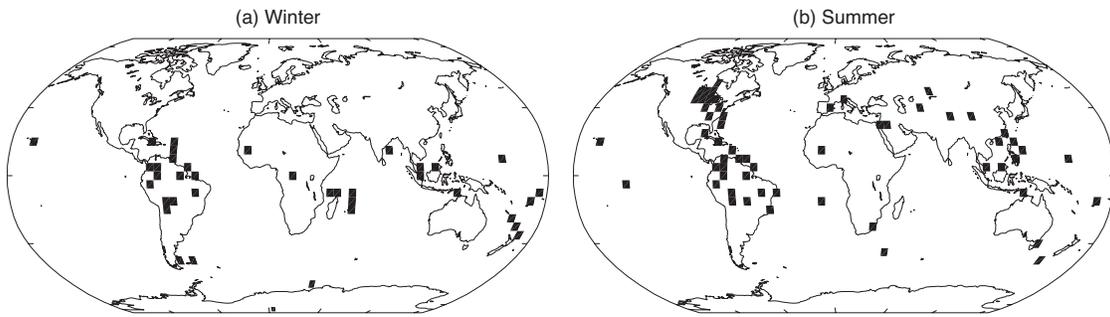


Figure 3.8: Grid boxes where TOE calculated from HadCRUT4 surface temperature observations fall outside the 95% bounds on TOE calculated for individual ensemble member trends and standard deviations. Only 5% of the observed TOE estimates lie outside the bounds given by the individual ensemble members.

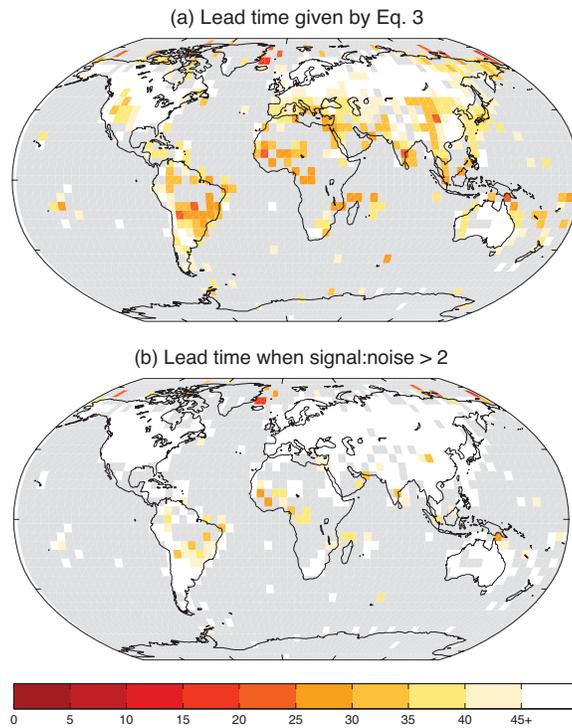


Figure 3.9: Comparison between lead times calculated using Equation 3.2 (top) and Equation 1.1 where $k = 2$ (bottom). In both cases, the forced signal is given as the ensemble mean temperature trends over 1970-2015, and the natural variability as the detrended observed interannual standard deviation. Grey denotes oceans and any regions of missing data. Note the top panel is reproduced from Figure 3.6a.

Chapter 4

Results: Part II

As demonstrated in the previous chapter, natural variability can make it difficult to identify the signal of anthropogenic forcing, especially in the extratropics where the variance in surface climate is largest. In this chapter, we examine various methods of quantifying and removing the dynamical contribution to internal variability in surface temperature trends. As introduced in Chapter 1, dynamical adjustment seeks to determine the role of temperature variability that is solely due to internal atmospheric circulation changes, all other factors being equal. The overall model for dynamical adjustment is as follows:

$$T = T_{fit} + \epsilon$$

where T is the surface temperature time series at a single grid box, T_{fit} is the temperature predicted by internal circulation dynamics, and ϵ is the error (or the non-dynamical contribution to temperature variability, i.e., all other contributions to natural variability as well as the radiative signal). We will specifically compare three dynamical adjustment methods used in previous studies by applying these methods to both the CESM preindustrial control run and the CESM large ensemble. The control run is useful because the only variations are due to internal variability, which allow us to separate the dynamical contribution from the total internal variability. The large ensemble simulations provide a simple and accurate way to test and quantify our results, since 1) there are 30 different realizations of model “reality,” and 2) we have a good estimate of the model forced signal (the ensemble-mean). Our study focuses on the Northern Hemisphere extratropics (20N to 90N) during the boreal winter months (December through February, or DJF), as the amplitude of internal climate variability is strongest during this period. All analyses are on monthly mean surface air temperature, with monthly mean sea level pressure (SLP) being used as a proxy for the circulation.

4.1 Comparing Methods with the CESM Control

The CESM preindustrial control run consists of 1800 years and has constant 1850 carbon dioxide levels. Because the control run does not include either anthropogenic forcings or natural external forcings (e.g., volcanic eruptions or solar irradiance changes), the only variability present within this run is internal variability. This makes the control run an easy way to compare different methods of estimating dynamical adjustment. In addition, the long simulation period allows us to accurately quantify the role of circulation dynamics in surface temperature variability. We first estimate the signature of stochastic circulation variability in surface temperature at a given grid box in the following ways: 1) As the linear least squares best fit of the leading PC time series of the SLP field to the grid box surface temperature (principal component regression, or PCR); 2) As the expansion coefficient time series of the SLP pattern most closely related to the grid box surface temperature, which is equivalent to Maximum Covariance Analysis (MCA) between the SLP field and the grid box temperature time series; and 3) Via the circulation analog method, as recently exploited in [28]. We first discuss the PCR method.

Figure 4.1 shows both SLP and temperature regressed onto the first 10 PCs of DJF SLP anomalies from the CESM control run. The SLP patterns of variability (i.e., EOF patterns) are contoured, while their corresponding temperature patterns associated with one standard deviation anomaly of the SLP are shaded. Positive (negative) SLP anomalies are shown in solid (dashed) lines, while positive (negative) temperature anomalies are shown in warmer (cooler) colors. The panels (especially PCs 1-3) show a strong temperature correspondence over the landmasses with the largest SLP variability. This shows that circulation dynamics have a large effect on both local and regional temperature. One caveat of the PCR method is that it is highly dependent on the number of PCs selected. In the case of the CESM-LE control run, we have 1700 years of data (first 100 years thrown out for spinup), which means 5100 timesteps when only looking at DJF months. If we use all 5100 PCs, it would explain 100% of the surface temperature data, leaving us with no residual error and making this method useless. If we choose too few PCs, we risk not capturing the entirety of the atmospheric circulation effects. Here we choose to use the first 10 EOFs as they explain a

large majority of the variance in the SLP field (90% in the CESM control run), and as a method of comparison based on previous studies (e.g., [26], [27]). However, there are other studies which also use a different number of EOFs for their dynamical adjustment calculations (e.g., [34], who only use the first 5 EOF patterns).

To estimate the total temperature contribution from internal circulation dynamics, we use multiple linear regression, where the predictors are the first 10 PCs. This is written as

$$T_{fit}(x) = \alpha_1(x) * PC_1 + \alpha_2(x) * PC_2 + \dots \alpha_{10}(x) * PC_{10} \quad (4.1)$$

where x is grid box and T_{fit} is the contribution to temperature from the combined effects of the first 10 PCs of SLP. Since each PC is orthogonal to the rest, there are no partial correlations between predictors, and thus each α_i can be calculated as the regression coefficient between the temperature time series and each PC time series at each grid box.

Figure 4.2 shows an example of this decomposition of temperature at a single grid box in central Siberia. The top time series (blue line) is the monthly DJF anomaly time series at the Siberia grid box from the CESM control run. The middle time series (red line) is the contribution of temperature from the first 10 PCs at this grid box, calculated as T_{fit} from Equation 4.1. This T_{fit} time series explains around 56% of the variance in the original (top) temperature time series. The bottom time series (orange line) is the residual time series ($T_{res} = T - T_{fit}$), obtained as the difference between the top two time series. We then apply this method to find a T_{fit} and T_{res} for each grid box in the Northern Hemisphere extratropics.

We next apply the MCA method to find the SLP hemispheric pattern that is most strongly coupled to temperature at each grid box. This is calculated as

$$A(i, x) = \frac{\overline{SLP(x, t)'T(i, t)'}}{\overline{T(i, t)'}} \quad (4.2)$$

where x and i are space (grid boxes) and t is time. Thus, at each grid box i , we have a temperature time series $T(i, t)$. $A(i, x)$ is then the regression map showing the hemispheric pattern of SLP most

strongly related to that grid box temperature time series. An expansion coefficient time series at each grid box can then be calculated by projecting SLP onto the respective regression maps. This can be written as

$$X_{ec}(i, t) = SLP(x, t) * A(i, x) * \cos(\theta) \quad (4.3)$$

where X_{ec} is the expansion coefficient time series and θ is the latitude at each grid box x . The dynamical contribution to temperature is then calculated as

$$T_{fit} = \alpha X_{ec} \quad (4.4)$$

where T_{fit} is analogous to the one in Equation 4.1, but is now calculated via the MCA method. α is solved as the regression coefficient between the original temperature time series and the expansion coefficient time series X_{ec} .

