SPATIAL CORRELATIONS IN GENERAL CIRCULATION MODELS AND OBSERVATION REANALYSIS

A Thesis

by

TAYLOR LEE SANSOM

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Gerald R. North
Committee Members,	Ramalingam Saravanan
	Ping Chang
Head of Department,	Ping Yang

May 2014

Major Subject: Atmospheric Science

ABSTRACT

The purpose of this study was to research the geographical structure and time evolution of spatial correlations of general circulation models, energy balance models, observations and reanalysis, as well as provide comparisons amongst them. This study analyzed five GCM pre-industrial control runs of at least 500 years, 132 years of observational data and 65 years of reanalysis data. After the seasons were removed the data were averaged in several different time scales and the correlation structures were calculated at several locations on the planet. Comparisons of the results revealed both similarities and differences between the models, stochastic theory and reanalysis data, with the largest differences occurring over ocean at long time averages. Several models show drastic differences between correlation structures and their evolution.

DEDICATION

This thesis is dedicated to the memory of my sister Autumn, whose endless optimism, courage and compassion still inspire me and countless others years after her death. To the memory of my father Mickey, who taught me to work hard, play hard and sing a good song, even if no one is listening. To my mother Fredell, whose strength is unparalleled in this world. She has been my rock, my confidant, my shoulder to cry on and my best friend through the good and bad. To my lifelong friends Tray and Nathan who are more like brothers than friends.

Finally, I would like to thank Dr. Jerry North for giving me this opportunity. It has been an absolute pleasure, both professionally and personally, to work alongside him.

ACKNOWLEDGEMENTS

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 3.1 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. NCEP Reanalysis data provided by the NOAA/OAR/ESRL PSD, Boulder, CO, USA, from their Web site at http://www.esrl.noaa.gov/psd/. Observational data provided by NOAA/NCDC, Asheville, NC, USA from their Web site at http://www.ncdc.noaa.gov/ghcnm/. This research was made possible by a grant from Los Alamos National Laboratory.

NOMENCLATURE

NCAR	National Center for Atmospheric Research
GFDL	Geophysical Fluid Dynamics Laboratory
MPI	Max Planck Institute
IPSL	Institut Pierre Simon LaPlace
CSIRO	Commonwealth Scientific and Industrial Research Organization
NCEP	National Centers for Environmental Predictions
NCDC	National Climatic Data Center
GHCN	Global Historical Climatology Network
ICOADS	International Comprehensive Ocean-Atmosphere Data Set
CCSM4	Community Climate System Model Version 4
CM3	Climate Model Version 3
Mk3	Mark 3.0
CM5A	Climate Model Version 5A
ESM	Earth System Model
OBS	Observations
REA	Reanalysis
picontrol	Pre-Industrial Control Run
GCM	General Circulation Model
NWP	Numerical Weather Prediction

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE Page

1. INTRODUCTION AND LITERATURE REVIEW

This study considers the second moment statistics, specifically spatial correlations in the surface temperature field, taken from simulations of several general circulation models (GCMs), an energy balance model (EBM), reanalysis based on observations and direct observational data. We are interested in the geographical structure of these second moments as well as how these structures are affected by different lengths of time averaging of the raw fields. Also, we are interested in how the second-moment statistics of GCMs compares to observations and reanalysis, stochastic EBMs and how the models compare with each other. Spatial correlation is the measure of how much one site "feels" another. There are large differences in the correlation structures depending on land or sea surface of the site-pairs. Thus, the type and size of the surrounding area for which a given station's data represents may provide significant information [5].

As shown by both theoretical considerations and by examples using data, when solving the various problems in the averaging of meteorological fields in space and time it is necessary to take into account the correlation structure of the fields. Failure to do so may lead to incorrect estimates of sampling characteristics obtained from averaging and to incorrect conclusions about their accuracy [10]. [15] argued that obtaining good agreement among second-order statistics, such as spatial correlations, is a necessary condition for trusting a model for various purposes. [11] gave several reasons for using second-moment statistics as a quantitative measure of the behaviour of the climate system. First, they provided an important test of our climate models beyond the conventional tests such as geographical distributions of selected monthly averaged snapshots or seasonal or annual means. Also, a reason for studying second-

moments in various frequency bands is that their characteristics might provide a clue as to the ability of models to predict the correct sensitivity of climate to small externally imposed forcings.

The statistical structure of meteorological variables has been studied over the past several decades in an effort to better understand and simulate the fluctuations in the dynamic atmosphere-ocean climate system. Variability and correlation (both spatial and temporal) in the climate are of paramount importance in the determination of long-term trends, large-scale averages and in general, predictability of the system. Whether the variability is a "natural" part of the system (North Atlantic Oscillation, Atlantic Multidecadal Oscillation, El Nino Southern Oscillation, etc.) or externally forced (such as anthropogenically induced climate change) needs to be differentiated [11].

The temporal structure of the climate system has long been better studied than the spatial structure for several reasons. First, temporal averaging is much simpler than spatial averaging because time intervals are usually equally spaced which cannot be said for the spatial coverage of the current observation network. [16] summarized this problem by saying that a function in time that describes a process in nature is known to depend only on the past but not on the future (causality). However, this notion of causality has no analog in space. A function describing a process in one dimensional space can depend on values from the left and right of a reference point. In higher dimensions even more freedom exists for the locations having influence on the value at a reference point. The greatest difficulties arise in higher dimensional averaging. These difficulties are to a considerable extent connected with the indeterminate nature of the areas to which the measured data at each point should be related. From this follows uncertainty in the method of selecting weights for the measured data, even if the territory is totally homogeneous [10]. Another

problem that arises in spatial averaging is that the measurements involve a variety of observation points that generally record a variety of conditions using a variety of instruments, and the data may relate to a climatically heterogeneous area. This is a stark difference from temporal averaging, for which the measured data used are from one and the same instrument registering the same conditions from the same location.

Near surface air temperature is the variable under investigation in this study. It is one of the most accurately measurable elements in meteorology and is an essential output variable of all climate models. Its spatial and temporal correlation are high so it has been the subject of detailed study and requires little averaging. However, a number of problems exist which do call for averaging. A particularly important role is in climatic theory, where it is required to discover the real significance of temperature changes of as small as a fraction of a degree. These problems call for spatial averaging of temperature over large areas extending over many millions of square kilometers, and averaging accuracy essentially depends on the selection of averaging method [10]. Although temperature is the most widely investigated variable, the methods described in this paper can and have been applied to pressure, height, wind, precipitation and any other consistently sampled variable. Multivariate correlation models could also be investigated to determine dependence of certain variables on others (e.g. temperature and precipitation).

[5] and [12] (herein referred to as HL87 and KNH96 respectively) provide the groundwork for this study. Although the motivations and methods used in each study are different, they are analogous and comparisons can be reasonably made. HL87 used annually averaged observed surface air temperature taken from available meteorological stations for the period 1880-1985 and calculated the spatial correlation coefficients for station pairs with at least 50 years of common records in certain

latitudinal bands. It should be noted that they did not discriminate between ocean sites and land sites which inherently makes their results different from those in this study where large uniform surface areas are investigated. Additionally, their study considered sites in the tropics which have not been included in this paper. HL87 described the problem of inconsistent spatial and temporal coverage of the observation network, which will be elaborated upon in Section 3. KNH96 used two coupled GCMs and an EBM along with observations, all of which were cast in Fourier-spherical harmonic space and filtered into various frequency bands. Because spherical harmonics were used, they needed a continuous series in both time and space. Missing values were replaced with a zero temperature anomaly for that time, which was meant to suppress spurious spatial variance, although it did introduce some spurious temporal variability and bias. KNH96 only used the last 100 years of observations (1890-1989) because they considered earlier records too sparse for analysis. In this study, in an attempt to mitigate the problems arising from an inconsistent observation network, observation reanalysis is included for comparison. Further comparisons between these studies will be presented in Section 5.

In the next section of this paper, the applications of spatial correlations are presented and the motivations of this study described. Optimal averaging and interpolation are discussed as well as the distinction between homogeneous, isotropic fields and heterogeneous, anisotropic fields.

Section 3 will detail the data used in this study, sources and any noteworthy characteristics. The problems involved with the global observation network will be described.

In Section 4 the methods used will be described. The filters that were used to categorize the data will be illustrated and the segmenting of the data into different boxes will be explained. There will be a brief note on stationary time series and how it

pertains to this study as well as assumptions used and their associated implications. The statistical theory as it applies to this study will be discussed in a stochastic framework.

Section 5 will contain the results and discussion. Comparisons will be made between the GCMs, EBM, observations and reanalysis.

The conclusions will be presented in Section 6 as well as possible future research topics and recommendations.

2. MOTIVATIONS

There have been great efforts in the history of meteorology to construct smooth fields of meteorological variables from the spatially and temporally uneven observation network. Climatologists have been hampered by the given locations of gauges comprising the observation network. These strongly favor the inhabited and more developed parts of the world and in particular, the oceans, the southern hemisphere and other poorly accessible areas are typically badly represented [6]. One problem lies in the fact that the gauges which are in place, are separated by finite spatial distances and that these spatial gaps lead to an inevitable "sampling error". Complicating the problem is the fact that the field has correlations from one point on the sphere to another and that these correlations depend on the length of temporal averaging employed [14]. Further complications can arise from instrument errors, topography differences and urban heat islands which cannot feasibly be extracted from the data [4]. With these difficulties in mind it is not hard to see why the topics of optimal averaging and interpolation have been so widely studied. Section 2.1 will describe the evolution of correlation functions and list some of their applications pertaining to modelling and data assimilation. Section 2.2 will introduce several questions as to what this study intends to find. These questions will be revisited in Section 6.

2.1 Correlation Functions

First consider how spatial correlations are computationally estimated. Most correlation functions of two dimensional fields, such as temperature that have been developed and studied assume that the data being modelled (temperature, precipitation, pressure, etc.) are homogeneous and isotropic at the length scales in question.

Some examples of these candidate correlation functions are described in [1], [9], [10] and [16] amongst others.

In these homogeneous, isotropic models the correlation function depends only on Euclidian distance (or great circle distance on a sphere) between two points in the domain and is directionally invariant. These univariate functions are the building blocks of correlation theory in meteorology and can be representative enough of the real atmosphere that residual errors will be minimum. The main advantage of these correlation functions is that they are computationally simple to integrate into data assimilation, however their simplicity is also a drawback that must be acknowledged.

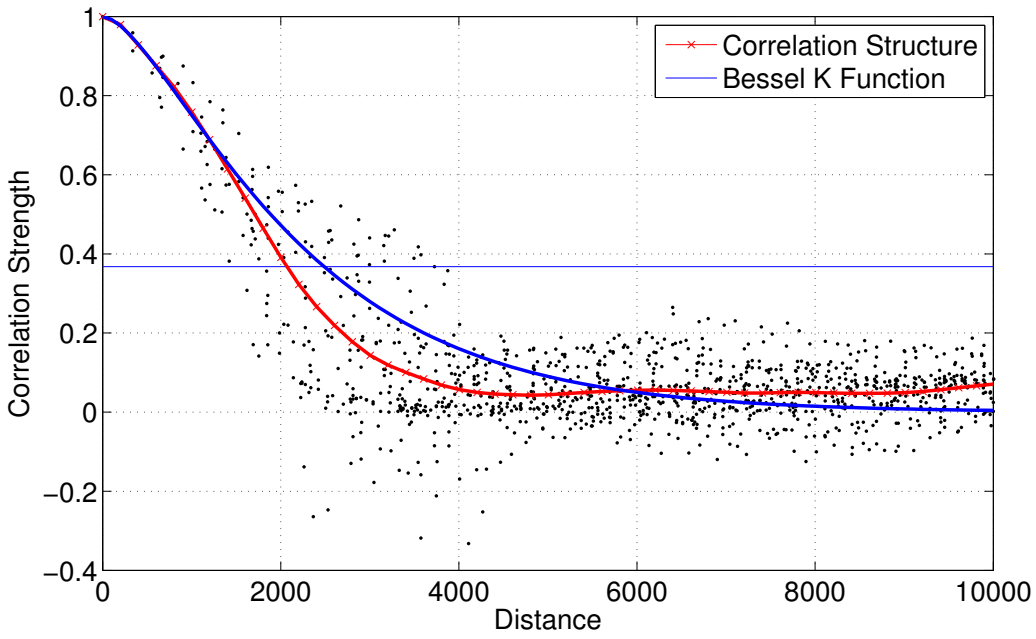


Figure 2.1: Correlation as a function of station separation for monthly averaged surface temperature observation data for Asia (red). The blue line is modified Bessel function $rK_1(r)$ using a decorrelation length of 1500 km.

In general, the assumptions of homogeneity and isotropy are clearly violated

for many, if not most, applications in the atmospheric sciences [3]. In the global temperature field, heterogeneity and anisotropy can arise for may reasons including: (1) directional dependency, for example, the prevalence of frontal phenomena (see [9]), (2) proximity to land-sea boundaries, (3) topography and (4) different surface types.

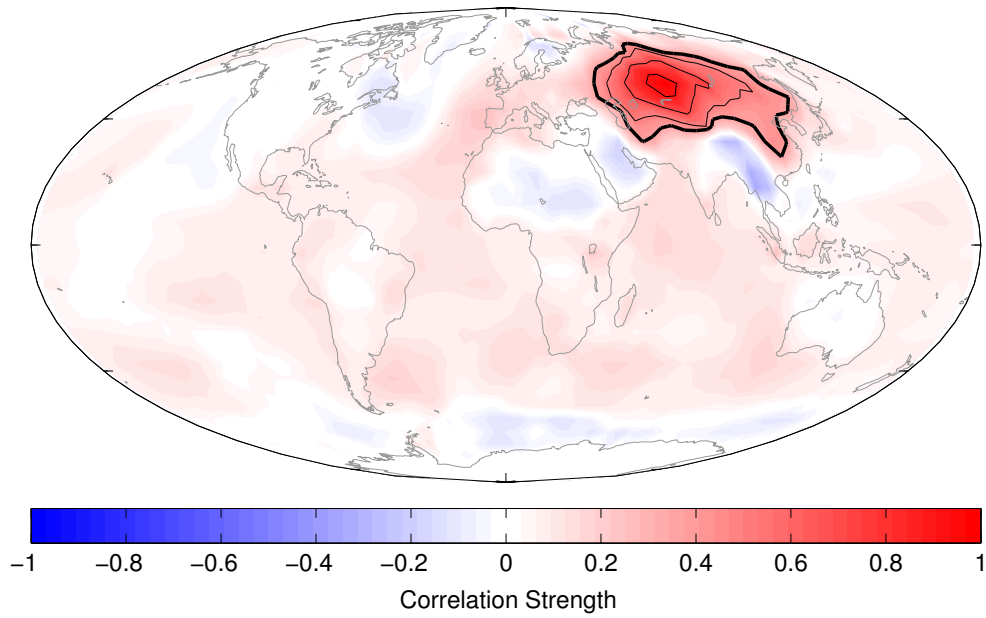


Figure 2.2: Contours of constant correlation of surface temperatues at various sites with a central point in Mid-Asia for monthly averaged surface temperature observation data for Asia. Thin contours are 0.9, 0.7, 0.5 and the thick contour is e^{-1}.

