

University of California Santa Cruz

Comparison of Climate Model Simulations

Master of Science
in
Statistics and Applied Mathematics

by
Sarah Jarvis

May 2019

The thesis of Sarah Jarvis is approved by:

Professor Bruno Sansò

Professor Raquel Prado

Contents

List of Figures	i
Abstract	iv
1	1
1.1 Introduction	1
1.2 Data	2
1.2.1 Climate model simulations	2
1.2.2 Observational Products	2
2	5
2.1 Statistical Models	5
2.1.1 Baseline and seasonal temperature estimation	5
2.1.2 Evolution Variance Specification	7
2.1.3 Assessment of Internal Variability	7
2.1.4 Comparing internal variability using total variation distance	8
3	9
3.1 Results	9
3.1.1 CanCM4 Results	9
3.1.2 HadCM3 Results	13
3.1.3 GFDL Results	14
3.1.4 Comparing Models	16
4	22
4.1 Conclusions	22

List of Figures

1.1 NCEP. ERA40. Decadal. Historical. Control. 10-year time series data of monthly-mean, spatially averaged 2 meter surface temperature from CancM4. X-axis is year, y-axis is average temperature (K). The different line types represent three replicates from each simulation. We plot NCEP for reference. Line colors represent simulation or observational product.	3
3.1 NCEP. ERA40. Decadal. Historical. Control. Baseline temperature estimates for CanCM4. Colors represent simulation types or observational products. Shaded regions represent 95% posterior intervals. Left column: Global and Northern Hemisphere. Right column: Tropical and Southern Hemisphere.	10
3.2 NCEP. ERA40. Decadal. Historical. Control. Annual and semi-annual posterior amplitude samples for CanCM4. Colors represent simulation type or observational product. Whiskers indicate maximum and minimum values while the boxes indicate the 95% posterior intervals.	11
3.3 NCEP. ERA40. Decadal. Historical. Control. CanCM4 AR \log_{10} spectra normalized with respect to white-noise with 95% posterior intervals. The x-axis is labeled at select years ($2\pi/12\omega$). Colors represent simulation type or observational product.	12
3.4 NCEP. ERA40. Decadal. Historical. Control. TVD with white-noise as the reference. Colors represent simulation type or observational product. Whiskers indicate the maximum and minimum values while boxes indicate 95% posterior intervals.	12
3.5 NCEP. ERA40. Decadal. Historical. Control. Baseline temperature estimates for HadCM3. Colors represent simulation types or observational products. Shaded regions represent 95% posterior intervals. Left column: Global and Northern Hemisphere. Right column: Tropical and Southern Hemisphere.	13
3.6 NCEP. ERA40. Decadal. Historical. Control. Annual and semi-annual posterior amplitude samples for HadCM3. Colors represent simulation type or observational product. Whiskers indicate maximum and minimum values while the boxes indicate the 95% posterior intervals.	14

3.7 NCEP. ERA40. Decadal. Historical. Control. HadCM3 AR \log_{10} spectra normalized with respect to white-noise with 95% posterior intervals. Colors represent simulation type or observational product. The x-axis is labeled at select years($2\pi/12\omega$).	15
3.8 NCEP. ERA40. Decadal. Historical. Control. TVD for HadCM3 with white-noise as the reference. Colors represent simulation type or observational product. Whiskers indicate the maximum and minimum values while boxes indicate 95% posterior intervals.	15
3.9 NCEP. ERA40. Decadal. Historical. Control. Baseline temperature estimates for GFDL. Colors represent simulation types or observational products. Shaded regions represent 95% posterior intervals. Left column: Global and Northern Hemisphere. Right column: Tropical and Southern Hemisphere.	16
3.10 NCEP. ERA40. Decadal. Historical. Control. Annual and semi-annual posterior amplitude samples for GFDL. Colors represent simulation type or observational product. Whiskers indicate maximum and minimum values while the boxes indicate the 95% posterior intervals.	17
3.11 NCEP. ERA40. Decadal. Historical. Control. GFDL AR \log_{10} spectra normalized with respect to white-noise with 95% posterior intervals. The x-axis is labeled at select years($2\pi/12\omega$). Colors represent simulation types or observational products.	17
3.12 NCEP. ERA40. Decadal. Historical. Control. TVD for GFDL with white-noise as the reference. Whiskers indicate the maximum and minimum values while boxes indicate 95% posterior intervals. Colors represent simulation types or observational products.	18
3.13 NCEP. ERA40. Decadal. Historical. Control. Baseline temperature estimates for all models and observations. Colors represent simulation types or observational products. Shaded regions represent 95% posterior intervals.	19
3.14 NCEP. ERA40. Decadal. Historical. Control. Annual and semi-annual posterior amplitude samples for all models and observations. Colors represent simulation or observational product. Whiskers indicate maximum and minimum values while the boxes indicate the 95% posterior intervals.	20
3.15 NCEP. ERA40. Decadal. Historical. Control. AR \log_{10} spectra normalized with respect to white-noise with 95% posterior intervals for all models and observations. The x-axis is labeled at select years($2\pi/12\omega$). Colors represent simulation type or observational product.	20
3.16 NCEP. ERA40. Decadal. Historical. Control. TVD for all models and observations with white-noise as the reference. Whiskers indicate the maximum and minimum values while boxes indicate 95% posterior intervals. Colors represent simulation type or observational product.	21

