ROBUST MULTI-YEAR PREDICTABILITY ON CONTINENTAL SCALES

by

Liwei Jia

A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Climate Dynamics

Committee:

Dr. Timothy DelSole, Dissertation Director

Dr. Bohua Huang, Committee Member

Dr. Barry Klinger, Committee Member

Dr. Edwin K. Schneider, Committee Member

Dr. James E. Gentle, Committee Member

Dr. David Straus, Department Chair

Dr. Richard Diecchio, Associate Dean
for Academic and Student Affairs,
College of Science

Dr. Vikas Chandhoke, Dean,
College of Science

Date: April 25, 2011

Spring Semester 2011
George Mason University
Fairfax, VA

Robust Multi-Year Predictability on Continental Scales

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Liwei Jia
Master of Science
Chinese Academy of Meteorological Sciences, 2006
Bachelor of Science
Lanzhou University, 2002

Director: Timothy DelSole, Professor
Department of Atmospheric, Oceanic and Earth Sciences

Spring Semester 2011
George Mason University
Fairfax, VA

Dedication

I dedicate this dissertation to my parents for their continued support and unconditional love throughout the course of my study.

Acknowledgments

I would like to show my gratitude to the following people who made this thesis possible.

First and foremost, I am heartily thankful to my supervisor, Dr. Timothy DelSole, whose invaluable guidance and inspirational encouragement I will never forget. His profound knowledge and patient explanation enabled me to develop a good understanding of the subject. Without him, this thesis would not have been completed.

I am grateful to my committee members Dr. Bohua Huang, Dr. Edwin Schneider, Dr. Barry Klinger and Dr. James Gentle for their constructive suggestions.

I would like to acknowledge Jennifer Adams and Thomas Wakefield for their technical support, and many research scientists in COLA for the scientific communications. Also, I have been blessed with a friendly and cheerful group of fellow students in climate dynamics program.

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of my study.

Table of Contents

List of Tables

List of Figures

Abstract

ROBUST MULTI-YEAR PREDICTABILITY ON CONTINENTAL SCALES

Liwei Jia, PhD

George Mason University, 2011

Dissertation Director: Dr. Timothy DelSole

This study identifies natural, unforced predictable components of surface air temperature (SAT) and precipitation in six continents from pre-industrial control runs of the Coupled Model Intercomparison Project phase 3 data set. The externally forced components of continental SAT also are identified by maximizing the variance ratio in the 20th century runs to the control runs. The leading unforced predictable components can be predicted in independent control runs with statistically significant skill for 3-6 years in SAT and 1-3 years in precipitation, depending on continent, using a linear regression model with global sea surface temperature (SST) as predictor. The leading unforced predictable components of SAT are related to ENSO and the persistence of SSTs near the continent itself. The only exception is Europe, which has no significant ENSO relation. The leading unforced predictable components of precipitation are significantly correlated with an ENSO-like SST pattern. No unforced predictability of land precipitation could be found in Europe. There is only one significant forced pattern of SAT in each continent. The largest amplitudes of these forced patterns concentrate in high latitudes. No significant forced pattern of continental precipitation could be identified on a multi-model basis. Although the forced and unforced patterns of SAT are identified in model simulations, they are not separable in the observations, presumably because of the large similarity between them.

Chapter 1: Summary of the Thesis

The purpose of this research thesis is to identify predictable components in the climate system. In chapter 2, we first define predictability in rigorous terms and show that most previous measures of predictability are fundamentally equivalent. Predictability in the climate system is known to vary on a wide range of time scales, from day-to-day weather variations to centennial scale climate changes. In this thesis, we are primarily interested in predictability on multi-year time scales of surface air temperature (SAT) and precipitation over land. On this time scale, two kinds of predictability are expected to be important, one kind arising from initial condition information and another kind arising from changes in human-induced greenhouse gas and aerosol concentrations, and natural solar insolation and volcanism. We review the scientific basis for expecting these kinds of predictability to exist in the climate system. However, most studies on this topic focus primarily on the ocean, leaving multi-year predictability of land variables unclear. In addition, most previous studies apply various univariate techniques that are not optimally designed to detect predictability. Although several multivariate techniques have been used in predictability study, they all have some limitations. The limitations of previous studies motivate us to develop new statistical optimization techniques to identify the predictable components over land. We identify both kinds of predictability in an optimal manner from multiple state-of-the-art climate models. Also, we attempt to separate these two kinds of predictability in observations.

In chapter 3, we develop a new statistical optimization method for finding components that are predictable on the longest possible time scales. We apply this method to the output of multiple model simulations with fixed greenhouse gas concentrations, solar radiation and so on, so that the resulting predictability, called unforced predictability, is due to dynamical processes in the climate system. To study the unforced land predictability on continental

scales, we diagnose the predictable components of SAT and precipitation in six individual continents. To ensure that the identified components are in fact predictable, we estimate the components in half of the data and formally verify the predictability in the other half. The results reveal that the most predictable components can be predicted in independent model runs with statistically significant skill for 3-6 years in SAT, depending on continent, using a linear regression model with global sea surface temperature (SST) as predictor. The patterns of the most predictable components of SAT in all six continents are of single sign, implying continental scale warming or cooling as a whole. To investigate the sources of such predictability, we computed the lagged correlations between the most predictable component and global SST in each continent. The correlation maps show that the predictability of land SAT in all continents, except Europe, arises from an oscillatory ENSO-like SST pattern and a persistent SST pattern near the continent in question. The predictability of land SAT in Europe is only correlated with the persistent SSTs in North Atlantic near Europe. The correlation patterns are reproducible in individual models, suggesting the results are robust across models.

In the case of precipitation, the most predictable component in each continent except Europe can be predicted 1-3 years ahead in independent model runs, depending on continents, using a multiple linear regression model with global SST as predictor. No multi-year predictability of precipitation in Europe is verifiable in independent model runs. The lagged correlation maps between global SST and the most predictable components reveal that the multi-year predictability of land precipitation is closely related to ENSO.

In chapter 4, we identify land predictability arising from changes in greenhouse gas concentrations, solar radiations and other changes in agents external to the climate system, known as external forcings. A statistical optimization method, called discriminant analysis, is used to identify predictable components of continental SAT and precipitation due to external forcings in multiple model runs. This technique optimizes the ratio of variance in the simulations forced by external forcings to the variance in the unforced simulations. The results show that there is only one predictable component in each continent that is caused by

external forcings. The time series corresponding to this component show increasing trends in each continent, and the spatial patterns show large amplitudes in high latitudes and are of single sign. However, no statistically significant component of continental precipitation due to external forcings could be identified on a multi-model basis.

In chapter 5, we attempt to separate the observed SAT changes into a component of unforced natural variability, and a component due to external forcings. To this end, another statistical optimization method, called optimal fingerprinting, is used. This method fits the observations to a linear combination of forced and unforced spatial patterns. The amplitudes and their standard errors associated with these two patterns are computed for each year. The results reveal that the SAT changes due to external forcings are not clearly detected in observations in any continent at a 5% significance level, implying that the forced pattern is not significantly distinguishable from the unforced pattern on continental scales, using spacial information alone.

Chapter 2: Introduction

2.1 Definition of predictability

Predictability is the extent to which an event can be predicted. Any forecast of nature is uncertain because the forecast model and initial/boundary conditions of the model have uncertainties. Therefore, the most complete statement of a forecast is its probability distribution. Let $P(\mathbf{y}|\mathbf{x})$ denote a forecast probability distribution conditioned on $\mathbf{x}$, where $\mathbf{y}$ is the vector of variables to be predicted at a future time and $\mathbf{x}$ represents all previous variables that are available. If the forecast $\mathbf{y}$ is independent of previous information $\mathbf{x}$, then

$$P(\mathbf{y}|\mathbf{x}) = P(\mathbf{y}), \tag{2.1}$$

and $\mathbf{y}$ is said to be unpredictable (DelSole, 2004). Therefore, a necessary condition for $\mathbf{y}$ to be predictable is that $P(\mathbf{y}|\mathbf{x}) \neq P(\mathbf{y})$. Loosely speaking, the greater the difference between $P(\mathbf{y}|\mathbf{x})$ and $P(\mathbf{y})$, the greater the predictability. Some standard measures of predictability are discussed in section 2.2.

In the study of climate predictability, there are two kinds of predictability to distinguish. One is called unforced predictability, in which $\mathbf{x}$ can be identified with antecedent observations/initial conditions of the system, and the predictability arises from internal dynamical processes, including coupled interactions between climate components (e.g., atmosphere, land, ocean, and sea ice). The climate variability generated by internal dynamical processes in the climate system is known as "internal variability". The other kind of predictability is called forced predictability, in which $\mathbf{x}$ can be identified with the time history of external forcings. The external forcings refer to natural external forcings occurring naturally such as the changes in solar radiation and volcanism, as well as anthropogenic forcings induced by

human activity such as the changes in greenhouse gas concentrations and human-induced aerosols (Hegerl et al., 2007)[sec. 9.1.1]. The external forcings change radiative energy budget of the Earth's climate system, and consequently cause climate change. The climate change generated in this way is usually known as climate response to external forcings. The probability distribution $P(\mathbf{y}|\mathbf{x})$ describes the characteristic response of the climate system to external forcings. To the extent that $\mathbf{x}$ is predictable, $\mathbf{y}$ is predictable.

2.2 Measures of predictability

The above definition implies that testing whether a variable is predictable is equivalent to testing whether that variable is independent of other variables. However, testing independence in full generality is not possible with finite samples. Accordingly, there are different methods for testing independence that take advantage of the specific data structure (e.g., existence of ensembles) or physical hypotheses about the relevant predictors (e.g., relation to selected variables). Although there is only one definition of predictability, there are several measures of predictability, including signal-to-noise ratio, mean square error, correlation between ensemble members, multiple correlation and autocorrelation. We will show that these measures are fundamentally the same, and are connected through the "law of total variance". These connections do not appear to be explicitly stated in the literature. In addition, understanding the limitations of these measures motivates us to propose a "better" method to measure predictability.

We first introduce the concept of ensemble forecasts, which are required in some measures of predictability. Fig. 2.1 illustrates the ensemble forecasts, initialized at three distinct initial states $t1, t2$ and $t3$. In this experimental setup, there exist three ensemble members generated by slightly different initial conditions at each initial state t . The data produced by these ensemble forecasts is a function of initial state t, forecast lead time τ, ensemble member e. At a given lead time τ, any single ensemble forecast can be written as y_{en}, where the index $n = 1, 2, ..., N$ denotes the initial state, index e=1, 2, ..., E denotes the ensemble member initialized at one state t, N is the total number of initial states, and E is the total

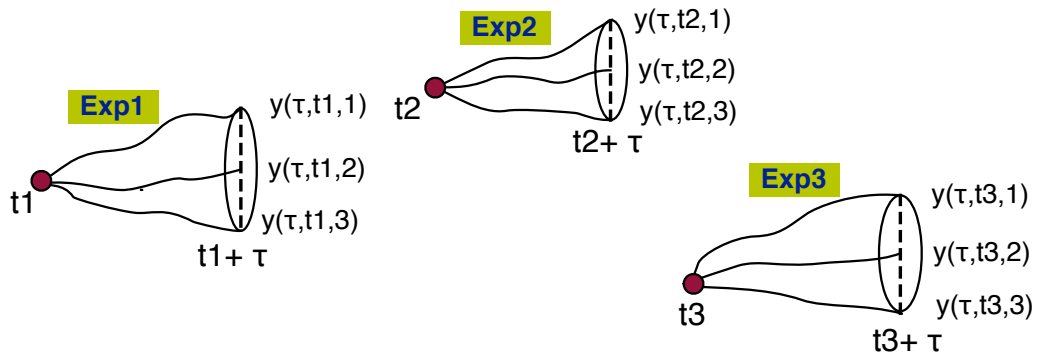

Figure 2.1: Schematic diagram of ensemble forecasts initialized at three initial states $t1, t2$ and $t3$. There are three ensemble members associated with each initial state. The forecast uncertainty, denoted by the solid red dot, increases with lead time τ.

number of ensemble members at one initial state. The initial uncertainty, denoted by the red solid dot, increases with lead time τ.

2.2.1 Signal-to-noise ratio

According to the definition of predictability, $\mathbf{y}$ is said to be unpredictable if $P(\mathbf{y}|\mathbf{x}) = P(\mathbf{y})$, which requires at least the means and variances of the two distributions to be identical, i.e.,

$$\mathrm{E}[\mathbf{y}|\mathbf{x}] = \mathrm{E}[\mathbf{y}] \tag{2.2}$$

and

$$\mathrm{var}(\mathbf{y}|\mathbf{x}) = \mathrm{var}(\mathbf{y}), \tag{2.3}$$

where E[.] denotes the expectation, and var(.) denotes the variance. The conditional expectation $\mathrm{E}[\mathbf{y}|\mathbf{x}]$ is the mean of $\mathbf{y}$ for fixed $\mathbf{x}$. The conditional variance $\mathrm{var}(\mathbf{y}|\mathbf{x})$ is the variance of $\mathbf{y}$ for fixed $\mathbf{x}$.

Consider the ensemble forecasts described above, the finite variance of random variable $\mathbf{y}$ is given according to the law of total variance in probability theory

$$\mathrm{var}(\mathbf{y}) = \mathrm{E}[\mathrm{var}(\mathbf{y}|\mathbf{x})] + \mathrm{var}\left(\mathrm{E}[\mathbf{y}|\mathbf{x}]\right), \tag{2.4}$$

where $\mathbf{x}$ denotes the vector of initial state of the ensemble forecasts. The law of total variance states that the total variance (i.e., climatological variance) of $\mathbf{y}$ can be decomposed into two parts as shown in the right hand side of (2.4). The first term is the average of variance about ensemble mean over all initial states, called "unexplained variance" or "noise variance". The second term is the variance of the ensemble mean, called "explained variance" or "signal variance". If $\mathbf{y}$ is unpredictable, then the "signal variance" vanishes, which implies $\text{var}(\mathbf{y}) = \text{E}[\text{var}(\mathbf{y}|\mathbf{x})]$. This suggests a measure of predictability as signal-to-noise ratio (SNR)

$$SNR = \frac{\text{var}\left(\text{E}[\mathbf{y}|\mathbf{x}]\right)}{\text{E}[\text{var}(\mathbf{y}|\mathbf{x})]} \tag{2.5}$$

or signal-to-total ratio (STR)

$$STR = \frac{\text{var}\left(\text{E}[\mathbf{y}|\mathbf{x}]\right)}{\text{var}(\mathbf{y})} = \frac{SNR}{SNR+1}. \tag{2.6}$$

No predictability implies that $SNR = 0$ and $STR = 0$. Perfect predictability implies that SNR is infinity and $STR = 1$. Equation (2.6) shows that SNR and STR are fundamentally the same. One can be derived from the other.

The measures of predictability SNR and STR can be estimated from samples drawn from the ensemble forecasts. We assume that the ensemble forecast is of the form

$$y_{en} = y_{\cdot n} + \epsilon_{en}, \tag{2.7}$$

where a 'dot' is used to indicate an average over an index, $y_{\cdot n}$ specifies the population mean for a fixed initial state, called the conditional mean, or ensemble mean or "signal", and ϵ_{en} represents "noise", which is assumed to be independently, identically and normally distributed with zero mean and variance σ_N^2. These assumptions imply that the "noise" is independent of the "signal". The "signal variance", which measures the variability in $y_{\cdot n}$,

can be estimated by

$$\hat{\sigma}_S^2 = \frac{1}{N}\sum_{n=1}^{N}(\hat{y}_{.n} - \hat{y}_{..})^2, \tag{2.8}$$

where the hat symbol denotes a sample quantity estimated from data, $\hat{y}_{.n}$ is an estimate of the ensemble mean

$$\hat{y}_{.n} = \frac{1}{E}\sum_{e=1}^{E} y_{en} \quad , \tag{2.9}$$

and $\hat{y}_{..}$ is an estimate of the unconditional mean or grand mean

$$\hat{y}_{..} = \frac{1}{NE}\sum_{n=1}^{N}\sum_{e=1}^{E} y_{en} \quad . \tag{2.10}$$

The "noise variance" is estimated by

$$\hat{\sigma}_N^2 = \frac{1}{NE}\sum_{n=1}^{N}\sum_{e=1}^{E}(y_{en} - \hat{y}_{.n})^2. \tag{2.11}$$

The total variance is estimated by

$$\hat{\sigma}_T^2 = \frac{1}{NE}\sum_{n=1}^{N}\sum_{e=1}^{E}(y_{en} - \hat{y}_{..})^2. \tag{2.12}$$

It can be shown that

$$\hat{\sigma}_T^2 = \hat{\sigma}_S^2 + \hat{\sigma}_N^2 \quad . \tag{2.13}$$

The SNR is then estimated by

$$\hat{SNR} = \hat{\sigma}_S^2/\hat{\sigma}_N^2 \quad . \tag{2.14}$$

The STR is estimated by

$$\hat{STR} = \hat{\sigma}_S^2/\hat{\sigma}_T^2 = \frac{S\hat{N}R}{S\hat{N}R+1}, \tag{2.15}$$

where (2.13) and (2.14) have been used.

If the null hypothesis of no predictability is true, and the noise is normally distributed, then the statistic

$$F = \frac{\hat{\sigma}_S^2}{\hat{\sigma}_N^2}\frac{N(E-1)}{N-1} = S\hat{N}R\frac{N(E-1)}{N-1} \tag{2.16}$$

has an F distribution with $N-1$ and $N(E-1)$ degrees of freedom. Large values of F favor rejection of the null hypothesis.

2.2.2 Mean square error (MSE)

Another measure of predictability is the MSE of ensemble forecasts. We now show that MSE is related to SNR for a "perfect" model. By saying "perfect" model, we assume that the forecast model and the true model are the same.

Consider the classical ensemble forecasts generated in the same way as described in the beginning of section 2.2, the "truth" can be identified with a randomly selected member from the ensemble forecasts under the "perfect" model assumption. The mean square difference between the forecast and the "truth" is therefore the mean square difference between all possible pairs of ensemble members. An estimate of the mean square difference at a given lead time is

$$\begin{aligned} \hat{MSE} &= \frac{1}{NE^2}\sum_{n=1}^{N}\sum_{e=1}^{E}\sum_{e'\neq e}(y_{en}-y_{e'n})^2 \\ &= \frac{1}{NE^2}\sum_{n=1}^{N}\sum_{e=1}^{E}\sum_{e'\neq e}[(y_{en}-\hat{y}_{\cdot n})-(y_{e'n}-\hat{y}_{\cdot n})]^2, \end{aligned} \tag{2.17}$$

where N specifies the number of initial states, E specifies the number of ensemble members

at a fixed initial state, $\hat{y}_{\cdot n}$ is an estimate of the ensemble mean defined in (2.9). By using the fact that

$$\sum_{e' \neq e} (y_{e'n} - \hat{y}_{\cdot n}) = (E\hat{y}_{\cdot n} - y_{en}) - (E-1)\hat{y}_{\cdot n} = \hat{y}_{\cdot n} - y_{en} \tag{2.18}$$

and

$$\sum_{n=1}^{N}\sum_{e=1}^{E}\sum_{e' \neq e}(y_{e'n} - \hat{y}_{\cdot n})^2 = (E-1)\sum_{n=1}^{N}\sum_{e'=1}^{E}(y_{e'n} - \hat{y}_{\cdot n})^2, \tag{2.19}$$

$\hat{MSE}$ in (2.17) is simplified as

$$\hat{MSE} = \frac{2}{NE}\sum_{n=1}^{N}\sum_{e=1}^{E}(y_{en} - \hat{y}_{\cdot n})^2 = 2\hat{\sigma}_N^2, \tag{2.20}$$

where (2.11) has been used. This equation shows that MSE also can be estimated as twice the mean square difference between ensemble member and ensemble mean, and this in turn equals twice the estimated "noise". Typically, $\hat{MSE}$ increases with lead time and eventually reaches a saturation value as lead time approaches infinity. This saturation value can be understood as follows. As the lead time approaches infinity, the estimated "noise variance" $\hat{\sigma}_N^2$ approaches climatological variance. Based on (2.20), the saturation value of $\hat{MSE}$ is twice the climatological variance, estimated by $2\hat{\sigma}_T^2$. When $\hat{MSE}$ reaches the saturation value, the system becomes unpredictable since the forecast provides no better a prediction than a randomly chosen state from the system. However, the absolute $\hat{MSE}$ is not a measure of predictability, because its value is strongly dictated by the natural variance of the variables in question. We anticipate that $\hat{MSE}$ saturates at $2\hat{\sigma}_T^2$. Therefore, predictability is determined by how far $\hat{MSE}$ is relative to $2\hat{\sigma}_T^2$. This suggests the measure

of predictability P estimated by

$$\hat{P} = 1 - \frac{\hat{MSE}}{2\hat{\sigma}_T^2} = 1 - \frac{\hat{\sigma}_N^2}{\hat{\sigma}_T^2} = \frac{\hat{SNR}}{1 + \hat{SNR}}, \tag{2.21}$$

where (2.11), (2.12), (2.13), (2.14) and (2.20) have been used. A perfect forecast corresponds to $P = 1$, while the loss of all predictability corresponds to $P = 0$. Testing the null hypothesis of no predictability is tantamount to testing $P = 0$, which implies the MSE equals twice the climatological variance. Equation (2.21) shows that $\hat{P}$ (or $\hat{MSE}$) has a one-to-one relation with $\hat{SNR}$. Testing the statistical significance of $\hat{P}$ (or $\hat{MSE}$) is equivalent to testing statistical significance of $\hat{SNR}$.

2.2.3 Correlation between ensemble members

Another measure of predictability that has been used in the literature is the correlation between one ensemble member and all the other ensemble members. Consider the classical ensemble forecasts described in the beginning of section 2.2. The squared correlation at a given lead time is computed for all possible pairs of ensemble members, which is

$$\hat{r}^2 = \frac{\frac{1}{NE(E-1)} \sum_{n=1}^{N} \sum_{e=1}^{E} \sum_{e' \neq e} (y_{en} - \hat{y}_{..}) (y_{e'n} - \hat{y}_{..})}{\frac{1}{NE} \sum_{n=1}^{N} \sum_{e=1}^{E} (y_{en} - \hat{y}_{..})^2}, \tag{2.22}$$

where N specifies the number of initial states, E specifies the number of ensemble members in a fixed initial state, $\hat{y}_{..}$ is the grand mean defined in (2.10). By making use of the fact that

$$\sum_{e' \neq e} (y_{e'n} - \hat{y}_{..}) = E (\hat{y}_{.n} - \hat{y}_{..}) - (y_{en} - \hat{y}_{..}), \tag{2.23}$$

$\hat{r}^2$ can be simplified as

$$\begin{aligned}
\hat{r}^2 &= \frac{1}{NE(E-1)\hat{\sigma}_T^2} \sum_{n=1}^{N} \sum_{e=1}^{E} (y_{en} - \hat{y}_{..}) \left(E\left(\hat{y}_{.n} - \hat{y}_{..}\right) - (y_{en} - \hat{y}_{..}) \right) \\
&= \frac{1}{NE(E-1)\hat{\sigma}_T^2} \left(E^2 \sum_{n=1}^{N} \left(\hat{y}_{.n} - \hat{y}_{..}\right)^2 - \sum_{n=1}^{N} \sum_{e=1}^{E} \left(y_{en} - \hat{y}_{..}\right)^2 \right) \\
&= \frac{E\hat{\sigma}_S^2 - \hat{\sigma}_T^2}{(E-1)\hat{\sigma}_T^2} = \frac{S\hat{N}R - \frac{1}{E-1}}{S\hat{N}R + 1},
\end{aligned} \tag{2.24}$$

where (2.8), (2.9), (2.10), (2.11), (2.12), (2.13) and (2.14) have been used. Equation (2.24) shows that $\hat{r}^2$ is monotonically related to $S\hat{N}R$. Testing the statistical significance of $\hat{r}^2$ is equivalent to testing the statistical significance of $S\hat{N}R$. It is noteworthy that the three measures of predictability $S\hat{N}R$, $M\hat{S}E$ and $\hat{r}^2$ are equivalent as shown in (2.21) and (2.24).

2.2.4 Multiple correlation

The measures of predictability discussed above are applicable to ensemble forecasts. When ensemble forecasts are not available, multiple correlation is often used as a measure of predictability. To define the multiple correlation between a single predictand $\mathbf{y}$ and multiple predictors in $\mathbf{X}$, we introduce the linear regression model

$$\begin{array}{ccccccc}
\mathbf{y} & = & \mathbf{X} & \mathbf{b} & + & \boldsymbol{\epsilon}, \\
[N \times 1] & & [N \times K] & [K \times 1] & & [N \times 1]
\end{array} \tag{2.25}$$

where $\mathbf{y}$ is the centered predictand vector, $\mathbf{X}$ is the centered predictor matrix, which consists of a set of N-dimensional predictors $\mathbf{x}_1$, $\mathbf{x}_2$, ... , $\mathbf{x}_K$, $\mathbf{b}$ is the vector of regression coefficients and $\boldsymbol{\epsilon}$ denotes the noise. The vector of regression coefficients can be estimated by classical

least squares as

$$\hat{\mathbf{b}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}, \tag{2.26}$$

where the superscript T denotes the transpose operation. The multiple correlation coefficient is defined as the correlation between predictand $\mathbf{y}$ and its predicted value $\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}}$, which is generally written as

$$\hat{R} = \frac{(\mathbf{X}\hat{\mathbf{b}})^{\mathrm{T}}\mathbf{y}}{\sqrt{(\mathbf{y}^{\mathrm{T}}\mathbf{y})\left((\mathbf{X}\hat{\mathbf{b}})^{\mathrm{T}}(\mathbf{X}\hat{\mathbf{b}})\right)}}. \tag{2.27}$$

Invoking (2.26), $(\mathbf{X}\hat{\mathbf{b}})^{\mathrm{T}}(\mathbf{X}\hat{\mathbf{b}})$ can be written as

$$(\mathbf{X}\hat{\mathbf{b}})^{\mathrm{T}}(\mathbf{X}\hat{\mathbf{b}}) = \hat{\mathbf{b}}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\hat{\mathbf{b}} = \hat{\mathbf{b}}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y} = (\mathbf{X}\hat{\mathbf{b}})^{\mathrm{T}}\mathbf{y}. \tag{2.28}$$

Substituting (2.28) into the numerator of (2.27) and then squaring the multiple correlation gives

$$\hat{R}^2 = \frac{(\mathbf{X}\hat{\mathbf{b}})^{\mathrm{T}}(\mathbf{X}\hat{\mathbf{b}})}{\mathbf{y}^{\mathrm{T}}\mathbf{y}} = \frac{\hat{\mathbf{\Sigma}}_{xy}\hat{\mathbf{\Sigma}}_{xx}^{-1}\hat{\mathbf{\Sigma}}_{xy}^{\mathrm{T}}}{\hat{\mathbf{\Sigma}}_{yy}}, \tag{2.29}$$

where $\hat{\mathbf{\Sigma}}_{xx} = \frac{1}{N}\mathbf{X}^{\mathrm{T}}\mathbf{X}$, $\hat{\mathbf{\Sigma}}_{yy} = \frac{1}{N}\mathbf{y}^{\mathrm{T}}\mathbf{y}$, $\hat{\mathbf{\Sigma}}_{xy} = \frac{1}{N}\mathbf{y}^{\mathrm{T}}\mathbf{X}$, and (2.26) has been used. If there is no predictability, then $\mathbf{y}$ is independent of $\mathbf{X}$ in the linear regression model (2.25), which implies $\mathbf{b} = 0$. To test the null hypothesis $\mathbf{b} = 0$, we calculate the statistic F

$$F = \frac{\hat{R}^2}{1-\hat{R}^2}\frac{N-K-1}{K}. \tag{2.30}$$

If the null hypothesis of no predictability is true, the statistic F has an F distribution with K and $N-K-1$ degrees of freedom. Large values of F favor rejection of the null hypothesis.