The same grid box in central Siberia from Figure 4.2 is shown in Figures 4.3 and 4.4. Following the MCA method, the SLP (Figure 4.3a) and temperature (Figure 4.3b) patterns most strongly related to the Siberia grid box temperature time series are calculated using Equation 4.2. Unsurprisingly, the temperature regression map shows an inverse relationship between the regression coefficient and the distance from the grid box, with the strongest values nearest the grid box. The SLP regression map shows a strong negative relationship in the region of northern Siberia extending towards the North Pole, and a positive relationship with SLP over western Europe. Figure 4.4 is analogous to Figure 4.2, showing surface temperature over the same Siberia grid box decomposed into an internal dynamical component and a residual component, with the dynamical component now calculated by fitting the expansion coefficient time series from MCA to the temperature time series (Equation 4.4). For this method, the variance in temperature explained by T_{fit} is only 40%, lower than that obtained by using the first 10 PCs. This also means that the residual time series (T_{res}) comprises 60% of the variance of the original temperature time series. The time series in Figures 4.2 and 4.4 show that: 1) circulation dynamics play a large role in temperature variability

over central Siberia, and 2) dynamical adjustment drastically decreases the variance in the residual temperature time series.

We now extend our results from both the PCR and MCA method to all grid boxes in the Northern Hemisphere extratropics. Figure 4.5 compares the percentage of temperature variance explained by internal circulation dynamics using these two methods. Figure 4.5a shows the variance in monthly surface air temperature that can be explained by the first 10 principal components (PCs) of SLP. The spatial average over the Northern Hemisphere is around 40%, but in some regions the percentage is as high as 80%. Figure 4.5b shows analogous results, but for the MCA method. The two plots show similar regions of larger variance (e.g., over central Siberia, central and western Canada), but overall the PCR method yields a larger fraction of variance explained. Over the entire NH extratropics, the percent variance in temperature explained by dynamics calculated from the MCA method is around 32%, which is a 20% decrease compared to the variance explained by the first 10 PCs. This is perhaps not surprising because, as noted earlier, the PCR method is highly dependent on the number of PCs chosen. The variance explained from the MCA method is about the same as the PCR method using 7 PCs. The additional PCs allows us to account for a large fraction of the variance in temperature, and thus we focus on the PCR method for further analysis in the large ensemble.

We also extend our analysis to examine how well the PCR method works to reduce the uncertainty in 50 year trends from the control run. First, we take the 1700-year monthly DJF anomaly temperature data from the control run and split it into 34 chunks of 50-year periods. We then form 3-month winter averages from the DJF monthly anomalies, and then calculate the linear trend over each 50-year period using ordinary least squares regression. The top panel of Figure 4.6 shows the standard error on the 34 chunks of 50-year trends from the surface air temperature DJF anomalies (left) and the interannual standard deviation of wintertime temperature over the 1700-year period (right). Based on the results of [1], the two plots should look very similar, as evident here.

We can then compare the variability of the original temperature to those of T_{fit} and T_{res} as calculated via the PCR method. T_{fit} and T_{res} of the control run are calculated as in Equation 4.1

and then each split into 34 chunks of 50-year periods. We form 3-month winter averages and then calculate the linear trend in the same manner. The middle and bottom panels in Figure 4.6 show the corresponding plots (the standard error of 50-year trends from the control and interannual standard deviation from the control) for T_{fit} and T_{res} from the control run. Since the control run has no anthropogenic forcing, all trends are due to internal variability. By using 50 year periods we are able to capture these trends from internal variability and circulation dynamics (which would be ~ 0 if we used all 1700 years) and examine the ranges of such trends. The similarities between these figures (each left panel with each right panel) affirm the results from [1], but they also give us insight into the dynamically induced temperature for the control run. The T_{fit} trends have their largest variances over land and closer to the pole, especially over central and northern Siberia and western Canada and Alaska. This suggests that circulation dynamics play a large role in those areas and thus induce large variations in overall temperature trends (this is consistent with areas of high r^2 values in Figure 4.5), but a minor role over the oceans and regions south of around 40N. In contrast with the original temperature trends, the standard error on T_{res} trends show: 1) much smaller amplitude across the entire hemisphere, and 2) a much closer range of errors between individual regions. Thus, by removing the effects of circulation dynamics on temperature, the range of uncertainty in the surface temperature trends decreases noticeably.

4.2 Applying Dynamical Adjustment Methods to the Large Ensemble

Studies that have examined dynamical adjustment in a single model or in observations provide valuable feedback in showing the contribution of circulation dynamics to surface temperature. For instance, [23] observed that the temperature trends of the 1970s to 1990s were mostly due to internal dynamics rather than anthropogenic forcing. However, these studies cannot estimate the forced trend given only one realization. To quantify the role of circulation dynamics in internal variability, we extend our dynamical adjustment methods to the CESM large ensemble simulations for the period 1955-2004. This time period was chosen because the CESM-LE is run with historical

forcings until 2005, which minimizes any uncertainty due to future emission scenarios. We first form 3-month winter averages for temperature at each grid box in each ensemble member for the period 1955-2004. Then for each ensemble member we calculated the trend across the 50 year period. These temperature trends are shown in figure 4.7 for all 30 ensemble members. Note that the differences among ensemble members are solely due to internal variability.

For each ensemble member, we calculate the first 10 PCs of monthly DJF anomaly SLP data from 1955-2004 and fitted the temperature time series of that ensemble member to its corresponding first 10 PCs. Residuals are obtained to result in 30 different T_{fit} and T_{res} time series. The temperature timeseries are again DJF averaged to form 50 years of 3-month winter temperatures, and trends are then calculated for each ensemble member. Figure 4.8 shows the 50 year winter trends of the residual temperature time series for each of the 30 ensembles. This can be directly compared, ensemble member by ensemble member, to Figure 4.7. When we remove the effects of circulation on temperature via the PCR method, the spread of the temperature trends across the 30 ensemble members is greatly decreased. The trends are more similar between individual ensemble members, and their amplitudes are decreased as well.

The reduction in the spread of the trends afforded by the PCR method can be quantified by calculating the standard error in the trends across the 30 ensemble members, as shown in Figure 4.9. This quantifies the variability amongst the 30 ensemble members shown in Figures 4.7 and 4.8. The first panel shows the standard error on 1955-2004 winter temperature trends from all 30 ensemble members (i.e., the variance of the ensemble results in Figure 4.7), the second panel shows the standard error on the T_{fit} trends as calculated via the PCR method, and the third panel shows the standard error on the T_{res} trends (i.e., the variance of the ensemble results in Figure 4.8). These are analogous to the left panels from Figure 4.6. Two key points can be gained from Figure 4.9: 1) The error on the trends of T_{res} is much smaller than the error on the trends of the original temperature data. This general pattern could be seen from Figures 4.7 and 4.8, but Figure 4.9 allows a direct mathematical comparison across the 30 ensemble members and for all grid boxes. 2) The standard errors of the trends from the large ensemble look very similar to their respective

values from the control run, even though the control run is only a single simulation. Thus, the role of the circulation in surface temperature and its variability amongst individual ensemble members can be estimated simply from a coupled control run, instead of needing to run a large ensemble of simulations.

Given that the uncertainty in T_{res} is reduced compared to the uncertainty in the original temperature data, how many ensemble members do we need to get a good estimate of the forced signal? We have previously defined the “true” forced signal as the ensemble mean across the 30 ensemble members. Figure 4.10a shows the spatial correlation between 50 year temperature trends of individual ensemble members and the ensemble mean for the original temperature data. The x-axis shows the number of ensemble members used for that calculation. For example, for 5 ensemble runs, we randomly select 5 of the 30 ensemble members, average their trends, and calculate the spatial correlation between that average and the ensemble mean 50 year trend. We then do this random selection 100 times. The blue line shows the average correlation of the 100 Monte Carlo runs, while the red dots show the distribution given by the 100 runs. By construction, at 30 ensemble members the correlation is 1. As expected, the uncertainty between ensemble members decreases as more ensemble members are used. For one ensemble run, the correlation between any one single randomly selected ensemble member and the ensemble mean (the true forced signal) ranges from 0.25 to 0.73. The real world, being only one realization, is simply one dot within this range.