Abstract. We compare large-scale spatial averages of surface temperature in climate model simulations and in observational products with Dynamic Linear Models (DLM) and autoregressive (AR) models. We draw from the methods presented in Barata, R., Prado, R., & Sansò, B. (2019), where this approach was used to compare MIROC5 to observational products, and compare three additional climate models; CanCM4, HadCM3, and GFDL. The output from the models analyzed in this paper are from phase 5 of the Coupled Model Intercomparison Project (CMIP5): preindustrial control runs, simulations of historic climate change, and decadal predictions. The variable of interest is monthly surface temperature averaged over four spatial domains (global, tropical, Northern Hemisphere, and Southern Hemisphere) from January 1981 to December 2012. The observational products (NCEP-2 and ERA-Interim) are two different reanalysis data sets. Our results show differences in seasonal temperature changes both in model-versus-model and model-versus-data comparisons. We were able to evaluate the climate models with the data-driven approach described in this project.

Chapter 1

1.1 Introduction

The climate model simulations analyzed here have relevance when studying human influence on global climate change. It is important for the models to accurately represent key features of present-day climate because they can help identify any influences on historical climate as well as try to predict any future climate changes. In this paper we draw on a statistical model-based approach presented in Barata, R., Prado, R., & Sansò, B. (2019) to compare the observational record with three additional models each with three different types of CMIP5 simulations. The data-driven model diagnostic method will show if there are statistically significant differences in aspects of the variability between the model simulations and the observational record, between the different models, and within the different types of simulation. The methods used present a Bayesian approach to model evaluation.

The climate models analyzed here were developed in three different countries. We have the Fourth Generation Coupled Global Climate Model developed by the Canadian Centre for Climate Modelling and Analysis (CanCM4), the Hadley Centre Coupled Model, version 3 (HadCM3) developed at the Hadley Centre in the United Kingdom, and a coupled ocean-atmosphere general circulation climate model developed by the Geophysical Fluid Dynamics Laboratory (GFDL) in the United States National Oceanic and Atmospheric Administration Office of Oceanic and Atmospheric Research. The analysis for the Model for Interdisciplinary Research on Climate (MIROC5) can be found in the Barata, R., Prado, R., & Sansò, B. (2019) paper previously mentioned. We examine three types of model simulation: 1) decadal prediction, run for 30 years and initialized in 1980 from a specific observational state; 2) historical run, uninitialized but driven by historical forcings both natural and from human activity; and 3) control run with no changes in external forcings.

Each model has data for these three simulation types although GFDL does not have a 30 year decadal run, so we only analyze the historical and control run for this model. We focus on monthly-mean surface temperature taken 2 meters above ground level over four spatial domains: global, tropical, Northern Hemisphere, and Southern Hemisphere. This allows us to look at spatial differences between the models as well as internal variability.

The approach used extracts changes in baseline temperature, seasonality, and estimates of internal variability. We compare between the three models and model simulations to the corresponding components in the observational products. The goal being to spot whether model-versus-model and model-versus-observed differences are significant. We also investigate whether there are differences between the spectral properties of the models and simulations.