In Figures 2.1 and 2.2 the effects of (1), (2) and (3) can be clearly seen with the biggest contribution to anisotropicity coming from the presence of the Himalayas along the southern boundary of the correlation contours. The impact of (4) will be seen in the size, shape and evolution of e-folding contours for land and ocean sites, which will be presented fully in Section 5.

2.2 Why Are We Doing This?

The main motivation of this study is to serve as a comparison point among several current GCMs. We are attempting to quantify the second moment statistical structure of the surface temperature field (more specifically the e-folding spatial autocorrelation distances) at various time averages on the raw data. The next section will describe the body of data that was used for this study, including sources and descriptions, and Section 4 will outline the methods used. Sections 5 and 6 will provide evidence and attempt to answer the following questions:

1. Is there a relation between these correlation length scales and GCM equilibrium sensitivity?

2. Do the second moment statistics of recent GCMs:

   (a) behave similar to observations and Reanalysis?

   (b) behave according to stochastic EBM theory?

   (c) agree with each other?

3. DATA SOURCES AND BACKGROUND INFORMATION

For this study 5 different GCMs were chosen from the Coupled Model Intercomparison Project Phase 5 (CMIP5). Established under the World Climate Research Programme (WCRP) and the Working Group on Coupled Modelling (WGCM), CMIP is a standard experimental protocol for studying the output of coupled atmosphere ocean GCMs. It provides a community-based infrastructure in support of climate diagnosis, validation, intercomparison, documentation and data access allowing scientists to systematically analyze GCMs and aid in model improvement. In this study models from the National Center for Atmospheric Research (NCAR), Geophysical Fluid Dynamics Laboratory (GFDL), Max Planck Institute for Meteorology (MPI), Institute Pierre-Simon Laplace (IPSL) and the Commonwealth Scientific and Industrial Research Organization (CSIRO) were used. The models are described in Table 3.1.

Observations were taken from The National Climatic Data Center (NCDC). Formerly known as the National Weather Records Center (NWRC), NCDC is the world's largest active archive of weather data archiving 99 percent of all National Oceanic and Atmospheric Administration (NOAA) data. The data are received from a variety of sources including satellites, radar, automated airport weather stations, National Weather Service (NWC) observers, aircraft, ships, radiosondes, wind profilers, rocketsondes and solar radiation networks.

Reanalysis data were taken from the National Centers for Environmental Prediction's (NCEP) Climate Prediction Center (CPC) division. Formerly known as the Climate Analysis Center (CAC), the CPC consolidated NOAA's climate diagnostics, monitoring and extended range forecasts.

Table 3.1: Model Information

Source	Model Name	Resolution ($\Delta\phi, \Delta\lambda$)	Time Length (Years)
NCAR	CCSM4	$1^\circ \times 1.25^\circ$	1300
GFDL	CM3	$2^\circ \times 2.5^\circ$	500
MPI	ESM	$1.9^\circ \times 1.9^\circ$	1000
IPSL	CM5A	$1.9^\circ \times 3.75^\circ$	1000
CSIRO	MK3	$3.2^\circ \times 5.6^\circ$	1000
NCDC	OBS	$5^\circ \times 5^\circ$	-
NCEP	REA	$2.5^\circ \times 2.5^\circ$	65

3.1 General Circulation Models

GCMs are useful tools for understanding the roles of the major climate system components. Analyses of GCM simulations also play a key role in the decision making process for environmental policy at all scales (local, state, national, etc.). Water resource management, agriculture, transportation and urban planning are just a few of the areas in which examination of the behavior of GCMs can be helpful. The main components of coupled GCMs are the atmosphere, land, ocean and sea ice. The advantage of GCMs is that they allow researchers to perform experiments based on synthetic external conditions or "forcings" that cannot be directly performed on Earth. Continued experience with these models has led to advancement in our understanding of climate variables and their evolution at various time scales (monthly, seasonal, annual, decadal, etc.).

Each GCM inherently has its own unique characteristics owing to its particular historical evolution. Some of the operational differences between models include different numerical methods, grid spacing (resolution), scale parameterizations and initialization techniques especially for the oceans. The main source of discrepancy between models however, is how they simulate the various feedback mechanisms

(water vapor, clouds, radiation, ice albedo, dynamics of air-sea interaction, etc.). For this reason, different GCMs may simulate different sensitivities to the same forcing simply because of the way certain processes and feedbacks are modelled. On the other hand, all models are based on the same fundamental set of equations and variables and are designed to simulate the dynamics and processes of the atmosphere and ocean. Hence, they are not independent of one another, with many sharing common components and algorithms, which can be seen in climate model responses sharing common patterns [8].

For this study, the raw data for the GCMs under investigation were downloaded from The Program for Climate Model diagnosis and Intercomparison (PCMDI) data access portal, part of the Earth System Grid data distribution project. All models are coupled atmosphere-ocean models and only pre-industrial control runs (picontrol) were used for their respective experiment. Picontrol runs contain no anthropogenic or natural forcings (such as volcanic dust veils) and are usually run for at least 100 years after spin-up. We used the picontrol runs because we wanted to test the underlying spatial statistics to see how the models compare at simulating the second moments of the unforced climate system. Presumably such a time series meets criteria for stationarity after removal of the seasonal cycle for monthly averages. Models were chosen with picontrol runs of at least 500 years.

3.2 Energy Balance Model

At the other end of the complexity spectrum of climate models we have EBMs, which are the simplest models of the global climate system. The zero dimensional models represent the entire climate system with just one number, the global average surface temperature, that researchers use as a basic indicator of the Earth's climate. The total rate of radiation absorbed must equal the rate of emission to space. To

simulate the climate in more dimensions on the spherical surface horizontal heat transport must be added in the form of diffusion. The basic physical driver for the two dimensional EBMs is the balance between incoming solar radiation absorbed by a column of air over a particular small grid box at the surface being balanced by the rate of release of radiation to space from the same column added to the net flux density of heat leaving the column to enter surrounding columns. EBMs treat two dimensional geography explicitly by using a different (uniform) heat capacity over land than over ocean [11].

Currently there is no working two dimensional EBM that is available at Texas A&M for research purposes. We will rely on published EBM results. On the one hand, we try to simulate by capturing as much of the dynamics as we can in comprehensive numerical models. On the other hand, we try to understand by simplifying and capturing the essence of a phenomenon in idealized models [7]. Presently, GCMs are the primary tools by which theory confronts observations and there is less emphasis on simple stochastic models.

The two dimensional EBMs we refer to are linear stochastic models. The linear model is forced by uniform white noise in space and time over the spherical surface. Our thinking here is that the noise represents "weather." This kind of model works well in mid and polar latidudes, the so-called "storm belts." Obviously, it is not expected to work in the tropics (where "weather" is not white); hence, we restrict this study to the extra-tropics.

The lowest order of the so-called Matern correlation model (see [15]) leads to the $rK_1(r)$ function shown in Figure 2.1. This happens to be the exact solution for the stochastic linear EBM on a uniform plane if the raw data are averaged over times longer than the relaxation time of the model (i.e. the low frequency limit)[15]. this model then provides an intuitive basis for the intercomparison of different GCM

performances.

Ideally for this research, EBM output would have been examined alongside the GCMs, observations and reanalysis under the same methodology presented in Section 4. In the absence of EBM output, e-folding distances have been estimated from [12] (see Figure A.1).

3.3 Observations and Reanalysis

The observations were taken from the NCDC. The data set is a merged land, air and sea surface temperature anomaly analysis based on surface temperature data from The Global Historical Climatology Network (GHCN) version 3.1.0 and sea surface temperature data from The International Comprehensive Ocean-Atmosphere Data Set (ICOADS). This dataset is on a $5^\circ \times 5^\circ$ grid starting in January 1880 and ending in December 2011.

Reanalysis data were taken from NCEP. It is a joint product from NCEP and NCAR which incorporates observations and numerical weather prediction (NWP) model output to produce a retroactive record of more than 60 years of global analyses of atmospheric fields in support of the needs of the research and climate monitoring communties. This effort involved the recovery of land surface, ship, rawinsonde, pibal, aircraft, satellite, and other data [13]. The dataset is on a continually updated $2.5^\circ \times 2.5^\circ$ grid.

3.3.1 Some Uncertainties

The principal limitation of the observational dataset for global temperature analysis is the incomplete spatial and temporal coverage. Figure 3.1 shows the total number of missing monthly average temperature observations globally with the red dots indicating grid sites with no missing data for the time series.

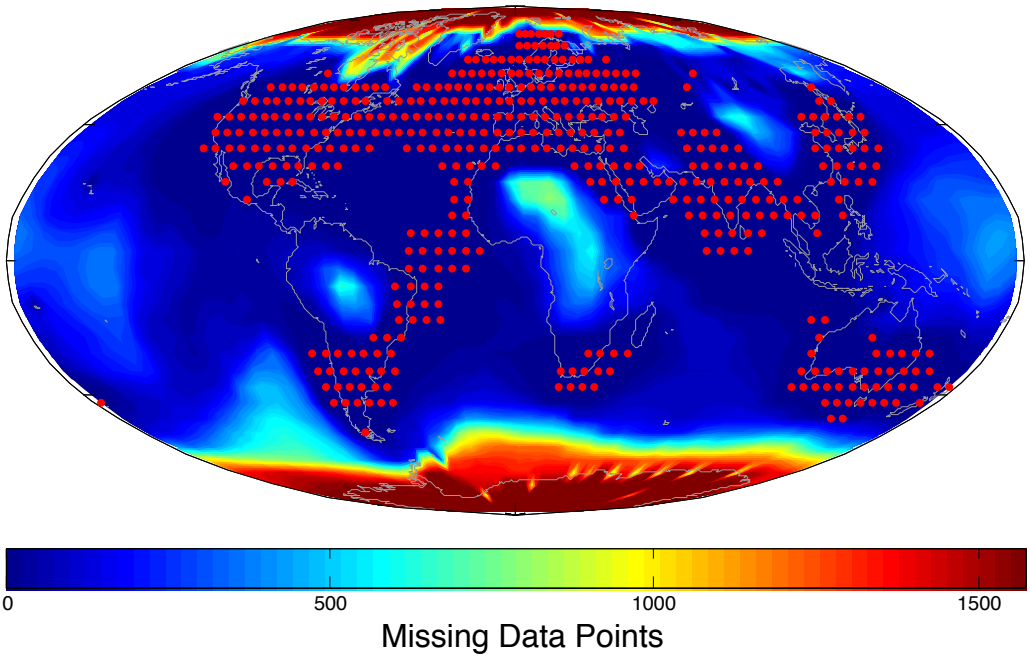


Figure 3.1: Missing data (color) from observation dataset. Sites with no missing data are marked with (•)

The observational dataset employed in this study has a total of 1584 entries (132 years). It can be seen that the poles are very poorly sampled with some sites having no observations on record. The interior area of Africa, Asia and South America as well as portions of the South Pacific Ocean are better sampled than the poles but still have long periods of time before observations were taken. In general, the Northern Hemisphere is better sampled than the Southern and land is better sampled than ocean.

To facilitate the computation of continuous spatial correlation structures, missing observation values were replaced with a (local) temperature anomaly of zero for that time step as in KNH96. This was intended to suppress ambiguities involved with comparing time series of different lengths (i.e. incomplete contours, underestimation,

etc.). However, this procedure does introduce some spurious overestimation of the correlation coefficients in poorly sampled areas. Unfortunately, it is impossible to suppress them both.

Reanalysis data were included in this study because of the intrinsic problems with the global observation network. Although it is a much shorter time series (65 years) than the longest observation record, it is both spatially and temporally complete. It should be noted that Reanalysis data were initialized with a mix of observations, hindcasting and forecasting so some of the drawbacks in the observation dataset "leak" into the Reanalysis data. These data are likely to be smoother (without false or "ghost" correlations) than a data set at all locations.

Gridded variables used in the reanalysis have been classified into three classes: *type A variables*, including upper-air temperatures, rotational wind and geopotential height, are generally strongly influenced by the available observations and are therefore the most reliable product of the reanalysis. *Type B variables*, including moisture variables, divergent wind and surface parameters, are influenced both by the observations and by the model, and are therefore less reliable. *Type C variables*, such as surface fluxes, heating rates and precipitation, are completely determined by the model (subject to the constraint of the assimilation of other observations). They should be used with caution and whenever possible compared with model-independent estimates. It is frequently noted that even when the model estimates are biased, the interannual variability tends to be correlated with independent observations [13].

4. METHODOLOGY

This section details the formulas and methods used in the study.