Figure 4.10b shows analogous results, but now for T_{res} . The major difference between the two figures is the narrower range of uncertainty exhibited by the ensemble members. This is useful for several reasons. First, a smaller range of uncertainty among the ensemble members means each ensemble member is closer to the forced signal. This helps us identify the forced signal that otherwise was obscured by circulation-induced variability. This would, for instance, improve our TOE calculation from Chapter 3. This is also important given that the real world is only a single realization, and reducing the uncertainty would make for better climate prediction. Second, the

number of ensemble members needed for a reasonable estimate of the true forced signal decreases. This lessens the computational burden and expense needed to run large ensembles.

To quantitatively assess the differences between the two panels in figure 4.10, we calculate the probability density distributions of the correlation coefficient when 1, 2, and 5 ensemble members are used (Figure 4.11). The left panels show the distribution from the trends of the original temperature data (Figure 4.10a), while the right panels show the distribution from the trends of T_{res} (Figure 4.10b). As more ensemble members are used, the correlation becomes closer to 1, and the spread in those correlations become much smaller. For each of the three cases, the spread for the T_{res} trends are narrower than the spread in the original temperature trends. It is interesting to note that at 10 ensemble members (not shown), the mean correlation in T_{res} starts to become smaller than the correlation from the original data. This is because the maximum correlation for T_{res} trends is 0.94 instead of 1. This difference suggests that there is a forced component in the circulation (at least in the PCR method) that we have removed, resulting in a different forced signal when we average the 30 T_{res} trends together. In this case, when using less than 10 ensemble members T_{res} provides a higher average correlation with the ensemble mean of the surface air temperature, but for more than 10 ensemble members the current PCR method is unable to do a better job at estimating the forced signal. Figure 4.10 therefore claims that dynamical adjustment alone is not enough. This method can narrow the uncertainty exhibited by internal variability. However, it cannot separate the circulation dynamics from unforced internal variability and circulation changes due to climate change.

While the PCR and MCA methods can be calculated separately for both the control run and the large ensemble, the analog method offers an alternative approach. Instead of using the ensemble SLP data from 1955-2004, as done previously in this chapter, the analog method described by [28] makes use of the control run to estimate the internal dynamical contribution to temperature in the large ensemble. This is a variation of the constructed analog method used in [29]. For this method, we take monthly SLP data for all ensemble members for the years 1955-2004, and for each month and each ensemble member we search all the corresponding months in the last 1700 years of

the control run and find the closest 150 months (from Euclidean distance). Then we randomly select 100 of the 150 and calculate the optimal linear combination of those analogs that best fit the ensemble SLP field. For example, for January 1955 we search all January values in the control run and find the 150 closest values to that of our 1955 field. We randomly select 100 of the 150 closest months, then calculate a set of linear regression coefficients β that best explain the original (January 1955) field. We can write the definition of a constructed analog as follows:

$$X_{CA} = \beta X_C \quad (4.5)$$

where X_{CA} is the constructed analog, X_C is the matrix containing the 100 analogs, and β is the optimal linear combination that best fits the ensemble SLP field.

This results in a basic linear equation $Ax = B$. However, because the elements of X are not orthogonal states no unique solution exists for the values of β . Instead, we estimate the β coefficients using the Moore-Penrose pseudoinverse of A, defined as $A^\dagger = (A^T A)^{-1} A^T$. In this case $A^\dagger A = I_n$. We can then solve for β as:

$$\beta = [(X_C^T X_C)^{-1} X_C^T] X_0 \quad (4.6)$$

where X_0 is the original monthly dataset.

In our example, X_0 is January 1955, and X_C are the 100 SLP analogs we have selected. After calculating β , we plug β back into Equation 4.5 and solve X_{CA} for the estimated temperature field. By using β and the closest analogs from the SLP field, the analogous calculation for temperature estimates in the temperature that is solely due to circulation dynamics. For 100 analogs, X_C has size $[M \times 100]$, β has size $[100 \times 1]$, and X_{CA} has size $[M \times 1]$, where M is the number of grid boxes (space) of the region you are looking at (e.g., Northern Hemisphere extratropics). We then do the random selection of the 100 analogs 50 times, and average the resulting 50 constructed analogs together. The average of these 50 constructed analogs removes any sampling error that may occur from using a single constructed analog.

An advantage of the constructed analog method is that it can be applied to any field, including the SLP field itself. Thus, we can calculate how much of the anomaly SLP field is reproduced by the linear combination of constructed analogs themselves. However, the analog method assumes that there are no changes in the SLP field itself due to climate change, and any differences in the control run SLP and large ensemble SLP due to external forcing will be (mistakenly) thought of as internal variability. Furthermore, it also requires a longer period to draw analogs from (the CESM1 1700 year control run here), and thus would not be as appropriate if applied to a short observational dataset.

Finally, Figure 4.12 offers a comparison of all three methods in the CESM-LE. The left columns show the root mean square error (RMSE) between the ensemble mean trend and each of the 30 ensemble trends, while the right columns show the same RMSE for the dynamically adjusted trends. All three dynamical adjustment methods show a significant decrease in both the mean RMSE as well as the variance amongst the ensemble members. The maximum RMSE is around 1.4 in the unadjusted temperature trends, but goes down to just over 1 for all three methods. Since the real world is only one realization (any single ensemble member), we need to look at the variance and outliers in addition to the mean RMSE values.

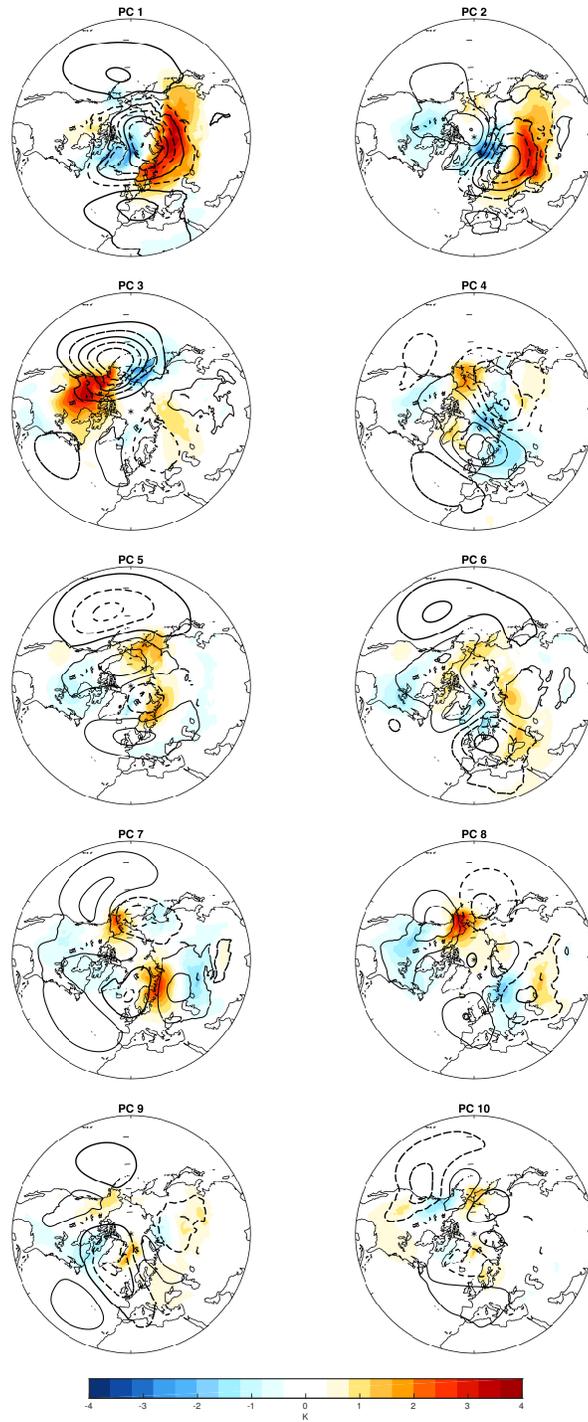


Figure 4.1: SLP (contours) and surface air temperature (shading) regressed onto the first 10 standardized principal components of monthly mean SLP data from the CESM control run. Contours are every 1.5 hPa; negative contours are shown in the dashed lines. DJF values only.