This paper is organized in the following way. Section 2 describes the models as well as the observational products used to compare. The modelling approach is described in Section 3, introducing the DLM that is used to estimate the baseline and seasonal components and the AR model used to estimate the internal variability of each simulation. The results of applying the DLM and AR models are shown in Section 4 over each climate model, simulation type, and spatial domain.

1.2 Data

1.2.1 Climate model simulations

CMIP5 is a collaborative international modelling plan designed to gain climate change insight. It consists of many simulations with several different climate models. We analyze the decadal, historical, and control runs for CanCM4, HadCM3, and only the historical and control runs for GFDL. The decadal and historical runs are implemented with external forcings, used to reflect the effects of human and natural influences on the climate system. A more detailed description of the three types of climate simulations can be found in Barata, R., Prado, R., & Sansò, B. (2019).

Each model (CanCM4, HadCM3, GFDL) has 10 replicates of each simulation type. We only have historical simulations from January 1981 to December 2005 so to obtain a 30 year run we concatenated the simulations corresponding to the predicted Representative Concentration Pathway (RCP). This method did not produce any discontinuities or jumps in the time series data. The control runs do not have any direct correspondence with actual time and are typically used to simulate long periods of internal variability. We extract 10 non-overlapping monthly-mean temperature series from the control run data, thus creating 30 year time series of 10 replicates comparable to the other simulation types. Figure 1.1 shows the time series of the three simulations as well as the observational products over the four spatial domains for CanCM4. The data for HadCM3 and GFDL look similar. Further details of CanCM4, HadCM3, and GFDL can be found in Chylek et al. (2011), Valdes et al. (2017), and Delworth et al. (2012) respectively.

1.2.2 Observational Products

In order to assess the accuracy of the climate models, we compare them to reanalysis data sets which we will refer to as observations. The reanalyses rely on numerical weather prediction (NWP) and use