We can show that the squared multiple correlation $\hat{R}^2$ is related to $S\hat{N}R$ too. Although

there is no ensemble forecasts, we can conceive of generating an "ensemble" from the linear regression model (2.25). According to the the law of total variance described in (2.4), the "signal variance" is the variance of $\mathrm{E}[\mathbf{y}|\mathbf{x}]$. Based on the linear regression model (2.25), $\mathrm{E}[\mathbf{y}|\mathbf{x}]$ is estimated by $\mathbf{X}\hat{\mathbf{b}}$, hence the "signal variance" is estimated by

$$\hat{\sigma}_S^2 = \frac{(\mathbf{X}\hat{\mathbf{b}})^{\mathrm{T}}(\mathbf{X}\hat{\mathbf{b}})}{N} = \hat{\mathbf{\Sigma}}_{xy}\hat{\mathbf{\Sigma}}_{xx}^{-1}\hat{\mathbf{\Sigma}}_{xy}^{\mathrm{T}}. \tag{2.31}$$

The total variance is the variance of $\mathbf{y}$, estimated by

$$\hat{\sigma}_T^2 = \frac{\mathbf{y}^{\mathrm{T}}\mathbf{y}}{N} = \hat{\mathbf{\Sigma}}_{yy}. \tag{2.32}$$

The $\hat{R}^2$ in (2.27) can be written as

$$\hat{R}^2 = \frac{\hat{\sigma}_S^2}{\hat{\sigma}_T^2} = \frac{S\hat{N}R}{S\hat{N}R + 1}, \tag{2.33}$$

where (2.13), (2.14), (2.31) and (2.32) have been used. Equation (2.33) shows that $\hat{R}^2$ has a one-to-one relation with $S\hat{N}R$. The predictability measured by $\hat{R}^2$ is equivalent to the predictability measured by $S\hat{N}R$, provided the ensembles are generated by a linear regression model that predicts $\mathbf{y}$ given $\mathbf{X}$.

2.2.5 Autocorrelation

Another measure used in the predictability literature is autocorrelation. The autocorrelation of a variable measures the dependence of the variable with itself at two distinct times. It is useful for measuring predictability due to "memory" or "persistence". Let $\mathbf{y}_t = [y_1\ y_2\ \ldots\ y_N]$ denote a N-dimensional centered time series at time t, and $\mathbf{y}_{t+\tau} = [y_{1+\tau}\ y_{2+\tau}\ \ldots\ y_{N+\tau}]$ denote the centered time series at time $t+\tau$, where τ is the lead time. The autoregressive

model is defined as

$$\mathbf{y}_{t+\tau} = b\mathbf{y}_t + \boldsymbol{\epsilon}, \tag{2.34}$$

where b denotes the regression coefficient, and $\boldsymbol{\epsilon}$ denotes noise. If the time series is stationary, the autocorrelation between $\mathbf{y}_{t+\tau}$ and $\mathbf{y}_t$ can be estimated by

$$\hat{\rho}_\tau = \frac{\mathbf{y}_t^{\mathrm{T}}\mathbf{y}_{t+\tau}}{\mathbf{y}_t^{\mathrm{T}}\mathbf{y}_t}. \tag{2.35}$$

If $\mathbf{y}_{t+\tau}$ is independent of $\mathbf{y}_t$ (i.e., no predictability), then the regression coefficient $b = 0$. To test the null hypothesis $b = 0$, we use the statistic

$$F = \frac{\hat{\rho}_\tau^2}{1 - \hat{\rho}_\tau^2}(N - 2), \tag{2.36}$$

which has an F distribution with 1 and $N - 2$ degrees of freedom.

The autoregressive model (2.34) is a special case of the multiple regressive model (2.25). It can be shown that $\hat{\rho}_\tau^2 = \hat{R}^2$ for this particular model. It follows that $\hat{\rho}_\tau$ also is related to $\hat{SNR}$ of an autoregressive model.

All measures of predictability reviewed in the above subsections are fundamentally related to $\hat{SNR}$. In other words, all these measures are measuring "the same thing". Fundamentally, there is only one measure of predictability, and any measure can be derived from any other measure (for fixed number of ensemble members and number of initial states).

2.3 Predictability on various time scales

Predictability measure $\hat{SNR}$ is found empirically to decay with lead time. However, different components within the climate system are found to decay at different rates. Generally, predictability of climate variables can be categorized as weather predictability, seasonal predictability, multi-year to decadal predictability and centennial predictability based on

the decaying rates.

Weather predictability refers to the predictability of day-to-day changes of weather system, which is induced by the non-linear, chaotic dynamics in the atmosphere. Given an initial perturbation in the weather system, the uncertainties in the initial perturbation will grow and contaminate the numerical simulation although the forecast model is "perfect". The weather system is then unpredictable when the forecast error saturates (Lorenz, 1963; Meehl et al., 2009). It is generally accepted that the upper limit of weather predictability is at most one month (Shukla, 1981).

However, monthly mean or seasonal mean atmospheric conditions are predictable several months ahead due to the coupled interactions with the slowly varying components in the climate system, including ocean, land, sea ice and so on. The slowly varying climate components integrate weather variability such that the climate is able to produce long time-scale variations of considerable magnitude without external forcings (as defined in section 3.1). The slowly varying climate components play a critical role for seasonal atmospheric predictability. The variability produced by coupled interactions is thought of as internal variability in the climate system.

As for the climate predictions on multi-year to decadal time scales, the climate response to external natural and anthropogenic forcings emerges. The response becomes more dominant for longer time scales. Therefore, decadal predictability depends on external forcings in addition to internal dynamical processes in the climate system.

The primary focus of this study is the predictability on multi-year time scales. We argue that there is no clear separation between multi-year and decadal predictability. For example, decadal predictability that exists in the 10-year mean might come from predictability in the first 3 years, say. This situation is analogous to seasonal predictability, in which predictability in the first two to four weeks dominates the predictability of the seasonal mean. In this sense, multi-year predictability could indicate decadal predictability, for some measures of decadal predictability. This point is relevant to the interpretation of the literature. For instance, Pohlmann et al. (2006) suggested that surface air temperature in Europe has

decadal predictability. However, this conclusion was based on analysis of 10-year means. It is conceivable that the true time scale of this predictability is on the order of 3-5 years, but that predictability on this time scale can cause 10-year means to be predictable. Conversely, predictability identified on 3-5 year time scales could imply predictability of 10-year means.

2.4 Importance of decadal predictability

Decadal predictability, especially over land, is important because decadal climate fluctuations have significant impacts on society and economy. Therefore, decadal predictability is of particular interest to policy makers for their decisions on energy, agriculture, water, resource management and infrastructure investment (Cane, 2010). For example, the Sahel region of Africa experienced a severe drought from the 1950's to the 1980's, which affected environment and food production. Conceivably, a good decadal prediction of the Sahel rainfall could have helped the policy makers to allocate resources to minimize the costs. Therefore, any skill in decadal predictions would be precious.

2.5 Scientific basis for atmospheric decadal predictability

Observations over the past century show that surface temperatures have warmed in a statistically significance sense over most of the Earth's surface (Trenberth et al., 2007, sec. 3.2.2.7). In addition, precipitation generally has increased over most land areas in the extratropics and decreased in the tropics over the past century (Trenberth et al., 2007, sec. 3.3.2.2). A basic question is whether these long-term variations are predictable. It is generally accepted that the atmosphere alone (i.e., under fixed boundary conditions) is predictable for at most one month (Shukla, 1981). Therefore, any atmospheric predictability beyond one month must arise from slowly varying components of the climate system or from predictable external forcings.

2.5.1 Predictability due to response to predictable external forcing

An example of the atmospheric response to external forcings is illustrated in Fig.2.2. This figure shows continental and global scale changes in surface temperature from observations and simulations in climate models. Blue shaded bands denote the 5-95% range for 19 simulations from five climate models using only natural forcings of solar activity and volcanoes. Red shaded bands denote the 5-95% range for 58 simulations from 14 climate models including both anthropogenic and natural forcings. The black line indicates the observed temperature change. In this figure, the temperature changes are unlikely due to natural forcings if the blue band does not encompass the black line at a 10% significance level. The temperature changes are attributable to anthropogenic forcings if the red band encompasses the black line at a 10% significance level. Generally, "detection" is defined as the process of demonstrating that climate has changed in some defined statistical sense without providing a reason for that change, and "attribution" of climate change is the process of establishing the most likely causes for the detected changes with some defined level of confidence (Hegerl et al., 2007). Fig.2.2 figure indicates that models with only natural forcings fail to reproduce the warming observed in recent decades. However, models can simulate the observed 20th century changes in temperature when including both anthropogenic and natural forcings. Therefore, the temperature changes are detectable and attributable to the anthropogenic forcings. If the anthropogenic forcings are known, then the long-term changes in temperature are predictable.

According to the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (Hegerl et al., 2007), greenhouse gas forcing has very likely caused most of the observed global warming over the past few decades. The surface temperature increases in each continent except Antarctic since the middle of the 20th century is likely due to external forcing, especially anthropogenic forcing. The climate change is predictable to the extent that external forcing is predictable.

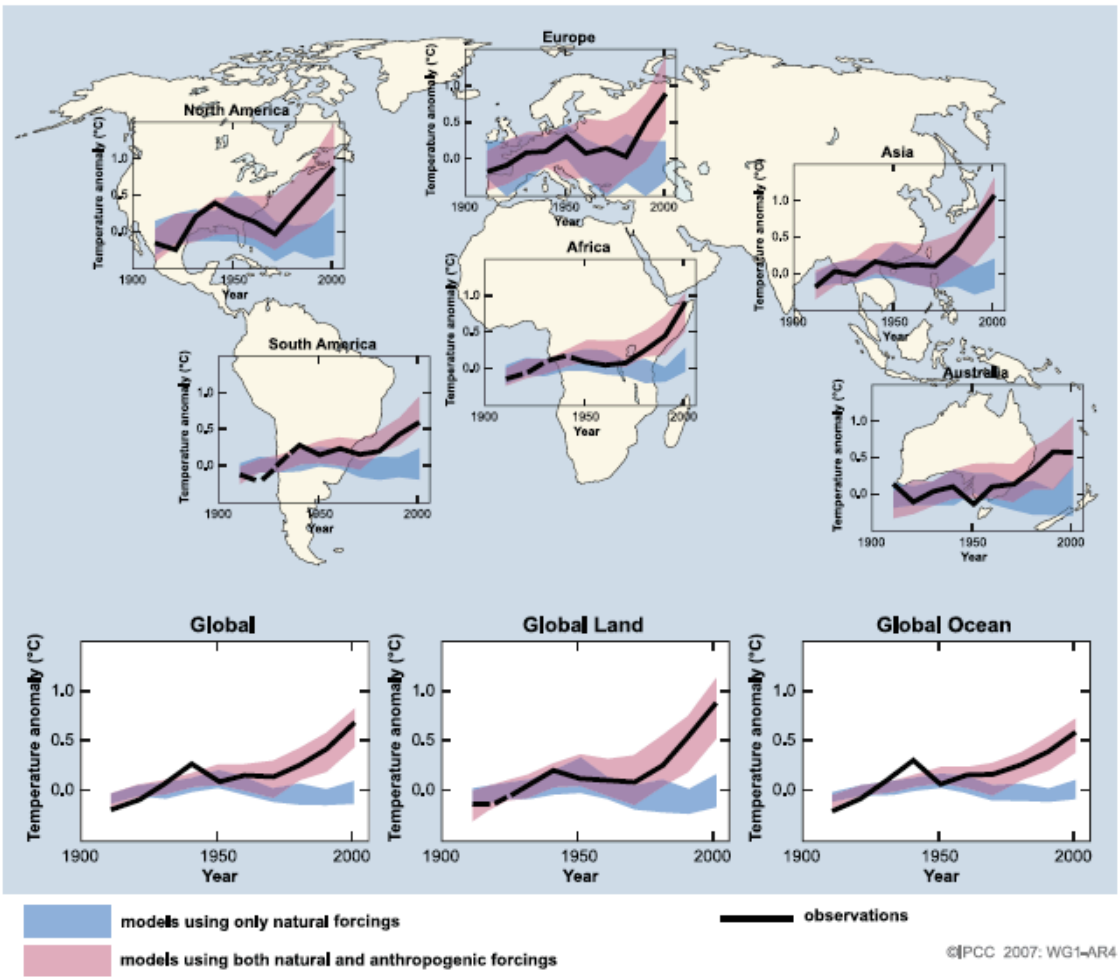

Figure 2.2: Temperature changes relative to the corresponding average for 1901-1950 (°C) from decade to decade from 1906 to 2005 over the Earth's continents, as well as the entire globe, global land area and the global ocean (lower graphs). The black line indicates observed temperature change, while the coloured bands show the combined range covered by 90% of recent model simulations. Red indicates simulations that include natural and human factors, while blue indicates simulations that include only natural factors. Dashed black lines indicate decades and continental regions for which there are substantially fewer observations. From Hegerl et al. (2007), Fig.1 of FAQ 9.2.

2.5.2 Predictability due to slowly varying climate components

In addition to the predictability due to the responses to external forcings, the atmospheric predictability also can be caused by responding to (or coupling with) the predictable slowly varying climate components. Among the climate components, ocean has been shown to be the primary factor for atmospheric predictability on decadal time scales (Compo and Sardeshmukh, 2009; Held et al., 2005; Latif and Barnett, 1994).

Various studies have shown that internal oceanic variations are potentially predictable on decadal time scales, and the atmosphere responds to these oceanic variations. By "potential", we mean that the predictability was identified in models and not confirmed in observations. The predictability of Atlantic meridional overturning circulation (AMOC) has been widely studied because it significantly contributes to the oceanic northward heat transport, and consequently plays a key role in maintaining the mean climate of the Earth (Trenberth and Caron, 2001). The AMOC is defined as the part of the ocean's circulation in which warm, saline surface water flows northward in the Atlantic basin and the dense deep water flows southward out of the Atlantic basin and into the Southern Ocean (Wunsch, 2002). The variations in the AMOC were found to be predictable out to a decade or more in atmosphere-ocean general circulation models(Collins and Sinha, 2003; Collins et al., 2006; Msadek et al., 2010), and have a significant influence on decadal atmospheric variability (Pohlmann et al., 2006; Stouffer et al., 2006, 2007; Timmermann and Latif, 1998). A particularly dramatic example of how AMOC changes can lead to decadal changes in the atmosphere is described in Stouffer et al. (2006). In this study, an extremely large freshwater flux was applied to the North Atlantic in multiple atmosphere-ocean general circulation models, which weakened and eventually shut down the AMOC. The mean response of global surface air temperature (i.e., temperature of air above land or ocean surface, abbreviated to SAT) after AMOC shut down (Fig.2.3) shows a cooling of 10° in North Atlantic and about 3° warming in tropical Atlantic. In addition, Stouffer et al. (2007) found the ITCZ shifted southward and precipitation decreased in extratropical Northern hemisphere (not

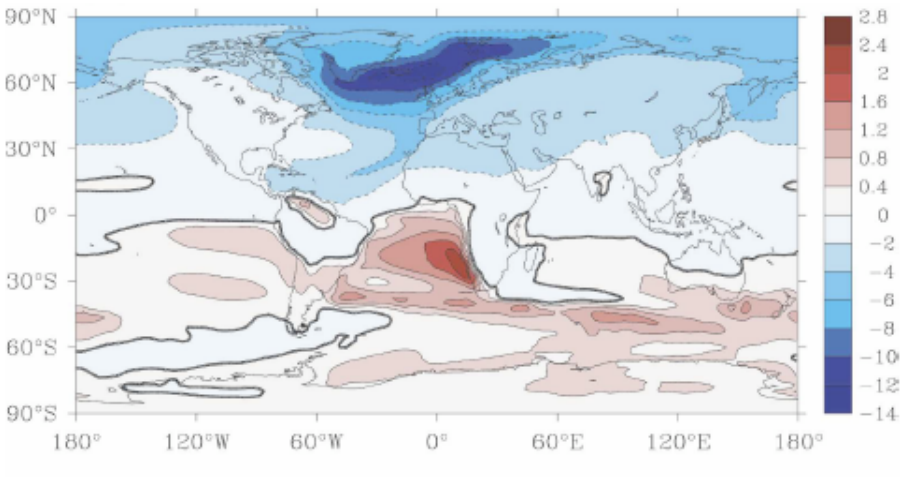

Figure 2.3: SAT anomalies after the shutdown of AMOC in ensemble mean of atmosphere-ocean general circulation models. From Stouffer et al. (2006), Fig. 14.

shown). Although the experiments in these studies are idealized, they support the hypothesis that variations in the AMOC cause variations in the atmosphere. Since the AMOC is predictable, these atmospheric responses to changes in AMOC are predictable.

The predictability of other oceanic variables and the atmospheric responses to oceanic variations have been found. For instance, Griffies and Bryan (1997) showed that dynamic topography and sea surface salinity in North Atlantic are predictable on the order of 10-20 years in numerical simulations with a coupled atmosphere-ocean general circulation model (AOGCM). DelSole et al. (2011) showed that SSTs in North Atlantic and North Pacific are predictable for a decade or so in AOGCMs using a statistical optimization method that will be discussed in chapter 3. The role of SSTs in the decadal variability of Sahel droughts, Atlantic hurricane activity and precipitation anomalies in the United States has been found (Held et al., 2005; Knight et al., 2006; Schubert et al., 2004; Zhang and Delworth, 2006). These decadal variations are predictable provided the SSTs are predictable.

In addition to the atmospheric response to the variations in the ocean, the coupled interactions between ocean and atmosphere also provides the scientific basis for decadal predictability. A striking example of mid- and high-latitude ocean-atmosphere coupling was described by Latif and Barnett (1996) using an AOGCM and observations. They suggested a cycle of unstable air-sea interactions involving the North Pacific subtropical gyre and the

Aleutian low-pressure system, although this coupled interactions have not been robustly reproduced by other models.

2.6 Previous studies on decadal predictability of atmospheric internal variability

We now review the studies on decadal predictability of atmosphere internal variability (i.e., not due to external forcing), but omitting discussion of studies on externally forced variability, because the externally forced variability has been well documented.

Many studies have shown evidence of predictability of atmospheric internal variability on decadal time scales, although the specific mechanisms were not clarified. Boer (2004) and Boer and Lambert (2008) diagnosed the potential predictability of SAT and precipitation in control simulations of multiple coupled models with fixed external forcings, and found potential predictability of 5-, 10-, and 25-year means predominately in high latitude oceans. The "potential predictability" in those studies refers to the fraction of long-term variability that may be distinguished from "noise", as estimated from analysis-of-variance-type calculations, modified to account for autocorrelation in the time series. The long-term variability is predictable given sufficient knowledge of physical mechanism, the availability of initial conditions and a significant amount of variability (Boer, 2000). Fig.2.4 from Boer and Lambert (2008) shows the percent of potential predictable variance of 5-year average for SAT and precipitation. In temperature, the most predictable regions are the North Atlantic, North Pacific and Southern Ocean, while there is weak predictability over land. The predictability of precipitation is an "attenuated" version of temperature predictability, with little to no predictability over land. A similar analysis has been applied to 500-year-long control simulations of an AOGCM, and high potential predictability of 10- and 20-year means of SAT was found in North Atlantic (Pohlmann et al., 2004). Predictability was also found for 925-hPa and 850-hPa temperature over North Atlantic in the first two decades in an AOGCM using classical ensemble methods (Pohlmann et al., 2004), namely by perturbing

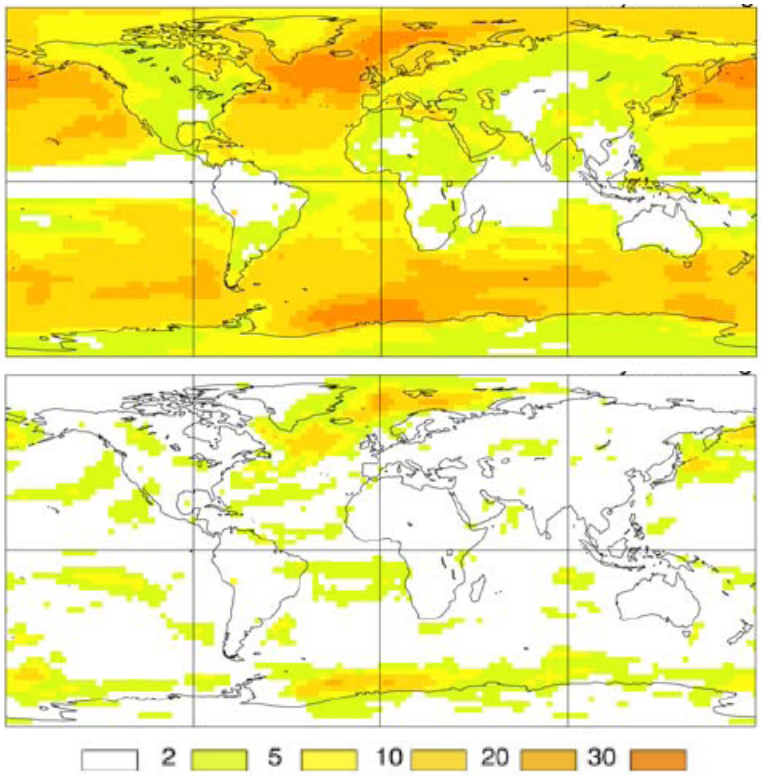

Figure 2.4: Percent of potential predictable variance of 5-year average for SAT (upper) and precipitation (bottom). From Boer and Lambert (2008), Fig. 4.

atmospheric initial conditions of control runs to produce ensemble forecasts. Collins (2002) studied the predictability in an AOGCM using classical predicability measure, and showed that 10-year means of SAT anomalies are predictable over North Atlantic and Southern Ocean, and 10-year average sea level pressure anomalies are predictable over large areas of the globe.

In spite of the evidence of atmospheric decadal predictability, there is little to no evidence for atmospheric predictability of internal variability over *land* on decadal time scales. In the studies described above, Pohlmann et al. (2004) found decadal predictability of SAT only in small maritime-influenced regions of Europe based on classical predictability measure. Boer and Lambert (2008) found very weak potential predictability of temperature in limited areas over land, and almost no predictability of precipitation over land on 5- and 10-year time scales in the control runs of multiple coupled models. Collins (2002) found little

potential predictability beyond seasonal time scales in extratropical land temperature and precipitation. Although Pohlmann et al. (2006) suggested decadal predictability of SAT in Europe in the control simulation of an AOGCM, it was based on 10-year means. As discussed at the end of section 2.3, the use of long time averaging of data obscures the exact time scale of predictability.

2.7 Motivation of this study

Previous studies on atmospheric decadal predictability are limited in many ways. Most studies are limited by the use of a single model, which raises questions about whether the predictability can be verified in other models. Temporal averaging obscures the exact time scale of predictability. Another potential limitation of previous studies is that the measures of predictability are *univariate*. Conceivably, some large-scale spatial structure of temperature might be predictable over land, but this predictability cannot be detected using a univariate analysis because unpredictable noise dominates on grid scales. To identify large scale predictable structures, spatial filters must be constructed to remove unpredictable noise. If the predictable structure is spatially uniform and the noise is white in space and time, then spatial averaging can reduce the small-scale noise without affecting the large-scale predictable signal, thereby allowing predictability to be detected. However, simple spatial averaging does not necessarily improve the signal-to-noise ratio, especially if the component has dipole structure. Construction of optimal spatial filters is straightforward when the predictable component is known (Hasselmann, 1979), but unfortunately predictable components over land are unknown.

The limitations in previous studies raise the question as to whether there exists a better method that identifies patterns over land that are predictable on decadal time scales. Several *multivariate* methods have been used to diagnose spatial structures of climate variables such as empirical orthogonal function (EOF) analysis, singular spectrum analysis and extended EOF analysis. These methods maximize particular aspects of climate variability. However, in predictability studies, we would like to maximize some measure of predictability.

Recently, DelSole and Tippett (2009) proposed a statistical optimization method for finding components that are predictable on the longest possible time scales. This method maximizes a quantity, called average predictability time (APT), and involves fitting a multivariate regression model to the leading principal components and then estimating the predictability of the resulting linear model. It turns out that naive application of this method fails to identify predictability over land. We will discuss this fact in the next chapter and develop a generalization of the method that overcomes the problem. The connections between this method and some other multivariate methods will be discussed in the end of section 3.1.

Chapter 3: Identification of Unforced Predictability over Land on Continental Scales

3.1 Introduction to generalized average predictability time

DelSole and Tippett (2009) proposed a new method for determining components that maximize predictability. To determine the components that maximizes predictability, one must first define a measure of predictability. Consider a set of ensemble forecasts generated in the same way as described in section 2.2. The only difference is that the data produced by the ensemble forecasts is not only a function of initial state t, lead time τ and ensemble member e, but also a function of space. In this context, a standard measure of predictability is

$$P(\tau) = \frac{\sigma_\infty^2 - \sigma_\tau^2}{\sigma_\infty^2}, \tag{3.1}$$

where σ_τ^2 is the forecast variance at lead time τ, averaged over all initial states t. As lead time increases, the forecast variance tends to increase and approaches the climatological variance σ_∞^2 as lead time τ approaches infinity. The measure $P(\tau)$ typically is close to one initially and decreases with lead time until it vanishes when the forecast variance equals the climatological variance.

We can show that the predictability measure $P(\tau)$ is related to SNR. In the context of a set of ensemble forecasts, the forecast variance σ_τ^2 at lead time τ can be estimated by the "noise variance" $\hat{\sigma}_N^2$, and the climatological variance can be estimated by the total variance

$\hat{\sigma}_T^2$. Therefore, the predictability measure $P(\tau)$ can be estimated by

$$\hat{P}(\tau) = \frac{\hat{\sigma}_T^2 - \hat{\sigma}_N^2}{\hat{\sigma}_T^2} = 1 - \frac{\hat{\sigma}_N^2}{\hat{\sigma}_T^2} = \frac{S\hat{N}R}{1 + S\hat{N}R}, \tag{3.2}$$

where (2.13) and (2.14) have been used.