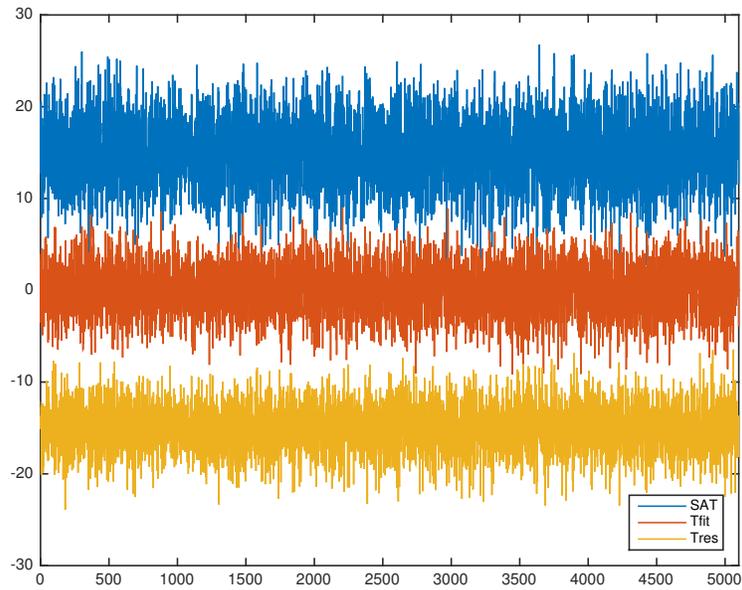


Figure 4.2: Time series for a single grid box in central Siberia of temperature (top, blue), T_{fit} from the first 10 PCs of SLP (middle, red), and T_{res} (bottom, orange), obtained as the difference between the first two time series.

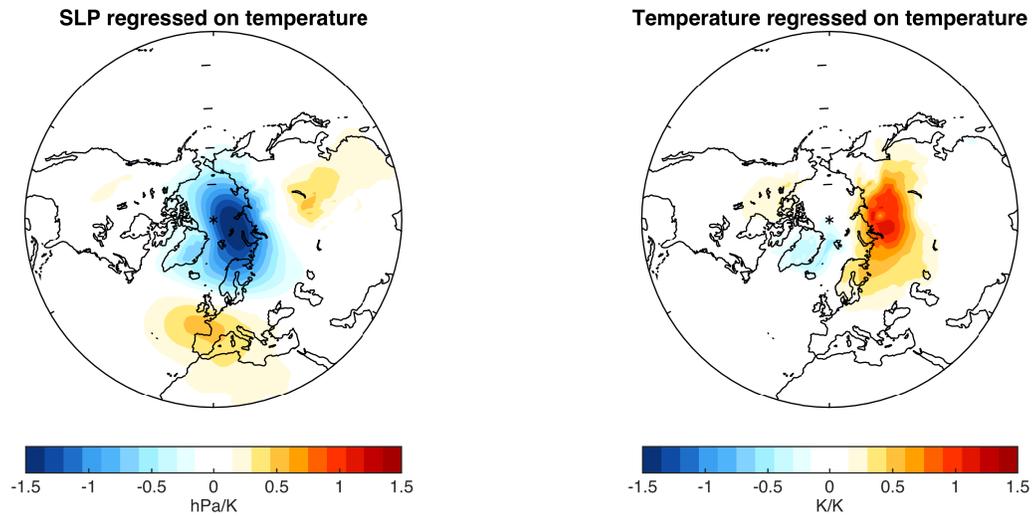


Figure 4.3: Regression maps for (a) SLP, and (b) temperature on the temperature timeseries at a single grid box in central Siberia. Data is for DJF monthly anomalies from the CESM-LE control run.

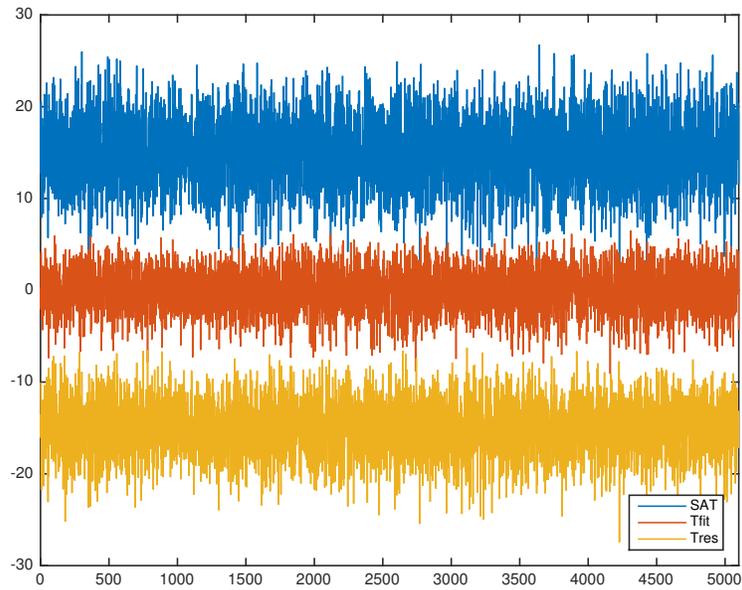


Figure 4.4: Same as Figure 4.2, but with T_{fit} and T_{res} from the MCA method.

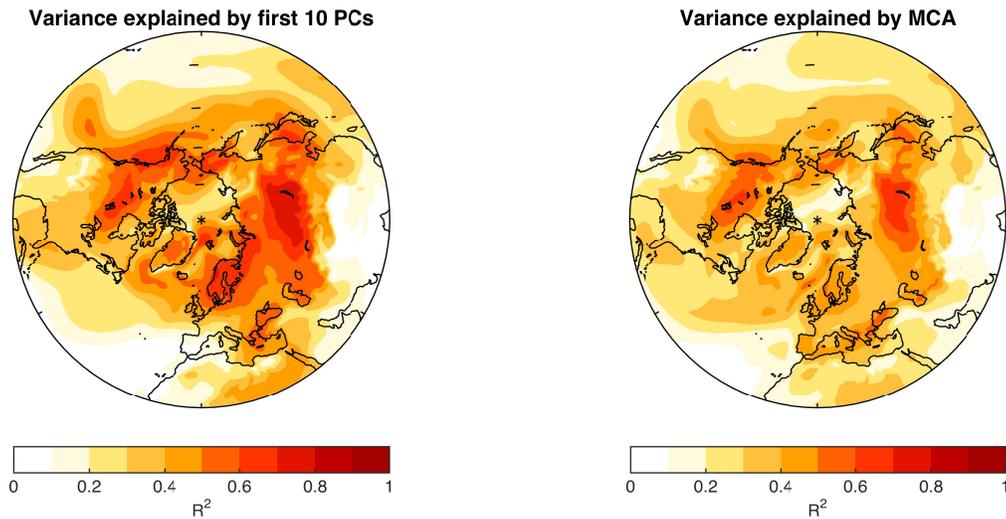


Figure 4.5: Variance in temperature for DJF months explained by (a) the first 10 PCs of SLP, and (b) the expansion coefficient timeseries, both calculated from DJF monthly anomalies of SLP from the CESM control run. PCR does a better job overall, explaining an average of 40% of the variance in the winter temperature field, while MCA explains an average of 32% of the variance over the Northern Hemisphere extratropics.