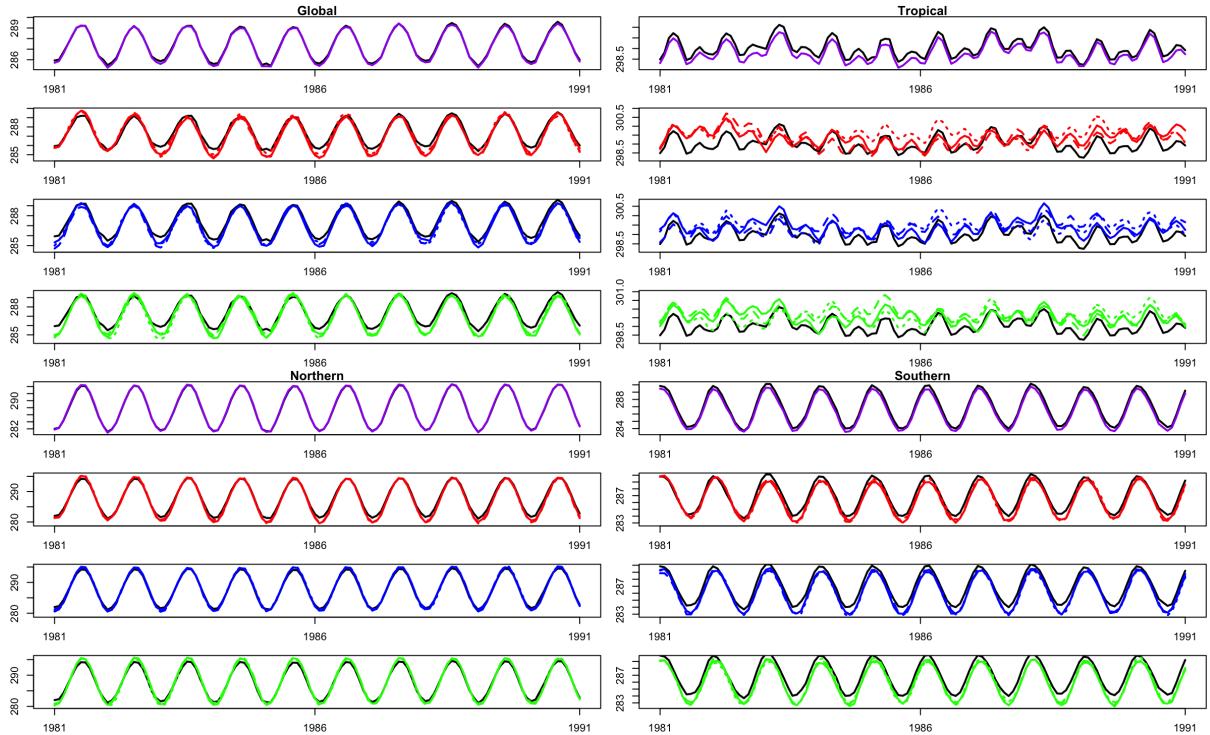


Figure 1.1: NCEP. ERA40. Decadal. Historical. Control. 10-year time series data of monthly-mean, spatially averaged 2 meter surface temperature from CancM4. X-axis is year, y-axis is average temperature (K). The different line types represent three replicates from each simulation. We plot NCEP for reference. Line colors represent simulation or observational product.

raw satellite observations to generate assimilated data. We use two reanalysis products as our observations for this analysis: Version 2 of the reanalysis performed by the National Centers for Environment Prediction (NCEP-2) and ERA-Interim (ERA-I) generated by the European Centre for Medium-Range Weather Forecasts (ECMWF). We only consider data from these observational products for our time frame of interest, the data can be seen in Figure 1.1. Refer to Fujiwara et al. (2017) for a more thorough explanation of the reanalyses.

All temperature data used in this analysis is available online and is characterized by longitude and latitude over the whole globe. We calculate the spatial averages over the four domains: global, tropical, Northern Hemisphere, and Southern Hemisphere. These domains are bounded by the following latitudes respectively: 90°S to 90°N, 20°S to 20°N, 0° to 90°N, 90°S to 0°. Figure 1.1 shows the first 10 years of the spatially averaged data for which the following statistical analysis focuses on.

Chapter 2

2.1 Statistical Models

We use Dynamic Linear Models (DLMs) to decompose each temperature time series into a baseline and a seasonal component. The baseline component is intended to capture long-term changes in temperature and the seasonal component captures annual and semi-annual temperature cycles. We extract these two components and compare the results between models and observations. We then obtain a residual time series from the DLM that represents the internal climate variability and fit an autoregressive (AR) model to investigate the spectral densities and thus compare the internal variability of the models and observations. Section 3.1 will describe the strategy for extracting the baseline and seasonality component with a multivariate DLM, which is required for dealing with multiple realizations of the model simulations, section 3.2 will discuss the strategy for selecting a discount factor, section 3.3 will discuss fitting the AR model to the DLM residuals. To compare the spectra of the DLM residuals we will use the total variation difference (TVD) presented in section 3.4.

2.1.1 Baseline and seasonal temperature estimation

DLMs are a Bayesian modeling approach used to analyze non-stationary time series. The approach to estimating time-series baseline and seasonality components was adapted from West and Harrison (1999) and Prado and West (2010).

First consider a univariate model with just one realization of temperature data. Let y_t denote the univariate average temperature at time t for $t = 1, \dots, T$. In our case, T is 360 for our 30-year time series. The model used to decompose the baseline and seasonality components is,

$$y_t = \eta_{1,t} + \sum_{k=1}^K \alpha_{1,t}^k + \nu_t, \quad \nu_t \sim N(0, V),$$

where $\eta_{1,t}$ is the baseline temperature, $\alpha_{1,t}^k$ are the seasonal components for harmonics $k = 1, \dots, K$ of a fundamental period p , and V is the unknown variance. (West and Harrison, 1999). We assume that the errors are independent. We then assume that the baseline component has structure,

$$\begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_{1,t-1} \\ \eta_{2,t-1} \end{pmatrix} + \boldsymbol{\omega}_t^\eta, \quad \boldsymbol{\omega}_t^\eta \sim N_2(\mathbf{0}, V\mathbf{W}_t^\eta)$$