The predictability measure $P(\tau)$ is a function of lead time τ. To avoid the ambiguity as to the proper choice of lead time, we integrate (3.1) over lead times to construct a measure of predictability that is independent of lead time. Accordingly, we define APT by

$$APT = 2\int_0^\infty \left(\frac{\sigma_\infty^2 - \sigma_\tau^2}{\sigma_\infty^2}\right)\, d\tau. \tag{3.3}$$

In general, the integral of a predictability measure is called an integral time scale. In our case, multiplying the integrated measure by two yields a time scale that agrees with the usual e-folding time for an exponentially decaying forecast signal. One could define an alternative time scale as the time beyond which the prediction error exceeds some predefined threshold (Lorenz, 1965), but this definition is problematic because it depends on an arbitrary threshold, and the time at which this threshold is exceeded can be very sensitive to sampling variability. An advantage of defining time scale by APT is that it is simple to understand and can be optimized by standard methods.

For discrete time, APT can be written as

$$APT = 2\sum_{\tau=1}^{\infty} \left(\frac{\sigma_\infty^2 - \sigma_\tau^2}{\sigma_\infty^2}\right) \Delta\tau. \tag{3.4}$$

The APT has units of time step. The summation in (3.4) starts from lead time $\tau = 1$ rather than $\tau = 0$, because that even unpredictable white noise has an APT of 2 if the summation starts at $\tau = 0$. It makes more sense to let APT vanish at $\tau = 0$ for unpredictable white noise. For arbitrary stochastic processes, APT may be unbounded. But in practice, the

summation has a finite limit, which implies that the estimated APT is finite. Theoretically, infinite APT implies $STR \neq 0$ at $\tau = \infty$, which in turn implies $E[\mathbf{y}|\mathbf{x}] \neq E[\mathbf{y}]$, and hence $\mathbf{y}$ is predictable at infinite time.

To maximize APT, we would like to find the projection vector $\mathbf{q}$ such that $\mathbf{q}^{\mathrm{T}}\mathbf{y}(\tau, t, e)$ maximizes APT, where the superscript T denotes the transpose operation, $\mathbf{y}(\tau, t, e)$ denotes the state vector at fixed lead time, initial state and ensemble member. Each element of the vector $\mathbf{y}(\tau, t, e)$ specifies a spatial parameter. $\mathbf{q}^{\mathrm{T}}\mathbf{y}(\tau, t, e)$ can be interpreted as the linear combination of variables in $\mathbf{y}(\tau, t, e)$, with weighting coefficients $\mathbf{q}$. The component $\mathbf{q}^{\mathrm{T}}\mathbf{y}(\tau, t, e)$ has forecast variance at lead time τ

$$\sigma_\tau^2 = \mathbf{q}^{\mathrm{T}}\overline{(\mathbf{y}(\tau, t, e) - \langle \mathbf{y}(\tau, t, e) \rangle)\,(\mathbf{y}(\tau, t, e) - \langle \mathbf{y}(\tau, t, e) \rangle)^{\mathrm{T}}}\mathbf{q} = \mathbf{q}^{\mathrm{T}}\hat{\mathbf{\Sigma}}_\tau \mathbf{q}, \tag{3.5}$$

where the angle brackets denote the average over ensemble members, the overline denotes the average over initial states, and $\hat{\mathbf{\Sigma}}_\tau$ denotes an estimate of the forecast covariance matrix at lead time τ. The climatological variance is just the forecast variance as the lead time approaches infinity. It is denoted as

$$\sigma_\infty^2 = \mathbf{q}^{\mathrm{T}}\hat{\mathbf{\Sigma}}_\infty \mathbf{q}, \tag{3.6}$$

where $\hat{\mathbf{\Sigma}}_\infty$ is an estimate of the climatological covariance matrix. Substituting (3.5) and (3.6) into (3.4) gives

$$APT = 2\sum_{\tau=1}^{\infty}\left(\frac{\mathbf{q}^{\mathrm{T}}(\hat{\mathbf{\Sigma}}_\infty - \hat{\mathbf{\Sigma}}_\tau)\mathbf{q}}{\mathbf{q}^{\mathrm{T}}\hat{\mathbf{\Sigma}}_\infty \mathbf{q}}\right)\Delta\tau. \tag{3.7}$$

This is a Rayleigh quotient. We show in appendix B that maximizing the Rayleigh quotient leads to the eigenvalue problem

$$2\sum_{\tau=1}^{\infty}(\hat{\mathbf{\Sigma}}_\infty - \hat{\mathbf{\Sigma}}_\tau)\Delta\tau\mathbf{q} = \lambda\hat{\mathbf{\Sigma}}_\infty \mathbf{q}. \tag{3.8}$$

The matrices $\hat{\mathbf{\Sigma}}_\infty - \hat{\mathbf{\Sigma}}_\tau$ and $\hat{\mathbf{\Sigma}}_\infty$ are symmetric. It is shown in appendix C that the eigenvectors of (3.8) produce components that are uncorrelated. Recall that eigenvectors are unique up to a multiplicative constant. To fix the amplitude of the eigenvectors, we normalize the climatological variance of the component to unit, i.e., $\sigma_\infty^2 = \mathbf{q}^{\mathrm{T}} \hat{\mathbf{\Sigma}}_\infty \mathbf{q} = 1$. The eigenvalues of (3.8) are the APT values corresponding to each component. It is convention to order the eigenvalues and their associated eigenvectors in decreasing order. It can be shown that the first component maximizes APT, the second component maximizes APT subject to being uncorrelated with the first, and so on. The above eigenvalue problem (3.8) gives a set of projection vectors $\mathbf{q}_1$, $\mathbf{q}_2$, ..., $\mathbf{q}_M$. We write the set of eigenvectors into a matrix

$$\mathbf{Q} = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \dots \quad \mathbf{q}_M]. \tag{3.9}$$

According to this notation, the component corresponding to each eigenvector can be written as $\mathbf{Q}^{\mathrm{T}}\mathbf{y} = [\mathbf{q}_1^{\mathrm{T}}\mathbf{y} \quad \mathbf{q}_2^{\mathrm{T}}\mathbf{y} \quad \dots \quad \mathbf{q}_M^{\mathrm{T}}\mathbf{y}]$.

Recall that the climatological variance of the component is normalized to unit variance, which implies

$$\mathbf{Q}^{\mathrm{T}} \hat{\mathbf{\Sigma}}_\infty \mathbf{Q} = \mathbf{I}. \tag{3.10}$$

To derive the spatial pattern, we would like to decompose $\mathbf{y}(\tau, t, e)$ in terms of components that maximize APT. Then we seek the matrix $\mathbf{P}$ that minimizes the mean square difference between $\mathbf{y}(\tau, t, e)$ and the linear combination of components $\mathbf{P}\mathbf{Q}^{\mathrm{T}}\mathbf{y}(\tau, t, e)$:

$$\overline{\langle ||\mathbf{y}(\tau, t, e) - \mathbf{P}\left(\mathbf{Q}^{\mathrm{T}}\mathbf{y}(\tau, t, e)\right)||^2 \rangle}, \tag{3.11}$$

where $||.||^2$ denotes the square of the Frobenius norm of a matrix, the angle brackets denote an average over ensemble members, and the overline indicates an averaging over initial states. The minimum value of (3.11) is obtained for the choice of

$$\mathbf{P} = (\mathbf{Q}^{\mathrm{T}})^{-1}. \tag{3.12}$$

Multiplying both sides of (3.10) with $(\mathbf{Q}^{\mathrm{T}})^{-1}$ gives

$$\hat{\mathbf{\Sigma}}_{\infty}\mathbf{Q} = (\mathbf{Q}^{\mathrm{T}})^{-1} = \mathbf{P}. \tag{3.13}$$

Each column of matrix $\mathbf{P}$ represents a spatial pattern associated with a particular component. $\mathbf{P}$ can be written as

$$\mathbf{P} = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \ldots \quad \mathbf{p}_M] = [\hat{\mathbf{\Sigma}}_{\infty}\mathbf{q}_1 \quad \hat{\mathbf{\Sigma}}_{\infty}\mathbf{q}_2 \quad \ldots \quad \hat{\mathbf{\Sigma}}_{\infty}\mathbf{q}_M]. \tag{3.14}$$

Since two distinct components are uncorrelated, the state vector $\mathbf{y}$ can be decomposed as

$$\mathbf{y} = \mathbf{p}_1(\mathbf{q}_1^{\mathrm{T}}\mathbf{y}) + \mathbf{p}_2(\mathbf{q}_2^{\mathrm{T}}\mathbf{y}) + \ldots + \mathbf{p}_M(\mathbf{q}_M^{\mathrm{T}}\mathbf{y}). \tag{3.15}$$

This decomposition based on APT is analogous to principal component analysis, except that instead of decomposing variance we decompose predictability.

For the data sets in which only one forecast is available for each initial date, we follow DelSole and Tippett (2009) and fit the data to a multivariate linear regression model and then use the forecast variance of the linear regression model to estimate APT. However, in contrast to the standard APT analysis (DelSole and Tippett, 2009), we generalized it by allowing the predictor to differ from the predictand in the regression model. Letting a centered state vector $\mathbf{y}_{t+\tau}$ denote the predictand at time $t+\tau$ and a centered state vector $\mathbf{x}_t$ denote the predictor at time t, the desired linear regression model is

$$\mathbf{y}_{t+\tau} = \mathbf{L}_{\tau}\mathbf{x}_t + \boldsymbol{\epsilon}_t, \tag{3.16}$$

where $\mathbf{L}_{\tau}$ is the regression operator, and $\boldsymbol{\epsilon}_t$ is the forecast error. Standard regression theory shows that the regression operator can be estimated by

$$\hat{\mathbf{L}}_{\tau} = \overline{\mathbf{y}_{t+\tau}\mathbf{x}_t^{\mathrm{T}}}\left(\overline{\mathbf{x}_t\mathbf{x}_t^{\mathrm{T}}}\right)^{-1}, \tag{3.17}$$

where the overline indicates an average over time. If the estimated forecast error is defined as $\hat{\boldsymbol{\epsilon}}_t = \mathbf{y}_{t+\tau} - \hat{\mathbf{L}}_\tau \mathbf{x}_t$, then the forecast covariance matrix is estimated by

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}_\tau = \overline{\hat{\boldsymbol{\epsilon}}\hat{\boldsymbol{\epsilon}}^{\mathrm{T}}} &= \overline{\left(\mathbf{y}_{t+\tau} - \hat{\mathbf{L}}_\tau \mathbf{x}_t\right)\left(\mathbf{y}_{t+\tau} - \hat{\mathbf{L}}_\tau \mathbf{x}_t\right)^{\mathrm{T}}} \\
&= \overline{\left(\mathbf{y}_{t+\tau}\mathbf{y}_{t+\tau}^{\mathrm{T}} - \left(\mathbf{y}_{t+\tau}\mathbf{x}_t^{\mathrm{T}}\right)\left(\mathbf{x}_t\mathbf{x}_t^{\mathrm{T}}\right)^{-1}\left(\mathbf{x}_t\mathbf{y}_{t+\tau}^{\mathrm{T}}\right)\right)} \\
&= \mathbf{C}_{yy} - \mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{yx}^{\mathrm{T}} \quad , \qquad (3.18)
\end{aligned}$$

where

$$\mathbf{C}_{yx} = \overline{\mathbf{y}_{t+\tau}\mathbf{x}_t^{\mathrm{T}}}, \quad \mathbf{C}_{xx} = \overline{\mathbf{x}_t\mathbf{x}_t^{\mathrm{T}}}, \quad \mathbf{C}_{yy} = \overline{\mathbf{y}_{t+\tau}\mathbf{y}_{t+\tau}^{\mathrm{T}}}, \qquad (3.19)$$

and (3.17) has been used. For zero mean stationary processes, $\mathbf{C}_{xx}$ and $\mathbf{C}_{yy}$ do not depend on τ. The climatological distribution is the forecast covariance as the lead time τ approaches infinity, in which case $\mathbf{y}$ and $\mathbf{x}$ are independent, hence $\mathbf{C}_{yx}$ is close to zero. The climatological covariance is then estimated by

$$\hat{\boldsymbol{\Sigma}}_\infty = \mathbf{C}_{yy}. \qquad (3.20)$$

Substituting (3.18) and (3.20) into (3.8) gives the generalized eigenvalue problem

$$\left(2\sum_{\tau=1}^{\infty}\mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{yx}^{\mathrm{T}}\Delta\tau\right)\mathbf{q} = \lambda\mathbf{C}_{yy}\mathbf{q}. \qquad (3.21)$$

All quantities appearing in (3.21) can be estimated directly from data. The components that maximize APT of a multivariate linear regression model are then obtained by solving the generalized eigenvalue problem (3.21).

The connections between the APT of multivariate linear regressive model and standard statistical metrics deserve clarification. Invoking (3.18) and (3.20), the APT in (3.7) can

be written as

$$APT = 2\sum_{\tau=1}^{\infty}\left(\frac{\mathbf{q}^{\mathrm{T}}\mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{yx}^{\mathrm{T}}\mathbf{q}}{\mathbf{q}^{\mathrm{T}}\mathbf{C}_{yy}\mathbf{q}}\right)\Delta\tau = 2\sum_{\tau=1}^{\infty}R_{\tau}^{2}\Delta\tau. \tag{3.22}$$

The R_{τ}^{2} in (3.22) stands for the squared multiple correlation in regression theory. In fact, APT is an integral of the squared multiple correlation over lead times. It can be shown that APT also has a connection with canonical correlation analysis (CCA). CCA is a procedure for finding components in two data sets that are maximally correlated. In our case, the two data sets are predictor $\mathbf{x}_t$ and the predictand $\mathbf{y}_{t+\tau}$. The main difference between APT and CCA is that CCA maximizes the multiple correlation at a fixed lead time τ, while APT analysis maximizes the sum of squared multiple correlations at all lead times. Another technique worth mentioning is predictable component analysis (PrCA), which finds components that minimize the normalized error variance (Déqué, 1988; Renwick and Wallace, 1995; Schneider and Griffies, 1999). DelSole and Chang (2003) have shown that PrCA of a multivariate linear regression model is precisely equivalent to CCA. Thus, for multivariate linear regression models, the only distinction between PrCA, CCA and APT analysis is that the first two optimize predictability at fixed lead time, while the last optimizes predictability over all lead times.

3.2 Model data

The data used here are the pre-industrial control runs from the World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3). These control runs are intended to represent the equilibrium climate in the pre-industrial period, and thus the external anthropogenic and natural forcing agents are fixed to their pre-industrial values. Since the external forcings are fixed, the control runs contain only internal climate variability. The CMIP3 data set includes the output of 25 models from 18 groups in the world. The fully coupled models consists of atmosphere, ocean, sea ice and land components. There is only one control run in most of models, so we use only one control

run from each model. The number of years in control runs vary with models. Table 3.1 lists the 17 CMIP3 models whose control runs are at least 300 years long and for which the variables SAT, SST and precipitation are available. The annual means of the last 300 years of each control run are analyzed. We use 300-year data because most models have more than 300 years of control runs and a 300-year-long sample size is considered to be long enough for studies on decadal time scales. The reason to choose annual mean resolution is that it is standard for studying predictability on multi-year time scales (Collins et al., 2006; Newman, 2007). The 300-year mean of each model is subtracted from each grid point of the respective control run. To facilitate model intercomparison, all fields are interpolated to a common 72×36 grid. The 300-year data from each individual model are lined up in temporal dimension to form a single multi-model data set. For example, if the data from each single model has a dimension of $M \times N$, where M denotes state dimension, and N denotes the number of time steps, then the multi-model data including n models has a dimension of $M \times nN$.

3.3 Excluding models based on variance and trend

As with all optimization techniques, APT is subject to overfitting when the number of parameters being estimated is not a small fraction of the sample size. To mitigate overfitting, we reduce the dimension of the data by projecting the data onto a few leading principal components (PCs). The PCs of global land SAT were computed using multi-model data from 17 control runs that are at least 300 years long. We first checked for outliers in terms of variance and found that model "GISS-ER" had a factor of 2 less variance in the first few leading PCs than most of the other models. Hence, this model was dropped from the multi-model data set.

Although control runs have constant forcing, some runs may exhibit significant trends due to adjustment to equilibrium ("spin up"). APT analysis is sensitive to such trends and including these models in the analysis tends to produce results that are dominated by these models. It is questionable whether the adjustment can be removed by regressing out

Table 3.1: Description of CMIP3 models with ≥ 300 years in SAT, SST and precipitation. The model number, originating group(s), model I.D., resolution, whether the model is included in the multi-model pool, number of ensemble members in the 20th century runs and the years in the 20th century runs are listed. Selected models are indicated by the check marks in the last column.

Number	Originating Group(s)	CMIP3 I.D.	Resolution	In multi-model?	# of ensemble	Years
1	Canadian Centre for Climate Modelling & Analysis	CGCM3.1(T47)	96 × 48	-	-	-
2	Canadian Centre for Climate Modelling & Analysis	CGCM3.1(T63)	128 × 64	-	-	-
3	Meteo-France / Centre National de Recherches Meteorologiques	CNRM-CM3	128 × 64	-	-	-
4	CSIRO Atmospheric Research (Australia)	CSIRO-Mk3.0	192 × 96	-	-	-
5	CSIRO Atmospheric Research (Australia)	CSIRO-Mk3.5	192 × 96	-	-	-
6	US Dept. of Commerce / NOAA / Geophysical Fluid Dynamics Laboratory	GFDL-CM2.0	144 × 90	✓	3	1861-1999
7	US Dept. of Commerce / NOAA / Geophysical Fluid Dynamics Laboratory	GFDL-CM2.1	144 × 90	✓	3	1861-1999
8	NASA / Goddard Institute for Space Studies (USA)	GISS-EH	72 × 46	-	-	-
9	NASA / Goddard Institute for Space Studies (USA)	GISS-ER	72 × 46	-	-	-
10	LASG / Institute of Atmospheric Physics (China)	FGOALS-g1.0	128 × 60	-	-	-
11	Institute for Numerical Mathematics (Russia)	INM-CM3.0	72 × 45	-	-	-
12	Institut Pierre Simon Laplace (France)	IPSL-CM4	96 × 72	✓	1	1860-1999
13	Center for Climate System Research- The University of Tokyo (Japan)	MIROC3.2(medres)	128 × 64	✓	3	1850-1999
14	Meteorological Institute of the University of Bonn, Meteorological Research Institute of KMA, and Model and Data group (Germany/Korea)	ECHO-G	96 × 48	✓	5	1860-1999
15	Meteorological Research Institute (Japan)	MRI-CGCM2.3.2	128 × 64	✓	5	1851-1999
16	National Center for Atmospheric Research (USA)	CCSM3	256 × 128	✓	5	1870-1998
17	Hadley Centre for Climate Prediction and Research / Met Office (UK)	UKMO-HadCM3	96 × 73	✓	2	1860-1998

the linear trend – for example, even linear models follow an exponential (i.e., nonlinear) time evolution when relaxing back to equilibrium. To avoid such questions, we remove any model with significant trends in the control run. To identify such models, we determined the components that maximize the squared canonical correlation between a linear trend and the data. DelSole and Yang (2011) reviewed this technique and showed that the maximum squared canonical correlation between a linear trend and the data is given by

$$\rho^2 = \frac{\mathbf{t}^{\mathrm{T}}\mathbf{Y}(\mathbf{Y}^{\mathbf{T}}\mathbf{Y})^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{t}}{\mathbf{t}^{\mathrm{T}}\mathbf{t}}, \tag{3.23}$$

where $\mathbf{Y}$ is the centered principal component matrix, with time varying in the first dimension and component varying in the second, and $\mathbf{t}$ is a linear function of time with zero mean. The detailed derivation of ρ^2 is shown in appendix A. The ρ^2 values between a linear trend and 10 PCs of global land SAT for each model is shown in Fig. 3.1. The models with the nine smallest canonical correlations, indicated by dark shading, appear to be separated from the other models in terms of their canonical correlations, and hence are selected. Among these nine models, model "GISS-ER" also was eliminated due to its small variance, as mentioned earlier. The remaining eight models that were selected are indicated in Table 3.1. In the field of precipitation, model runs have weaker trends than SAT (not shown), and there are no significant trends in the selected eight models. Although models show stronger trends in SST than in SAT, trends in the selected eight models are weak relative to the other models. For simplicity, the same set of models were used for maximizing APT of precipitation and SST.

3.4 Statistical details in APT analysis

The eight selected control runs were pooled to construct a multi-model data set, from which the PCs of SAT and precipitation in each continent were computed. The PCs of global SST also were computed from this eight-model data set. Six continents were chosen

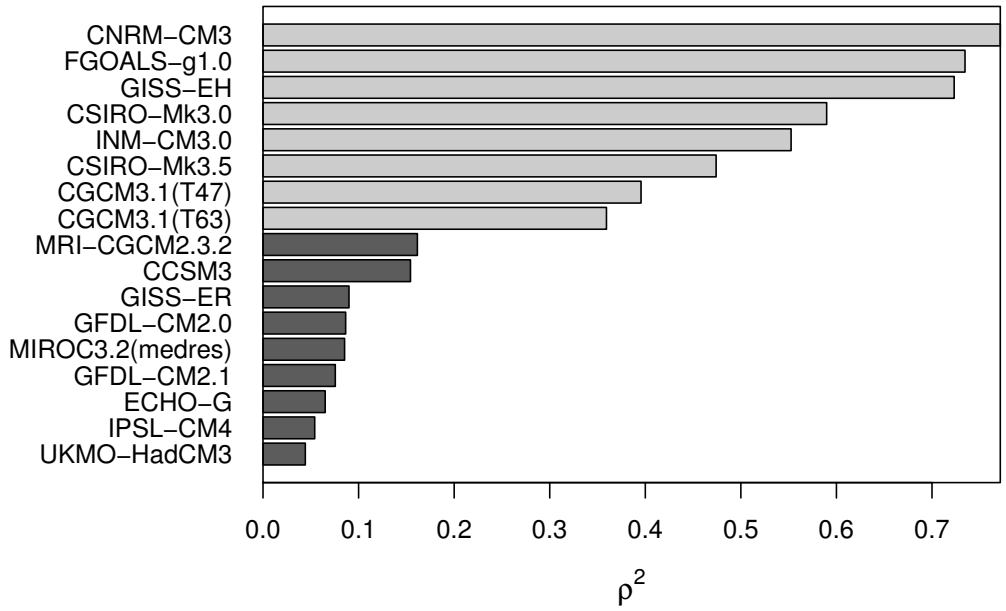

Figure 3.1: The maximum squared canonical correlation (see (3.23)) between a linear trend and 10 PCs of global land SAT for each model. The models selected for predictability analysis are indicated by dark shading.

with boundaries specified in Table 3.2. The resulting PCs of each model were split into two halves. The first 150 years of each control run were used to maximize APT, and the second 150 years were reserved for verification. To maximize APT, sample covariance matrices estimated from training data were substituted into the eigenvalue problem (3.21). We chose a maximum time lag of 20 years and 30 PCs to maximize APT, but the APT values were not sensitive to either the number of PCs or the choice of time lag, presumably because of the large sample size of the multi-model data set.

To validate a predictable component, the projection vector $\mathbf{q}$ estimated from training data was applied to the verification data. The squared multiple correlation between the component time series and the first 30 PCs in the verification data at lead time τ was calculated as

$$R_\tau^2 = \frac{\mathbf{q}^{\mathrm{T}}\mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{yx}^{\mathrm{T}}\mathbf{q}}{\mathbf{q}^T\mathbf{C}_{yy}\mathbf{q}}, \tag{3.24}$$

where $\mathbf{q}$ is computed from training data and $\mathbf{C}_{yx}, \mathbf{C}_{xx}, \mathbf{C}_{yy}$ are computed from verification data. R_τ^2 can be interpreted as the variance of the component time series explained by a linear regression prediction at lead time τ. APT is in fact the integral of R_τ^2 over lead times (see (3.22)), hence a slowly decreasing R_τ^2 with lead time implies large APT. The statistical significance of R_τ^2 can be determined by standard methods since $\mathbf{q}$ was determined independently of the covariance matrices $\mathbf{C}_{yx}, \mathbf{C}_{xx}, \mathbf{C}_{yy}$.

Table 3.2: Domains of six continents.

Continent	Longitude	Latitude
North America (NA)	170°W - 20°W	15°N - 90°N
South America (SA)	90°W - 25°W	65°S - 15°N
Asia	170°W - 20°W	15°N - 90°N
Africa	25°W - 55°E	40°S - 40°N
Australia	110°E - 160°E	45°S - 10°S
Europe	5°W - 55°E	40°N - 70°N

3.5 Statistical significance test of APT

The statistical significance of APT is assessed relative to the null hypothesis that the time series is unpredictable – that is, that the time series is white noise. Since the results of APT analysis are invariant to nonsingular linear transformation (DelSole and Tippett, 2009), the covariance matrix of the process can be assumed to be the identity matrix. For M spatial dimensions and N time steps, we generated MN independent random numbers drawn from a Gaussian distribution with zero mean and unit variance. APT analysis was then applied to this $M \times N$ data set to produce an ordered sequence of optimized APT values. In the present case, $M = 30$ and $N = 150 \times 8$. This procedure was then repeated 100 times and the 95th percentile of each eigenvalue from (3.8), ordered by decreasing value, was selected. The null hypothesis of no predictability is rejected if the APT value computed from the training data exceeds the 95th percentile computed from the Monte Carlo sample.

3.6 Results

3.6.1 Standard APT analysis of global SST

We first identify predictable components of global SST using standard APT analysis (i.e., the predictand and predictor in (3.16) are global SST at two different times). This analysis is similar to DelSole et al. (2011), except that the components are computed using only eight models on a global grid (as opposed to 14 models on a grid with missing values masked out). The leading components have statistically significant APT values, hence are predictable. The spatial pattern of the leading predictable component, shown in Fig. 3.2, has amplitudes concentrated in the Southern Ocean, North Pacific and North Atlantic. The regression coefficients between this component and global SAT and precipitation are shown in Fig. 3.3. Not surprisingly, the regression pattern for SAT resembles the spatial pattern of the leading predictable component of global SST. The regression pattern for precipitation has large coefficients only in limited areas of tropical Indo-Pacific. Importantly, the regression coefficients for both SAT and precipitation are small over land, implying little to no response of land SAT and precipitation to the leading predictable component of SST. A similar lack of land response was found for a few other secondary SST components (not shown). The fact that little to no land predictability was found using this standard APT analysis and the regression technique consistents with previous studies.