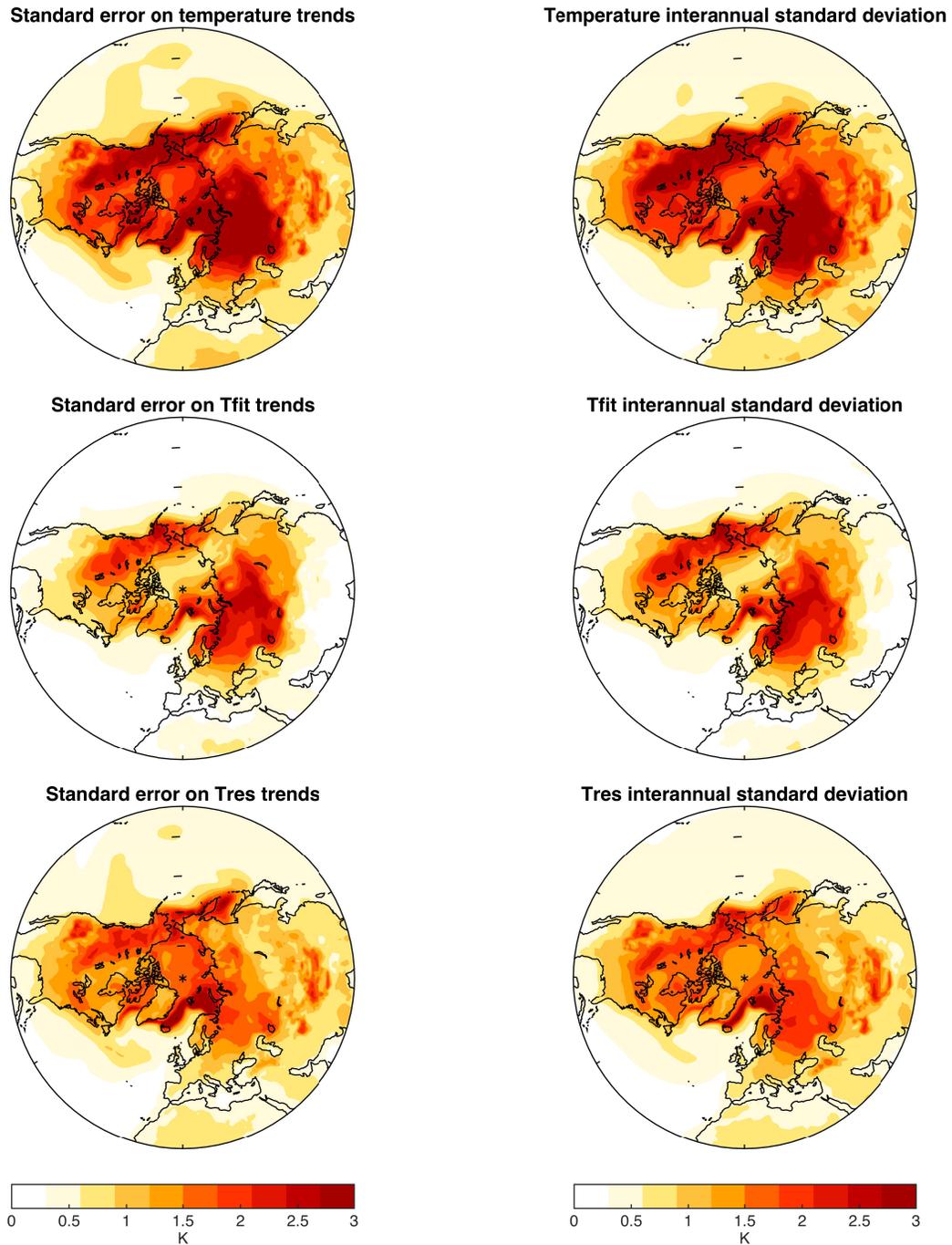


Figure 4.6: Plots showing (left) the standard error on 50 year DJF averaged trends, calculated as (left) two times the standard deviation across all 34 chunks of 50 year trends, and (right) the interannual standard deviation of DJF averaged anomaly data. The calculations are shown for (top) temperature, (middle) dynamically induced temperature (T_{fit} via PCR), and (bottom) the dynamically adjusted temperature (T_{res}). From the CESM control run.

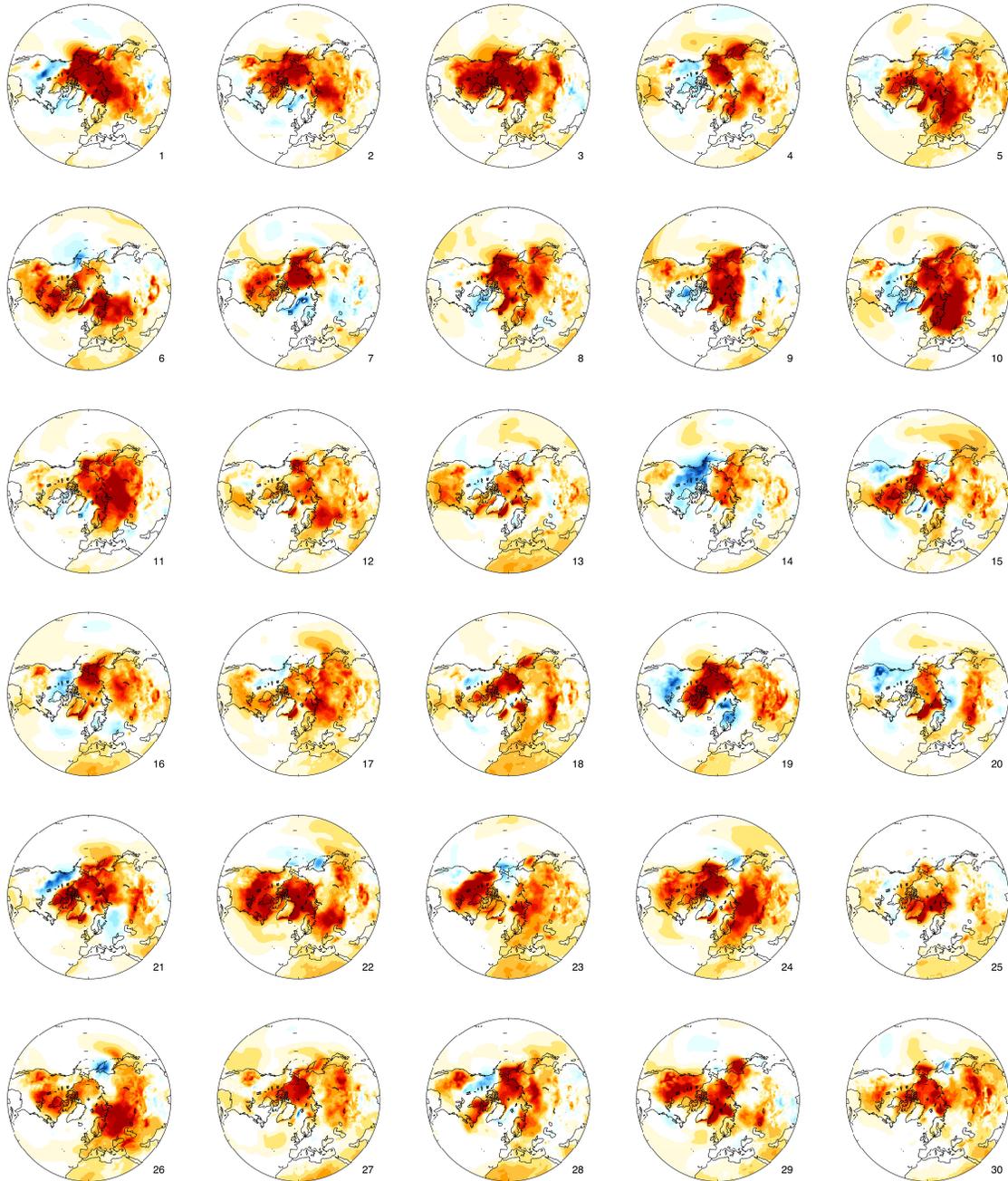


Figure 4.7: 50-year DJF surface temperature trends for each ensemble member of the CESM-LE (labeled 1-30) for the period 1955-2004.

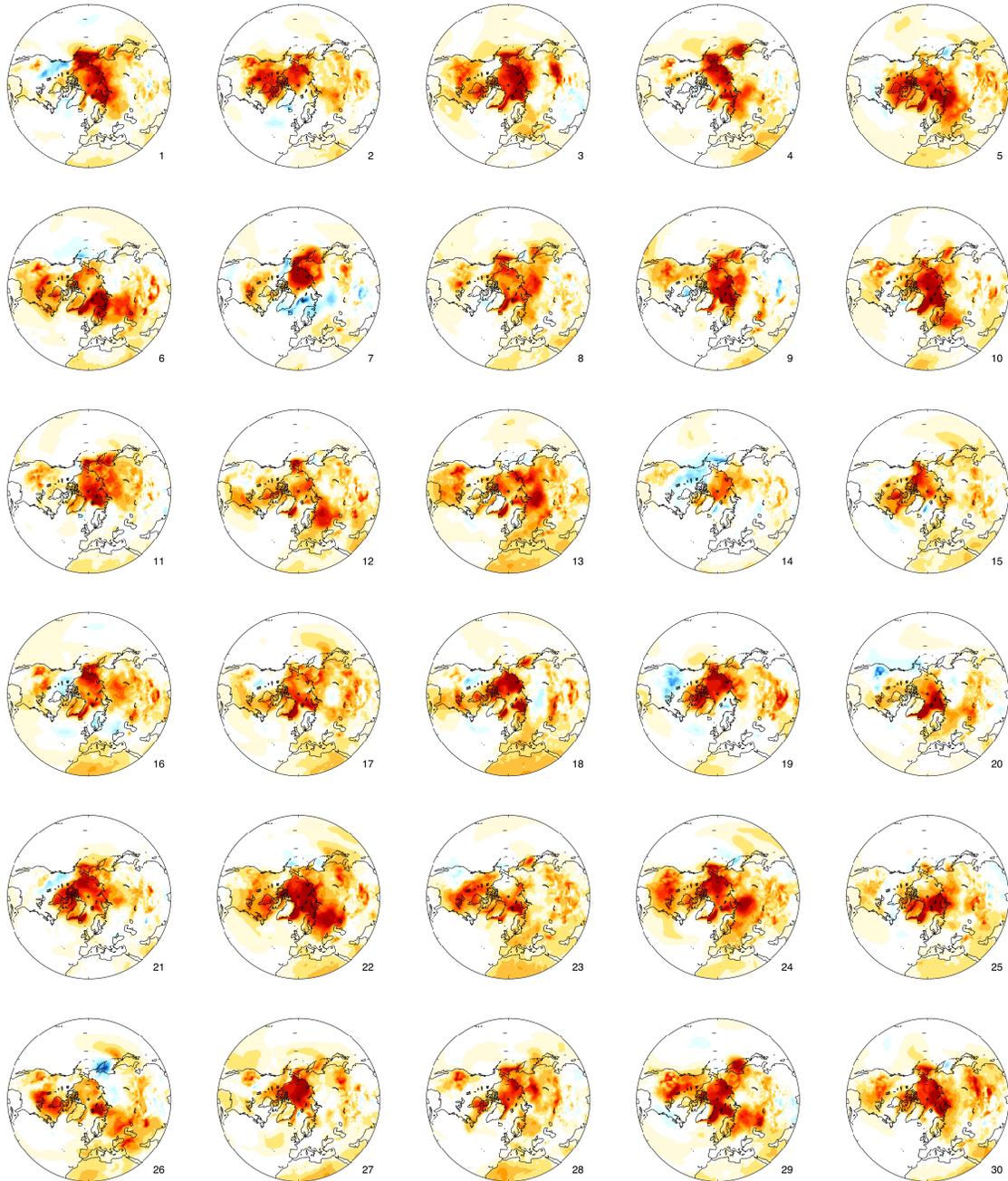


Figure 4.8: 50-year dynamically-adjusted DJF temperature trends for each ensemble member of the CESM-LE (labeled 1-30) for the period 1955-2004. The dynamically-adjusted temperature was calculated as the residual obtained by subtracting the temperature fitted to first 10 PCs for each ensemble member from the original temperature data.