4.1 Filtering

The data for each model were run through two filters for standardizing. First, the seasonal cycle was removed from the data by the methods discussed below. This insures that the time series is as temporally stationary as possible. The seasonal cycle will induce much higher correlation calculations if not removed because of the repetitive pattern. For this, the data (which are sampled monthly) was partitioned into months ($i = 1..12$). Each month will have $tdim/12$ ($j = 1..tdim/12$) data points for each latitude (ϕ) and longitude (λ) combination, where $tdim$ is the total time dimension (e.g. for 1000 years sampled monthly, $tdim = 1,000 \times 12 = 12,000$).

$$\text{Month}_{1,j,\phi,\lambda} = Jan_{j,\phi,\lambda}$$

$$\text{Month}_{2,j,\phi,\lambda} = Feb_{j,\phi,\lambda}$$

$$\vdots$$

The mean for the entire time series at each point is then calculated.

$$\text{Mean}_{\phi,\lambda} = \frac{\sum_{k=1}^{tdim} \text{Temp}_{k,\phi,\lambda}}{tdim}$$

Next, the mean for each month is calculated for each location.

$$\text{MonthMean}_{i,\phi,\lambda} = \frac{\sum_{j=1}^{\frac{tdim}{12}} \text{Month}_{i,j,\phi,\lambda}}{\frac{tdim}{12}}$$

The anomaly away from the mean time series is then calculated for each month at each location by subtracting the full mean from the monthly mean.

$$\text{MonthAnom}_{i,\phi,\lambda} = \text{MonthMean}_{i,\phi,\lambda} - \text{Mean}_{\phi,\lambda}$$

This monthly anomaly is then subtracted away from the original data to remove the seasonal cycle.

$$\text{Temp}_{k,\phi,\lambda} = \text{Temp}_{k,i,\phi,\lambda} - \text{MonthAnom}_{i,\phi,\lambda}$$

The next filter was used to partition the data and find the anomaly from the local long term mean at each point for each time step. The partitioning of the data was done by taking different time average intervals of the original data (1 month, 3 months, 6 months, 1 year, 5 years, 10 years, 20 years). For most of the data sets we will only be looking in the 1 month to 5 year range, although the long term averages (10 and 20 years) will be useful when doing model comparisons to see how they deal with low-frequency climate variations. To find the anomalies, the time mean of each newly created dataset is found for each location. Using the one month average as an example:

$$\text{TimeMean}_{\phi,\lambda} = \frac{\sum_{k=1}^{tdim} \text{Temp}_{k,\phi,\lambda}}{tdim}$$

$$\text{Temp}_{k,\phi,\lambda} = \text{Temp}_{k,\phi,\lambda} - \text{TimeMean}_{\phi,\lambda}$$

To estimate the correlation coefficient, a sample covariance matrix must be constructed based on the two time series being compared. The covariance is defined as the (sample) expected value of the product of their deviations from their means [2]:

$$\mathrm{Cov}(x,y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n} \tag{4.1}$$

Showing that the covariance is non-zero suggests (subject to sampling error) that the two time series are not independent. The sample correlation coefficient, r, is then calculated as the covariance divided by the product of the standard deviations of the two distributions:

$$r = \frac{\frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{n}}{\sqrt{\frac{\sum(x_i-\overline{x})^2}{n}\frac{\sum(y_i-\overline{y})^2}{n}}} = \frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum(x_i-\overline{x})^2(\sum(y_i-\overline{y})^2}}$$

Or simply:

$$r(x,y) = \frac{\mathrm{Cov}(x,y)}{S_x S_y} \tag{4.2}$$

where S is the standard deviation of the respective time series. This will give a dimensionless quantity which can therefore be used, with certain reservations, as an absolute measure of the relationship between the two variables [2]. If x and y were independent, r would be zero. If large values of x are accompanied by large values of y they are positively correlated with the maximum being +1 for a perfect correlation. If large values of x are accompanied by negative values of y they are negatively correlated with a minimum being -1 for an exact opposite correlation. P-values were also calculated to test the significance of our calculations, all of which fall in the 0.01 significance level. This assumes no autocorrelation in the two time series. Roughly speaking the time series length has to be shortened by a factor of $\frac{1}{2\tau}$, where

τ is the autocorrelation time (Figure 4.1). This reduces the number of independent samples. It makes the significance level larger.

Autocorrelation functions were also calculated for each site and model. This is the cross-correlation of a time series with itself as a function of the lag time (τ) between observations. Assuming x_t is a stationary process the mean ($\overline{x}$) and variance (σ) are time-independent and autocorrelation (R) is expressed as a function of time lag only

$$R(\tau) = \frac{\langle (x_t - \overline{x})(x_{t+\tau} - \overline{x}) \rangle}{\sigma^2} \tag{4.3}$$

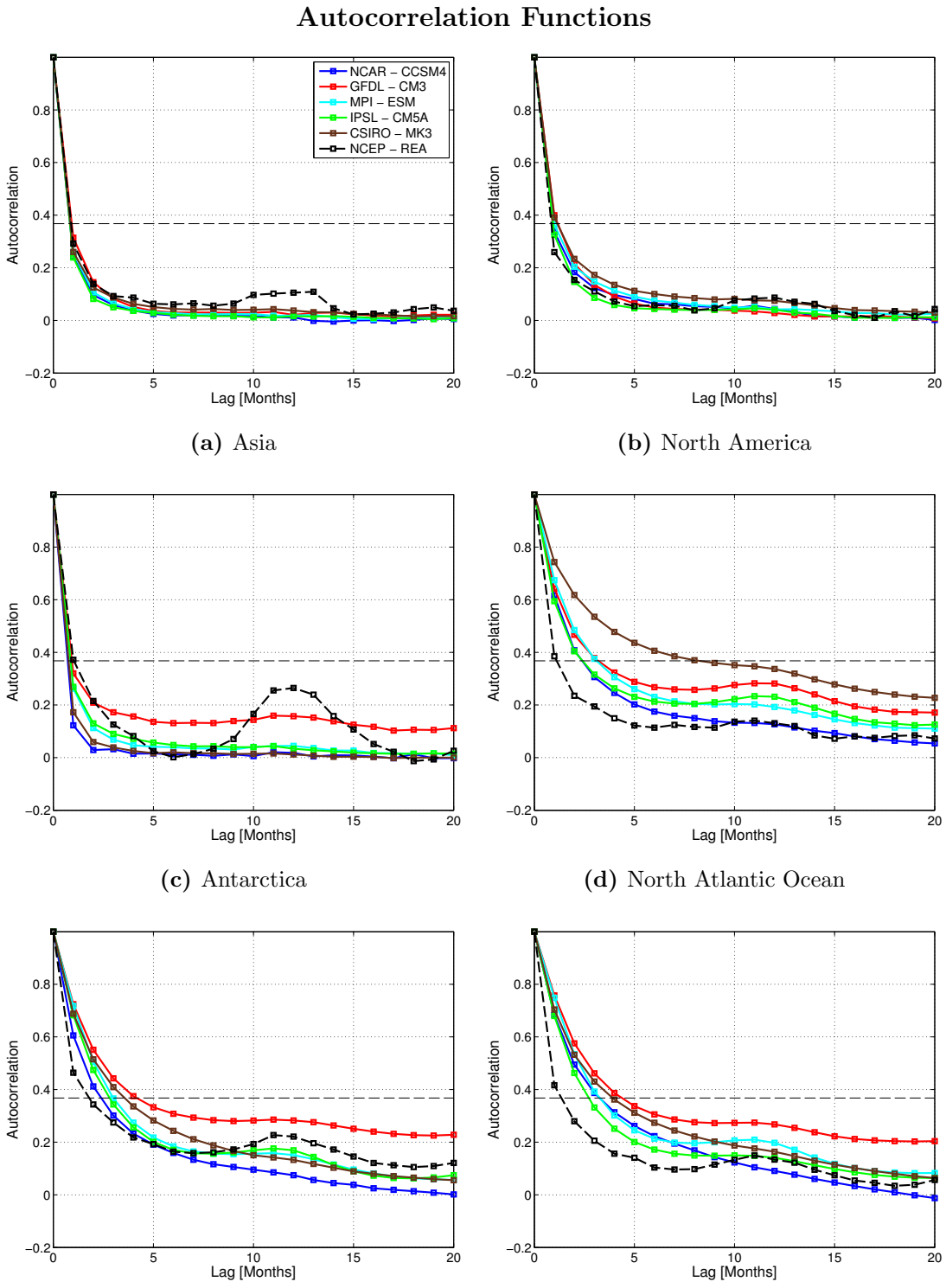
where $\langle \cdot \rangle$ is the expected value. Here lag time is expressed in months. Autocorrelation was calculated for each point inside the reference boxes for each site (Figure 4.2) up to a lag of 20 months and averaged to produce the curves seen in Figure 4.1. Autocorrelation time as discussed herein refers to the number of lags it takes for the autocorrelation function to fall to a value of e^{-1}. Note the great spread of lagged correlation times. This will show up again in interpreting the frequency dependence later.

The absolute distance, Δd, between two latitude-longitude pairs, (ϕ_1, λ_1) and (ϕ_2, λ_2) was calculated using the great circle length

$$\Delta d = R_e * \arccos(\cos(\phi_1)\cos(\phi_2)\cos(|\Delta\lambda|) + \sin(\phi_1)\sin(\phi_2)) \tag{4.4}$$

where $|\Delta\lambda| = |\lambda_1 - \lambda_2|$ and $R_e = 6374$ km.

The linear warming trend of observations and reanalysis was not removed from the data because of the uncertainty of its magnitude. This trend means that the data will be slightly non-stationary, which causes a minimal overestimation of correlation structure.



(a) Asia

(b) North America

(c) Antarctica

(d) North Atlantic Ocean

(e) Indian Ocean

(f) South Pacific Ocean

Figure 4.1: Autocorrelation functions for all models

4.2 Correlation Calculations

In this study the data were examined in two separate ways. First, a rectangular grid was taken around the area of interest in a way that maximized the amount of coverage of the location under examination. Within this box, 10,000 random latitude-longitude pairs were created and the correlation between the two time series was calculated as well as the distance separating the two sites. This was done to see what the large scale representative structure of correlation (or decorrelation length) is for a relatively uniform land or ocean surface. The second and more interesting evaluation of the data was constructed by choosing a point as close as possible to the center of the uniform land or ocean mass and correlating the time series of that site in all directions to find where the correlation falls to the e-folding threshold (e^{-1}) at which the correlation is no longer considered significant. For this study we are only interested in the statistics of the large uniform land or ocean surfaces, so we are only interested in the first instance of the e-folding contour and everything outside of that was ignored.

4.2.1 Random-Pair Correlations

The limits for the boxes used in the random-pair correlation calculations are represented in Table 4.1 and Figure 4.2. The bounding boxes were made as close as possible to these limits. However, because each model has a different grid spacing the actual limits will vary around these values by about $\pm 5^{\circ}$.

After the data were read in respecting these boundaries, 10,000 random latitude-longitude pairs were generated to get a representative sample of the mostly uniform surface. Inevitably, in every grid box there will be a combination of land and water surfaces. However, these usually lie on the outside of these boxes so their contribution to the whole will be minimal.

Table 4.1: Limits for random-pair correlation boxes and central point

Site [Abbreviation]	Latitude	Longitude	Central Point
Asia [ASIA]	25N - 70N	35E - 130E	50N 90E
North Atlantic Ocean [NAT]	35N - 70N	60W - 5E	50N 30W
North America [NAMC]	20N - 60N	125W - 75W	40N 100W
Indian Ocean [IND]	70S - 25S	30E - 120E	40S 80E
South Pacific Ocean [SPAC]	70S - 25S	130E - 80W	45S 140W
Antarctica [ANT]	90S - 70S	All	90S 0E

For each of these 10,000 combinations the correlation coefficient and distance were calculated as detailed above. Duplicate combinations were removed from the calculations so the maximum iteration of any combination would be one. The data were then scatter-plotted with the sample correlation coefficient on the ordinate and distance on the abscissa (see Figure 2.1). For the average line, the data was binned by distance into 100 bins and averaged. Next, the e-folding distance was found by evaluating at what distance the average line fell to a value of e^{-1}. This was done for each model, site and time average to provide a representative correlation curve and e-folding distance for the uniform areas of interest.

4.2.2 Central Correlations

For the central correlations the procedure was different. Here, we try to show how the correlation falls off with respect to distance from the central site. It may be easier to think of this as a surface or dome of correlation with the maximum of +1 at the center point and falling off in each direction until it reaches the e^{-1} contour. A point was chosen at the most central point possible in the uniform areas mentioned above. These points were chosen as close to the coordinates listed in Table 4.1 as possible.

The time series for each point inside the gridded area was then compared to

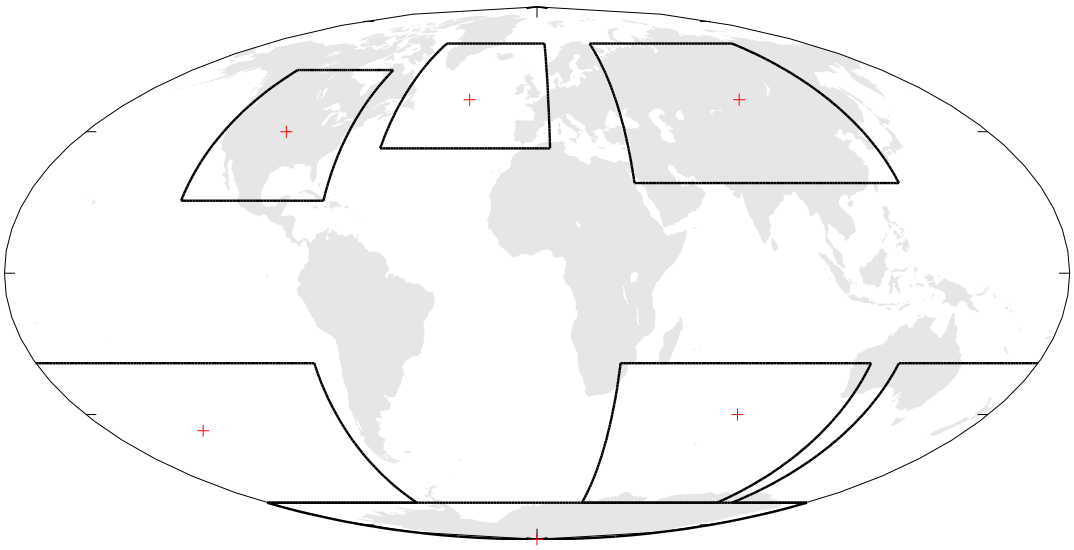


Figure 4.2: Limits for random-pair correlation boxes (black) and central reference point (+)

the central point to find the correlation coefficient and distance as described earlier, however, because each point is being compared to a central point the results are stored in a grid matching the bounding box for the site in question. With this gridded data, the correlation coefficient was then plotted as contours on a map at the intervals of 0.9, 0.7, 0.5 and a thicker contour at the e-folding distance, e^{-1}. The Mollweide projection was used for the maps in this study because accurate representation of area should take precedence over shape.