3.6.2 Generalized APT analysis of continental SAT

The above result merely implies that the *leading* predictable component of global SST is not related to land, it does not preclude the possibility that some other SST patterns (with less predictability) induce strong predictability over land. Even if we can find some predictable SST components that cause land predictability, this does not indicate the limit of land predictability. Thus, traditional regression is not a definitive method to identify land predictability. The same argument applies to the method of regressing principal components of SST on land variables. In order to definitely settle the question of whether land

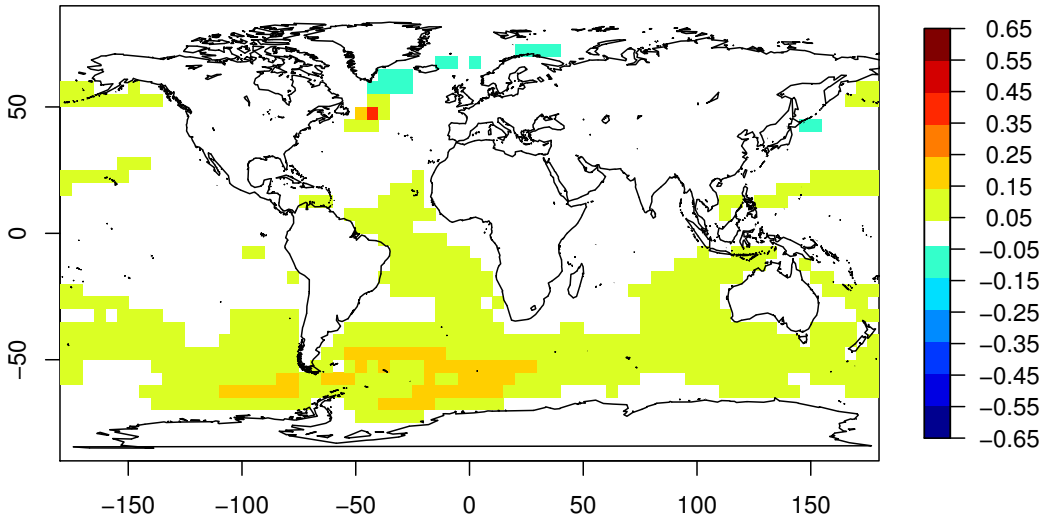

Figure 3.2: Spatial pattern of the leading predictable component of global SST from eight CMIP3 control simulations. The spatial pattern has units of degrees Kelvin, and the corresponding time series has unit variance. The amplitudes are the deviations from time mean.

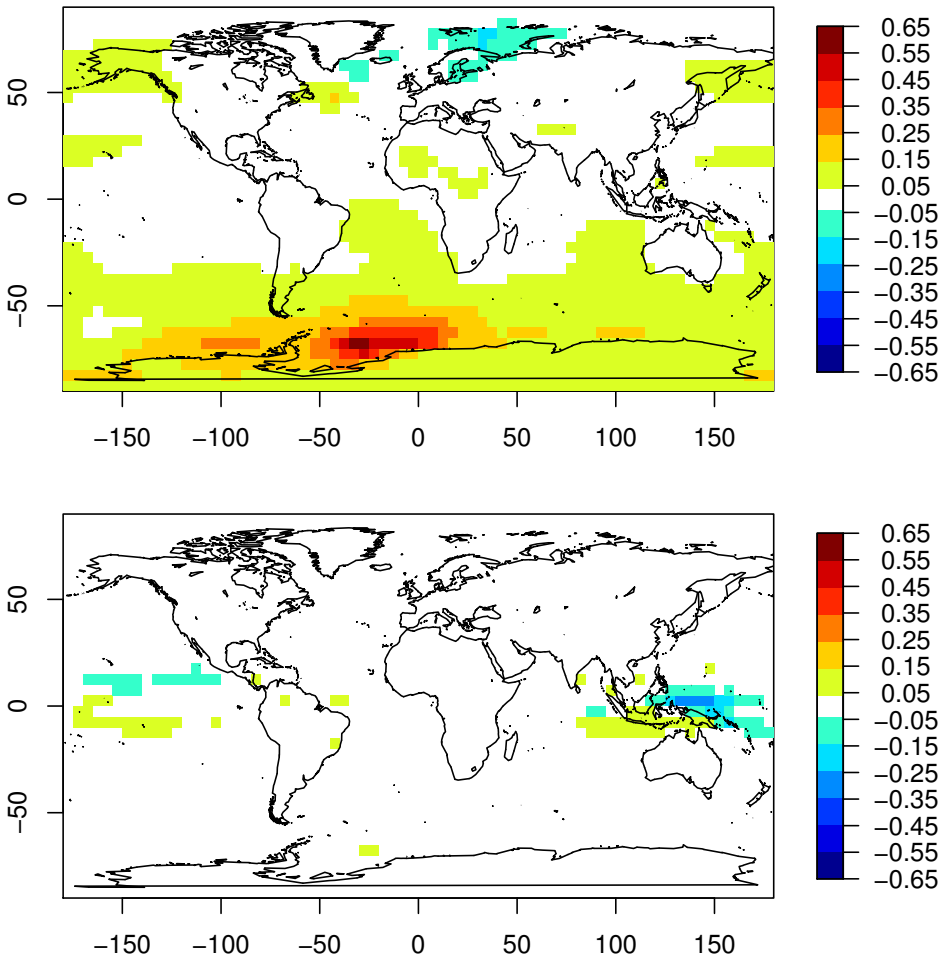

Figure 3.3: Regression coefficients between the leading predictable component of SST from eight CMIP3 control simulations, and global SAT (upper) and precipitation (bottom). The regression pattern of SAT has units of degrees Kelvin per unit predictable component time series, and the pattern of precipitation has units of mm day^{-1} per unit predictable component time series.

variables can be predicted on multi-year time scales, the technique must be modified to specifically isolate land predictability. The natural approach is to optimize APT over each continent individually. Therefore, we restrict the predictand variable to be over a specific continent. However, two types of predictors were investigated: (1) SAT or precipitation over the continent under investigation in the standard APT analysis, and (2) global SST in the generalized APT analysis. The reason why we chose global SST as predictor in the generalized analysis is that SST has been show to be the primary driver for land predictability on interannual-to-decadal time scales (Held et al., 2005; Hoerling et al., 2010). From these two analyses, we found that in all cases, except SAT in North America, the predictability derived from using global SST as predictor was comparable to or greater than the predictability derived from using land SAT or precipitation as predictor. Since the use of SST predictors gives equal or larger estimates of predictability, and clarifies the source of predictability, we present results only for the generalized analysis in this chapter. The exceptional case of North American temperature will be discussed in the summary section of this chapter.

The maximum APT values of SAT estimated from the multi-model training data using global SST as predictor are shown in Fig. 3.4 for all continents. This figure reveals that only the first half dozen or so components in each continent have APT values that differ significantly from white noise. Since our interest is predictability on the longest time scales, we focus only on the leading component in each continent. The leading component tends to be well separated from the secondary components. However, in South America, the first two components are relatively close, implying that these components are probably sensitive to sampling errors and therefore dangerous to interpret physically.

The physical patterns corresponding to the leading predictable component of SAT in the six individual continents (derived from (3.13)) are shown in Fig. 3.5. The spatial structure in North America has largest amplitudes in northeast of North America, Alaska and Greenland. The corresponding APT value is 2.52 years, which is the largest of any continent. Large amplitudes are found in northeast of Asia also, while the amplitudes in

other continents are smaller. Interestingly, the amplitudes are all single sign, which may present difficulties for separating these components from forced patterns (based on spatial information alone). The predictable variance derived by this method clearly is greater than the predictable variance over land implied by the regression patterns associated with global SAT shown in Fig. 3.3.

The multi-model R_τ^2 values of the leading predictable component of SAT derived from independent CMIP3 control simulations (i.e., "verification") are shown in Fig. 3.6 for each continent as a function of time lag. By definition, R_τ^2 gives the fraction of variance explained by a linear regression model with predictors based on 30 PCs of global SST. The R_τ^2 values in the verification and training data (not shown) are similar, suggesting that overfitting is not a problem. The R_τ^2 values remain significant 3-6 years, depending on continent, suggesting predictability time scales of 3-6 years. The one-year lag values of R_τ^2 are smaller in Europe and Australia than in the other four continents. The R_τ^2 for non-zero lags in each continent are less than 0.5, implying that less than 50% of variations in annual mean SAT can be predicted from global SST. We emphasize that the spatial patterns were estimated from control runs with no natural or anthropogenic forcing – the predictability found here arises from unforced, internal processes.

To gain insight into the model dependence of predictability time scales, the R_τ^2 values of the leading multi-model predictable component of SAT are calculated separately for each model. The result, shown in Fig. 3.7, shows that the R_τ^2 values of individual models exhibit more fluctuations than the multi-model R_τ^2 values, due to the shorter sample size, and that the predictability time scale can range from zero year, corresponding to no predictability, to ten years, depending on model and continent. The predictability among models is more consistent in South America and Africa than in other continents.

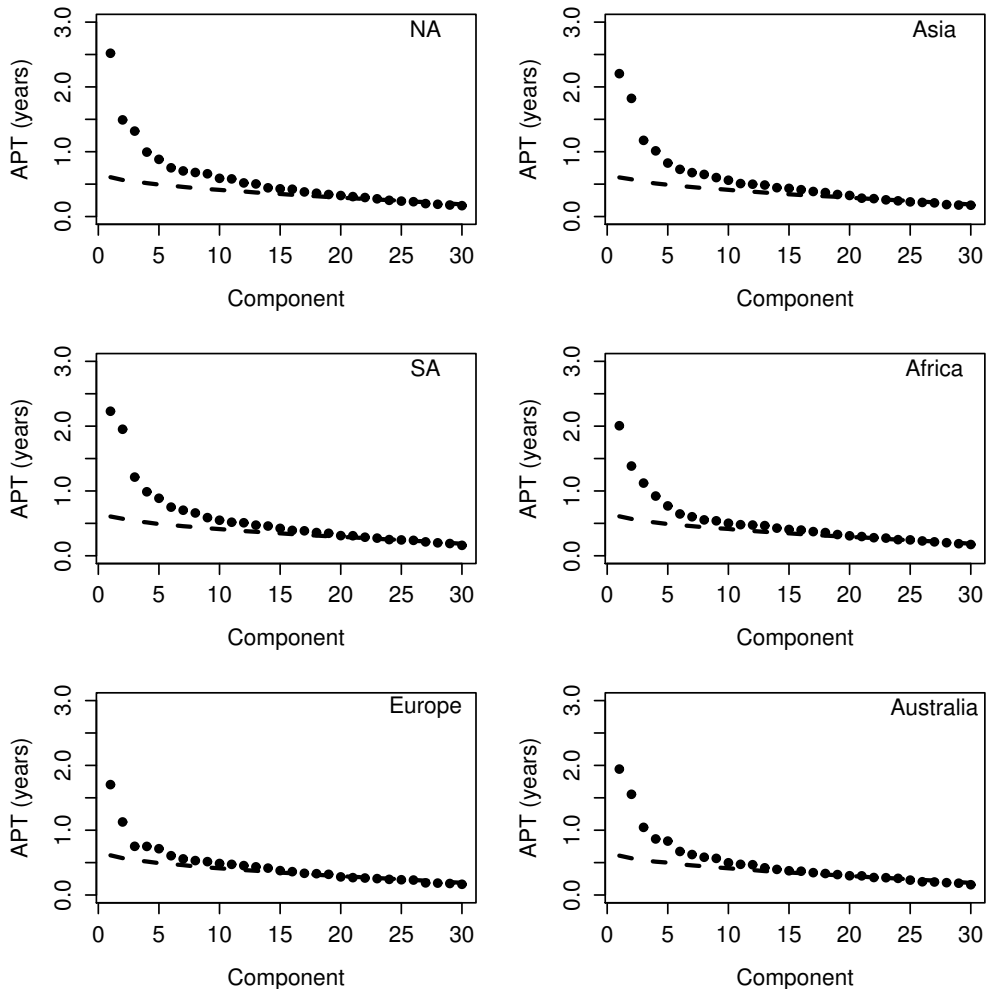

Figure 3.4: APT values for SAT in six continents, as determined by optimizing APT over eight CMIP3 control simulations. The domains of the six continents are listed in Table 3.2. The dashed lines in each panel indicate the 5% significance level estimated by Monte Carlo method.

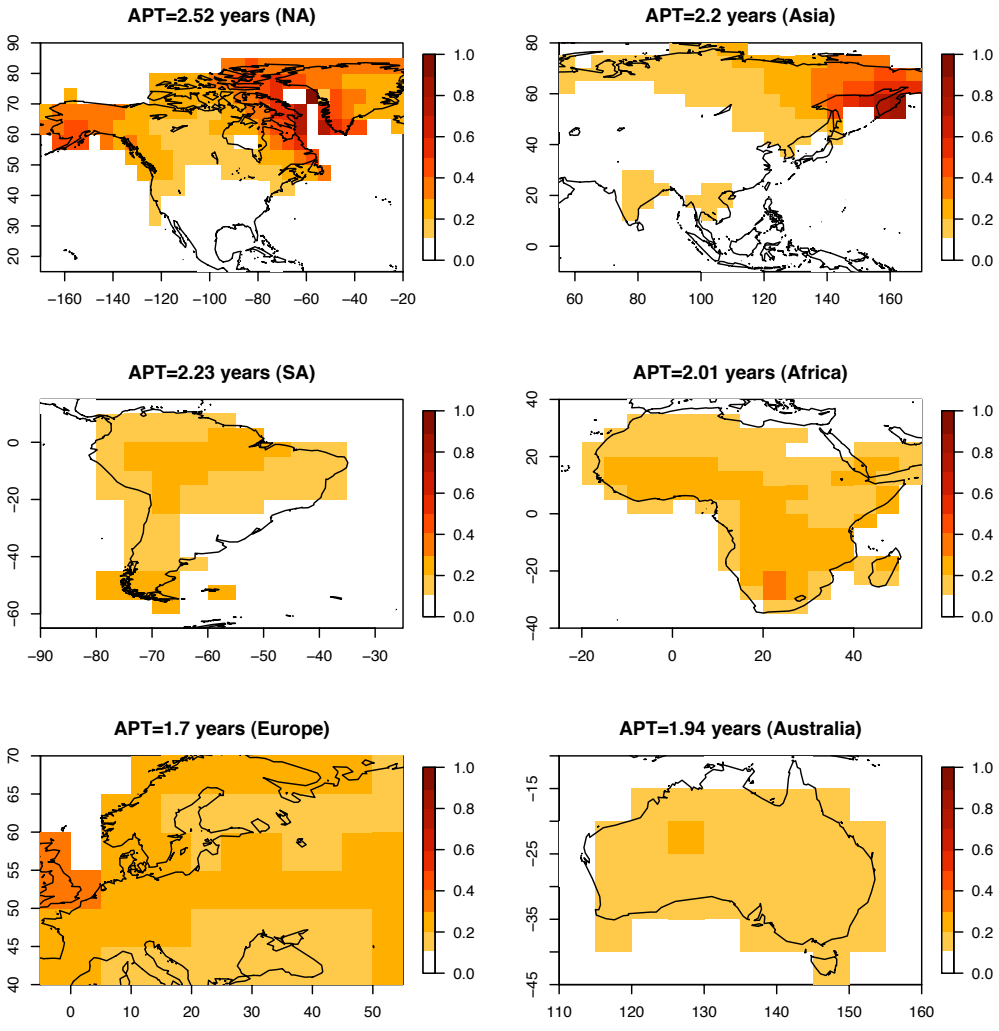

Figure 3.5: Spatial patterns of the leading predictable components of SAT in six continents derived from eight CMIP3 control simulations. The spatial patterns have units of degree Kelvin, and the time series for each component has unit variance. The amplitudes are the deviations from time mean.

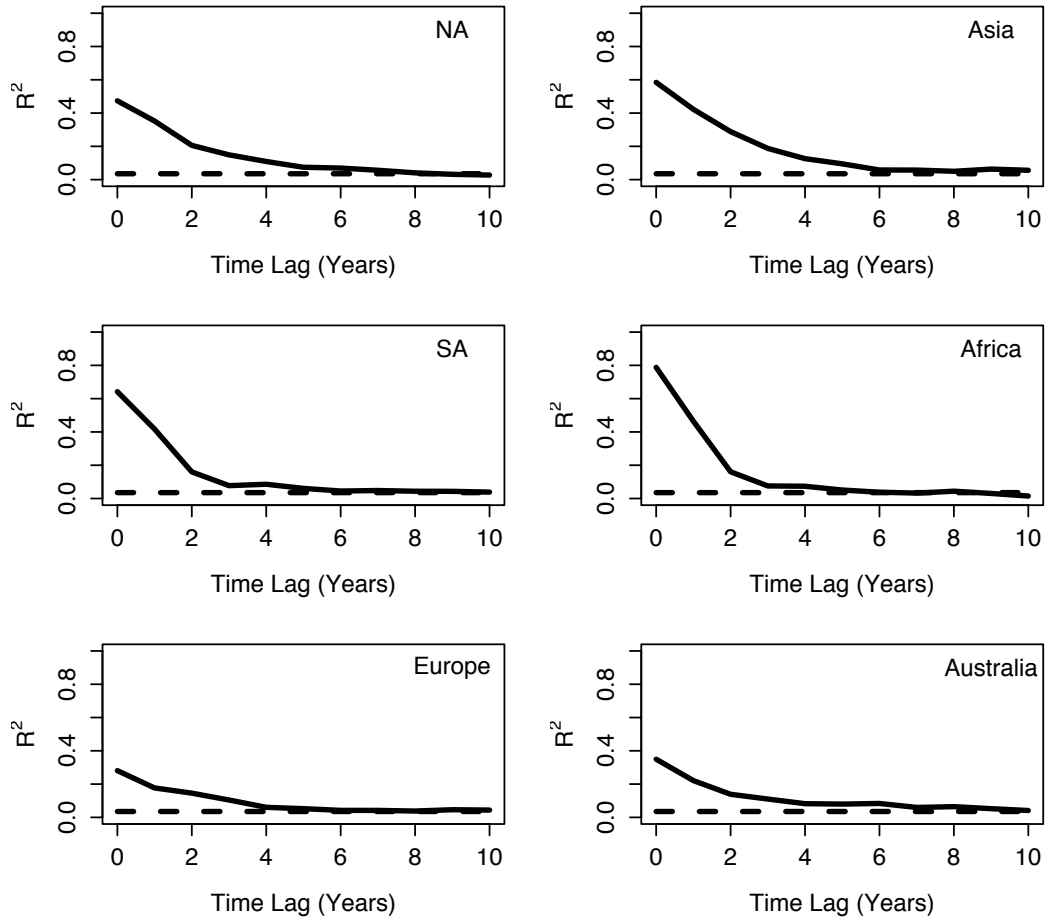

Figure 3.6: Multi-model R^2_τ values of the leading predictable component of SAT derived from eight independent CMIP3 control simulations (the simulations are independent in the sense that they were not used to derive the predictable component). The horizontal dashed lines in each panel indicate the 5% significance level.

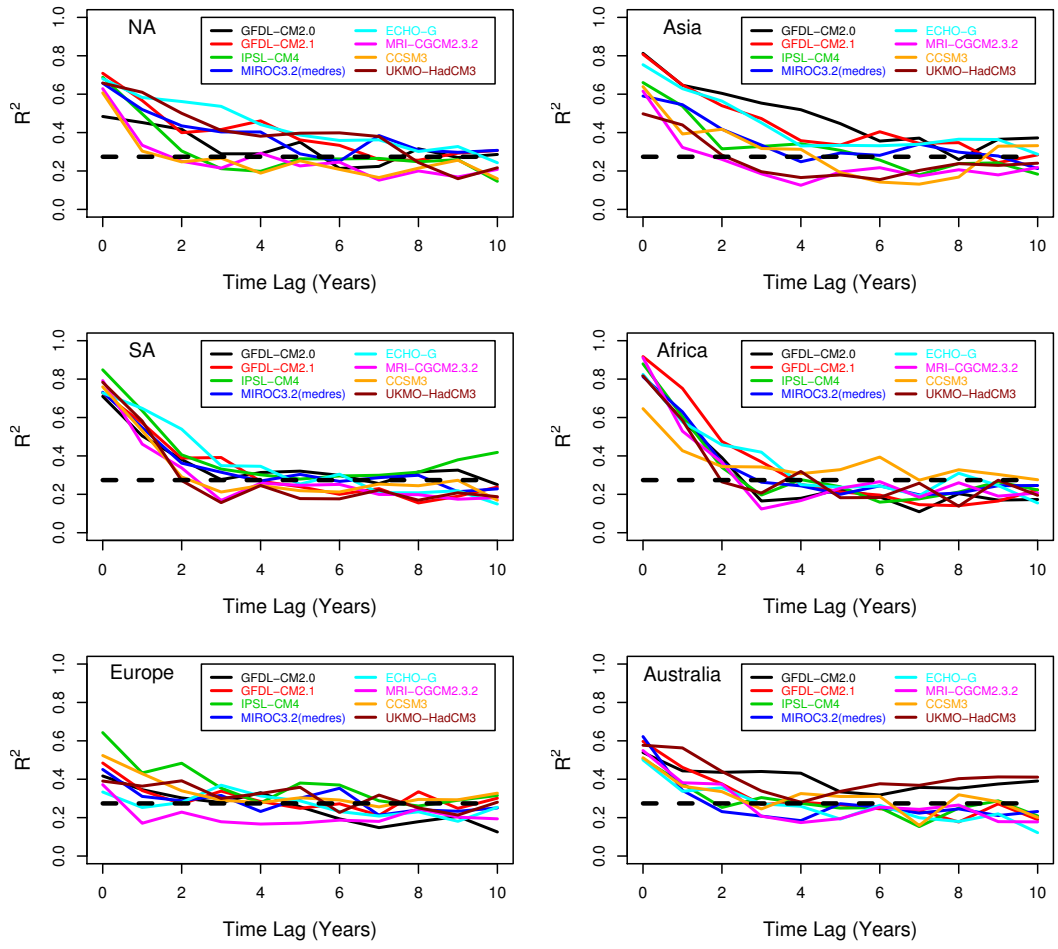

Figure 3.7: R_τ^2 values in individual models for the leading multi-model predictable component of SAT. R_τ^2 is calculated from CMIP3 control simulations independent of the data used to calculate the predictable component. Each colored line represents a particular model. The horizontal dashed lines in each panel indicate the 5% significance level.

3.6.3 Relation between SST and predictable SAT

To investigate the SST pattern from which SAT can be predicted, we calculated correlations between global SST and the leading predictable component at different leads. The results are shown in Fig. 3.8. Each row represents a particular continent and each column represents a particular lead time. Values that are not significant at a 5% level are masked out. In all continents except Europe, significant positive correlations are found in tropical oceans for leads zero and one. Not surprisingly, the leading predictable component of each continent (except Europe) is significantly related to the Niño 3 index in the control runs, suggesting that these components are ENSO-related. Loosely speaking, the correlation patterns in these five continents appear to be a superposition of two patterns, one ENSO pattern that changes structure with lead time, and one persistent pattern with relatively enhanced correlations near the continent under investigation. In Europe, no ENSO component is found, but the predictable component is correlated with persistent SSTs in northeast Atlantic near Europe.

In each continent, the correlation patterns for individual models are similar to each other and to the multi-model correlation pattern (not shown). To quantify this similarity, we calculated the anomaly pattern correlations of the correlation patterns between each particular model and the multi-model. We found that the anomaly correlations exceed 0.5 for most of the models, and are typically in the range 0.6-0.8. The consistency of correlation patterns among models suggests that a genuine component of predictability that arises from the same SST structure in each model has been identified. These results are consistent with the hypothesis that the variability in SST induces changes in the atmospheric circulation that are manifested in land climate variability.

3.6.4 Generalized APT analysis of continental precipitation

The maximum APT values of precipitation with global SST as predictor are shown in Fig. 3.9 for each continent. In general, the APT values for land precipitation are less than those for SAT, with values over Europe being the smallest overall. To study the predictability

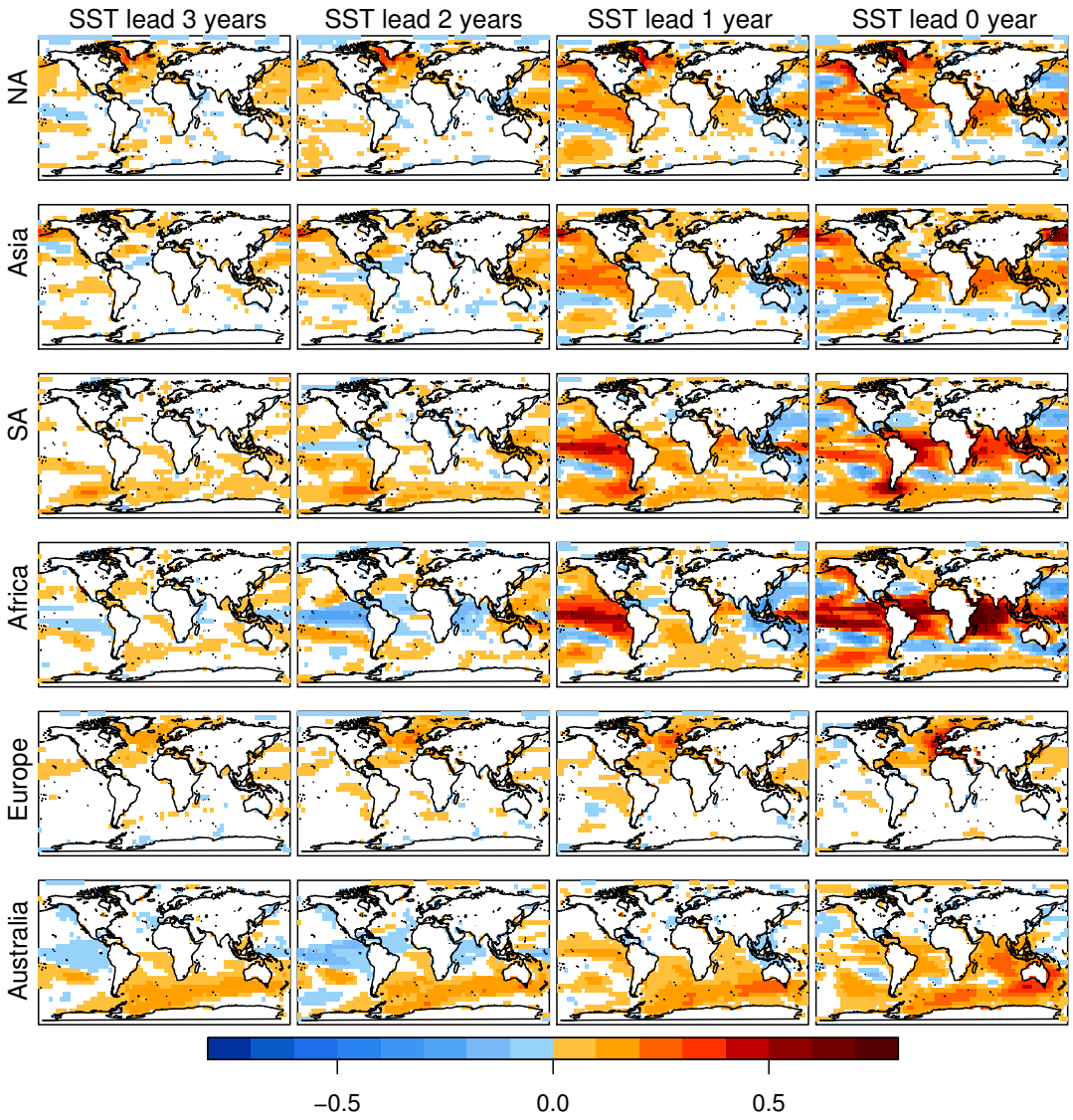

Figure 3.8: Correlation patterns of global SST with the leading predictable component of SAT in six individual continents at 3 years (first column), 2 years (second column), 1 year (third column) and 0 year (last column) lead. Each row represents a particular continent and each column represents a particular lead time. Insignificant areas are masked out based on Student's t-test at a 5% significance level.

of precipitation on long time scales, we focus on the first predictable component in each continent. Note that the first predictable components in North America and Europe are not well separated from the other components, suggesting that these components are probably sensitive to sampling errors and therefore dangerous to interpret physically.