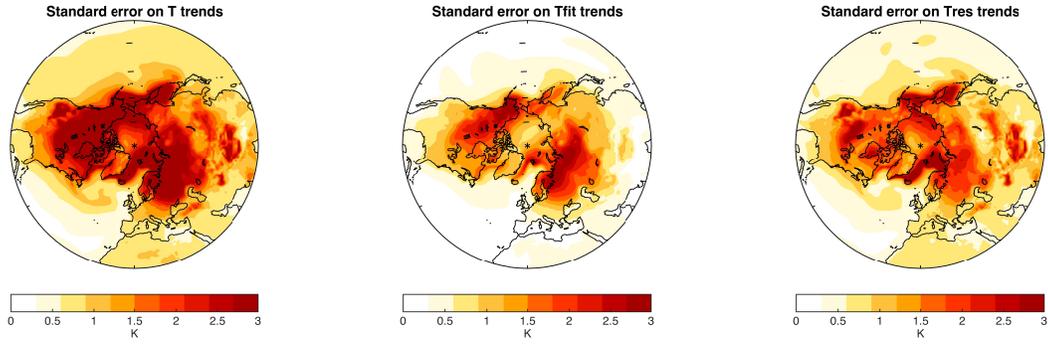


Figure 4.9: 95% standard error on DJF averaged trends across all ensemble members for 1955-2004 of (a) surface air temperature (b) T_{fit} from internal circulation dynamics found by applying the PCR method, and (c) T_{res} , the residual temperature not associated with internal dynamics.

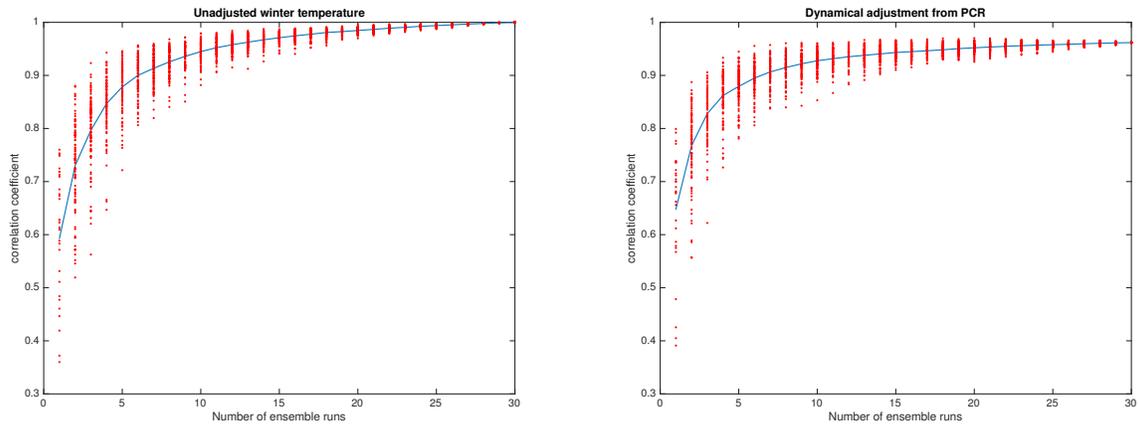


Figure 4.10: The correlation between varying numbers of ensemble members used and the ensemble-mean DJF temperature trend. The left panel shows the spread given by the ensemble members for the actual DJF temperature trends for the period 1955-2004, while the right panel shows the same spread for the temperature with circulation dynamics removed via the PC regression method. For the ensemble mean, the residual temperature trends and actual temperature trends have a correlation of 0.94.

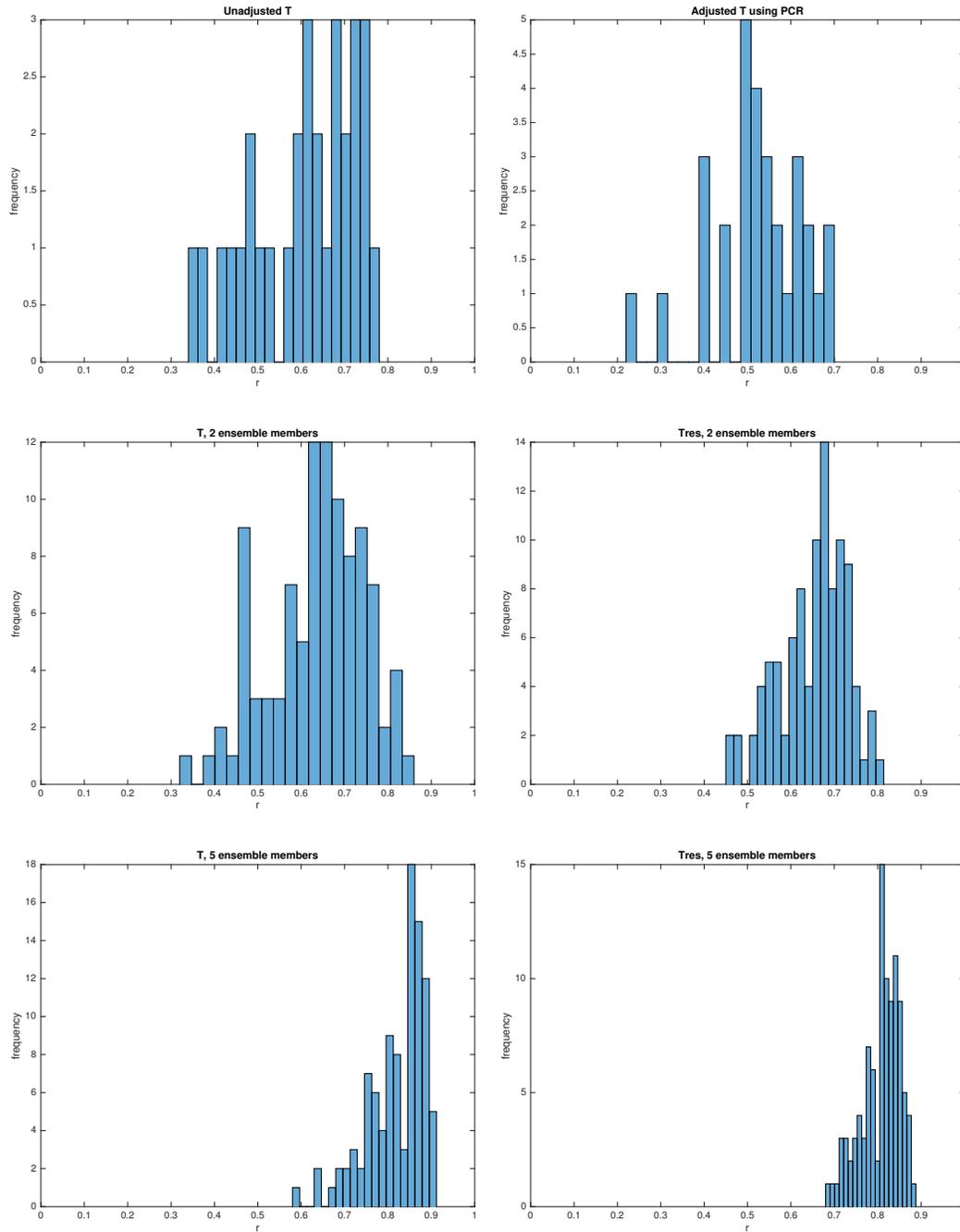


Figure 4.11: Probability density distributions of the spatial correlations shown in Figure 4.10. The left panels show the correlations for surface air temperature data, while the right panels show the correlations for the dynamically-adjusted temperature. Comparisons are for 100 runs of random averaged values for 1, 2, and 5 ensemble members.