The shape and features of these contours are unique to each model and will therefore be a good means of comparison. Some features that are not unique are the flattening of the contours at coastal boundaries such as in North America showing a decoupling of temperature between land and ocean. This happens because of the difference in heat capacity between land and ocean, being the determining factor in the spatially lagged autocorrelation. Also there is, in most models, a flattening of contours near the Himalayan Mountains (also in [15] and [12]). Over land the

autocorrelation time is on the order of less than a month, while over the ocean it is on the order of months or a few years (longer if the frequency is low enough to activate the deeper ocean). Because of this, mostly the one month, one year and five year averages will be discussed in this study when comparing models and observations, although, the remaining averaging options were calculated and included in several figures in Section 5.

4.3 More About Stochastic EBMs

[11] showed that stochastically forced EBMs were successful at modelling the spatial patterns of variability when compared to observations with a few caveats. This section will use the simple defining equations of a noise-driven EBM as a mathematically convenient analog to describe the climate system and model simulations.

The surface temperature $T(\hat{\mathbf{r}}, t)$ ($\hat{\mathbf{r}}$ is a unit vector pointing from the center of the Earth to a point on the surface and t is time), satisfies the noise-forced damped diffusive equation

$$C\frac{\partial T}{\partial t} - D\nabla^2 T + BT = F \tag{4.5}$$

where C is a uniform heat capacity per unit area of surface, B is a radiative-damping coefficient and D is a thermal diffusion coefficient. $F = F(\hat{\mathbf{r}}, t)$ is a stochastic function which is white in time and whose statistics are rotationally invariant on the sphere.

$$\langle F(\hat{\mathbf{r}}, t)F(\hat{\mathbf{r}}', t')\rangle = \sigma_F^2 \rho(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}')\delta(t - t') \tag{4.6}$$

where $\langle\cdot\rangle$ denotes ensemble averaging and ρ is the spatial autocorrelation function which depends on the great circle distance or opening angle $\theta = \arccos(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}')$, separating the two points; σ_F^2 is the total variance of the forcing noise F at a point and $\delta(\cdot)$ is the Dirac delta function.

The e-folding distance described in this study is reasonably close to the EBM low-frequency diffusive length scale ($\lambda_L \sim \sqrt{D/B}$). The correlation length scale, λ_L, is smaller for higher frequency (shorter time average) temporal fluctuations and is larger over uniform land masses than over the ocean. This is because the very definition of high- versus low-frequency depends on the relevant relaxation time of the surface medium. Typically for EBMs it is about one month for a uniform all-land area and about 5 years over the equivalent open mixed-layer ocean [11]. This is much longer than the length scales seen in Figure 4.1 due to the fact that for EBMs the correlation length scales were calculated on an all land or all ocean planet while the models use a more realistic distribution of geography. Also, the models use multiple vertical layers which introduces multi-level advection effects.

The correlation length scale increases upon averaging the surface temperature field over longer times. This is simply expressed in the EBM as the lengthening of the correlation length scale towards its low frequency limit $\sqrt{D/B}$. Once a site has reached its corresponding low frequency limit, longer time averages (lower frequencies) will not increase λ_L further. This will be an important factor when comparing the model performance in Section 5.

5. RESULTS AND DISCUSSION

5.1 Results

5.1.1 Correlation of Random-Pairs

For the random-pair correlation calculations the results for each model, site and time average are shown in this section. Note that for the random correlation calculations, no distinction is made on direction of separation between the two sites, although some anisotropy in the fields is known to exist. For example, two sites that are taken at the same latitude and are only separated in the east-west direction will usually have a higher correlation coefficient than two sites on the same longitude band and are separated by the same distance only in the north-south direction. This arises due to the shape and prevalence of frontal phenomena which usually have a southwest to northeast orientation and move in a southeasterly direction in the Northern Hemisphere. Latitudinal dependence of solar insolation also plays a small role in the anisotropic structure of correlation.

Table 5.1: Length scales (km) (e-folding correlation distances), monthly averaged data relating random-pairs.

	Land Sites			Ocean Sites		
Model	ASIA	NAMC	ANT	NAT	IND	SPAC
NCAR - CCSM4	1653	1475	1235	1014	1233	1514
GFDL - CM3	1585	1383	913	944	1551	1746
MPI - ESM	1575	1318	881	720	836	1048
IPSL - CM5A	1607	1486	1030	968	1166	1336
CSIRO - MK3	1155	1187	764	787	893	1148
NCEP - REA	1793	1578	1530	1314	1580	1557

Table 5.2: See Table 5.1, yearly averaged data relating random-pairs.

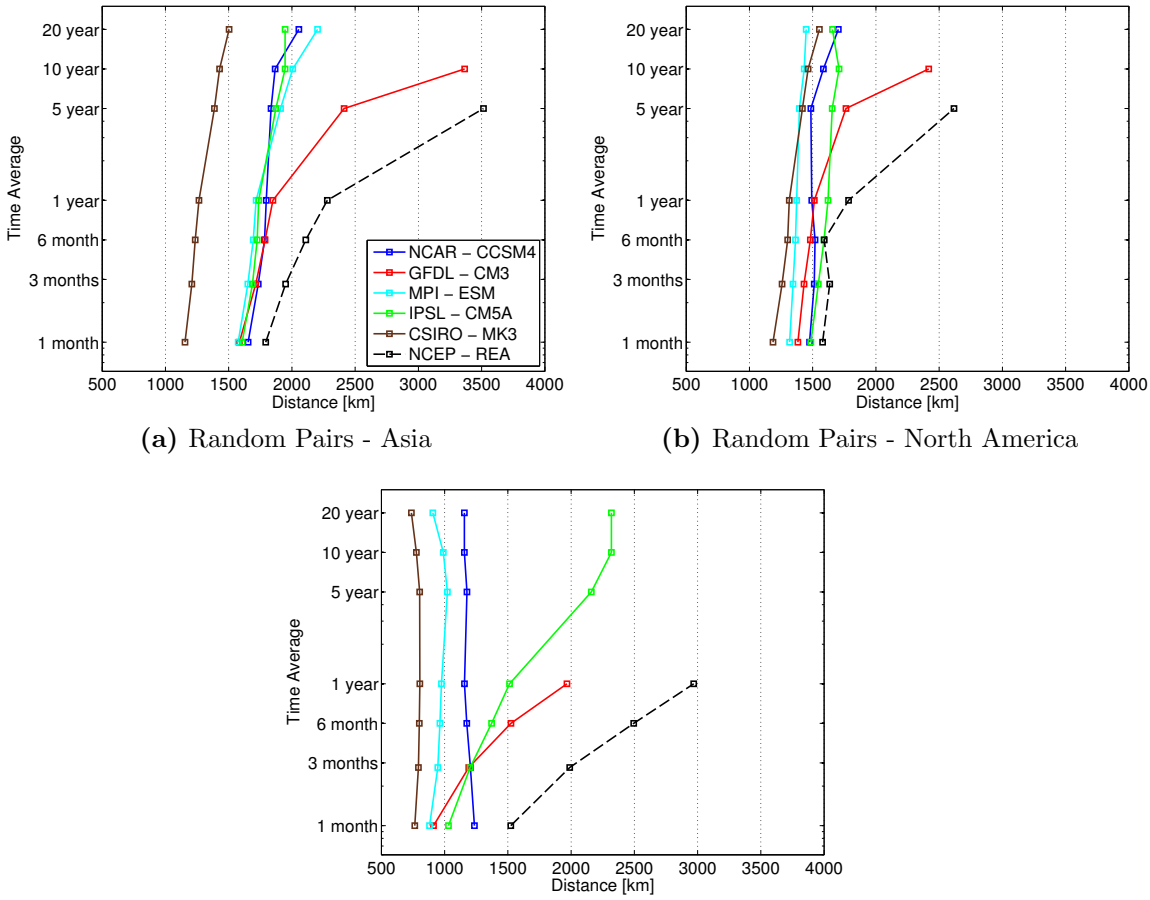
	Land Sites			Ocean Sites		
Model	ASIA	NAMC	ANT	NAT	IND	SPAC
NCAR - CCSM4	1797	1493	1157	1278	1574	1854
GFDL - CM3	1849	1513	1967	1209	2753	2549
MPI - ESM	1717	1373	975	925	937	1098
IPSL - CM5A	1737	1621	1513	1149	1498	1499
CSIRO - MK3	1265	1315	804	979	1066	1277
NCEP - REA	2280	1784	2243	1619	2278	1763

Table 5.3: See Table 5.1, five year averaged data relating random-pairs.

	Land Sites			Ocean Sites		
Model	ASIA	NAMC	ANT	NAT	IND	SPAC
NCAR - CCSM4	1835	1485	1177	1685	1949	2356
GFDL - CM3	2412	1763	-	1746	-	-
MPI - ESM	1910	1395	1022	1290	1062	1129
IPSL - CM5A	1873	1654	2160	1321	1649	1614
CSIRO - MK3	1388	1419	802	1253	1222	1348
NCEP - REA	3515	2617	-	2624	3280	1833

In Tables 5.1, 5.2 and 5.3 and Figures 5.1 and 5.2 the e-folding distances for the random-pair correlation boxes have been separated out into land sites (Figure 5.1) and ocean sites (Figure 5.2). Even in the 1-month average, which is about the autocorrelation time of a uniform land area and also the sampling rate of the models, there is quite a large spread in the e-folding distances ranging from around 700km to 1700km. This spread not only depends on the type of surface enclosing the site pairs but the model as well. The "genealogy" of each model differs, each using its own ocean-atmosphere coupling method, dynamics, deep ocean mixing, feedbacks, mixing layer depth, seasonality, etc.

Some initial inferences can be made simply by looking at these figures. (1) The



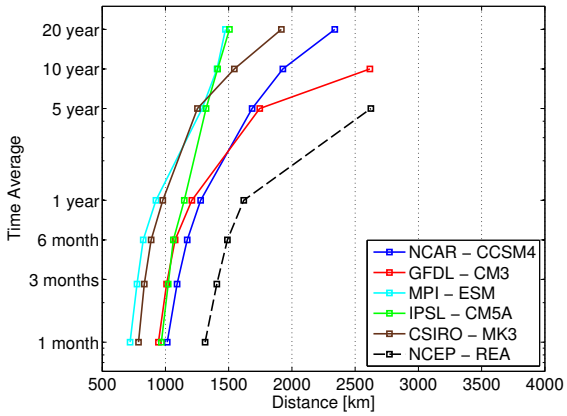
(a) Random Pairs - Asia

(b) Random Pairs - North America

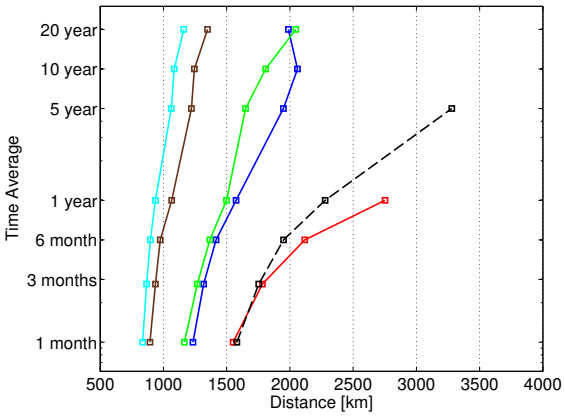
(c) Random Pairs - Antarctica

Figure 5.1: e-folding correlation distances for various time averages for land sites. NCEP - REA was only calculated up to 5 years because of the shorter time length. Lines that terminate before 20 years had an e-folding distance of more than 4000 km or did not fall to the value of e^{-1} at all.

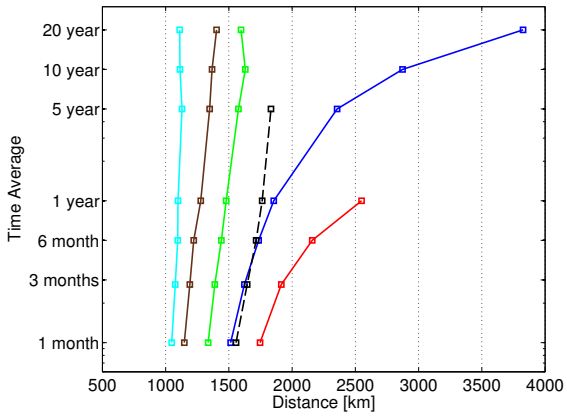
GFDL - CM3 consistently has the highest e-folding distances amongst the GCMs for most of the sites, especially the ocean sites (note the growth for long time averages). (2) The MPI - ESM and the CSIRO - MK3 exhibit mostly smaller e-folding distances for all sites especially in the longer time averages. (3) There is in general less agreement between models over ocean sites than over land sites with the largest spread in the Southern Hemisphere oceans. Keep in mind that these are random latitude-



(a) Random Pairs - North Atlantic Ocean



(b) Random Pairs - Indian Ocean



(c) Random Pairs - South Pacific

Figure 5.2: e-folding correlation distance for various time averages for ocean sites.