The spatial patterns of the first predictable component of precipitation in six individual continents are shown in Fig. 3.10. Southeast Asia, Amazon, Northern Australia and equatorial Africa show the largest amplitudes and hence are the most predictable regions. The multi-model R_τ^2 values of the leading component of precipitation in the verification data are shown in Fig. 3.11. The R_τ^2 values in all continents except Europe remain significant up to 1-3 years, depending on continent. Therefore, the leading component in these continents are predictable 1-3 years ahead, at least by a linear regression model. The R_τ^2 in Europe is insignificant at 1 year lag. Thus, although the Europe predictable component appears significant in the training data, this predictability cannot be confirmed in independent data. We conclude that there is no verifiable multi-year predictability of precipitation in Europe based on SST. This result is consistent with the result of using European precipitation itself as predictor, which showed no multi-year predictability of precipitation in Europe (not shown). The R_τ^2 values are less than 0.4 at non-zero time lags in all continents, implying that less than 40% of variations in annual mean precipitation can be predicted from global SST. The R_τ^2 values for individual models are shown in Fig. 3.12. Most models show multi-year predictability in Asia, South America, Africa and Australia, weak predictability in North America, and almost no predictability in Europe.

3.6.5 Relation between SST and predictable precipitation

The correlation patterns between global SST and the leading component of precipitation at different leads are shown in Fig. 3.13. The correlation patterns in Asia, South America, Africa and Australia resemble each other at all lead times with highest correlations in tropical oceans. Not surprisingly, the leading predictable component in these continents also are strongly correlated with the NINO3 index. Also, the ENSO-like SST pattern

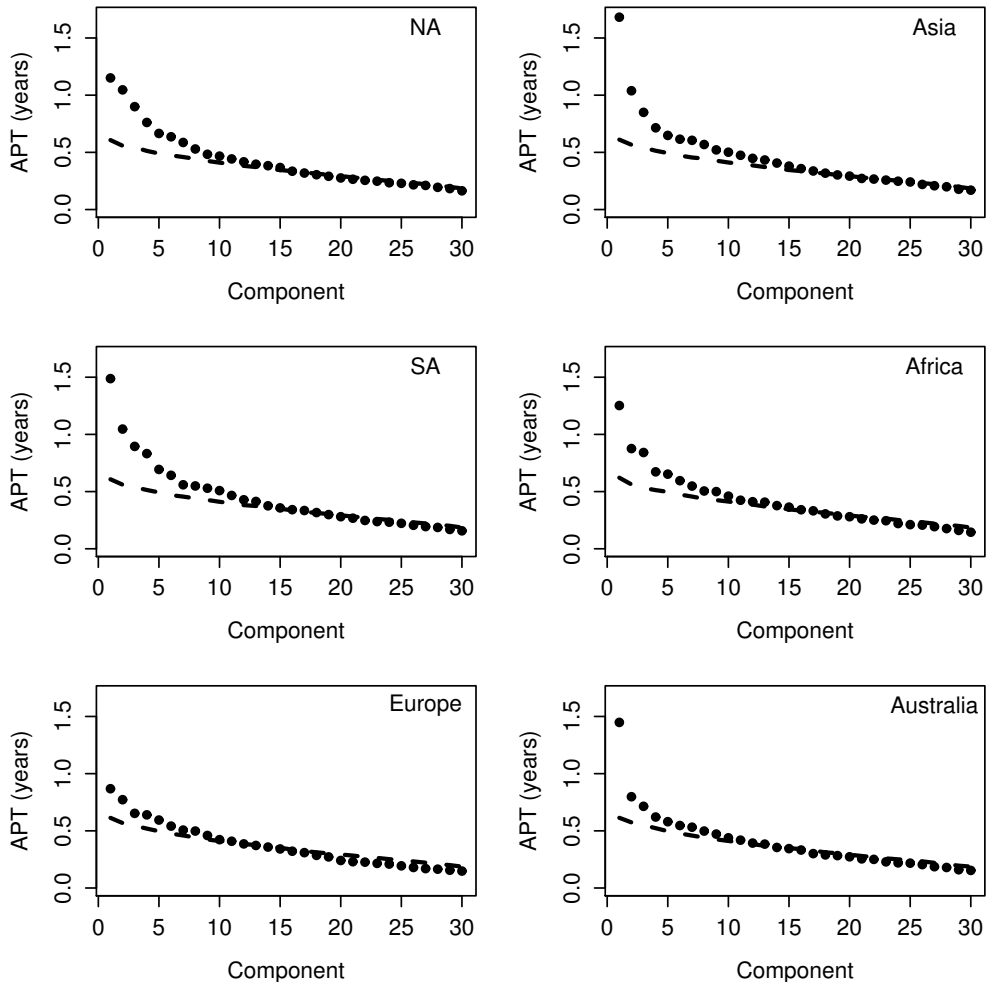

Figure 3.9: APT values for precipitation in six continents, as determined by optimizing APT over eight CMIP3 control simulations. The domains of the six continents are listed in Table 3.2. The dashed lines in each panel indicate the 5% significance level estimated by Monte Carlo method.

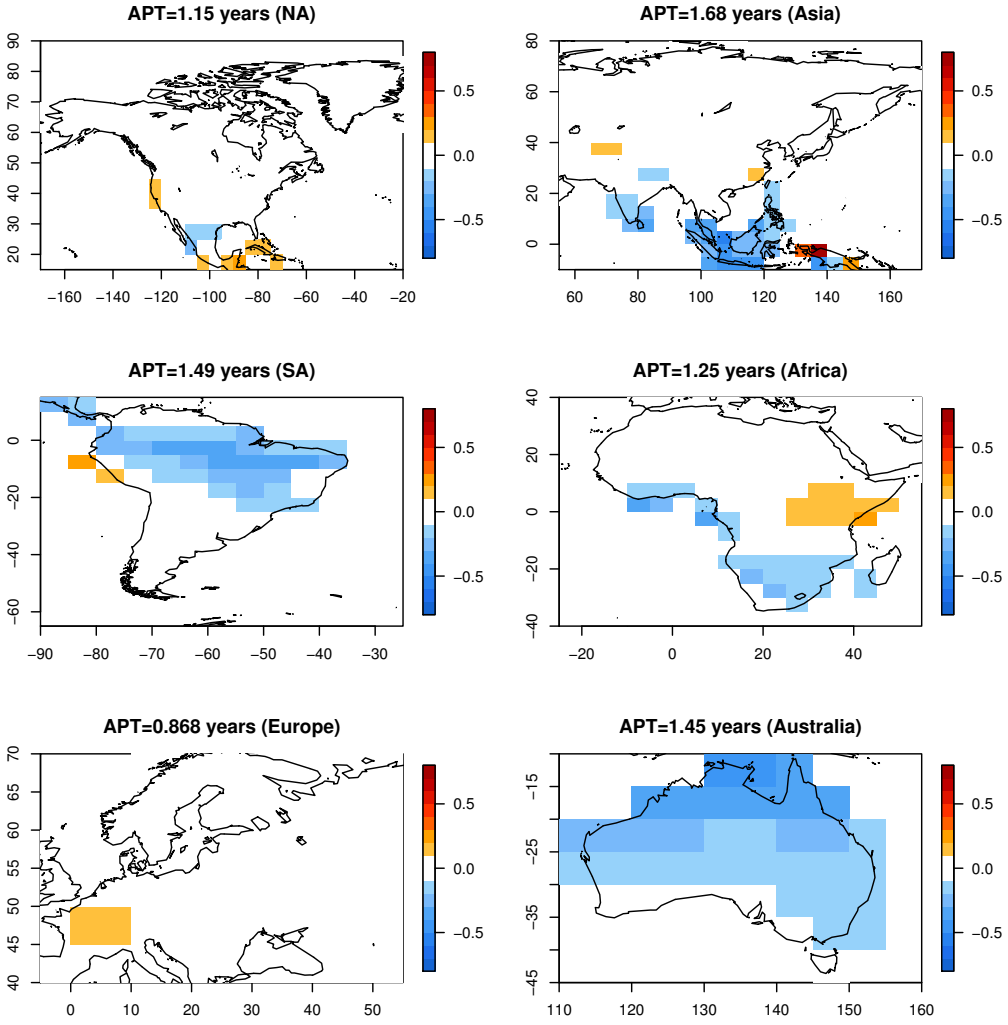

Figure 3.10: Spatial patterns of the leading component of precipitation in six continents derived from eight CMIP3 control simulations. The spatial patterns have units of mm day^{-1}, and the time series for each component has unit variance. The amplitudes are the deviations from time mean.

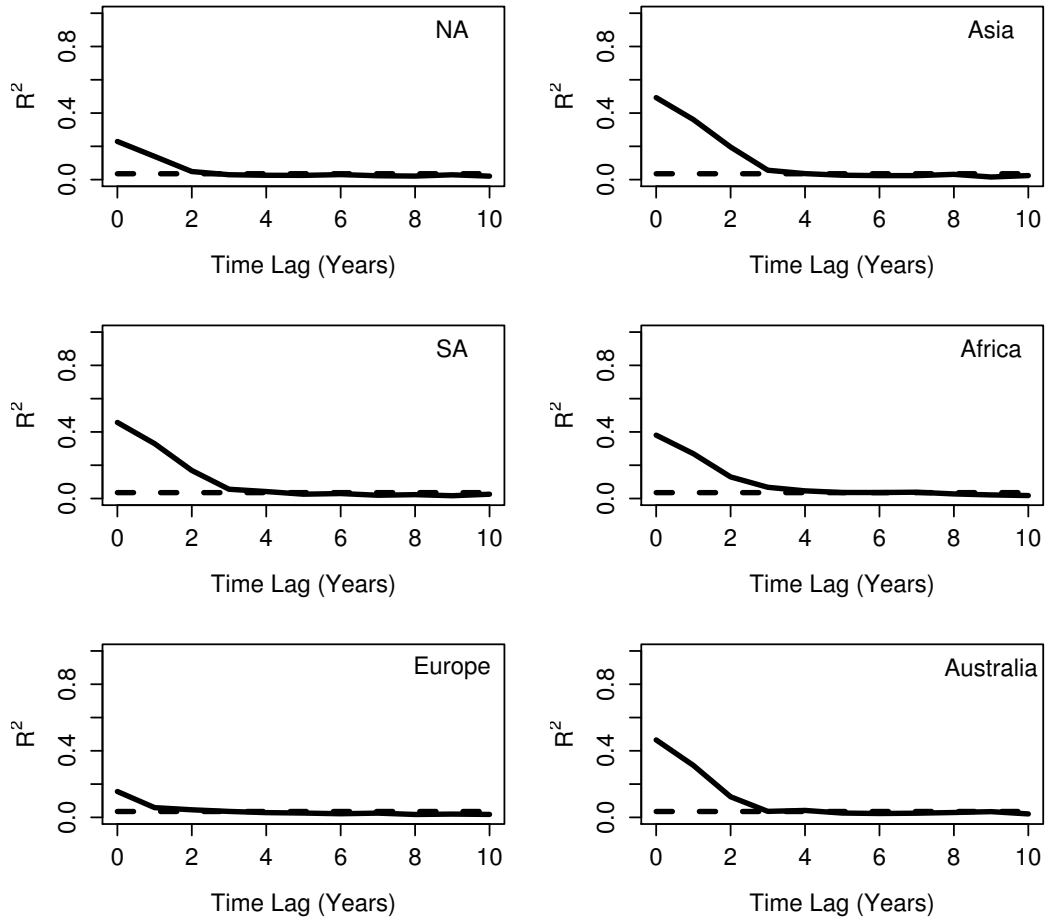

Figure 3.11: Multi-model R^2_τ values of the leading component of precipitation derived from eight independent CMIP3 control simulations (the simulations are independent in the sense that they were not used to derive the predictable component). The horizontal dashed lines in each panel indicate the 5% significance level.

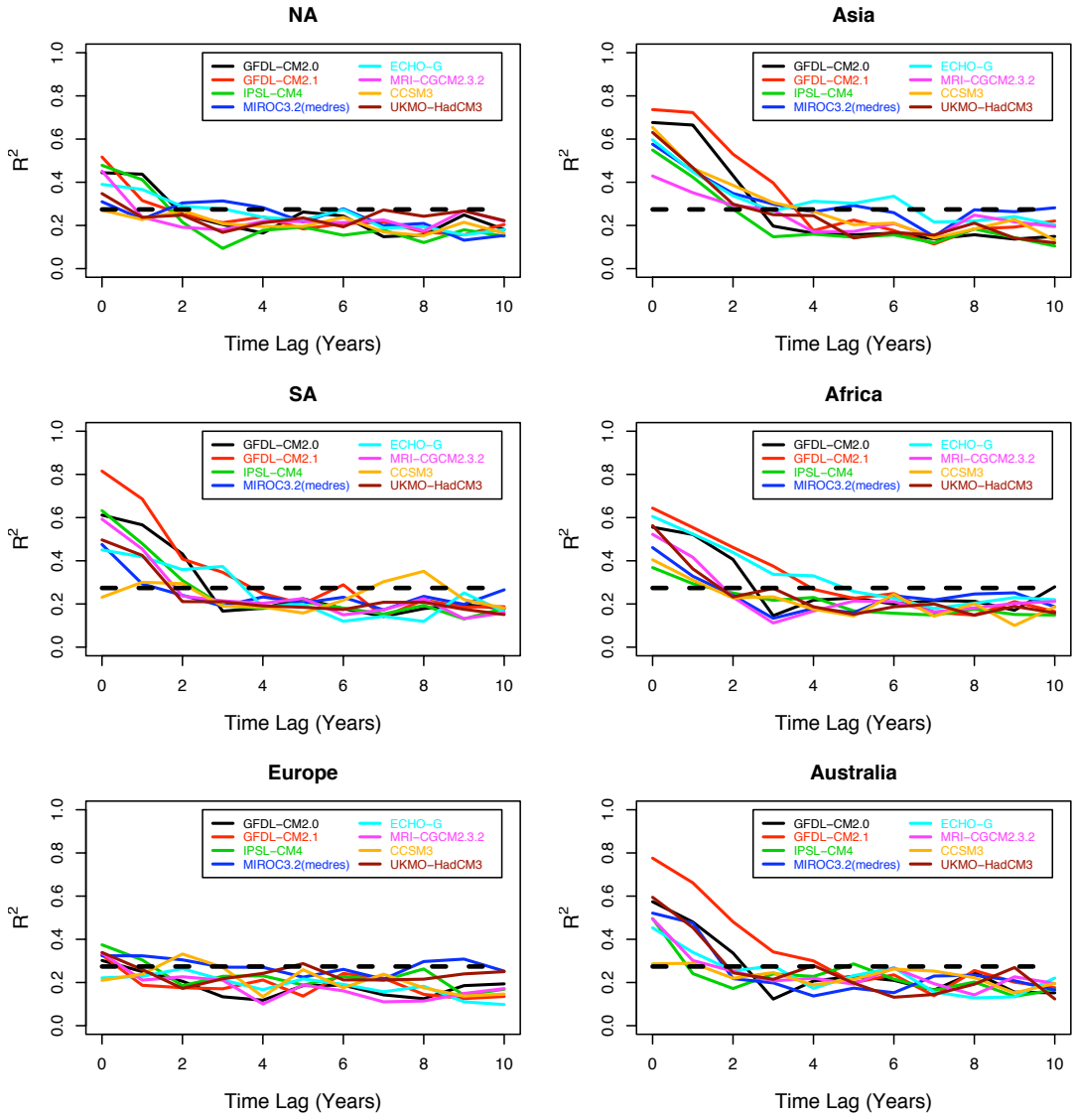

Figure 3.12: R^2_τ values in individual models for the leading multi-model component of precipitation. R^2_τ is calculated from CMIP3 control simulations independent of the data used to calculate the predictable component. Each colored line represents a particular model. The horizontal dashed lines in each panel indicate the 5% significance level.

evolves with lead time. Specifically, the negative correlations in eastern Pacific at lead 2 years become positive at lead 1 year, and the positive correlations appear in tropical Atlantic and Indian Ocean at lead zero. In Europe, the predictable component is coupled with simultaneous tropical SSTs, but has little correlation with topical SSTs at previous years. In North America, significant correlations are shown in tropical oceans, although the pattern does not look like the ENSO pattern in other continents due to opposite signs in some regions. Nevertheless, this predictable component is significantly correlated with the NINO3 index, suggesting the component also is ENSO-related.

3.7 Summary and discussion

This chapter used a new statistical optimization technique, called generalized APT analysis, to identify the most predictable components of annual mean SAT and precipitation in six continents on multi-year time scales. The generalized APT analysis maximizes a measure of predictability derived from linear regression model, in which global SST is the predictor and regional SAT (or precipitation) is the predictand. This method prevents the degradation of predictability that comes from reducing the size of the predictor domain, and clarifies the source of the predictability. The data analyzed are control simulations of fully-coupled models from the CMIP3 archive that are a least 300 years long, do not have significant trends, and have similar variances. The control runs do not contain natural or anthropogenic forcing, and hence the predictability detected here is unforced. APT was maximized by deriving a linear regression model from the leading 30 PCs of 150 years of eight selected control runs, and then maximizing the integrated prediction variance of the resulting regression model. The resulting components were then validated on independent data by testing the significance of the multiple correlation between the component and the leading 30 PCs.

One motivation for generalizing APT analysis was the failure of traditional regression methods to identify land predictability. That is, we maximized APT over the global SST and derived a significant predictable component, but regression analysis revealed that this

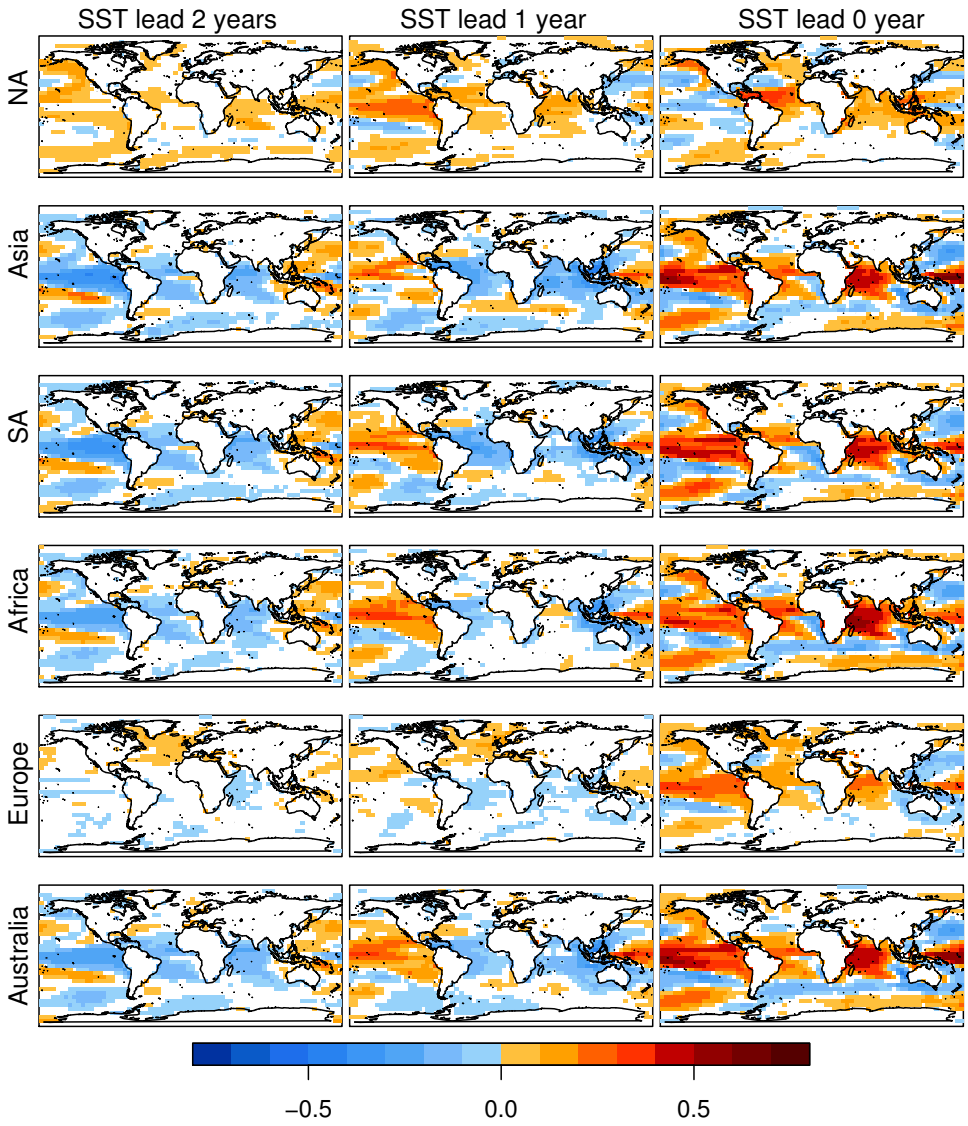

Figure 3.13: Correlation patterns of global SST with the leading component of precipitation in six individual continents at 2 years (left column), 1 year (middle column) and 0 year (right column) lead. Each row represents a particular continent and each column represents a particular lead time. Insignificant areas are masked out based on Student's t-test at a 5% significance level.

component had little to no influence on land temperature and precipitation. This failure to detect land predictability merely implies that the most predictable component in the ocean is unrelated to land predictability. However, it is possible that other ocean components (with less predictability) can lead to significant predictability over land.

The leading predictable component of SAT in each continent can be predicted in independent verification data by a linear regression model with statistically significant skill for 3-6 years, depending on continent. The most predictable components are temperatures over North America and Asia. Each continental pattern is of single sign, suggesting that each one may be difficult to distinguish from anthropogenically forced components based on spatial structure alone. Lagged correlation maps indicate that the predictability of SAT arises from two SST patterns: ENSO, and the persistence of SSTs near the continent under investigation. The only exception is Europe, which has no significant ENSO relation and where the predictability of SAT is related to the persistence of SSTs in North Atlantic. The correlation patterns are reproducible in individual models. In the case of precipitation, the leading component can be predicted by a linear regression model with statistically significant skill for 1-3 years in each continent except Europe, which has no verifiable multi-year predictability in precipitation. Also, the leading North American component in precipitation is not well separated from other components. The multi-year predictability of precipitation in each continent is related primarily to ENSO.

The fact that a significant portion of the multi-year predictability in continental SAT and precipitation appears to arise from an ENSO-like SST pattern is consistent with the well established fact that most seasonal-to-interannual predictability arises from ENSO variability. This consistency between our analysis and previous studies might lead one to assume that nothing new has been learned. However, this is not the case: this study is distinguished from previous studies in that predictability has been diagnosed by an optimal procedure. Even if the procedure produced a pattern that had been diagnosed previously, we still learn something new, namely that no other pattern can be found to have more predictability. The results of this study are definitive in the sense that no other analysis of PC-filtered,

annual mean CMIP3 control simulations can reveal more multi-model predictability than has been diagnosed here, at least if predictability is measured by APT and the same eight models are analyzed. Aside from this, our results do reveal something apparently new, namely that ENSO is not the only source of predictability in each continent – in addition, another component characterized by persistent SSTs near the continent under investigation also appears to be relevant. In general, the ENSO component appears to dominate in the first two lead years, while the persistent component dominates for leads three to six years, depending on continent and variable. The main exception is Europe, which has no significant ENSO component and no verifiable predictability in precipitation.

We found that the predictability derived from the generalized method using global SST as predictor was comparable to or greater than the predictability derived from land SAT (or precipitation) as the predictor. This result strongly suggests that the source of multi-year land predictability identified here arises from SST. Moreover, the fact that using land SAT (or precipitation) as predictors did not enhance the predictability beyond using SST as predictors suggests that other mechanisms such as land-atmosphere feedbacks do not contribute significantly to multi-year land predictability. The only exception to this rule was North America, where we found that predictors based on SAT lead to more predictability than predictors based on SST. We verified that the SST regression pattern associated with this predictable component could be represented by the leading PCs, indicating that the PC-filtering did not prevent this SST pattern from being detected in the generalized APT. Thus, these results suggest that an additional source of multi-year land predictability other than SST exists over North America, presumably related to land-atmosphere interactions.

Predictability identified in the pre-industrial control runs may not exist in forced runs or in the observed climate system. In this sense, any predictability identified by analyzing the control runs in this chapter is "potential predictability." However, the meaning of the word "potential" is stronger than in the sense normally used in predictability studies (e.g., Boer and Lambert 2008). Specifically, the predictability identified by our analysis can be predicted in the control runs with a linear regression model using SST as initial condition.

In contrast, "potential predictability" in most other studies refers to "extra variance" beyond some null hypothesis reference, but the source of the extra variance is usually not clear. In our case, the source of the predictability is clear, and an explicit model for making predictions is generated as part of the analysis. The unforced predictability detected in this chapter has not been verified with observations. We will investigate whether the predictability diagnosed in control runs can be detected in observations in chapter 5. The physical mechanisms that give rise to the other characteristic SST patterns, especially those in the North Atlantic and North Pacific, was not investigated and therefore needs further study.

Chapter 4: Identification of Externally Forced Predictability over Land on Continental Scales

Besides the internal variability in the climate system discussed in chapter 3, the climate system also responds to external radiative forcings. The response is predictable to the extent that the external forcings are predictable. To understand more about the regional climate changes to which societies need to adapt, we need to refine our understanding of the effects of external forcing and internal variability (Stott et al., 2010).

The key objective of this chapter is to identify externally forced predictable components (i.e., atmospheric response to external forcings) over land in multiple climate models. The presence of internal variability in the realistic climate models means that identifying forced components leads to a statistical "signal to noise" problem. Several approaches have been utilized previously to diagnose the forced components, including linear trend analysis and spatial average (Knight et al., 2006; Trenberth and Shea, 2006). However, these approaches have some limitations. For example, the linear trend analysis assumes that the forced trend is linear and uniform over time. In this study, we apply a statistical optimization approach, called discriminant analysis, to estimate the response to external forcings by maximizing the ratio of externally forced variance to natural internal variability. Discriminant analysis maximizes the detectability of the response, thus the response is more likely to be detected in observations.