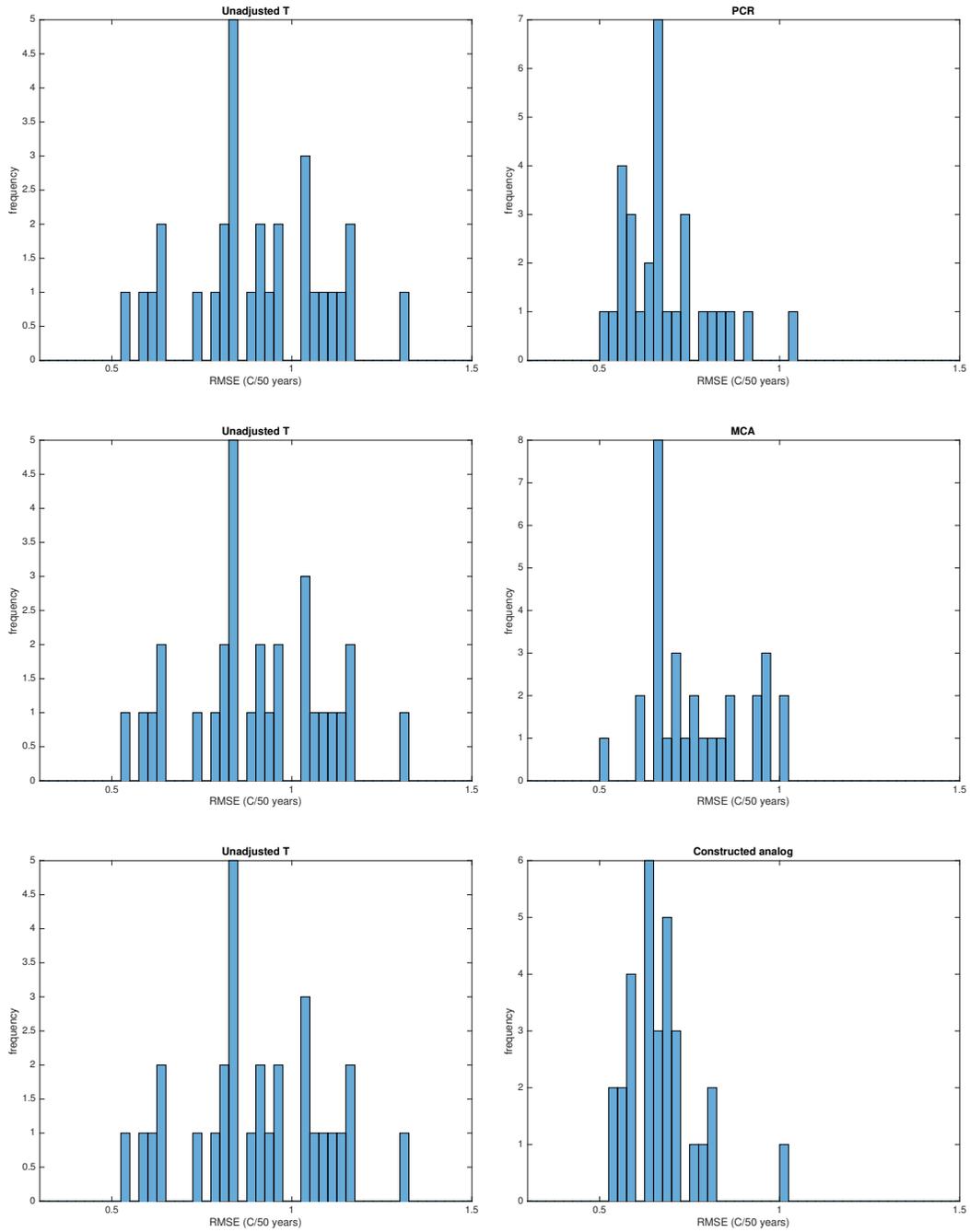


Figure 4.12: PDFs of (left) RMSE between the ensemble mean temperature trend and temperature trends from 30 ensemble members; and (right) RMSE between the ensemble mean temperature trend and temperature trends from 30 dynamically adjusted ensemble members.

Chapter 5

Conclusions

5.1 Time of Emergence

The impacts of anthropogenic climate change are felt locally. But the lead times when warming and related impacts emerge from the natural climate variability vary greatly from one location to the next. In this thesis we have derived an expression (Equation 3.2) that provides a simple analytic tool for estimating the lead time when regionally-dependent impacts of climate change emerge from the natural variability in a statistically significant sense. This expression is tested on surface temperature data from the CESM large ensemble of climate change simulations and found to be in good agreement with TOE predicted by the model. We then apply the expression to observations and find that the real world is warming faster than what the CESM-LE would predict. However, this result falls within the bounds of natural variability in the large ensemble, thus suggesting that natural variability has contributed to the additional warming in observations for the past 45 years.

We have focused on the application of Equation 3.2 to surface temperature, but the expression holds for any time series where the following three conditions are met: 1) the forced signal can be modeled as a linear trend; 2) the statistics of the natural variability (detrended values of the time series) are Gaussian; and 3) the standard deviation of the natural variability is stationary. These three assumptions derive from our use of the standard error of the regression. The bases for all three assumptions are discussed and justified in [1].

The methodology outlined here is derived from statistical tools that are used widely in climate change research. It is potentially useful for climate change research for three primary reasons. One, it provides an *analytic* estimate of the lead time required for a trend to emerge, and can thus be trivially calculated given a) the amplitude and autocorrelation of the observed natural variability and b) the simulated forced signal. Two, it provides an estimate of the time required for a linear trend to emerge in a *statistically significant* sense, rather than as a (statistically arbitrary) factor of

the internal variability. And three, the expression requires no treatment of the data, which renders the resulting lead times easy to compare across different model configurations, different forcing scenarios, and different estimates of the natural variability.

5.2 Dynamical Adjustment

Given the large amplitudes of natural and internal variability innate to the climate system, the signal of radiative forcing can be hard to distinguish and requires a long timescale to emerge in a statistically significant sense. There is a need to narrow the amplitude of natural variability by removing various aspects of natural variability that obscure the forced signal, and the climate literature includes many different methods for dynamical adjustment. In this thesis we have examined the role of circulation dynamics on surface temperature via several of these statistical methods using the SLP field. Two popular methods, PCR and MCA, were applied to a fully coupled control run. Both methods removed a significant amount of the internal variability within the model, but the PCR method explained a larger percent of the surface temperature variance in the Northern Hemisphere extratropics. We then applied 3 methods (PCR, MCA, and constructed analogs) to the 30 climate simulations in the CESM-LE over the period 1955-2004 to examine the specific effects of dynamical adjustment. The circulation effects in the large ensemble were found to be similar to those predicted by the control run. Since we know the forced signal in the large ensemble, we are able to quantify the effects of dynamical adjustment, something that would not be possible in observations.

Dynamical adjustment greatly reduces the uncertainty in the residual temperature trends, but there is an additional forced component of the circulation that was removed as well. This resulted in a slightly different forced signal estimate by the dynamically adjusted temperature. The dynamically adjusted temperature proved more useful in estimating the true forced signal when using a smaller number of ensembles (≤ 10), but the forced component in the circulation prevented the dynamically adjusted trends from doing better with more than 10 ensemble members. We found that all three methods reduces the average and spread of the uncertainty in the forced signal. However,

there appears to be a lower limit on the error reduction, which represents all additional aspects of natural variability.

5.3 Future Work

This thesis raises additional questions on the role of natural and internal variability and the obstacle it presents in estimating future climate change. The use of a large ensemble is invaluable in this regard, as it allows us to know the “true” forced signal and quantify the role of natural or internal variability. However, since the observations is only one realization, knowing where the real world falls within this range of natural variability is crucial in future climate projections. We have examined differences in TOE results between the CESM-LE and observations in surface temperature and found that observations fall within the bounds of natural variability in the large ensemble. Future work would benefit from extending the result further to more accurately assess where observations fall in the range of natural variability. We know that the observational and model trends differ, which result in different TOE values. However, we do not know how much other errors contribute, including: model error in the estimation of the forced signal, model error in the representation of natural variability, observational error due to sampling inconsistencies and missing data, and error in observational estimates due to the short length of the record used. By adjusting for all of the other uncertainties we can better pinpoint where observations fall within the natural variability of the climate system.