We assume the system evolution error vectors $\boldsymbol{\omega}_t^\eta$ are independent over time. We use $p = 12$ months and include harmonics $1, \dots, K$ with $K = 4$ to represent the annual, semi-annual, tri-annual, and quarterly cycles. Each harmonic k has the Fourier form representation of cyclical functions,

$$\begin{pmatrix} \alpha_{1,t}^k \\ \alpha_{2,t}^k \end{pmatrix} = \begin{pmatrix} \cos(\frac{2\pi}{p}k) & \sin(\frac{2\pi}{p}k) \\ -\sin(\frac{2\pi}{p}k) & \cos(\frac{2\pi}{p}k) \end{pmatrix} \begin{pmatrix} \alpha_{1,t-1}^k \\ \alpha_{2,t-1}^k \end{pmatrix} + \boldsymbol{\omega}_t^{\alpha,k}, \quad \boldsymbol{\omega}_t^{\alpha,k} \sim N_2(\mathbf{0}, V\mathbf{W}_t^{\alpha,k})$$

Where the k th seasonal evolution matrix is $\mathbf{G}^{\alpha,k} = \begin{pmatrix} \cos(\frac{2\pi}{p}k) & \sin(\frac{2\pi}{p}k) \\ -\sin(\frac{2\pi}{p}k) & \cos(\frac{2\pi}{p}k) \end{pmatrix}$ and we assume that $\boldsymbol{\omega}_t^{\alpha,k}$ are independent over time and independent of $\boldsymbol{\omega}_t^\eta$. We write the model as a hierarchy with an observation equation and a system equation,

$$y_t = \mathbf{F}'\boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim N(0, V)$$

$$\boldsymbol{\theta}_t = \mathbf{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N_n(\mathbf{0}, \mathbf{W}_t)$$

where $\mathbf{G} = \text{blockdiag}(\mathbf{G}^\eta, \mathbf{G}^{\alpha,1}, \dots, \mathbf{G}^{\alpha,K})$ and $\mathbf{W}_t = \text{blockdiag}(\mathbf{W}_t^\eta, \mathbf{W}_t^{\alpha,1'}, \dots, \mathbf{W}_t^{\alpha,K'})$. The state vector $\boldsymbol{\theta}_t = (\eta_{1,t}, \eta_{2,t}, \alpha_{1,t}^1, \alpha_{2,t}^1, \dots, \alpha_{1,t}^K, \alpha_{2,t}^K)$, and $\mathbf{F}' = (\mathbf{F}^{\eta'}, \mathbf{F}^{\alpha,1'}, \dots, \mathbf{F}^{\alpha,K'})$ with $\mathbf{F}^{\cdot\cdot\cdot'} = (1, 0)$ for all components.

For R realizations of model simulations, we consider a multivariate DLM. Let $\mathbf{Y}_t = \mathbf{F}'\boldsymbol{\theta}_t + \boldsymbol{\nu}_r$ denote the spatially averaged temperature for each realization. So $\mathbf{Y}_t = (y_{t,1}, \dots, y_{t,R})'$ is a vector of R realization values and $\boldsymbol{\nu}_r = (\nu_{t,1}, \dots, \nu_{t,R})'$, a vector of R i.i.d. error terms. Now we have the following hierarchical model,

$$\mathbf{Y}_t = \mathbf{F}'\boldsymbol{\theta}_t + \boldsymbol{\nu}_r, \quad \boldsymbol{\nu}_r \sim N_R(0, V\mathbf{I}_R)$$

$$\boldsymbol{\theta}_t = \mathbf{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N_n(0, V\mathbf{W}_t)$$

where \mathbf{F}' is a $R \times n$ dynamic regression matrix with identical rows $\mathbf{F}'_r = (\mathbf{F}^{\eta'}, \mathbf{F}^{\alpha,1'}, \dots, \mathbf{F}^{\alpha,K'})$ for $r = 1, \dots, R$. The posterior distributions for $\boldsymbol{\theta}_t$ can be updated using filtering and backwards smoothing (West and Harrison, 1999). We choose the following conjugate priors: for the initial state vector, $\boldsymbol{\theta}_0 \sim N_n(\mathbf{m}_0, V\mathbf{C}_0)$ for the unknown constant, $V \sim IG(n_0/2, n_0S_0/2)$ with $\mathbf{m}_0 = (285, 0, \dots, 0)'$, $\mathbf{C}_0 = \text{diag}(5, 2 \times 10^{-6}, 5, 1, \dots, 1)$, $n_0 = 1$, $S_0 = 0.01$.