4.1 Introduction to discriminant analysis

Discriminant analysis (DA) is a method for finding a linear combination of variables that maximizes some measure of difference between two data sets. In our case, the difference is measured by the ratio of variance between two data sets, which are the output of control

runs in the absence of external forcings and the output of the 20th century runs forced by external forcings, respectively. As we assume that internally unforced and externally forced variability in the 20th century runs are independent and additive, the variance in the 20th century runs consists of two parts. One is the variance due to internal unforced variability, written as σ_U^2. One is the variance due to forced variability, written as σ_F^2. In contrast, the variance in the control runs is induced only by the unforced variability, because the external forcings are fixed. The ratio of variance between the 20th century and control runs is $\frac{\sigma_F^2+\sigma_U^2}{\sigma_U^2} = 1 + \frac{\sigma_F^2}{\sigma_U^2}$, which is greater or equal to one. The goal of DA is to find a linear combination of variables that maximizes the variance ratio. Consider two centered data sets from the 20th century runs $\mathbf{X}$ and control runs $\mathbf{Y}$ with dimension of $N_x \times M$ and $N_y \times M$ respectively, where N_x and N_y are identified with the number of time steps (samples) in each run, and M is identified with the state dimension. The state dimension M in the two data sets should be equal, while the number of time steps can differ. Let the weights of the linear combination be $\mathbf{q}$, projecting $\mathbf{q}$ onto the data yields a time series. The two time series corresponding to each data set are

$$\mathbf{r}_x = \mathbf{X}\mathbf{q} \tag{4.1}$$

and

$$\mathbf{r}_y = \mathbf{Y}\mathbf{q}. \tag{4.2}$$

Based on (4.1) and (4.2), the variances of these two time series $\mathbf{r}_x$ and $\mathbf{r}_y$ are estimated as

$$\hat{\sigma}_x^2 = \frac{\mathbf{r_x^T r_x}}{N_x} = \mathbf{q^T}\left(\frac{\mathbf{X^T X}}{N_x}\right)\mathbf{q} = \mathbf{q^T}\hat{\mathbf{\Sigma}}_x\mathbf{q} \tag{4.3}$$

and

$$\hat{\sigma}_y^2 = \frac{\mathbf{r_y^T r_y}}{N_y} = \mathbf{q^T}\left(\frac{\mathbf{Y^T Y}}{N_y}\right)\mathbf{q} = \mathbf{q^T}\hat{\mathbf{\Sigma}}_y\mathbf{q}, \tag{4.4}$$

where $\hat{\mathbf{\Sigma}}_x$ and $\hat{\mathbf{\Sigma}}_y$ are sample covariance matrices of $\mathbf{X}$ and $\mathbf{Y}$, respectively. The hat symbol denotes sample quantities estimated from data. The ratio of these two variances is

$$\frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2} = \frac{\mathbf{q^T}\hat{\mathbf{\Sigma}}_x\mathbf{q}}{\mathbf{q^T}\hat{\mathbf{\Sigma}}_y\mathbf{q}}. \tag{4.5}$$

Notice that (4.5) is a Rayleigh quotient. As shown in appendix B, the problem of maximizing the Rayleigh quotient leads to the generalized eigenvalue problem

$$\hat{\mathbf{\Sigma}}_x\mathbf{q} = \lambda\hat{\mathbf{\Sigma}}_y\mathbf{q}. \tag{4.6}$$

The above equation (4.6) has more than one solution (i.e., more than one eigenvalue and eigenvector). The eigenvalues of (4.6) turn out to be variance ratios in (4.5) corresponding to each eigenvector $\mathbf{q}$. The eigenvectors are the projection vectors which maximize the variance ratios. Each eigenvector corresponds to a discriminant component. Let the complete set of eigenvectors be denoted as $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_M$, and collected in the matrix

$$\mathbf{Q} = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \ldots \quad \mathbf{q}_M]. \tag{4.7}$$

Then, the generalized eigenvalue problem (4.6) can be written equivalently as

$$\hat{\mathbf{\Sigma}}_x\mathbf{Q} = \hat{\mathbf{\Sigma}}_y\mathbf{Q}\mathbf{\Lambda}, \tag{4.8}$$

where $\mathbf{\Lambda}$ is the diagonal matrix whose diagonal elements equal the corresponding eigenvalues of (4.6). The matrices $\hat{\mathbf{\Sigma}}_x$ and $\hat{\mathbf{\Sigma}}_y$ are symmetric. It is shown in appendix C that the time series in (4.1) produced by two different eigenvectors are uncorrelated. In addition, the eigenvectors are unique up to a multiplicative constant, which can be chosen to normalize

the time series to unit variance. These properties can be summarized as

$$\mathbf{Q}^{\mathrm{T}}\hat{\mathbf{\Sigma}}_y\mathbf{Q} = \mathbf{I}. \tag{4.9}$$

Multiplying both sides of (4.8) by $\mathbf{Q}^{\mathrm{T}}$ gives

$$\mathbf{Q}^{\mathrm{T}}\hat{\mathbf{\Sigma}}_x\mathbf{Q} = \mathbf{\Lambda}. \tag{4.10}$$

It is convention to order eigenvectors in decreasing order, in which case the first eigenvector maximizes the variance ratio, the second eigenvector maximize the variance ratio subject to being uncorrelated with the first, and so on.

To derive the physical patterns associated with each component, we collect the time series for all components into a single matrix associated with both $\mathbf{X}$ and $\mathbf{Y}$ respectively as

$$\mathbf{R}_x = \mathbf{XQ} = [\mathbf{r}_{x1} \quad \mathbf{r}_{x2} \quad \dots \quad \mathbf{r}_{xM}] \tag{4.11}$$

and

$$\mathbf{R}_y = \mathbf{YQ} = [\mathbf{r}_{y1} \quad \mathbf{r}_{y2} \quad \dots \quad \mathbf{r}_{yM}], \tag{4.12}$$

where we have used (4.1) and (4.7). Invoking (4.11), the orthogonality and normalization constraints in (4.9) and (4.10) imply

$$\frac{1}{N_y}\mathbf{R}_y^{\mathrm{T}}\mathbf{R}_y = \mathbf{I} \quad \text{and} \quad \frac{1}{N_x}\mathbf{R}_x^{\mathrm{T}}\mathbf{R}_x = \mathbf{\Lambda}. \tag{4.13}$$

Since the original data can be decomposed into components corresponding to each eigenvector, the physical pattern $\mathbf{P}$ can be derived by minimizing

$$||\mathbf{Y} - \mathbf{R}_y\mathbf{P}^{\mathrm{T}}||, \tag{4.14}$$

where ||.|| denotes the Frobenius norm of a matrix. Minimization of (4.14) is a standard least squares problem with solution

$$\mathbf{P} = \mathbf{Y}^{\mathrm{T}}\mathbf{R}_y \left(\mathbf{R}_y^{\mathrm{T}}\mathbf{R}_y\right)^{-1} = \frac{1}{N_y}\mathbf{Y}^{\mathrm{T}}\mathbf{R}_y = \hat{\mathbf{\Sigma}}_y\mathbf{Q}, \tag{4.15}$$

where (4.11) and (4.13) have been used. The matrix $\mathbf{P}$ consists of all spatial patterns associated with each component, which can be written as

$$\mathbf{P} = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \ldots \quad \mathbf{p}_M] = [\hat{\mathbf{\Sigma}}_y\mathbf{q}_1 \quad \hat{\mathbf{\Sigma}}_y\mathbf{q}_2 \quad \ldots \quad \hat{\mathbf{\Sigma}}_y\mathbf{q}_M], \tag{4.16}$$

where $\mathbf{p}_k = \hat{\mathbf{\Sigma}}_y\mathbf{q}_k$ $(k = 1, 2, \ldots, M)$ indicates the spatial pattern of the kth component. Note that the patterns in (4.16) are derived by decomposing data $\mathbf{Y}$. It can be shown that the same spatial patterns can be derived by decomposing data $\mathbf{X}$, hence, only one spatial pattern corresponds to each component and there is no need to distinguish the pattern matrix $\mathbf{P}$ by a subscript x or y. Since the time series associated with two distinct components are uncorrelated, the original data $\mathbf{X}$ and $\mathbf{Y}$ can be decomposed as

$$\mathbf{X} = \mathbf{r}_{x1}\mathbf{p}_1^{\mathrm{T}} + \mathbf{r}_{x2}\mathbf{p}_2^{\mathrm{T}} + \ldots + \mathbf{r}_{xM}\mathbf{p}_M^{\mathrm{T}}, \tag{4.17}$$

$$\mathbf{Y} = \mathbf{r}_{y1}\mathbf{p}_1^{\mathrm{T}} + \mathbf{r}_{y2}\mathbf{p}_2^{\mathrm{T}} + \ldots + \mathbf{r}_{yM}\mathbf{p}_M^{\mathrm{T}}. \tag{4.18}$$

4.2 Model data

Models used in this chapter are identical to the models used in chapter 3. In addition to the 300-year runs of annual average SAT and precipitation from CMIP3 pre-industrial control simulations of each model, the annual average SAT and precipitation from CMIP3 20th century runs are used in this chapter. The 20th century runs are initialized from a specific point in the pre-industrial control runs, and forced by historic, varying concentrations

of well-mixed greenhouse gases and sulfate aerosols, and in some models by other anthropogenic (e.g., black carbon particulate or land-use patterns) and natural (solar radiation and volcanic aerosols) forcings (Biasutti et al., 2008). Because the 20th century runs are forced by external radiative forcings, they are simply called forced runs in this study to distinguish them from the pre-industrial control runs, called unforced runs. In the forced runs, the number of ensemble members and number of years vary with models. We use a maximum of five ensemble members in each model, and if the ensemble members are less than five in a model, we use all available ensemble members. The number of ensemble members used in each model and the number of years in each forced run are listed in Table 3.1.

The climatology of the period 1961-1990 in each ensemble member was subtracted from the forced runs to be consistent with the climatology in the observations that will be used in the next chapter. The data from each individual ensemble member of all models are lined up in temporal dimension to form a single multi-model data set. For example, if each ensemble member has a dimension of $N_k \times M$ ($k = 1, 2, \ldots, n$), where N_k denotes the number of time steps in the kth ensemble member, M denotes the state dimension, and n denotes the total number of ensemble members in all models (n=27 in our case), then the multi-model data set from the forced runs has a dimension of $(N_1 + N_2 + \ldots + N_n) \times M$. The multi-model data from eight selected control runs with 300 years in each control run has a dimension of $2400 \times M$. The two data sets required in DA are the multi-model data sets from control runs and forced runs in this study.

4.3 Statistical analysis

4.3.1 Statistical details in DA

Similar to APT analysis, DA is subject to overfitting if the state dimension is not a small fraction of the sample size. To mitigate overfitting, we reduce the dimension of the data by projecting the data onto a few leading PCs. The PCs of SAT and precipitation from

multi-model control runs in each continent have been computed in the APT analysis. The PCs of the forced runs in each continent are then derived by projecting the pseudo-inverse of EOFs from the control runs onto the multi-model forced runs. The multi-model sample covariance matrices from control runs and forced runs are computed respectively based on the leading 30 PCs, and then substituted into the eigenvalue problem (4.6) to compute eigenvector $\mathbf{q}$. In (4.6), $\hat{\mathbf{\Sigma}}_x$ represents the sample covariance matrix in the forced runs, and $\hat{\mathbf{\Sigma}}_y$ represents the sample covariance matrix in the control runs. Here, the first 150 years of PCs (i.e., "training data") in control runs have been used to compute the sample covariance matrix. We chose 150 years of PCs because this sample size is comparable to the sample size of each ensemble member in the forced runs. However, all available data in the forced runs are used to compute the forced sample covariance matrix. Once $\mathbf{q}$ is obtained, the corresponding physical pattern and the forced and unforced time series can be derived.

4.3.2 Statistical significance test in DA

To test the statistical significance of the variance ratio between forced and unforced runs, we use Monte Carlo methods to estimate the sampling distribution under the null hypothesis of no forced predictability. Specifically, we generated two data sets with the same *dimension* as the multi-model data sets from control runs and forced runs respectively, by drawing random numbers from a Gaussian distribution with zero mean and unit variance. We chose unit variance because the results are invariant to invertible linear transformation. The state dimension in multi-model control runs and forced runs is 30 in our case. The sample size is 150×8 in multi-model control runs, and is the sum over the time steps of all available ensemble members and all selected models in multi-model forced runs. These two data sets have been substituted into (4.6) to produce ratios of variance (i.e. eigenvalues of (4.6)). This procedure was repeated 100 times, and each ratio was sorted by decreasing order. The upper and lower fifth percentiles of the ratios associated with each component were selected to generate the 90% confidence interval. The null hypothesis of no significant forced predictability is rejected if the ratio lies outside the 90% confidence interval.

4.4 Results

4.4.1 Discriminant analysis of continental SAT

First, DA is applied to the multi-model forced runs and control runs of SAT to discriminate between the forced response and the unforced natural variability. The optimized ratios of forced-to-unforced variance for all 30 components in each continent are shown in Fig. 4.1. The shaded region in each panel shows the upper and lower fifth percentiles estimated from Monte Carlo methods, which is the 90% confidence interval. The confidence intervals are narrow in all continents, suggesting there is low susceptibility to sampling problems because of the large amount of data in our case. This figure shows that only the ratio of the first component is clearly separated from the others and is well outside the 90% confidence interval in all six continents, suggesting that only one forced component in each continent is statistically significant and hence predictable. Only one forced predictable component emerges in each continent indicating that the forced response to different natural and anthropogenic forcings are similar. There is no statistically significant forced response in the other components. Among all the continents, the variance ratio of the leading component in Europe is the smallest, which means the forced response in Europe is the weakest.

The time series of the leading forced predictable component in six continents are shown in Fig. 4.2. The thick black line in each panel indicates the ensemble mean forced time series, and each thin colored line represents the time series of an individual ensemble member. The time series of the ensemble mean in each continent shows an increasing trend. The weakest trend is found in Europe, which is consistent with the smallest variance ratio of forced runs to control runs in Europe shown in Fig. 4.1, suggesting that the response to external forcings is less predictable in Europe than in the other continents.

The physical patterns of the leading forced predictable component in all continents, shown in Fig. 4.3, are all single sign. The positive sign and increasing trend in each continent indicates warming on continental scales. The spatial structures are showing large similarity to the unforced predictable patterns from the control runs in Fig. 3.5. The similarities

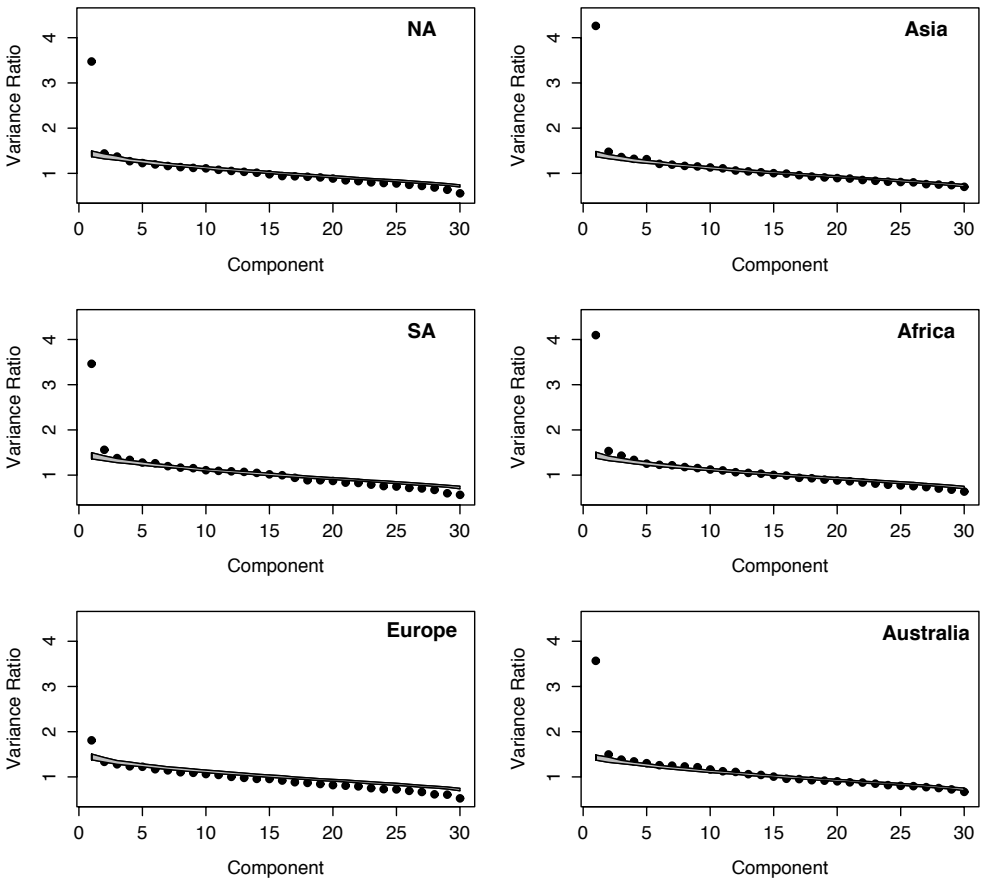

Figure 4.1: Optimized ratios of forced to unforced variance for SAT, as determined by discriminant analysis of the leading 30 multi-model PCs. The shaded region in each panel shows the upper and lower fifth percentiles estimated from Monte Carlo methods under the null hypothesis of no forced response.

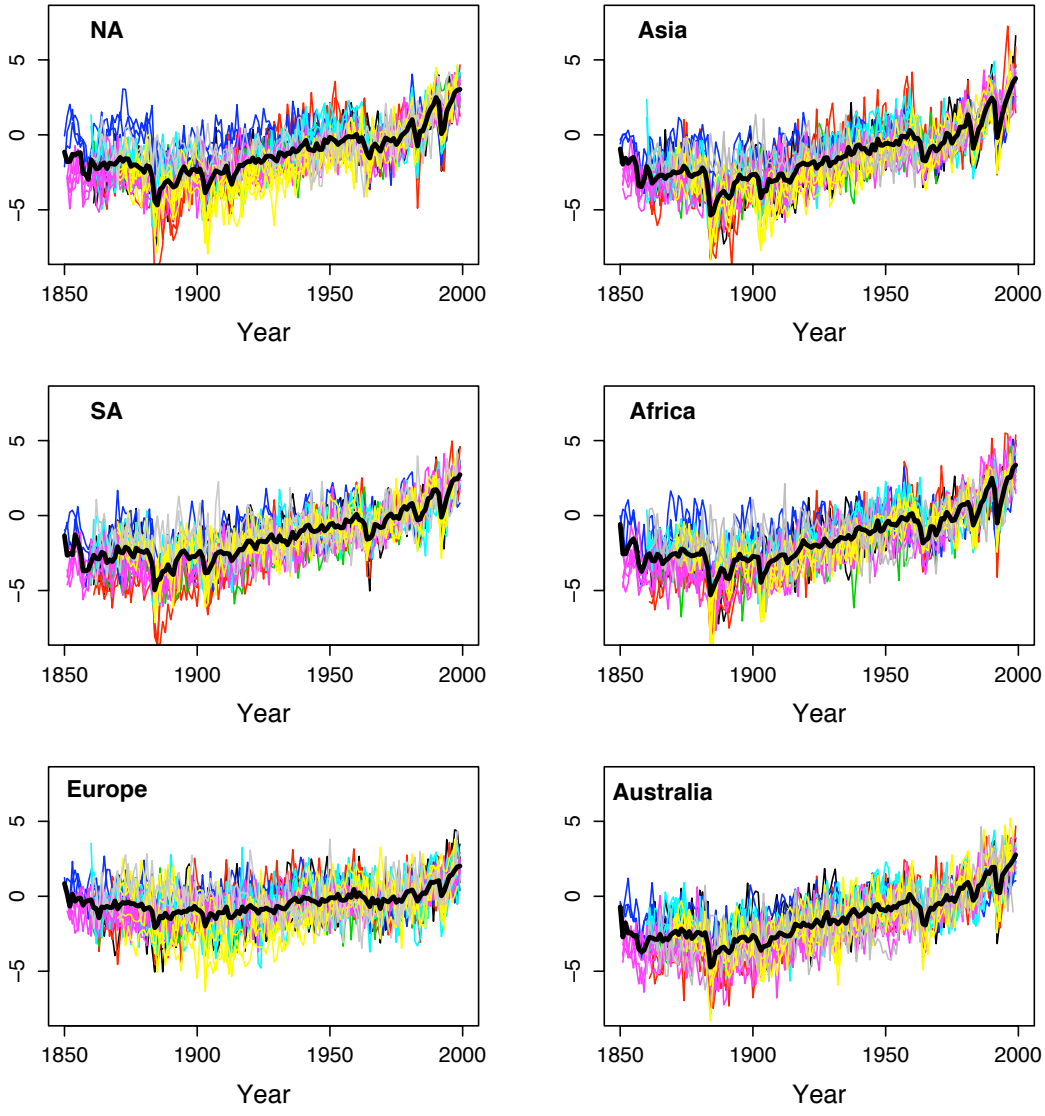

Figure 4.2: Time series of SAT for the forced predictable component for each ensemble member (thin colored lines) and ensemble mean (thick black lines) in six continents estimated from discriminant analysis (see (4.1)). The ensemble members of the same model are indicated by the same color.

between the forced and unforced patterns produce difficulties in distinguishing them in space. In these spatial patterns, largest amplitudes are concentrated in high latitudes such as Northern North America and Northeast of Europe.

4.4.2 Discriminant analysis of continental precipitation

The optimized variance ratios of precipitation for all 30 components in each continent are shown in Fig. 4.4. In contrast to SAT, none of the ratios are well outside of the 90% confidence interval. Although the variance ratios of the leading components in South America and Australia are barely statistically significant, we ignore them since they are so marginal. Because the variance ratios are close to the 90% confidence interval estimated from Monte Carlo methods under the null hypothesis of no forced response, we conclude that response of land precipitation to external forcings cannot be identified in model simulations.

4.5 Summary and discussion

This chapter diagnosed the forced components of SAT and precipitation in six continents using discriminant analysis on a multi-model basis. Both pre-industrial control runs and the 20th century runs of eight selected models from CMIP3 archive are used. The forced runs are assumed to have additional variability induced by the external forcings, which is independent of internal variability. The externally forced components can be diagnosed by maximizing the ratio of forced variability to internal variability.

In each continent, only one forced predictable component of SAT is identified. The time series of the forced components have increasing trends, with the smallest trend in Europe. The spatial patterns of the forced predictable components are all single sign and similar to the spatial patterns of the unforced predictable components described in chapter 3. The similarity between forced and unforced predictable patterns may cause difficulties in separating them in observations. Largest amplitudes of the forced patterns of SAT tend to concentrate in high latitudes. However, no significant forced components are identifiable in

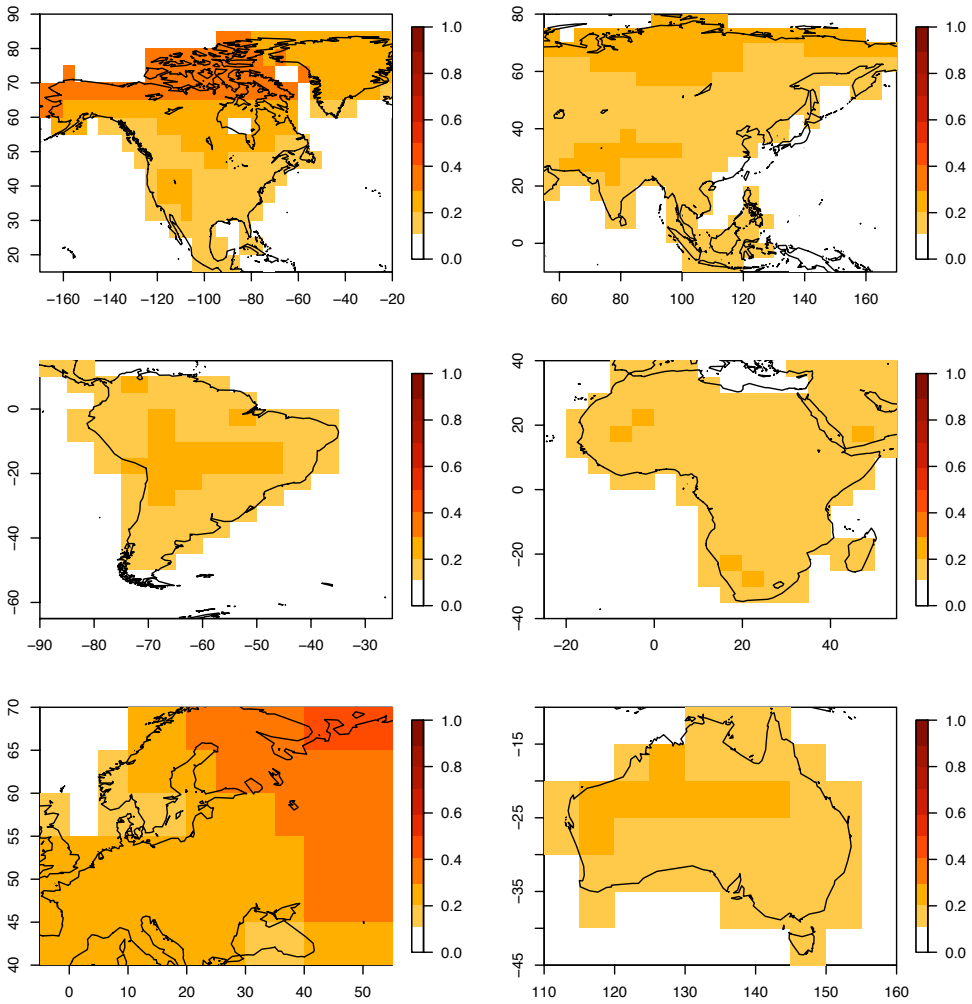

Figure 4.3: Spatial patterns of the forced response in all continents, obtained by maximizing the ratio of variances between forced and unforced runs (see (4.15)). The spatial pattern has units of degrees Kelvin. The amplitudes are the deviations from time mean.

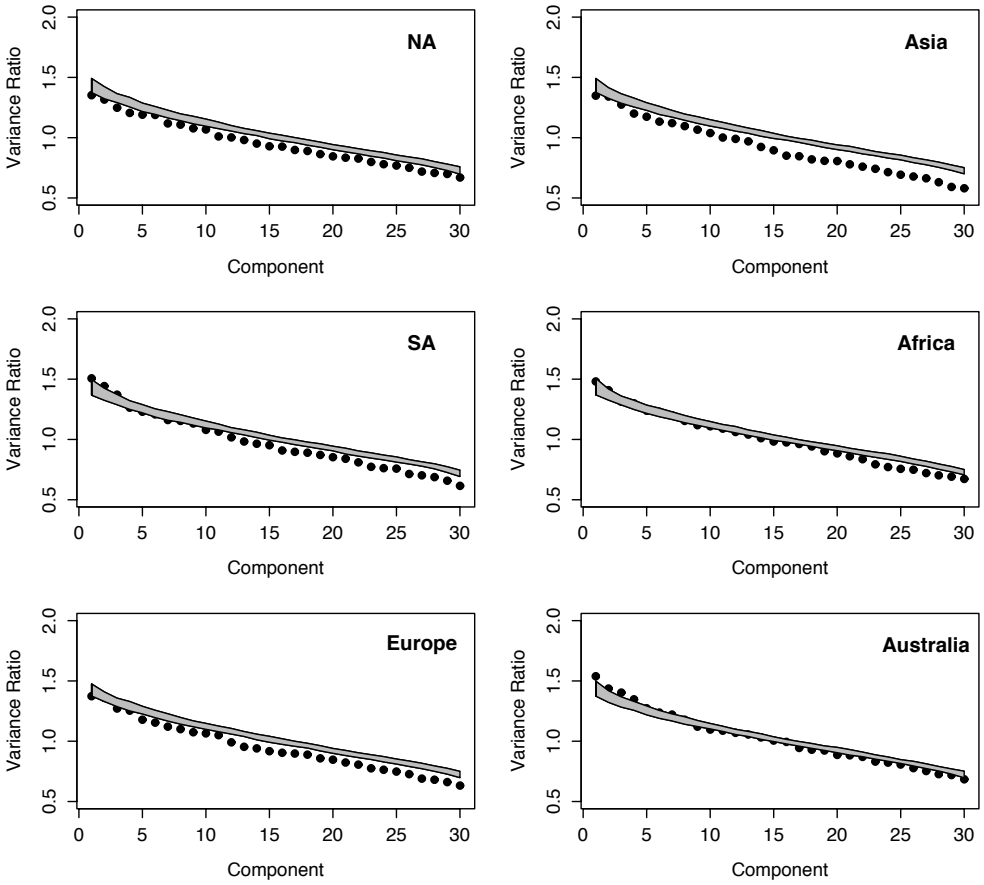

Figure 4.4: Optimized ratios of forced to unforced variance for precipitation, as determined by discriminant analysis of the leading 30 multi-model PCs. The shaded region in each panel shows the upper and lower fifth percentiles estimated from Monte Carlo methods under the null hypothesis of no forced response.

continental precipitation. It is not surprising that precipitation response to external forcings is difficult to be diagnosed since previous studies have suggested that the forced response of precipitation is weak in models (Held and Soden, 2006; Lambert et al., 2005, 2004; Min et al., 2008; Zhang et al., 2007). It might be even more difficult in our case because the anthropogenic and natural forcings are combined in models, the response to each kind of forcing could be diluted such that the overall response is not identifiable. The fact that forced components of land precipitation is not identifiable on a multi-model basis does not exclude the possibility that they are identifiable in individual models, because there is a wide range of precipitation response among models (Lambert et al., 2005).