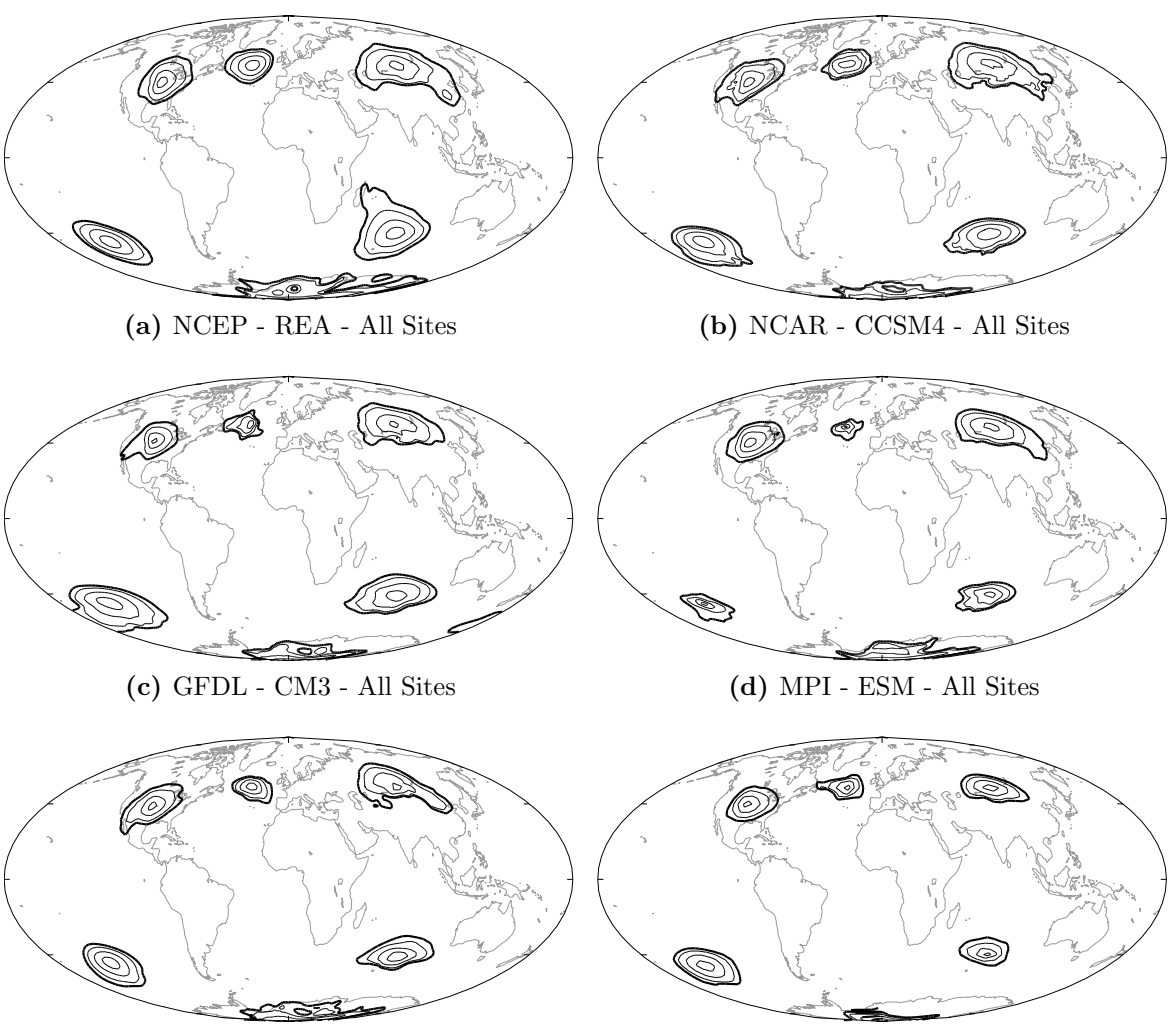
longitude pairs in a box of uniform surface type and are similar but not directly related to a central point which will be discussed in the next section. Because the random-pair method contains more data, there is less sampling error for estimating correlations. However, the use of random-pairs means that the correlation structure and evolution can be different in sequential tests.

Disregarding the magnitude of the random-pair e-folding distances for a moment and focusing on the shape of the lines made by connecting these points (as in Figure 5.1) the datasets can be roughly partitioned into two groups: (1) nearly vertical lines

indicate models whose e-folding distances do not increase with longer time averages and (2) curved lines indicate models whose e-folding distances do increase with longer time averages. *Group* (1) would consist of models for which their autocorrelation time is much less than the time averages employed while models in *Group* (2) have autocorrelation times greater than the time averages. This can also be seen in Figures 5.8, 5.9, 5.10 and 5.11.

5.1.2 Correlation Relative to Central Point

The central correlations will mostly be represented in maps with the correlation contoured about a central point described in Table 4.1. The results are grouped based on time average and separated into individual sites when overlapping of correlation contours would cause confusion.



(a) NCEP - REA - All Sites

(b) NCAR - CCSM4 - All Sites

(c) GFDL - CM3 - All Sites

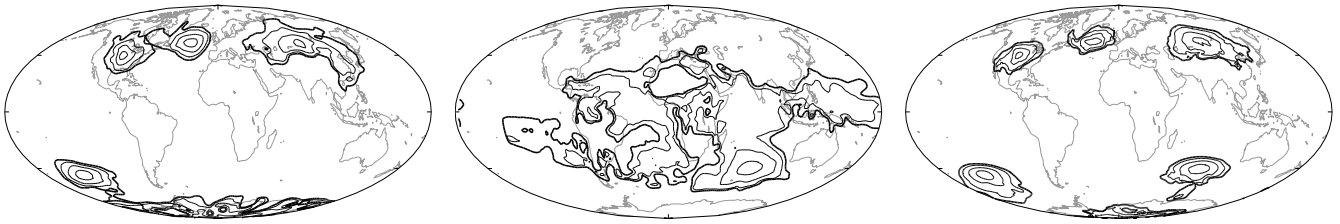
(d) MPI - ESM - All Sites

(e) IPSL - CM5A - All Sites

(f) CSIRO - MK3 - All Sites

Figure 5.3: Central correlations about a proto-site nearly centered in a reasonably uniform surface region for 1 month averaging. The contour lines are at 0.9, 0.7, 0.5 and the thick contour at e^{-1}

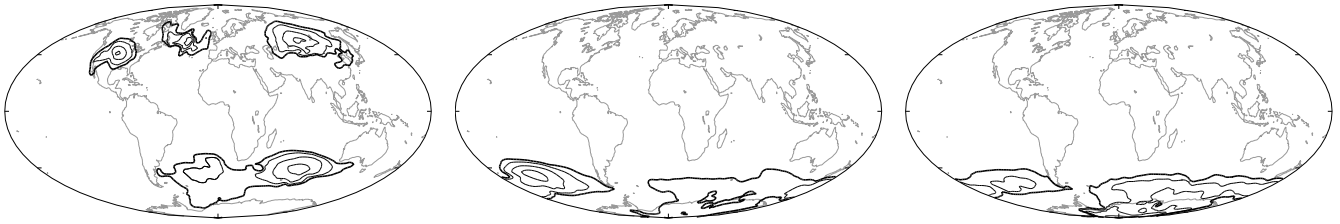
1 Year Averages - Central Correlations



(a) NCEP - REA - ASIA, NAMC, ANT, NAT, SPAC

(b) NCEP - REA - IND

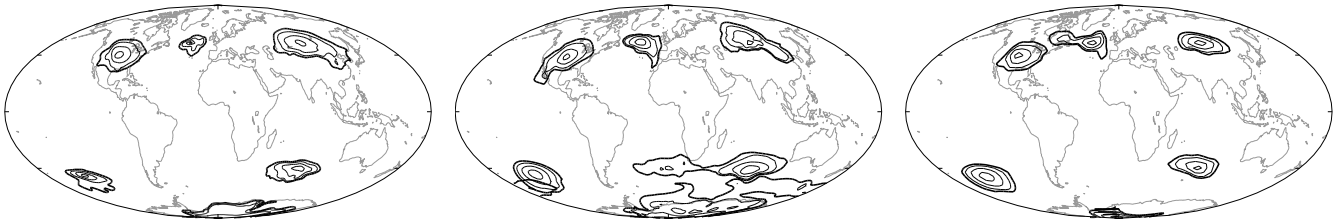
(c) NCAR - CCSM4 - All Sites



(d) GFDL - CM3 - ASIA, NAMC, NAT, IND

(e) GFDL - CM3 - SPAC

(f) GFDL - CM3 - ANT



(g) MPI - ESM - All Sites

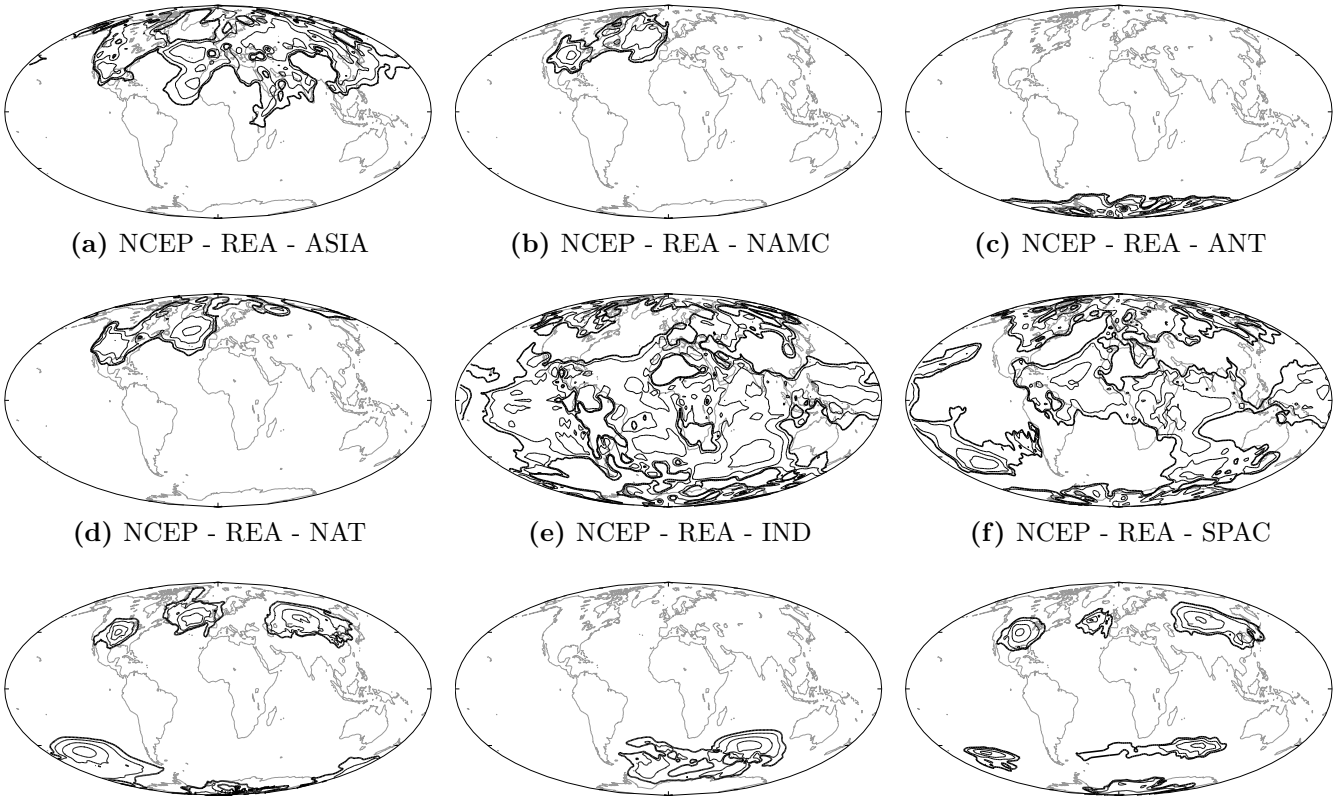
(h) IPSL - CM5A - All Sites

(i) CSIRO - MK3 - All Sites

Figure 5.4: Central correlations for 1 year averaging

In Figure 5.3 the spread of e-folding distances as discussed in the previous paragraph can be seen. For the 1-month average, the largest correlation contours are over land while the smallest are over ocean. All models and the reanalysis show a flattening of the contours south of the Asia central protopoint associated with the presence of the Himalayas as well as a southeastern tendency of "higher" correlations. The models seem to agree with one another on the general shape of these correlation contours as compared to the NCEP - REA except none mimic the northward tendency of the Southern Indian Ocean site. It should be emphasized that the NCEP - REA is an analysis/forecast based on observations, and although it can be shown to have a good fit to observations in areas of high sampling, no inferences can be made in places where observations are few or non-existent as in most of the southern hemisphere. Moreover, a good fit to raw observations need not lead to a good fit to statistical correlations.

Figures 5.4, 5.5 and 5.5 begin to show the expansion of these e-folding correlation contours in some of the models. Some noteworthy observations can be noted from comparing these maps. (1) The westward expansion of correlations for the Southern Indian Ocean site in the GFDL - CM3, IPSL - CM5A and to a lesser extent the NCAR - CCSM4 which has a smaller and more southerly expansion. Not until the 5-year average do the CSIRO - Mk3 and MPI - ESM begin to show this expansion. (2) The southern hemisphere sites for the GFDL - CM3 "blow up" to being very highly correlated to some distance north of the equator. (3) Correlation falls faster away from unity as distance increases over ocean sites than over land sites.



(a) NCEP - REA - ASIA

(b) NCEP - REA - NAMC

(c) NCEP - REA - ANT

(d) NCEP - REA - NAT

(e) NCEP - REA - IND

(f) NCEP - REA - SPAC

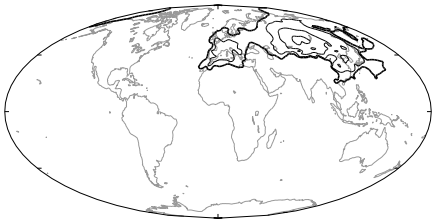
(g) NCAR - CCSM4 - ASIA, NAMC, ANT, NAT, SPAC

(h) NCAR - CCSM4 - IND

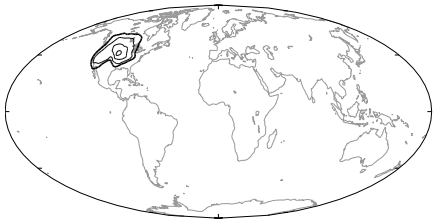
(i) MPI - ESM - All Sites

Figure 5.5: Central correlations for 5 year averaging

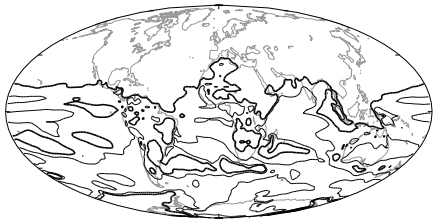
5 Year Averages - Central Correlations (cont'd)