This thesis also examined various methods of removing the temperature contribution from circulation dynamics. We would like to extend our results from Chapter 4 to create a comprehensive comparison and analysis between each of the methods used in the climate literature. While the methods all give fairly similar results, they are also very different and can be applied in different ways. It would be helpful for the climate community to understand these differences and know which methods are best for their respective needs. In addition, there still remains a large uncertainty in the residual temperature trends. The results from dynamical adjustment could be extended to remove additional contributions to internal variability. For instance, contributions to the uncertainty

in T_{res} include: sea ice dynamics, amount and location of snow cover, and soil moisture, among others. By removing these additional factors of internal variability we can further reduce the range of uncertainty. We would also like to extend our analysis to quantify the forced contribution to these various contributions to internal variability. Separating out the circulation dynamics, sea ice, snow cover, and soil moisture changes that are a result of the anthropogenic signal will allow us to more clearly identify the true forced signal. Further work would also benefit from extending these methods to observational datasets or reanalysis data. This would provide additional insights into historical temperature data trends, allowing us to examine the temperature contribution from internal variability versus the contribution from anthropogenic forcing.

Bibliography

- [1] D. W. J. Thompson, E. A. Barnes, C. Deser, W. E. Foust, and A. S. Phillips. Quantifying the role of internal climate variability in future climate trends. *J. Climate*, 28:6443–6456, 2015.
- [2] N. L. Bindoff, 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. 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 Intergovernmental Panel on Climate Change. Cambridge University Press*, 2013.
- [3] C. Deser, A. S. Phillips, V. Bourdette, and H. Teng. Uncertainty in climate change projections: the role of internal variability. *Climate Dyn.*, 38:527–546, 2012.
- [4] C. Deser, A. S. Phillips, M. A. Alexander, and B. V. Smoliak. Projecting North American Climate over the Next 50 Years: Uncertainty due to Internal Variability. *J. Climate*, 27:2271–2296, 2014.
- [5] B. D. Santer, J. F. Painter, C. Bonfils, C. A. Mears, S. Solomon, T. M. L. Wigley, P. J. Gleckler, G. A. Schmidt, C. Doutriaux, N. P. Gillett, K. E. Taylor, P. W. Thorne, and F. J. Wentz. Human and natural influences on the changing thermal structure of the atmosphere. *PNAS*, 110:17235–17240, 2013.
- [6] C. Deser, R. Knutti, S. Solomon, and A. S. Phillips. Communication of the role of natural variability in future North America climate. *Nat. Clim. Change*, 2:775–779, 2012.
- [7] J. E. Kay, C. Deser, A. Phillips, A. Mai, C. Hannay, G. Strand, J. Arblaster, S. Bates, G. Danabasoglu, J. Edwards, M. Holland, P. Kushner, J. F. Lamarque, D. Lawrence, K. Lindsay, A. Middleton, E. Munoz, R. Neale, K. Oleson, L. Polvani, and M. Vertenstein. 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. Meteor. Soc.*, 96:1333–1349, 2015.
- [8] R. Madden and V. Ramanathan. Detecting climate change due to increasing carbon dioxide. *Science*, 209:763–768, 1980.
- [9] B. D. Santer, U. Mikolajewicz, W. Brüggemann, U. Cubasch, K. Hasselmann, H. Höck, E. Maier-Reimer, and T. M. L Wigley. Ocean variability and its influence on the detectability of greenhouse warming signals. *J. Geophys. Res.*, 100:10693–10725, 1995.
- [10] E. C. Weatherhead, G. C. Reinsel, G. C. Tiao, X. Meng, D. Choi, W. Cheang, T. Keller, J. DeLuisi, D. J. Wuebbles, J. B. Kerr, A. J. Miller, S. J. Oltmans, and J. E. Frederick. Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *J. Geophys. Res.*, 103:17149, 1998.
- [11] J. H. Christensen, B. Hewitson, A. Busuioc, A. Chen, X. Gao, I. Held, R. Jones, R. K. Kolli, W.-T. Kwon, R. Laprise, V. Magana Rueda, L. Mearns, C. G. Menendez, J. Raisanen, A. Rinke, A. Sarr, and P. Whetton. 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. Cambridge University Press*, 2007.
- [12] F. Giorgi and X. Bi. Time of emergence (TOE) of GHG forced precipitation change hot-spots. *Geophys. Res. Lett.*, 39:L06709, 2009.
- [13] I. Mahlstein, R. Knutti, S. Solomon, and R. W. Portmann. Early onset of significant local warming in low latitude countries. *Environ. Res. Lett.*, 6:034009, 2011.
- [14] E. Hawkins and R. Sutton. Time of emergence of climate signals. *Geophys. Res. Lett.*, 39:L01702, 2012.
- [15] G. Zappa, B. J. Hoskins, and T. G. Shepherd. Improving Climate Change Detection through Optimal Seasonal Averaging: The Case of the North Atlantic Jet and European Precipitation. *J. Climate*, 28:6381–6397, 2015.

- [16] B. Kirtman, 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. 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. Cambridge University Press*, 2013.
- [17] N. S. Diffenbaugh and M. Scherer. Observational and model evidence of global emergence of permanent, unprecedented heat in the 20th and 21st centuries. *Climatic Change*, 107:615–624, 2011.
- [18] I. Mahlstein, G. Hegerl, and S. Solomon. Emerging local warming signals in observational data. *Geophys. Res. Lett.*, 39:L21711, 2012.
- [19] G. Walker and E. Bliss. World Weather V. *Mem. Roy. Meteor. Soc.*, 4:53–84, 1932.
- [20] G. Walker and E. Bliss. World Weather VI. *Mem. Roy. Meteor. Soc.*, 4:119–139, 1937.
- [21] J. M. Wallace, Y. Zhang, and J. A. Renwick. Dynamic contribution to hemispheric mean temperature trends. *Science*, 270:780–783, 1995.
- [22] A. J. Broccoli, N. Lau, and M. J. Nath. The Cold Ocean-Warm Land Pattern: Model Simulation and Relevance to Climate Change Detection. *J. Climate*, 11:2743–2763, 1998.
- [23] J. W. Hurrell. Influence of variations in extratropical wintertime teleconnections on northern hemisphere temperature. *Geophys. Res. Lett.*, 23:665–668, 1996.
- [24] R. Quadrelli and J. M. Wallace. A Simplified Linear Framework for Interpreting Patterns of Northern Hemisphere Wintertime Climate Variability. *J. Climate*, 17:3728–3744, 2004.
- [25] J. M. Wallace, Q. Fu, B. V. Smoliak, P. Lin, and C. M. Johanson. Simulated versus observed patterns of warming over the extratropical northern hemisphere continents during the cold season. *PNAS*, 109:14337–14342, 2012.

- [26] D. W. J. Thompson, J. M. Wallace, P. D. Jones, and J. J. Kennedy. Identifying Signatures of Natural Climate Variability in Time Series of Global-Mean Surface Temperature: Methodology and Insights. *J. Climate*, 22:6120–6141, 2009.
- [27] B. V. Smoliak, J. M. Wallace, P. Lin, and Q. Fu. Dynamical Adjustment of the Northern Hemisphere Surface Air Temperature Field: Methodology and Application to Observations. *J. Climate*, 28:1613–1629, 2015.
- [28] C. Deser, L. Terray, and A. S. Phillips. Forced and Internal Components of Winter Air Temperature Trends over North America during the past 50 Years: Mechanisms and Implications. *J. Climate*, 29:2237–2258, 2016.
- [29] H. M. Van den Dool, J. Huang, and Y. Fan. Performance and analysis of the constructed analogue method applied to U.S. soil moisture over 1981-2001. *J. Geophys. Res.*, 108:8617, 2003.
- [30] J. W. Hurrell, M. M. Holland, P. R. Gent, S. Ghan, J. E. Kay, P. J. Kushner, J.-F. Lamarque, W. G. Large, D. Lawrence, K. Lindsay, W. H. Lipscomb, M. C. Long, N. Mahowald, D. R. Marsh, R. B. Neale, P. Rasch, S. Vavrus, M. Vertenstein, D. Bader, W. D. Collins, J. J. Hack, J. Kiehl, and S. Marshall. The Community Earth System Model: A Framework for Collaborative Research. *Bull. Amer. Meteor. Soc.*, 94:1339–1360, 2013.
- [31] P. D. Jones, D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice. Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010. *J. Geophys. Res.*, 117:D05127, 2012.
- [32] B. D. Santer, T. M. L. Wigley, J. S. Boyle, D. J. Gaffen, J. J. Hnilo, D. Nychka, D. E. Parker, and K. E. Taylor. Statistical significance of trends and trend differences in layer average atmospheric temperature time series. *J. Geophys. Res.*, 105:7337–7356, 2000.
- [33] E. A. Barnes and R. J. Barnes. Estimating linear trends: Simple linear regression vs. epoch differences. *J. Climate*, 28:9969–9976, 2015.

- [34] C. Saffioti, E. M. Fischer, S. C. Scherrer, and R. Knutti. Reconciling observed and modeled temperature and precipitation trends over Europe by adjusting for circulation variability. *Geophys. Res. Lett.*, 43:8189–8198, 2016.