The results that forced components of land precipitation are not identifiable on continental scales might appear to contradict previous studies that suggested forced response in land precipitation (Min et al., 2008; Zhang et al., 2007). However, there is no real contradiction. First, the precipitation response to external forcing diagnosed in previous studies was very weak, and the statistical significance of the response was not tested. Second, since the response varies with models, selecting different sets of models could lead to different results.

Our results based on discriminant analysis do show something new compared to previous studies, namely that there is *only* one predictable forced component in land SAT in each continent (i.e., no other strong forced patterns can be diagnosed), and no significant forced pattern could be identified for land precipitation on continental scales on a multi-model basis.

The forced predictable components in land SAT diagnosed in this chapter were based on model simulations. The results were not confirmed in the observational data yet. In this sense, the forced predictability identified from model simulations also is considered as "potential" predictability. The detection of the forced predictability in observations will be discussed in the next chapter.

Chapter 5: Separating Unforced and Forced Predictability in Observations

5.1 Background

Numerous studies have suggested that the observed temperature changes in the 20th century are unlikely to be explained by natural causes (i.e., natural external forcing and internal variability) both on global scales and continental-to-subcontinental scales (Hegerl et al., 2007) [sec. 9.4.2]. There is increasing interest in climate detection and attribution on continental scales because of its critical role for local planners and regulators (Stott, 2003; Zhang et al., 2006; Zwiers and Zhang, 2003). However, there are limitations in climate change detection and attribution on continental scales. Namely, internal variability increases at smaller spatial scale, which decreases the signal-to-noise ratio; forcing that could be important at small scales such as land use change are uncertain and may not have been included in models used for detection; the credibility of small-scale details of climate simulated by models is lower than for large-scale features (Hegerl et al., 2007) [sec. 9.4.2.2].

Because of its importance, we attempt to separate the unforced and forced predictability identified in chapter 3 and 4 in observations using optimal fingerprinting technique. As the forced components of land precipitation cannot be identified, as discussed in chapter 4, we do not distinguish them from natural variability in observations. Considering that chapter 4 suggested that there is one single forced predictable component of land SAT in each continent, we do not separate the effects of various sources of forcings (e.g., anthropogenic and natural forcings).

5.2 Introduction to optimal fingerprinting

Optimal fingerprinting is generalised multivariate regression adapted to the detection of climate change and the attribution of change to externally-forced climate change signals (Allen and Tett, 1999; Hasselmann, 1997, 1979). General application of optimal fingerprinting is to detect and attribute climate change due to various sources of external forcings. DelSole et al. (2011) expanded the standard detection and attribution framework by including an unforced pattern in addition to the forced pattern, such that it diagnoses both forced and unforced components. We applied this technique to distinguish forced variability from unforced internal variability, instead of distinguishing different sources of external forcings. Note that our forced variability arises from both anthropogenic and natural forcings. The ability to distinguish unforced and forced variability depends on the extent to which these two patterns differ (DelSole et al., 2011). The linear regression has the form

$$\underset{[M\times T]}{\mathbf{O}} = \underset{[M\times 1]}{\mathbf{p}_F}\ \underset{[1\times T]}{\boldsymbol{\beta}_F} + \underset{[M\times 1]}{\mathbf{p}_N}\ \underset{[1\times T]}{\boldsymbol{\beta}_N} + \underset{[M\times T]}{\boldsymbol{\epsilon}}, \tag{5.1}$$

where $\mathbf{O}$ denotes observations, $\mathbf{p}_F$ denotes externally forced pattern obtained from forced-to-unforced discriminant in forced runs, $\mathbf{p}_N$ denote the unforced pattern derived from APT analysis in pre-industrial control runs, $\boldsymbol{\beta}_F$ and $\boldsymbol{\beta}_N$ denote amplitudes associated with forced pattern and unforced pattern, $\boldsymbol{\epsilon}$ denotes the "noise" term (i.e. unpredictable internal variability) varying in space and time. Importantly, the "noise" is not assumed to be independent and identically distributed in space. The M denotes the state dimension (i.e. the number of grid points) and T denotes the number of time steps. At a given time t, (5.1) can be written as

$$\underset{[M\times 1]}{\mathbf{o}_t} = \underset{[M\times 2]}{\mathbf{P}_t}\ \underset{[2\times 1]}{\boldsymbol{\beta}_t} + \underset{[M\times 1]}{\boldsymbol{\epsilon}_t}, \tag{5.2}$$

where the forced and unforced patterns are combined into a single matrix $\mathbf{P}_t$, whose two columns are the patterns $\mathbf{p}_F$ and $\mathbf{p}_N$, $\boldsymbol{\beta}_t$ includes the amplitudes corresponding to forced and unforced respectively and $\mathbf{o}$ is a state vector (i.e., a function of space only). Although patterns $\mathbf{p}_F$ and $\mathbf{p}_N$ are not a function of time, their representation on the observation grid after missing values are masked out does. If the "noise" has covariance matrix $\sigma_\epsilon^2 \hat{\boldsymbol{\Sigma}}_\epsilon$, then the generalized least squares estimate of $\boldsymbol{\beta}$ at a specific time t is

$$\hat{\boldsymbol{\beta}}_t = \left(\mathbf{P}_t^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_\epsilon^{-1} \mathbf{P}_t\right)^{-1} \mathbf{P}_t^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_\epsilon^{-1} \mathbf{o}_t, \tag{5.3}$$

where the hat symbol denotes an estimated quantity. $\hat{\boldsymbol{\Sigma}}_\epsilon$ is the noise covariance matrix estimated from control runs. Note that at a given time t, the matrix $\mathbf{P}_t$ in (5.2) includes merely spatial patterns that are independent of time. This is different from commonly used space-time optimal fingerprinting techniques, where the patterns include a time dimension. The estimated standard error of the i'th element of $\boldsymbol{\beta}_t$, denoted $\hat{se}(\boldsymbol{\beta})_i$, is

$$\hat{se}(\boldsymbol{\beta})_i = \sqrt{\hat{\sigma}_\epsilon^2 \left((\mathbf{P}_t^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_\epsilon^{-1} \mathbf{P}_t)^{-1}\right)_{ii}} \quad , \tag{5.4}$$

where $\hat{\sigma}_\epsilon^2$ is a sample estimate of σ_ϵ^2 given by

$$\hat{\sigma}_\epsilon^2 = \frac{1}{M-2} \left(\mathbf{o}_t - \mathbf{P}_t \hat{\boldsymbol{\beta}}_t\right)^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_\epsilon^{-1} \left(\mathbf{o}_t - \mathbf{P}_t \hat{\boldsymbol{\beta}}_t\right). \tag{5.5}$$

Matrix $\hat{\boldsymbol{\Sigma}}_\epsilon$ does not change with time. However, the pattern matrix $\mathbf{P}_t$ of observations with non-missing values depends on time, hence the covariance matrix $\hat{\sigma}_\epsilon^2 \hat{\boldsymbol{\Sigma}}_\epsilon$ depends on time. The standard error estimated by (5.4) is large if the forced pattern $\mathbf{p}_F$ and unforced pattern $\mathbf{p}_N$ in matrix $\mathbf{P}_t$ resemble each other, in which case $\mathbf{P}_t^{\mathrm{T}} \mathbf{P}_t$ is nearly singular.

5.3 Data

The observational data set used in this chapter is the HadCRUT3 data set of monthly mean global temperature anomalies on a $5^\circ \times 5^\circ$ grid-box basis. Its spatial resolution is the same as in the interpolated pre-industrial control and 20th century runs used in chapter 3 and chapter 4. This data set has been developed by Climatic Research Unit (CRU), in conjunction with Hadley Center of the UK Met Office, and is available online at http://www.cru.uea.ac.uk/cru/data/temperature/. The anomalies are relative to the climatology over 1961-1990. The data set covers the time period from 1850 to 2009, but a large amount of missing values occur before 1900. More details of this data set can be found from Rayner et al. (2006). The annual means are computed from the monthly means of HadCRUT3 data set only if at least 10 months of non-missing values exist in a year, otherwise the annual mean of this year is considered to be missing.

5.4 Statistical details in optimal fingerprinting

To apply the optimal fingerprinting method, we estimated the noise covariance matrix $\hat{\mathbf{\Sigma}}_\epsilon$ by averaging the sample covariance matrix of eight control runs, in which the leading 30PCs are used in each continent. The results may be sensitive to the number of PCs used to represent the dimension of the space. This sensitivity are investigated by choosing the number of PCs ranging from 8 to 40. The unforced spatial pattern $\mathbf{p}_N$ used is the leading unforced predictable component identified from APT analysis, and the forced spatial pattern $\mathbf{p}_F$ is the spatial pattern identified from discriminant analysis. We applied optimal fingerprinting technique to continental SAT.

5.5 Results

To assess the sensitivity of the forced and unforced amplitudes to the number of PCs, we computed the amplitudes as a function of PCs. The amplitudes of the forced predictable

component and the unforced predictable component for 8-40 PCs are shown in Fig. 5.1 and Fig. 5.2. The amplitudes in all continents are clustered together, indicating that the amplitudes are nearly insensitive to the number of PCs. However, more sensitivity is shown in Australia compared to the other five continents, presumably induced by the relative small size of the continent. Also, more uncertainty occurs in the period before 1900 (not shown) presumably because of more missing values in that period.

The amplitudes of the forced patterns estimated from (5.3), expressed as a 95% confidence interval, are shown by the shading area in Fig. 5.3. The amplitudes in each continent show a slight trend at least in recent years, except for Australia and South America. We emphasize that the forced amplitudes take into account both natural and anthropogenic forcings. Largest uncertainties are found in Australia among the six continents, presumably due to its small domain, which filters out less noise from internal variability. The similarity between forced and unforced patterns also could cause large uncertainty.

The amplitude of the forced pattern estimated by averaging the amplitude of each ensemble member from forced-to-unforced discriminant analysis is shown as the red curve in each panel in Fig. 5.3. The amplitudes estimated from discriminant analysis exhibit increasing trends in all six continents, which suggest warming because the spatial pattern of the forced components are in positive sign (see Fig. 4.3). Moreover, the amplitudes estimated from model simulations (i.e., from discriminant analysis) are smoother than those from observations, because the ensemble mean over all forced runs removes the fluctuations induced by internal variability. Checking the consistency between the observed amplitude and the simulated amplitude is equivalent to checking whether the simulated amplitude falls in the 95% confidence interval in each year. In our case, if the red curve lies within the shaded area, the observed amplitude is said to be consistent with simulated amplitude. This condition is satisfied in most years, implying the response to external forcing in observations agree with model simulations. However, the confidence intervals are large in South America, Europe and Australia, which contain zero in most years (i.e., consistent with no trends). One could argue that no trends occur in observations in these three continents.

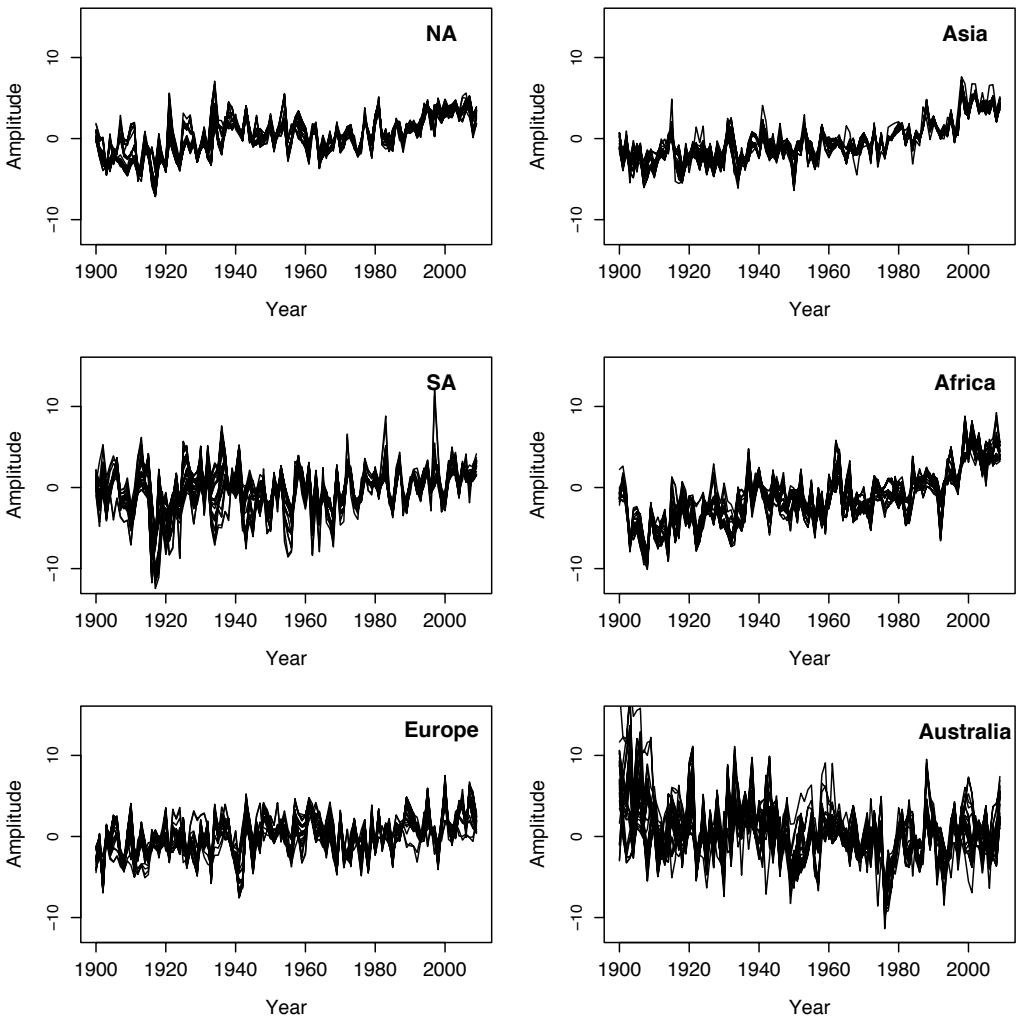

Figure 5.1: Generalized least squares estimate of the amplitudes of the forced predictable component of SAT in HadCRUT3 data set using 8-40 PCs. The result of each PC truncation is shown as a separate curve.

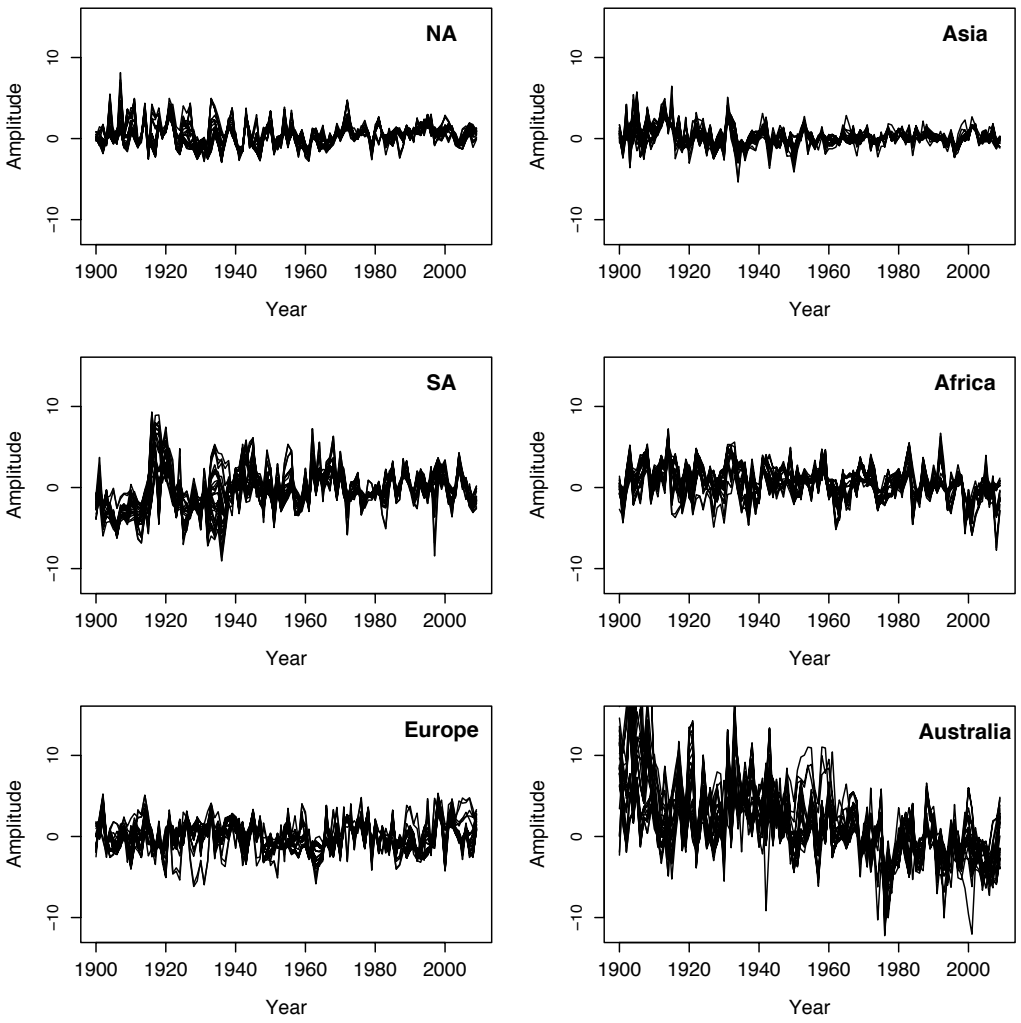

Figure 5.2: Generalized least squares estimate of the amplitudes of the unforced predictable component of SAT in HadCRUT3 data set using 8-40 PCs. The result of each PC truncation is shown as a separate curve.

To detect the forced response in observations, we check that the 95% confidence interval of the forced amplitude in observations does not include zero more than 5% of the time. In other words, we check that the shaded area in each panel contains zero less than 95% of the time. This condition cannot be satisfied in Australia, Europe and South America. However, this condition is marginally satisfied in North America, Asia and Africa. Therefore, we conclude that the forced responses are marginally detectable in North America, Asia and Africa, but not detectable in Australia, Europe and South America.

Fig. 5.4 shows the amplitudes of unforced components in six individual continents estimated from observations (see (5.3)). The shaded areas indicate the 95% confidence interval. The confidence internals in each continent contain zero more than 95% of the time. The time series have so much uncertainty that it is difficult to claim that the time series are predictable on multi-year time scales. We conclude that no multi-year natural variability is detected in any continent.

5.6 Summary and discussion

This chapter attempted to separate unforced and forced predictable components of continental SAT in observational data based on optimal fingerprinting method. By fitting the observations to a linear combination of forced and unforced patterns, the amplitudes and their standard errors corresponding to these two patterns are estimated in each year. If the 95% confidence interval of the forced amplitudes contains zero less than 95% of the time, the forced response is detectable. The results reveal that the forced response cannot be clearly detected on continental scales, suggesting that the forced and unforced patterns are not separable in observations, although the observed amplitudes of the forced pattern marginally agree with the simulated warming in North America, Asia and Africa. This could be attributable to the similarities in the spatial patterns of forced and unforced. However, the forced and unforced patterns are indistinguishable does not necessarily mean that the forced and unforced predictability does not exist in observations.

The fact that we could not distinguish forced and unforced patterns in observations

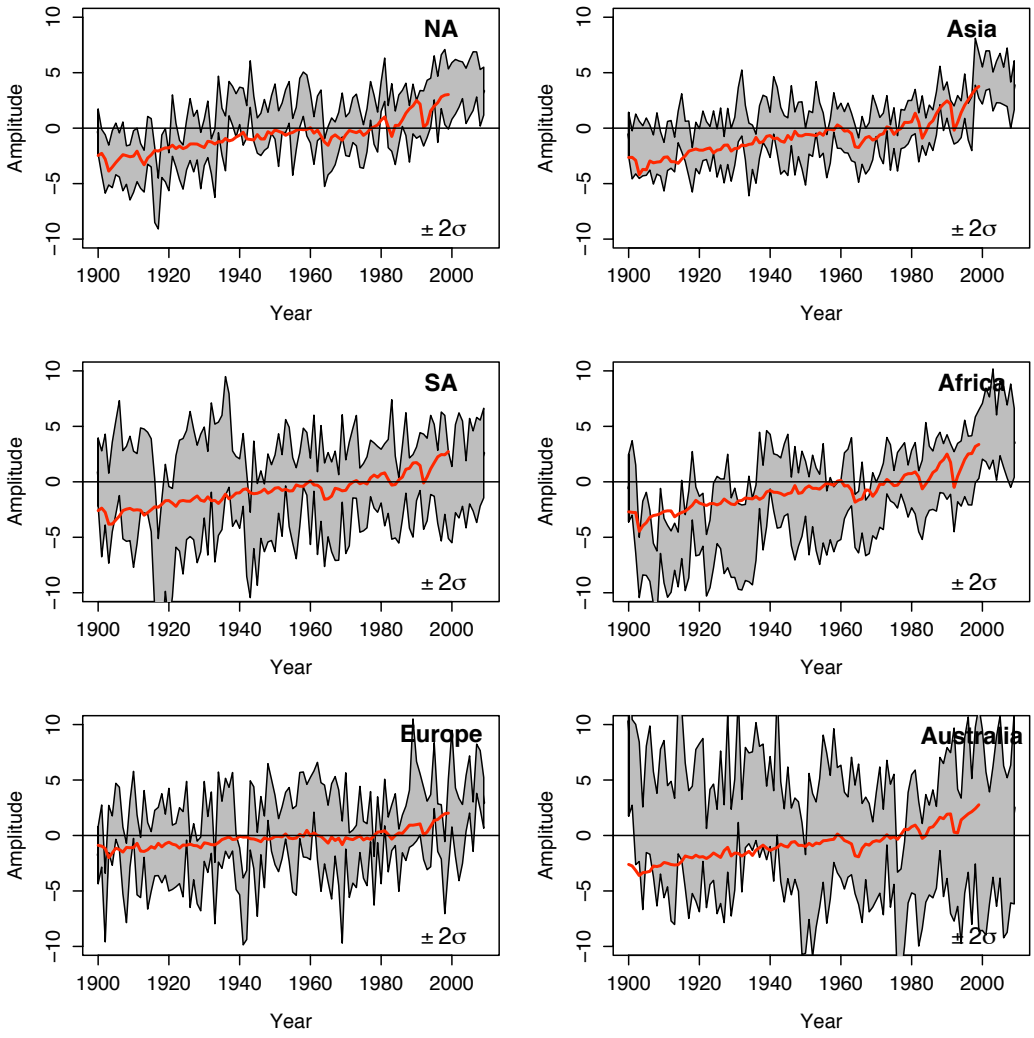

Figure 5.3: Generalized least squares estimate of the amplitude of the forced component of SAT in each continent when the forced component is determined from the forced-to-unforced discriminant. The shading indicate twice the standard error estimated from standard regression theory. The red curves in each panel indicate the ensemble mean time series from the forced-to-unforced discriminant in the forced runs.

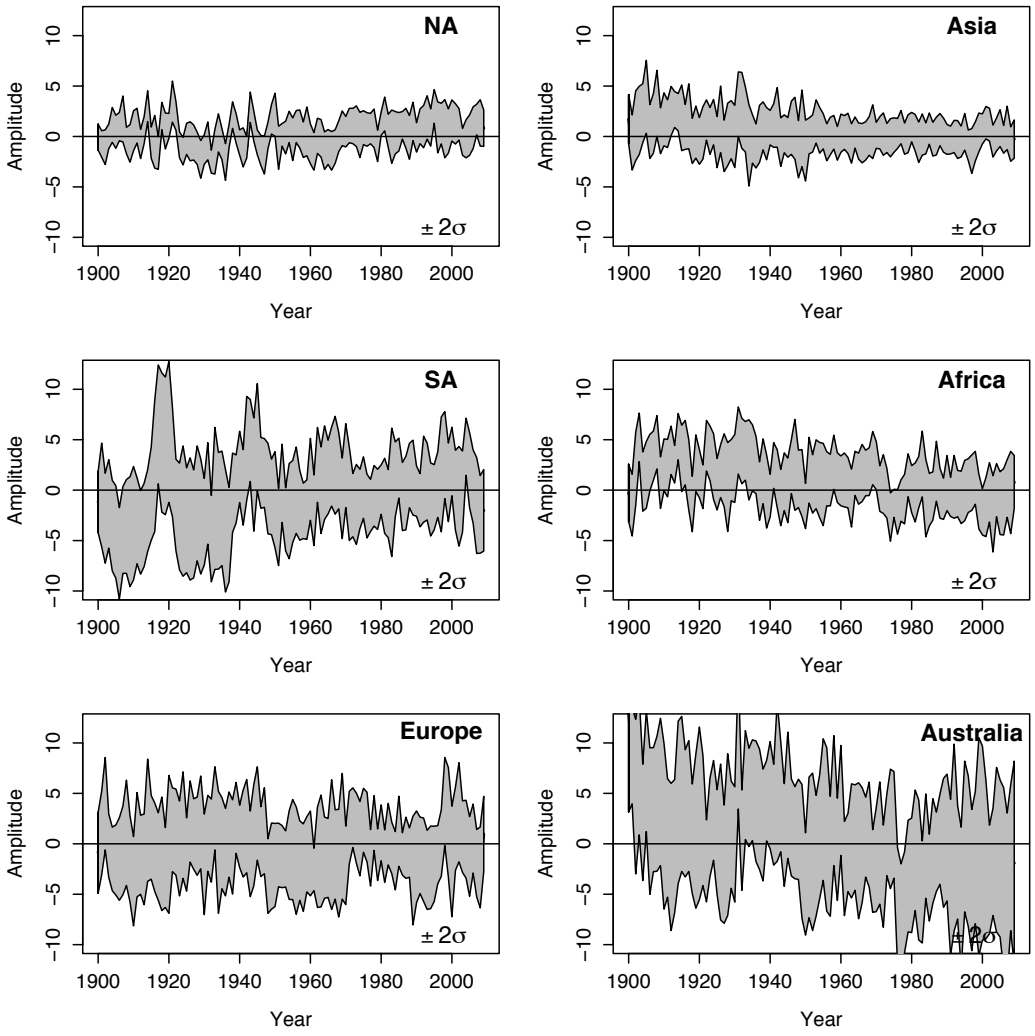

Figure 5.4: Generalized least squares estimate of the amplitude of the unforced component in each continent when the unforced component is determined from the generalized APT analysis. The shading indicate twice the standard error estimated from standard regression theory.

seems inconsistent with previous studies, which have shown that forced responses are detectable on continental scales (Stott, 2003; Zhang et al., 2006; Zwiers and Zhang, 2003). However, there is no real inconsistency. A major difference is that our optimal fingerprinting method only takes into account the spatial information of the patterns, no any temporal information is considered. Our results indicate the forced and unforced patterns are not clearly separable in observations in space alone.