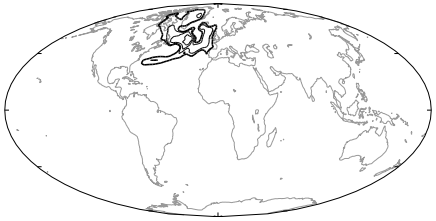
(j) GFDL - CM3 - ASIA



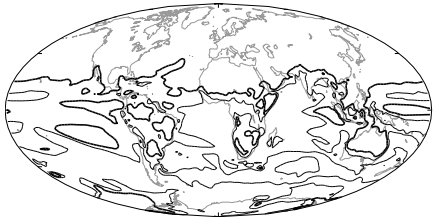
(k) GFDL - CM3 - NAMC



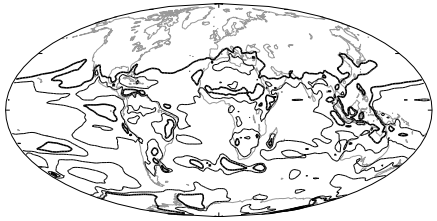
(l) GFDL - CM3 - ANT



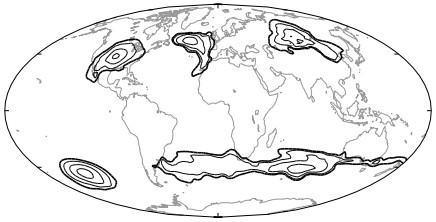
(m) GFDL - CM3 - NAT



(n) GFDL - CM3 - IND

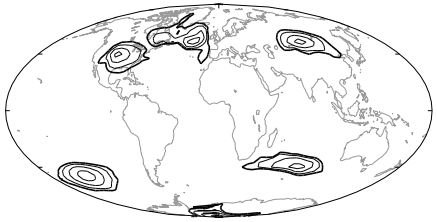


(o) GFDL - CM3 - SPAC



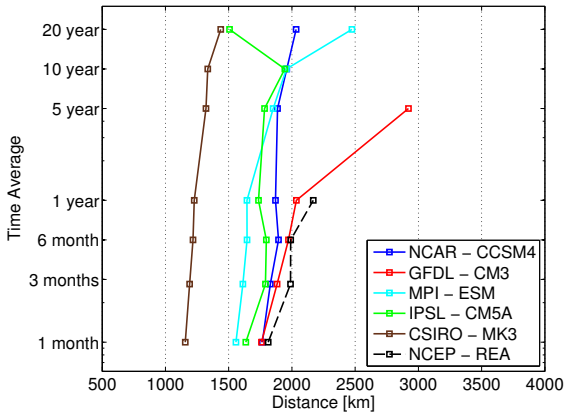
(p) IPSL - CM5A - ASIA, NAMC, NAT, IND, SPAC

(q) IPSL - CM5A - ANT

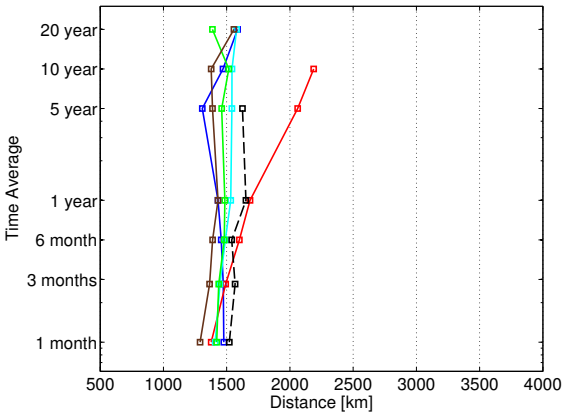


(r) CSIRO - MK3 - All Sites

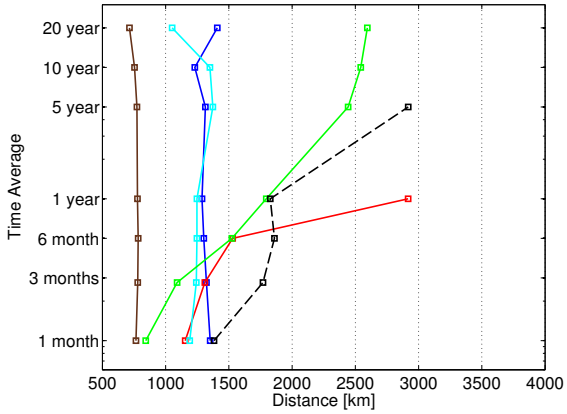
Figure 5.5: Continued



(a) Central Correlation - Asia



(b) Central Correlation - North America



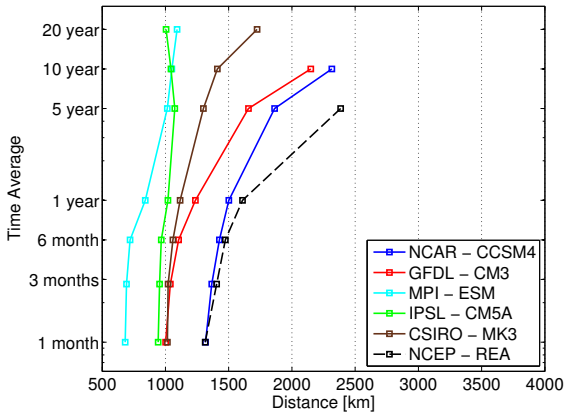
(c) Central Correlation - Antarctica

Figure 5.6: e-folding correlation distance for land sites. NCEP - REA was only calculated up to 5 years because of the shorter time length. Lines that terminate before 20 years had an e-folding distance of more than 4000 km or did not fall to the value of e^{-1} at all.

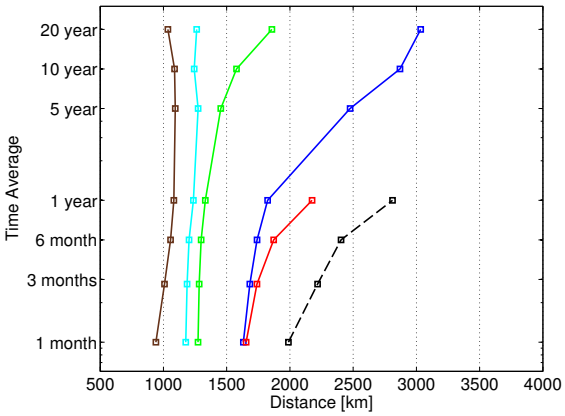
5.2 Discussion

This study attempted to quantify the second moment statistical structure of the surface temperature field (more specifically the e-folding spatial autocorrelation distances) at various time averages on the raw data. The remainder of this section will provide evidence that will be used to answer the questions proposed in Section 2.

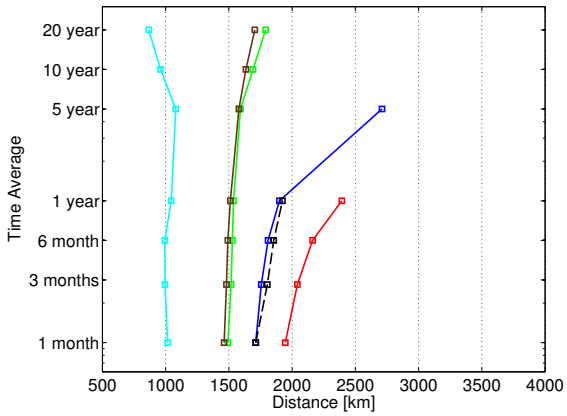
The interpretation of results of this study will be discussed from a stochastic



(a) Central Correlation - North Atlantic Ocean



(b) Central Correlation - Indian Ocean



(c) Central Correlation - South Pacific

Figure 5.7: e-folding correlation distance for ocean sites.

EBM point of view because of their simplicity and robustness as shown from previous studies. As for nomenclature, the terms time average lengths and frequencies will be used interchangeably with long time averages relating to low-frequencies and vice versa.

5.2.1 Sensitivity

Although no strong connection was found between the spatial correlation structure and equilibrium climate sensitivity, this does not mean that a relationship does

not exist. Transient sensitivity is the amount of warming after doubling CO_2 as it increases at a rate of one percent per year. The global climate system will not have come to equilibrium; hence it is less than the equilibrium sensitivity.

Table 5.4: Equilibrium climate sensitivity values.

Model	Sensitivity (K)
NCAR - CCSM4	3.20°
GFDL - CM3	3.97°
MPI - ESM	3.63°
IPSL - CM5A	4.13°
CSIRO - MK3	4.08°

Transient sensitivity is closely tied to vertical heat exchange in the ocean, since $\tau = C/B$ and the effective thermal inertia is C, the thermal inertia of the deep ocean is much greater than that of the upper ocean (mixed-layer). In the case that there is no heat exchange between the upper ocean and the deep ocean, a model can be integrated to a new equilibrium in a few tens of years. However, if there is heat exchanged between the upper ocean and deep ocean, this can delay equilibration by many decades or even centuries. Moreover, the depth of the upper ocean and upward/downward diffusion parameterizations are different for each model. It is not clear in this study whether the frequencies in question were low enough to "activate" the deeper ocean layers although looking at the e-folding times of the autocorrelation functions (Figure 4.1) one would expect a much longer autocorrelation time (on the order of a few years as opposed to several months) had there been deep ocean heat exchange.

5.2.2 GCMs vs. Observations/Reanalysis

Unfortunately, for the temporal and spatial scales used in this study, observational sampling was insufficient for an accurate comparison. The observational network is simply too sparse (in both time and space) to get meaningful results without introducing results that could be misleading. For this reason reanalysis data were included as an analog to to observations, however, its performance is also open to question. Hence, any comparison made to the statistical structure of the real climate should be considered as speculative, depending on one's confidence in reanalysis process.

At the 1-month average (which is also the sampling rate of model and reanalysis archived data), all models seem to agree with reanalysis on the general shape of contours when mapped as in Figures 5.3-5.5 with a few small exceptions. Over land these differences are smaller. In North America the NCAR - CCSM4, GFDL - CM3 and IPSL - CM5A have a southwesterly expansion of the contours along the western coastline that NCEP - REA does not show. Over Asia the southeasterly expansion which exists in NCEP - REA is shown in all the models except the GFDL - CM3 and CSIRO - MK3. In Antarctica the CSIRO - MK3 largely underestimates the size of the contours when compared with NCEP - REA, but all other models seem to be in good agreement. Over the North Atlantic Ocean all the models underestimate the magnitude of the contours except for the NCAR - CCSM4, with the biggest difference being in the MPI - ESM. None of the models show the northward expansion of contours over the Indian Ocean and all underestimate the size as compared to NCEP - REA, although the GFDL - CM3 comes closest. The GFDL - CM3 overestimates the e-folding distance for the Southern Pacific Ocean and the MPI - ESM underestimates it while all other models are close to one another.

Looking at 1-month data in Figures 5.6 and 5.7, there is a large disparity in the e-folding distance when compared to NCEP - REA with the largest disagreements occurring over the Southern Hemisphere oceans. The NCAR - CCSM4 and GFDL - CM3 are consistently the closest which is not a surprise since they are the models with the longest heritage.

At the 1-year and 5-year time averages larger differences between the models and reanalysis start to show. In the NCEP - REA correlation contours "blow up" to being highly correlated at large distances over Asia, the Indian Ocean, South Pacific Ocean and to a lesser extent the North Atlantic Ocean, although it is most prominent in the Southern Hemisphere. The magnitude of this expansion is not informative in itself, however, the occurrence of this expansion might be. The only model that seems to come close to imitating this phenomenon over the Southern Hemisphere ocean sites is the GFDL - CM3. Over Antarctica the IPSL - CM5A most closely resembles the NCEP - REA and none of the models match the expansion of contours shown over the North American or North Atlantic Ocean sites.

At first it was thought that this large expansion of the correlation contours in the NCEP - REA was related to the autocorrelation times shown in Figure 4.1 because the NCEP - REA had the shortest autocorrelation times over the Southern Hemisphere ocean sites (1-2 months). However, the GFDL - CM3, which most closely mimicked this expansion, had the longest autocorrelation times over the same sites (3-5 months) which leads to more questions as to the origin of this expansion.

5.2.3 GCMs vs. EBMs

For this study we would expect from stochastic EBM thinking that (1) as the time average gets longer for the e-folding distances to increase [15], approaching a limit related to the relative low-frequency limit and autocorrelation time. Similarly, we

would expect (2) the rate at which the correlation falls away from unity to decrease with ocean sites falling away from unity faster than land sites. We also expect that (3) the e-folding distances would increase faster over ocean than over land as the time averages increase. Figures 5.10 and 5.11 include the estimated e-folding correlation distances taken from [12] for comparison. Also taken from this study are the e-folding contour maps (Figure A.2).

As mentioned earlier, EBMs treat two dimensional geography explicitly by using a different (uniform) heat capacity over land than over ocean, which is why there is no flattening of the contours over Asia related to the Himalayas as seen in the models and reanalysis. Additionally, in KNH96 the e-folding correlation distances were calculated in Fourier-spherical harmonic space. The remainder of this section will focus on the shape and evolution of correlation structure over Asia, North America and the North Atlantic Ocean as it relates to stochastic EBM theory and will disregard magnitudes of e-folding distances.

Over Asia, the NCAR - CCSM4, IPSL - CM5A and CSIRO - MK3 have nearly vertical lines in Figure 5.6a meaning that they have reached their low-frequency limit and will not expand more with time averages of increasing duration. Interestingly, the IPSL - CM5A actually shows a large decrease in its e-folding distance at the twenty year average for unknown reasons (recall that with 20 year averages, there can be a sampling problem). The MPI - ESM shows a small increase after the 1-year average while the GFDL - CM3 shows the largest increase. These effects can also be seen in Figure 5.8 with tightly packed lines corresponding to small increases in the e-folding distances and widely spread lines corresponding to large increases. Over North America the lines from all the models are nearly vertical except for the GFDL - CM3 which shows a small increase. Over the North Atlantic (Figures 5.7a and 5.9) the IPSL - CM5A and the MPI - ESM exhibit nearly vertical lines while the rest of

the models show a large increase at lower frequencies.

The rate at which the correlation falls away from unity decreases in all the models that exhibit increasing e-folding distances as expected. Models that do not show an increase in e-folding distances have no change in the rate in which they fall away. In all the models the rate at which the correlation falls away from unity is faster over ocean than over land except for the CSIRO - MK3 which shows the same rate of decrease over ocean and land.

5.2.4 GCMs vs. GCMs

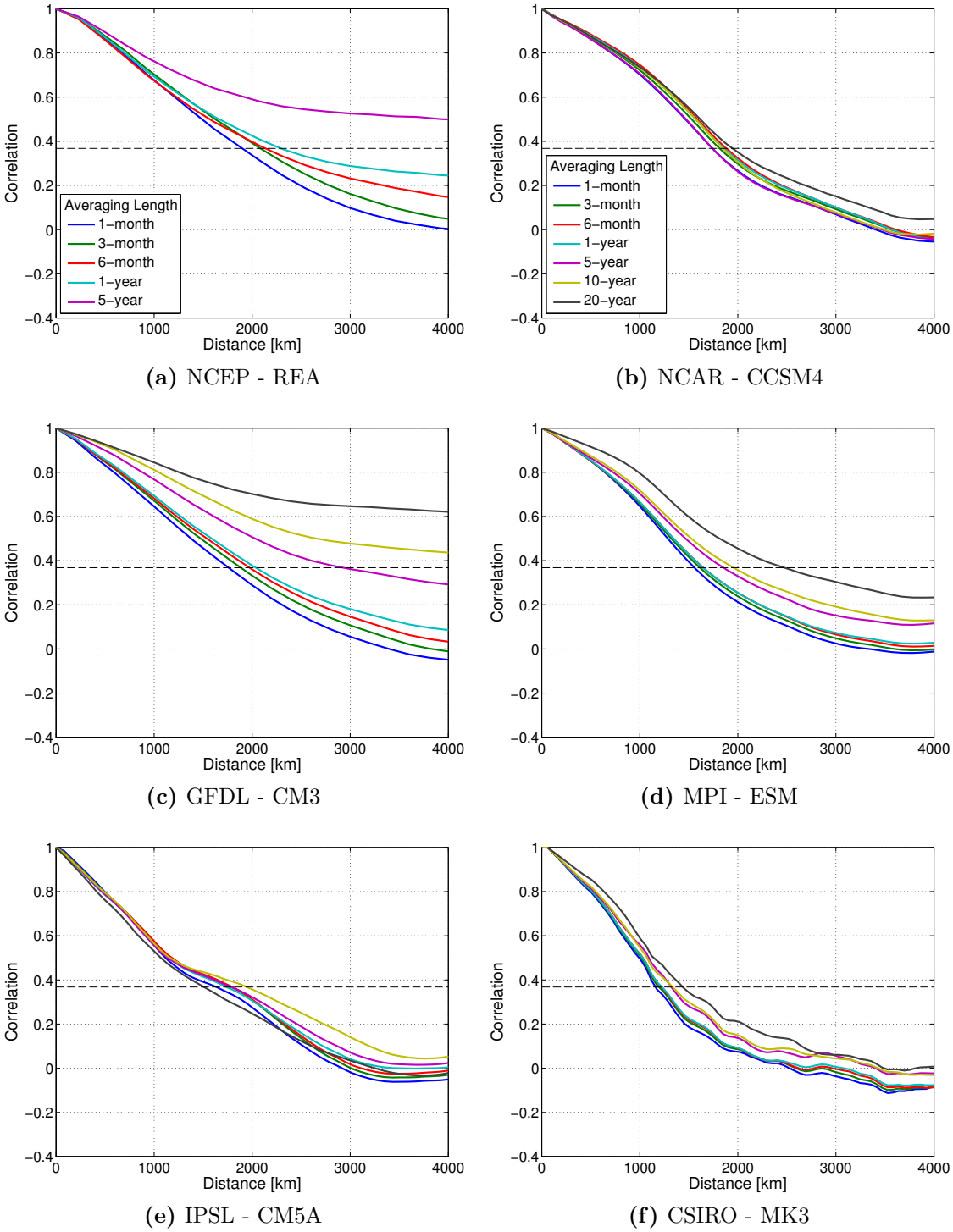
This section will highlight the main motivation for this study, a comparison among the second moment statistics of current GCMs from the recent CMIP5 simulations. Care has been taken to insure that all models have been treated objectively. No judgment has been made as to right or wrong only that there is a difference, sometimes very large, among the models. This study does not attempt to explain why such differences exist in the climate models. In general there is better agreement between models over land than over ocean. Secondly, there is greater agreement in the Northern Hemisphere than in the Southern Hemisphere. The GFDL - CM3 consistently has the largest e-folding distances at longer time averages, however it has the shortest run time of the models at 500 years so sampling may be an issue at low-frequencies.

The North American site has the best model agreement in this study (Figures 5.1b and 5.6b). Only the GFDL - CM3 shows an increase in the e-folding distance after the 1-year average. Over Asia, the model behavior is similar except that the CSIRO - MK3 has a much smaller magnitude of e-folding distance although the structure and evolution is similar. Also the MPI - ESM e-folding distances begin to increase after the 1-year average, but to a lesser extent than the GFDL - CM3.

The Antarctic site has the least agreement among land sites. The CSIRO - MK3 has the smallest e-folding distances but has the same structure as the NCAR - CCSM4 and MPI - ESM. The IPSL - CM5A and GFDL - CM3 e-folding distances exhibit large increases after the 1-month average, but the IPSL - CM5A line begins to go vertical after the 5-year average while the GFDL - CM3 continues to increase. The autocorrelation times (Figure 4.1) are very similar for all land sites.

The North Atlantic Ocean site has the best model agreement for ocean sites (Figures 5.2 and 5.9). The IPSL - CM5A is the only model that has a nearly vertical line. The MPI - ESM begins to exhibit a small expansion after the 6-month average but it is not as drastic as the CSIRO - MK3, GFDL - CM3 or NCAR -CCSM4. The South Pacific and Indian Oceans show the least amount of model agreement. Over the Indian Ocean the CSIRO - MK3 and MPI - ESM lines structures are nearly identical, differing only in magnitude. The IPSL - CM5A is similar except that it exhibits a small increase in e-folding distance after the 5-year average. The NCAR - CCSM4 and GFDL - CM3 have much larger e-folding distances than the other models and show a greater increase at lower frequencies with the GFDL - CM3 having the largest increase. For the South Pacific Ocean site the MPI - ESM has the smallest e-folding distances by a large margin and they begin to decrease after the 5-year average. The IPSL - CM5A and the CSIRO - MK3 are nearly identical and vertical with a very small increase after the 5-year average. The NCAR - CCSM4 shows a large increase in e-folding distances after the 1-year average but not as drastic as the GFDL - CM3.

Asia



(a) NCEP - REA

(b) NCAR - CCSM4

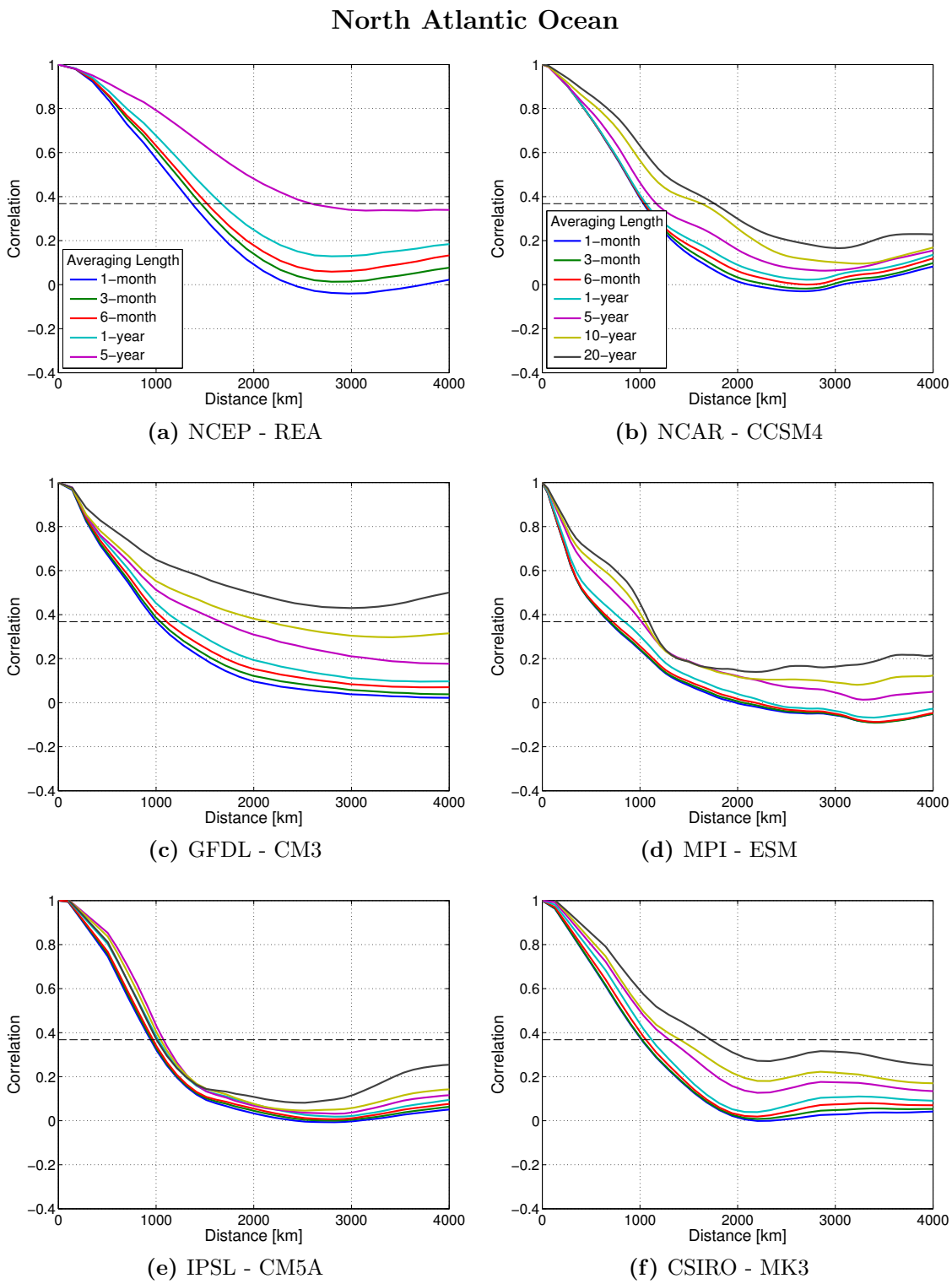
(c) GFDL - CM3

(d) MPI - ESM

(e) IPSL - CM5A

(f) CSIRO - MK3

Figure 5.8: Correlation structures for various time averages for Asia.



(a) NCEP - REA

(b) NCAR - CCSM4

(c) GFDL - CM3

(d) MPI - ESM

(e) IPSL - CM5A

(f) CSIRO - MK3

Figure 5.9: Correlation structures for various time averages for the North Atlantic Ocean.

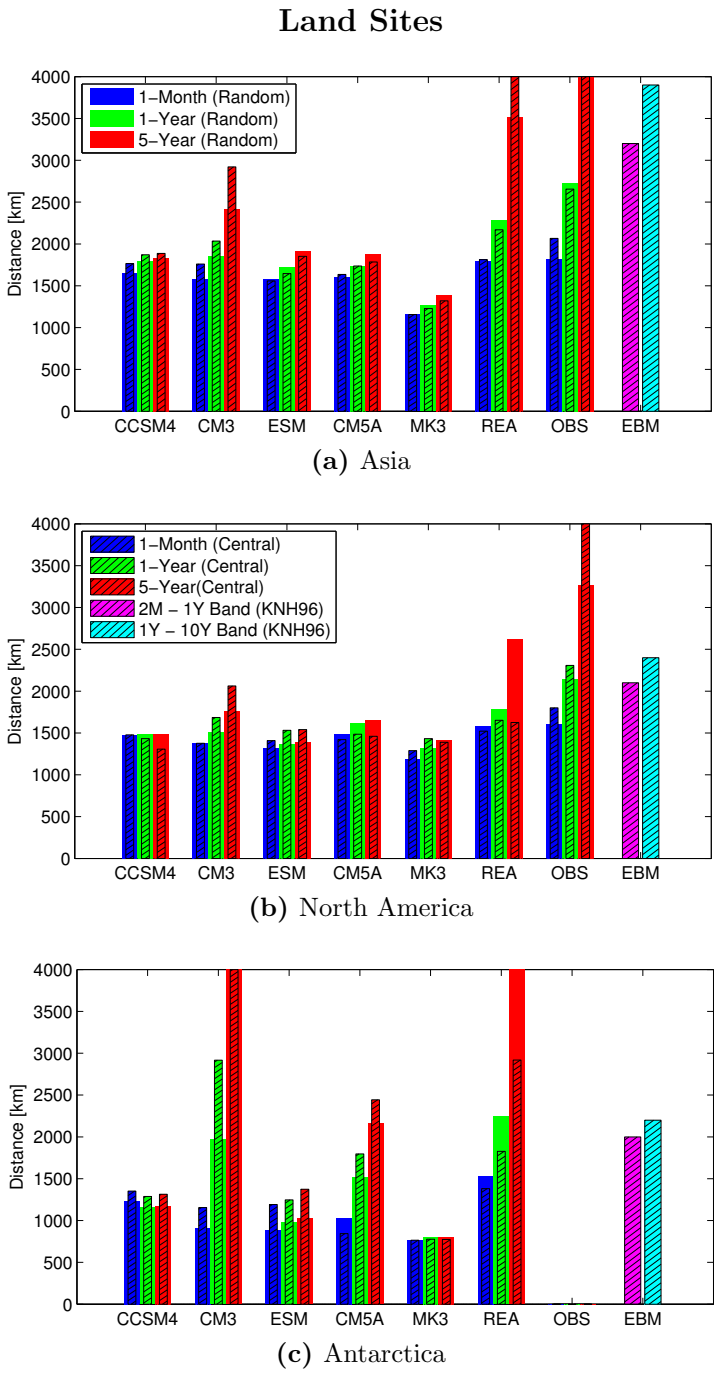
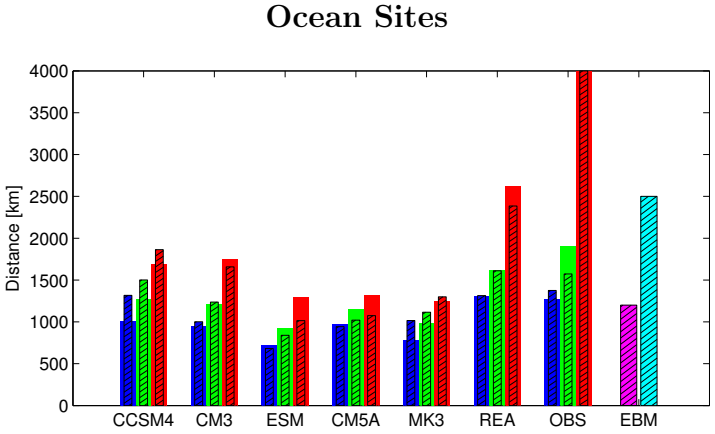
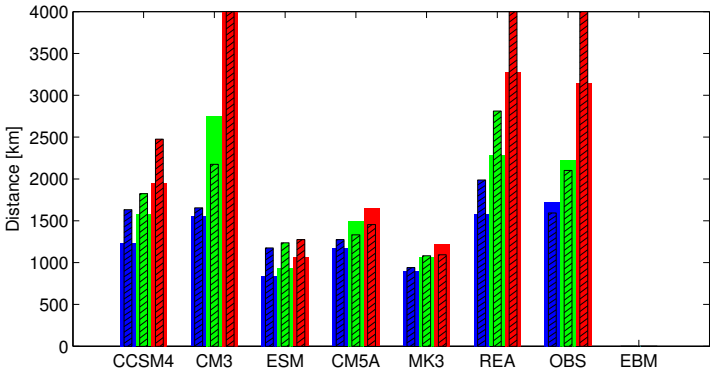


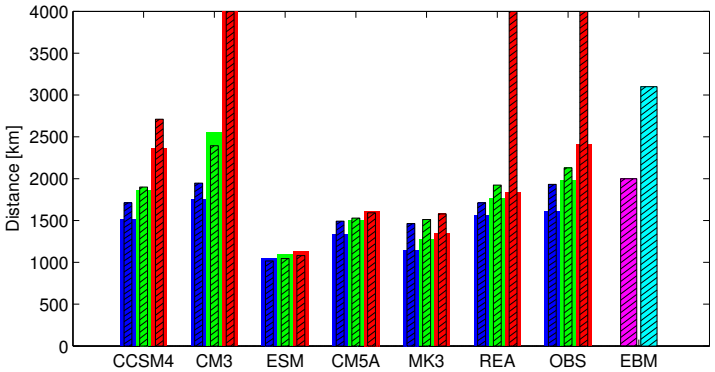
Figure 5.10: Summary of the 1-month (blue), 1-year (green) and 5-year (red) e-folding distances for land sites. Central correlation distances are hatched with black lines and laid over the random-pair correlations, which are solidly colored. Also included in these figures is the estimated e-folding correlation length for the EBM used in *Kim, North and Hegerl* [1996], which were calculated in a Fourier-spherical harmonic space in the 2-month to 1-year frequency band (magenta) and the 1-year to 10-year frequency band (cyan).