Chapter 6: Implications and Limitations of This Study

In this study, we identified the multi-year predictability over land on continental scales, and attempted to separate them in the real world. Our results have a couple of implications.

The unforced predictable components of continental SAT and precipitation provide a scientific rationale for regional prediction on multi-year time scales. The actual linear regression model in the APT analysis makes empirical predictions possible. The unforced components over land are predictable only for a couple of years, indicating that the expensive decadal prediction experiments, say 30-year predictions, may be unnecessary for land prediction. Considering that the unforced predictability of land precipitation is significantly correlated with ENSO, we expect land precipitation to be well predicted if models are able to simulate ENSO accurately. However, such models also would need to capture the persistent component of SST that affects the SAT in each continent.

Our study identifies only one predictable forced pattern in continental SAT, presumably because the pattern of the response to different forcings resemble each other. This implies that it is impossible for us to attribute the forced response of land SAT to different sources of forcings on continental scales using annual mean data. The fact that no significant forced pattern in continental precipitation was found does not mean there is no response to external forcings, but rather indicates that the forced response are not consistent among individual models (recall that we pool all model output in our analysis). It would be interesting to investigate the precipitation response in individual models to see if strong responses can be identified. The observed precipitation trend is quite strong, especially in the tropics. So it would be interesting and perhaps useful for modelers to identify the model that best matches observations. The unforced and forced spatial patterns of land SAT are not significantly distinguishable in observations. Perhaps more discrimination may be obtained from either more detailed knowledge of initial conditions or predictable external forcings.

These results could offer some instructions for local governments to make policies. The unforced components are predictable on multi-year time scales, indicating that it is reasonable for local planners to plan at most 6 years ahead for temperature and 3 years ahead for precipitation, while they are not suggested to make a long-term plan more than a decade in advance. To make decadal plans in temperature, the response to external forcings need to be considered.

There are many limitations in this study. First, all results were obtained from a multi-model data set consisting of eight models, which were selected according the variance and trend. The predictability found in these eight models might not exist in another set of models. Second, the predictability was identified in the space spanned by the leading 30 PCs in each continent. Maybe the leading 30 PCs of SST prevent land predictability from being identified if the predictability does not arise from the leading 30 PCs of SST. Third, the unforced predictability identified from APT analysis was based on a 'linear' regression model. Conceivably, the predictability arising from nonlinear processes in the system, if there is any, is missing. Fourth, the fingerprinting technique used in this study only includes the spatial information of the forced and unforced patterns regardless of temporal information. This could raise difficulties for separating the forced response with natural variability in observations. Last, our results are subject to model errors and missing values in observations.

Appendix A: Detection of Trend

A method developed by DelSole and Yang (2011) is applied to diagnose the trend of the data. Consider a centered data set specified by a $N \times M$ matrix $\mathbf{Y}$, where the first dimension N specifies time and the second dimension M specifies space. Consider a linear regression model

$$\underset{[N \times M]}{\mathbf{Y}} = \underset{[N \times 1]}{\mathbf{t}} \underset{[1 \times M]}{\mathbf{z}^{\mathrm{T}}} + \underset{[N \times M]}{\mathbf{E}}, \tag{A.1}$$

where $\mathbf{t}$ is a pre-specified linear trend with zero mean, $\mathbf{z}$ is the vector of regression coefficient, $\mathbf{E}$ is the residual noise. The least squares estimate of $\mathbf{z}$ is

$$\hat{\mathbf{z}} = \mathbf{Y}^{\mathrm{T}}\mathbf{t}\left(\mathbf{t}^{\mathrm{T}}\mathbf{t}\right)^{-1}. \tag{A.2}$$

The sample covariance matrix of $\mathbf{Y}$ can be written as

$$\hat{\mathbf{\Sigma}}_T = \hat{\mathbf{\Sigma}}_S + \hat{\mathbf{\Sigma}}_N, \tag{A.3}$$

where

$$\hat{\mathbf{\Sigma}}_T = \frac{1}{N}\mathbf{Y}^{\mathrm{T}}\mathbf{Y}, \tag{A.4}$$

$$\hat{\mathbf{\Sigma}}_S = \left(\frac{\mathbf{t}^{\mathrm{T}}\mathbf{t}}{N}\right)\hat{\mathbf{z}}\hat{\mathbf{z}}^{\mathrm{T}}, \tag{A.5}$$

$$\hat{\mathbf{\Sigma}}_N = \frac{1}{N}\hat{\mathbf{E}}^{\mathrm{T}}\hat{\mathbf{E}} = \frac{1}{N}\mathbf{Y}^{\mathrm{T}}\left(\mathbf{I} - \frac{\mathbf{t}\mathbf{t}^{\mathrm{T}}}{\mathbf{t}^{\mathrm{T}}\mathbf{t}}\right)\mathbf{Y}. \tag{A.6}$$

Equation (A.3) shows that the total variability of $\mathbf{Y}$, measured by $\hat{\mathbf{\Sigma}}_T$, can be split into two parts: variability of the linear trend $\mathbf{t}$, measured by $\hat{\mathbf{\Sigma}}_S$, and the variability of residual noise $\hat{\mathbf{E}}$, measured by $\hat{\mathbf{\Sigma}}_N$. The purpose of detecting trend is to test the statistical significance of the linear trend. If data $\mathbf{Y}$ does not have a trend, the regression coefficients vanish, which means $\mathbf{z} = 0$. The null hypothesis is $\mathbf{z} = 0$, or equivalently,

$$z_1 = z_2 = z_3 = \cdots = z_M = 0. \tag{A.7}$$

The standard method for testing the hypothesis $\mathbf{z} = 0$ in multivariate regression is the likelihood ratio test under the assumption that the columns of residual noise $\mathbf{E}$ are independent and identically distributed as a normal distribution with zero mean. The significance test of $\mathbf{z} = 0$ leads to the statistic

$$\lambda = \frac{|\hat{\mathbf{\Sigma}}_N|}{|\hat{\mathbf{\Sigma}}_T|}, \tag{A.8}$$

where $\hat{\mathbf{\Sigma}}_T$ and $\hat{\mathbf{\Sigma}}_N$ are defined in (A.4) and (A.6). If the null hypothesis $\mathbf{z}$=0 is true, then the statistic

$$F = \frac{\lambda}{1-\lambda}\frac{N-M-1}{M} \tag{A.9}$$

has a F distribution with M and $N - M - 1$ degree of freedom. Large values of F favor rejection of the null hypothesis. If F is larger than a critical value at a given significance level, the null hypothesis $\mathbf{z} = 0$ is rejected, which means the linear trend is statistically significant.

The statistic λ in (A.8) can be derived without calculating the determinants as follows.

$$|\hat{\mathbf{\Sigma}}_N| = |\hat{\mathbf{\Sigma}}_T - \hat{\mathbf{\Sigma}}_S| = |\hat{\mathbf{\Sigma}}_T - \left(\frac{\mathbf{t}^{\mathrm{T}}\mathbf{t}}{N}\right)\hat{\mathbf{z}}\hat{\mathbf{z}}^{\mathrm{T}}| = |\hat{\mathbf{\Sigma}}_T|\left(1 - \left(\frac{\mathbf{t}^{\mathrm{T}}\mathbf{t}}{N}\right)\hat{\mathbf{z}}^{\mathrm{T}}\hat{\mathbf{\Sigma}}_T^{-1}\hat{\mathbf{z}}\right) = |\hat{\mathbf{\Sigma}}_T|(1-\rho^2), \tag{A.10}$$

where

$$\rho^2 = \frac{\mathbf{t}^{\mathrm{T}}\mathbf{Y}\left(\mathbf{Y}^{\mathrm{T}}\mathbf{Y}\right)^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{t}}{\mathbf{t}^{\mathrm{T}}\mathbf{t}}. \tag{A.11}$$

It follows from (A.8) and (A.10) that $\lambda = 1 - \rho^2$. Substitute this into (A.9) gives

$$F = \frac{\rho^2}{1-\rho^2}\frac{N-M-1}{M}. \tag{A.12}$$

Appendix B: Proof of Maximizing a Rayleigh Quotient is An Eigenvalue Problem

In mathematics, given a real symmetric matrix $\mathbf{A}$ with a dimension of $M \times M$ and a M-dimensional nonzero vector $\mathbf{q}$, the Rayleigh quotient R is defined as

$$R = \frac{\mathbf{q}^{\mathrm{T}}\mathbf{A}\mathbf{q}}{\mathbf{q}^{\mathrm{T}}\mathbf{q}}. \tag{B.1}$$

We prove in this appendix that the vectors maximize a Rayleigh quotient are given by the eigenvectors of the eigenvalue problem

$$\mathbf{A}\mathbf{q} = \lambda\mathbf{q}. \tag{B.2}$$

To prove this, we note that matrix $\mathbf{A}$ is symmetric and it can be decomposed as

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}}, \tag{B.3}$$

where $\mathbf{U}$ is a unitary matrix ($\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I}$) with each column representing a eigenvector of $\mathbf{A}$. $\mathbf{S}$ is a diagonal matrix with real diagonal elements

$$\mathbf{S} = \begin{pmatrix} s_1 & 0 & 0 & \dots & 0 \\ 0 & s_2 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & \dots & s_M \end{pmatrix}.$$

The diagonal elements s_1, s_2, ..., s_M are the eigenvalues of $\mathbf{A}$ and can be ordered in decreasing order such that $s_1 \geq s_2 \geq \ldots \geq s_M$. The corresponding eigenvectors in $\mathbf{U}$ are ordered accordingly. Substituting (B.3) into (B.1) gives,

$$R = \frac{\mathbf{q}^{\mathrm{T}}\mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}}\mathbf{q}}{\mathbf{q}^{\mathrm{T}}\mathbf{q}}. \tag{B.4}$$

Let $\mathbf{p} = \mathbf{U}^{\mathrm{T}}\mathbf{q}$, where $\mathbf{p}$ is a M-dimensional vector. Equation (B.4) can be written as

$$R = \frac{\mathbf{p}^{\mathrm{T}}\mathbf{S}\mathbf{p}}{\mathbf{p}^{\mathrm{T}}\mathbf{p}} = \frac{p_1^2 s_1 + p_2^2 s_2 + \ldots + p_M^2 s_M}{p_1^2 + p_2^2 + \ldots + p_M^2}, \tag{B.5}$$

where $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}$ has been used and p_1, p_2, ..., p_M denote the M elements of $\mathbf{p}$. Because of $s_1 \geq s_2 \geq \ldots \geq s_M$, the numerator of (B.5) cannot be decreased by replacing all s_k $(k = 1, 2, \ldots, M)$ with s_1, since $s_1 \geq s_k$ for all k. Therefore, if we substitute s_2, s_3 ... s_M with s_1 in (B.5), we obtain the bound

$$R < \frac{p_1^2 s_1 + p_2^2 s_1 + \ldots + p_M^2 s_1}{p_1^2 + p_2^2 + \ldots + p_M^2} = s_1. \tag{B.6}$$

But, this upper bound can be achieved by choosing $p_1 = 1$, $p_2 = p_3 = \ldots = p_M = 0$. Invoking $\mathbf{p} = \mathbf{U}^{\mathrm{T}}\mathbf{q}$, $\mathbf{q}_1 = \mathbf{U}\mathbf{p}_1$, which implies that $\mathbf{q}_1$ is the first column of $\mathbf{U}$. In other words, $\mathbf{q}_1$ is the leading eigenvector of $\mathbf{A}$ with maximum eigenvalue. Similarly, the second vector $\mathbf{q}_2$ that maximizes R subject to being uncorrelated with $\mathbf{q}_1$ can be obtained by choosing $p_1 = 0$, $p_2 = 1$, $p_3 = 0 = \ldots = p_M = 0$, such that $\mathbf{q}_2 = \mathbf{U}\mathbf{p}_2$. It implies that $\mathbf{q}_2$ is the second leading eigenvector of $\mathbf{A}$, hence the second largest eigenvalue s_2 is the maximum ratio R. Following the same procedure, the third vector $\mathbf{q}_3$ that maximizes R with respect to being uncorrelated with $\mathbf{q}_1$ and $\mathbf{q}_2$ can be derived, and so on.

The above proof can be extended to show that maximizing the generalized Rayleigh

quotient $R = \frac{\mathbf{q}^{\mathrm{T}}\mathbf{A}\mathbf{q}}{\mathbf{q}^{\mathrm{T}}\mathbf{B}\mathbf{q}}$, where $\mathbf{A}$ and $\mathbf{B}$ are real, symmetric and positive-definite matrices, leads to the generalized eigenvalue problem $\mathbf{A}\mathbf{q} = \lambda\mathbf{B}\mathbf{q}$. To prove this, recall that $\mathbf{B}$ is symmetric and hence can be decomposed as $\mathbf{B} = \mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}}$, where $\mathbf{U}$ is unitary, and $\mathbf{S}$ is a diagonal matrix. Let $\mathbf{p} = \mathbf{S}^{\frac{1}{2}}\mathbf{U}^{\mathrm{T}}\mathbf{q}$, the generalized Rayleigh quotient can be written as

$$R = \frac{\mathbf{p}^{\mathrm{T}}\mathbf{A}^{'}\mathbf{p}}{\mathbf{p}^{\mathrm{T}}\mathbf{p}}, \tag{B.7}$$

where $\mathbf{A}^{'} = \mathbf{S}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{U}\mathbf{S}^{-\frac{1}{2}}$. (B.7) becomes the classical Rayleigh quotient as in (B.1). Therefore, based on the proof above, vectors $\mathbf{p}$ that maximize Rayleigh quotient in (B.7) are the eigenvectors of

$$\mathbf{A}^{'}\mathbf{p} = \lambda\mathbf{p}. \tag{B.8}$$

Inverting the transformation gives the genralized eigenvalue problem $\mathbf{A}\mathbf{q} = \lambda\mathbf{B}\mathbf{q}$.

Appendix C: Proof: if $\mathbf{Aq} = \lambda\mathbf{Bq}$ and matrices A and B are symmetric, then $\mathbf{q}_i^{\mathrm{T}}\mathbf{Bq}_j = 0$ and $\mathbf{q}_i^{\mathrm{T}}\mathbf{Aq}_j = 0 (i \neq j)$.

Let $\mathbf{q}_i$ and $\mathbf{q}_j$ be eigenvectors of the generalize eigenvalue problem $\mathbf{Aq} = \lambda\mathbf{Bq}$ with two distinct eigenvalues λ_i and λ_j, respectively. The eigenvalue problem associated with these two eigenvectors can be written as

$$\mathbf{Aq}_i = \lambda_i \mathbf{Bq}_i \tag{C.1}$$

and

$$\mathbf{Aq}_j = \lambda_i \mathbf{Bq}_j. \tag{C.2}$$

Multiplying both sides of (C.1) with $\mathbf{q}_j$, and multiplying both sides of (C.2) with $\mathbf{q}_i$ gives

$$\mathbf{q}_j^{\mathrm{T}}\mathbf{Aq}_i = \lambda_i \mathbf{q}_j^{\mathrm{T}}\mathbf{Bq}_i \tag{C.3}$$

and

$$\mathbf{q}_i^{\mathrm{T}}\mathbf{Aq}_j = \lambda_j \mathbf{q}_i^{\mathrm{T}}\mathbf{Bq}_j. \tag{C.4}$$

Since $\mathbf{q}_i^{\mathrm{T}}\mathbf{Aq}_j$ and $\mathbf{q}_i^{\mathrm{T}}\mathbf{Bq}_j$ are scalars, the transpose of a scalar is equal to itself, implying that

$$\mathbf{q}_i^{\mathrm{T}}\mathbf{Aq}_j = \left(\mathbf{q}_i^{\mathrm{T}}\mathbf{Aq}_j\right)^{\mathrm{T}} = \mathbf{q}_j^{\mathrm{T}}\mathbf{Aq}_i, \tag{C.5}$$

$$\mathbf{q}_i^{\mathrm{T}}\mathbf{Bq}_j = \left(\mathbf{q}_i^{\mathrm{T}}\mathbf{Bq}_j\right)^{\mathrm{T}} = \mathbf{q}_j^{\mathrm{T}}\mathbf{Bq}_i. \tag{C.6}$$

Subtracting (C.3) from (C.4) gives

$$0 = (\lambda_j - \lambda_i)\mathbf{q}_i^{\mathrm{T}}\mathbf{Bq}_j. \tag{C.7}$$

As λ_i and λ_j are two distinct eigenvalues, (C.7) implies that $\mathbf{q}_i^{\mathrm{T}}\mathbf{B}\mathbf{q}_j = 0$, and hence $\mathbf{q}_i^{\mathrm{T}}\mathbf{A}\mathbf{q}_j = 0$ based on (C.4).

Bibliography

Bibliography

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

Biasutti, M., I. M. Held, A. H. Sobel, and A. Giannini, 2008: sst forcings and sahel rainfall variability in simulations of the twentieth and twenty-first centuries. *J. Climate*, **21**, 3471–3486.

Boer, G. J., 2000: A study of atmosphere-ocean predictability on long time scales. *Clim. Dyn.*, **16**, 469–477.

Boer, G. J., 2004: Long time-scale potential predictability in an ensemble of coupled climate models. *Clim. Dyn.*, **23**, 29–44.

Boer, G. J. and S. J. Lambert, 2008: Multi-model decadal potential predictability of precipitation and temperature. *Geophys. Res. Lett.*, **35**, L05 706, doi:10.1029/2008GL033 234.

Cane, M. A., 2010: Climate science: decadal predictions in demand. *Nature Geoscience*, **3**, 231–232.

Collins, M., 2002: Climate predictability on interannual to decadal time scales: the initial value problem. *Clim. Dyn.*, **19**, 671–692.

Collins, M. and B. Sinha, 2003: Predictability of decadal variations in the thermohaline circulation and climate. *Geophys. Res. Lett.*, **30**, 1306, doi:10.1029/2002GL016 504.

Collins, M., et al., 2006: Interannual to decadal climate predictability in the North Atlantic: A multimodel-ensemble study. *J. Climate*, **19**, 1195–1203.

Compo, G. P. and P. D. Sardeshmukh, 2009: Oceanic influences on recent continental warming. *Clim. Dyn.*, **32**, 333–342.

DelSole, T., 2004: Predictability and information theory part I: Measures of predictability. *J. Atmos. Sci.*, **61**, 2425–2440.

DelSole, T. and P. Chang, 2003: Predictable component analysis, canonical correlation analysis, and autoregressive models. *J. Atmos. Sci.*, **60**, 409–416.

DelSole, T. and M. K. Tippett, 2009: Average predictability time. Part II: Seamless diagnoses of predictability on multiple time scales. *J. Atmos. Sci.*, **66**, 1188–1204.

DelSole, T., M. K. Tippett, and J. Shukla, 2011: A significant component of unforced multidecadal variability in the recent acceleration of global warming. *J. Climate*, **24**, 909–926.

DelSole, T. and X. Yang, 2011: Field significance of regression patterns. *J. Climate*, in press.

Déqué, M., 1988: 10-day predictability if the northern hemisphere winter 500-mb height by the ECMWF operational model. *Tellus*, **40A**, 26–36.

Griffies, S. and K. Bryan, 1997: A predictability study of simulated North Atlantic multidecadal variability. *Clim. Dyn.*, **13**, 459–487.

Hasselmann, K., 1997: Multi-pattern fingerprint method for detection and attribution of climate change. *Clim. Dyn.*, **13**, 601 – 611.

Hasselmann, K. F., 1979: On the signal-to-noise problem in atmospheric response studies. *Meteorology of the tropical ocean*, D. B. Shaw, Ed., Royal Meteorological Society, 251–259.

Hegerl, G. C., et al., 2007: Understanding and attributing climate change. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon, D. Qin,

M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller, Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Held, I. M. and B. J. Soden, 2006: Robust responses of the hydrological cycle to global warming. *J. Climate*, **19**, 5686–5699.

Held, I. M., T.L.Delworth, J.Lu, K.L.Findell, and T.R.Knutson, 2005: Simulation of Sahel drought in the 20th and 21st centruries. *PNAS*, **102 (50)**, 17 891–17 896.

Hoerling, M., J. Eischeid, and J. Perlwitz, 2010: Regional precipitation trends: Distinguishing natural variability from anthropogenic forcing. *J. Climate*, **23**, 2131–2145.

Knight, J. R., C. K. Folland, and A. A. Scaife, 2006: Climate impacts of the Atlantic multidecadal oscillation. *Geophys. Res. Lett.*, **33**, L17 706, doi:10.1029/2006GL026 242.

Lambert, F. H., N. P. Gillett, D. A. Stone, and C. Huntingford, 2005: Attribution studies of observed land precipitation changes with nine coupled models. *Geophys. Res. Lett.*, **32**, L18 704, doi:10.1029/2005GL023 654.

Lambert, F. H., P. A. Stott, M. R. Allen, and M. A. Palmer, 2004: Detection and attribution of changes in 20th century land precipitation. *Geophys. Res. Lett.*, **31**, L10 203, doi:10.1029/2004GL019 545.

Latif, M. and T. Barnett, 1994: Causes of decadal climate variability over the North Pacific and North America. *Science*, **266**, 634–637.

Latif, M. and T. Barnett, 1996: Decadal climate variability over the North Pacific and North America: dynamics and predictability. *J. Climate*, **9**, 2407–2423.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmosphere model. *Tellus*, **17**, 321–333.

Meehl, G. A., et al., 2009: Decadal prediction can it be skillful? *BAMS*, **90**, 1467–1485.

Min, S.-K., X. Zhang, and F. Zwiers, 2008: Human-induced Arctic moistening. *Science*, **320**, 518–520.

Msadek, R., K. W. Dixon, T. L. Delworth, and W. Hurlin, 2010: Assessing the predictability of the atlantic meridional overturning circulation and associated fingerprints. *Geophys. Res. Lett.*, **37**, L19 608, doi:10.1029/2010GL044 517.

Newman, M., 2007: Interannual to decadal predictability of tropical and North Pacific sea surface temperatures. *J. Climate*, **20**, 2333–2356.

Pohlmann, H., M. Botzet, M. Latif, A. Roesch, martin Wild, and P. Tschuck, 2004: Estimating the decadal predictability of a coupled AOGCM. *J. Climate*, **17**, 4463–4472.

Pohlmann, H., F. Sienz, and M. Latif, 2006: Influence of the multidecadal Atlantic meridional overturning circulation variability on European climate. *J. Climate*, **19**, 6062–6067.

Rayner, N., P. Brohan, D. Parker, C. Folland, J. J. Kennedy, M. Vanicek, T. Ansell, and S. F. B. Tett, 2006: Improved analyses of changes and uncertainties in marine temperature measured in situ since the mid-nineteenth century: the HadSST2 dataset. *J. Climate*, **19**, 446–469.

Renwick, J. A. and J. M. Wallace, 1995: Predictable anomaly patterns and the forecast skill of northern hemisphere wintertime 500-mb height fields. *Mon. Wea. Rev.*, **123**, 2114–2131.

Schneider, T. and S. Griffies, 1999: A conceptual framework for predictability studies. *J. Climate*, **12**, 3133–3155.

Schubert, S. D., M. J. Suarez, and P. J. P. et al., 2004: On the cause of the 1930s dust bowl. *Science*, **303**, 1855–1859.

Shukla, J., 1981: Dynamical predictability of monthly means. *Mon. Wea. Rev.*, **38**, 2547–2572.

Stott, P., N. P. Gillett, G. C. Hegerl, D. J. Karoly, S. A. Stone, X. Zhang, and F. Zwiers, 2010: Detection and attribution of climate change: a regional perspective. *WIREs: Climate Change*, **1**, 192–211.

Stott, P. A., 2003: Attribution of regional-scale temperature changes to anthropogenic and natural causes. *Geophys. Res. Lett.*, **30**, 1728, doi:10.1029/2003GL017 324.

Stouffer, R., et al., 2006: Investigating the causes of the response of the thermohaline circulation to past and future climate changes. *J. Climate*, **19**, 1365–1387.

Stouffer, R. J., D. Seidov, and B. J. Haupt, 2007: Climate response to external sources of freshwater: North Atlantic versus the Southern Ocean. *J. Climate*, **20**, 436–448.

Timmermann, A. and M. Latif, 1998: Northern hemispheric interdecadal variability: A coupled air–sea mode. *J. Climate*, **11**, 1906–1931.

Trenberth, K., et al., 2007: Observations: Surface and atmospheric climate change. *Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*, S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller, Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Trenberth, K. E. and J. M. Caron, 2001: Estimates of meridional atmosphere and ocean heat transports. *J. Climate*, **14**, 3433–3443.

Trenberth, K. E. and D. J. Shea, 2006: Atlantic hurricanes and natural variability in 2005. *Geophys. Res. Lett.*, **33**, L12 704, doi:10.1029/2006GL026 894.

Wunsch, C., 2002: What is the thermohaline circulation? *Science*, **298**, 1179–1181.

Zhang, R. and T. L. Delworth, 2006: Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes. *Geophys. Res. Lett.*, **33**, L17 712, doi:10.1029/2006GL026 267.

Zhang, X., F. W. Zwiers, G. C. Hegerl, F. H. Lambert, N. P. Gillett, S. Solomon, P. A. Stott, and T. Nozawa, 2007: Detection of human influence on twentieth-century precipitation trends. *Nature*, **448**, 461–465.

Zhang, X., F. W. Zwiers, and P. A. Stott, 2006: Multi-model multi-signal climate change detection at regional scale. *J. Climate*, **19**, 4294–4307.

Zwiers, F. W. and X. Zhang, 2003: Toward regional-scale climate change detection. *J. Climate*, **16**, 793–797.

Curriculum Vitae

Liwei Jia was born in Northeast of China. She graduated from Hegang High School, Hegang, China, in 1998. She received her Bachelor of Science from Lanzhou University, China in 2002. She worked as a weather forecaster in Harbin Meteorological Bureau from 2002 to 2003. She received her Master of Science from Chinese Academy of Meteorological Sciences in 2006.