(a) North Atlantic Ocean



(b) Indian Ocean



(c) South Pacific Ocean

Figure 5.11: Summary of the 1-month (blue), 1-year (green) and 5-year (red) e-folding distances for ocean sites

6. CONCLUSIONS

6.1 Summary

In this study we estimated the second-moment statistics (specifically the spatial correlation structure) of five GCMs, observations and reanalysis. The five coupled models we considered were the NCAR - CCSM4, GFDL - CM3, MPI - ESM, IPSL - CM5A and the CSIRO - MK3. Only pre-industrial model control runs of at least 500 years were examined with the longest run being 1,300 years. The NCDC observational dataset was determined to be too sparse in space and time for a fair comparison so reanalysis data from NCEP were considered as an alternative. The reanalysis dataset included 65 years of monthly data on a $2.5^\circ \times 2.5^\circ$ grid. After removal of the seasonal cycle the datasets were assumed to be stationary which is a prerequisite for calculating spatial correlation.

As demonstrated, comparison of the second-moment statistics at various time averages revealed similarities and differences between the models and reanalysis data, stochastic EBM theory and also among the climate models themselves. The largest differences occurred at long time averages (lower frequencies), especially in the Southern Hemisphere oceans. Northern Hemisphere sites showed better model agreement with the best agreement over the North American site.

6.2 Questions Revisited

Returning to the questions proposed in Section 2, this section will attempt to answer these questions based on the evidence provided in Section 5.

Is there a relation between the correlation length scales found in this study and GCM equilibrium sensitivity? In this study there was no discernible relation found between a model's correlation length scale and its equilibrium sensitivity. This does

not mean that a relationship definitely does not exist, only that the procedures used in this study did not uncover one. Perhaps at even longer time averages (50-years, 100-years or even 1,000-years) the relation would be evident. Unfortunately, no current GCMs have picontrol runs that have a long enough time scale to study this.

Do the second moment statistics of recent GCMs behave similar to observations and Reanalysis? Because the observational network was too spars in both time and space to have meaningful results in this study, Reanalysis data was included in its place. Although some of the drawback in the observational network "leak" into the Reanalysis data it is a complete dataset in both space and time which means fewer assumptions must be made while doing calculations.

In general the GCMs do not behave similar to Reanalysis with a few exceptions. Most models underestimate the correlation structure as compared to the NCEP - REA except for the GFDL - CM3 in North America, Antarctica and the South Pacific Ocean.

Do the second moment statistics of recent GCMs behave according to stochastic EBM theory? At each site there are models that evolve in time exactly as a stochastic EBM would. This means that as the time average gets longer the e-folding correlation distance increases, with larger increases happening over the ocean. However there are also models at each site whose e-folding distances show no change at longer time averages. No generalization can be made for all models for this question.

Do the second moment statistics of recent GCMs agree with each other? With only a few exceptions it can be said that the second moment statistics of the GCMs used in this study do not agree with each other. Neither the magnitude of e-folding distances, evolution of these distances at various time averages or structure of the calculated correlation functions is consistent between the models.

6.3 Future Research

Although this study did not attempt to describe the reason for differences in the models, continuing this research could give insight as to why the second-moment statistics of models differ so greatly. First, since the open Southern Hemisphere oceans showed the greatest differences between the models, it would stand to reason that the treatment of the mixed-layer and deep oceans as well as mixing between the two affect results greatly. Investigation into how different mixing schemes are employed and the various atmosphere-ocean coupling methods could hold the greatest value. Additionally, various grid resolutions should be tested within the same model to see how a finer or courser resolution affects results and different variables (pressure, precipitation, wind, etc.) should be included in a multivariate test similar to the one described in this paper. Also, analyzing upper air temperature data would yield different results due to the lack of diffusive damping that is present at the surface. Likely there would be oscillations not seen in this study. Repeating the methods used in this study on model data at shorter time averages (e.g. hourly, daily or weekly) would most likely show even larger differences among the models than was shown in this study.

6.4 Recommendations

There are opportunities for understanding the Earth's dynamic climate system by developing simple stochastic energy balance models. In Isaac Held's 2005 essay *The Gap between Simulation and Understanding in Climate Modeling* he highlighted the need for a model hierarchy in which the gap between comprehensive models and more idealized models is less biased towards complexity. What does it mean, after all, to understand a system as complex as the climate, when we cannot fully understand idealized nonlinear systems with only a few degrees of freedom [7]?

REFERENCES

[1] C. E. Buell. Correlation Functions for Wind and Geopotential on Isobaric Surfaces. *Journal of Applied Meteorology*, 11:51–59, February 1972.

[2] M. G. Bulmer. *Principles of statistics*. New York : Dover Publications, New York, 1979.

[3] Tilmann Gneiting. Correlation functions for atmospheric data analysis. *Quarterly Journal of the Royal Meteorological Society*, 125(559):2449–2464, 1999.

[4] R. F. Gunst. Estimating spatial correlations from spatial- temporal meteorological data. *Journal of Climate*, 8(10), 1995.

[5] James Hansen and Sergej Lebedeff. Global trends of measured surface air temperature. *Journal of Geophysical Research: Atmospheres*, 92(D11):13345–13372, 1987.

[6] J. W. Hardin, G. R. North, and S. S. Shen. Minimum error- estimates of global mean temperature through optimal arrangement of gauges. *Environmetrics*, 3(1):15–27, 9203.

[7] I. M. Held. The Gap between Simulation and Understanding in Climate Modeling. *Bulletin of the American Meteorological Society*, 86:1609–1614, November 2005.

[8] Gareth S. Jones, Peter A. Stott, and Nikolaos Christidis. Attribution of observed historical near surface temperature variations to anthropogenic and natural causes using cmip5 simulations. *Journal of Geophysical Research: Atmospheres*, 118(10):4001–4024, 0527.

[9] Paul R. Julian and H. J. Thiebaux. On some properties of correlation functions used in optimum interpolation schemes. *Monthly Weather Review*, 103(7):605–616, 7507.

[10] Ruvim Leizerovich Kagan. *Averaging of meteorological fields*. Dordrecht; Boston: Kluwer Academic Publishers, Dordrecht ; Boston, 1997.

[11] K. Y. Kim and G. R. North. Surface temperature fluctuations in a stochastic climate model. *Journal of Geophysical Research: Atmospheres*, 96:18573–18580, 9101.

[12] K. Y. Kim, G. R. North, and G. C. Hegerl. Comparisons of the second- moment statistics of climate models. *Journal of Climate*, 9(9):2204–2221, 9609.

[13] R. Kistler. The ncep-ncar 50-year reanalysis: Monthly means cd-rom and documentation. *Bull. Am. Meteorol. Soc.*, 2001.

[14] G. R. North, S. S. Shen, and J. W. Hardin. Estimation of the global mean temperature with point gauges. *Environmetrics*, 3(1):1–14, 9203.

[15] G. R. North, J. Wang, and M. G. Genton. Correlation models for temperature fields. *Journal of Climate*, 24(22):5850–5862, 1115.

[16] R. O. Weber and P. Talkner. Some remarks on spatial correlation- function models. *Monthly Weather Review*, 121(9):2611–2617, 9309.

APPENDIX A

ADDITIONAL FIGURES

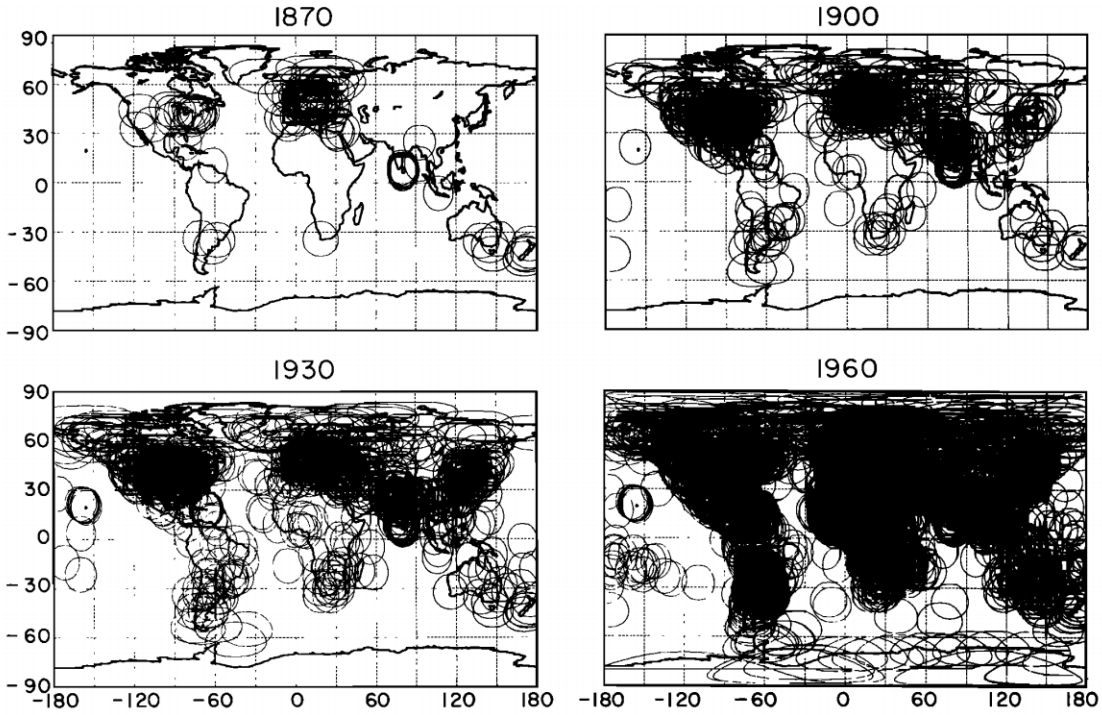
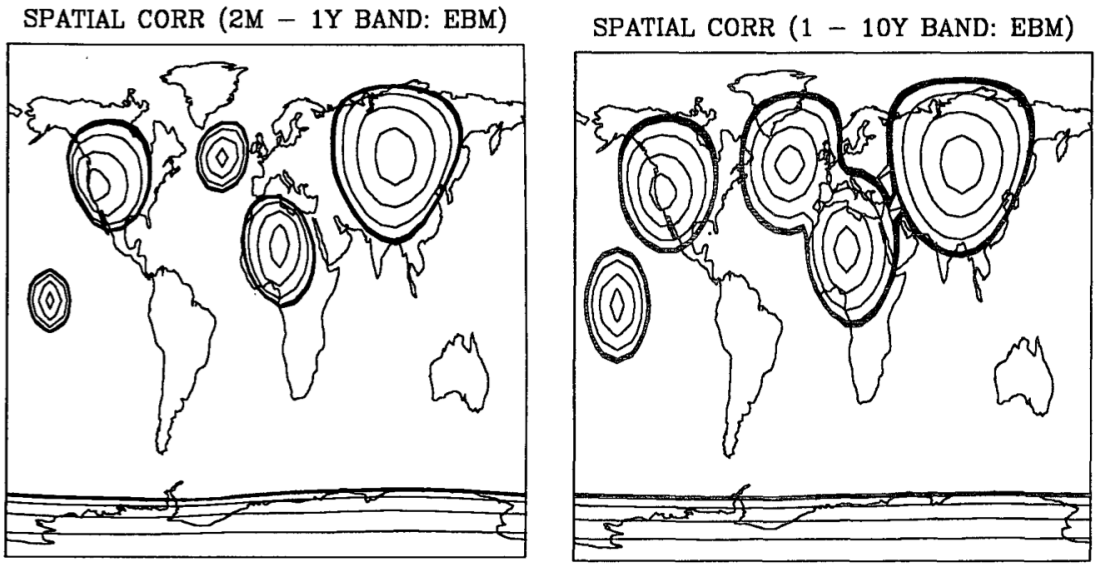


Figure A.1: Global distribution of meteorological stations with surface air temperature records for the four indicated dates. A circle of 1200-km radius is drawn around each station. Taken from [5].



(a) **(b)**

Figure A.2: Spatial correlation functions at six selected sampling points in the 2-mo to 1-yr (a) and 1-yr to 10-yr (b) frequency bands. The test points are (50°N, 90°E), (40°N, 120°W), (20°N, 10°E), (0°, 150°W), (90°S, 0°). The contour lines are at 0.9, 0.7, 0.5, and e^{-1} (thick contour). Taken from [12].