2.1.2 Evolution Variance Specification

If the posterior variance of the state vector θ_{t-1} is denoted as $Var(\theta_{t-1} | \mathbf{Y}_{t:(t-1)}) = \mathbf{C}_{t-1}$, the sequential updating equations produce prior variance $\mathbf{R}_t = Var(\theta_t | \mathbf{Y}_{t:(t-1)}) = \mathbf{G}\mathbf{C}_{t-1}\mathbf{G}' + \mathbf{W}_t$. We can then write $\mathbf{R}_t = \mathbf{G}\mathbf{C}_{t-1}\mathbf{G}'/\delta \geq \mathbf{G}\mathbf{C}_{t-1}\mathbf{G}'$, where $0 < \delta \leq 1$ is a discount factor. This results in an evolution variance matrix $\mathbf{W}_t = \frac{1-\delta}{\delta} \mathbf{G}\mathbf{C}_{t-1}\mathbf{G}'$. If $\delta = 1$ the model is static and the parameters do not change over time (West and Harrison, 1999).

We use one discount factor for the baseline and one for the seasonal component (δ_{base} and δ_{seas}). To select the optimal value for the discount factor, we maximize the likelihood of the one-step-ahead forecast distribution over a grid of $(\delta_{base}, \delta_{seas})$ values in $(0.9, 1] \times (0.9, 1]$. We optimize the discount factor for the control runs for all simulation types in that region, choosing the same maximum value over all models. For our four spatial domains global, tropical, Northern and Southern Hemispheres, we obtained discount factors of $\delta = (0.94, 0.91, 0.99, 0.99)$, respectively. This approach allows for a clear comparison within each region and over the models.

2.1.3 Assessment of Internal Variability

As well as comparing the baseline and seasonality, we are also interested in comparing the internal variability between models. We can obtain the DLM residuals and this time series represents the natural internal climate variability. The process of fitting the residuals with an AR model is described as follows.

To obtain the DLM residuals we subtract the posterior mean of the univariate DLM at time t from the model or observational product data. We will refer to the residuals as $z_t = y_t - \mathbf{F}'\hat{\theta}_t$, where $\hat{\theta}_t$ represents the posterior mean mentioned above. To capture the temporal structure of z_t , we fit an autoregressive model of order q , AR(q):

$$z_t = \sum_{j=1}^q \phi_j z_{t-j} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2)$$

where $\phi = (\phi_1, \dots, \phi_q)$ is a vector of AR coefficients and ϵ_t are independent over time.

We need to extend this model to the multivariate case to reflect all R realizations for each model. We then have $z_{t,r} = \sum_{j=1}^q \phi_j z_{t-j,r} + \epsilon_{t,r}$. Each error term $\epsilon_{t,r} \sim N(\mathbf{0}, \sigma^2 I_R)$ is independent and normally distributed. We choose conjugate priors $\phi \sim N_q(\mathbf{0}, \mathbf{I}_q)$ and $\sigma^2 \sim IG(1, 0.01)$.

We make the assumption that the order q can vary over spatial domains but in one domain, we will use the same order over all simulation types. We choose the order that maximizes the log-predictive likelihood using the univariate time series of residuals for each simulation type, each domain, each individual realization. Similar to choosing the discount factor, we use the maximum order over every model. For global, tropical, Northern and Southern hemispheres, we obtain $q = (4, 7, 5, 5)$ respectively. Using the same order for every model ensures that the results are comparable within each domain.

We obtain the $AR(q)$ coefficients ϕ and fit the spectral density $f(\omega) = \frac{\sigma^2}{2\pi|1-\phi_1 e^{-i\omega} - \dots - \phi_q e^{-iq\omega}|^2}$. We can compare spectra with respect to differences in AR coefficients by normalizing the density with respect to the white-noise spectrum, $\sigma^2/2\pi$.

2.1.4 Comparing internal variability using total variation distance

To evaluate differences between the spectral densities we use the total variation distance (TVD). The TVD of two normalized spectral densities $f^*(\omega) = f(\omega)/\int_{\Omega} f(\omega)d\omega$ and $g^*(\omega) = g(\omega)/\int_{\Omega} g(\omega)d\omega$ is defined as $TVD(f^*, g^*) = 1 - \int_{\Omega} \min\{f^*(\omega), g^*(\omega)\} d\omega$. The distance measure takes values in $[0, 1]$ with 0 being the smallest possible discrepancy between spectra and 1 the largest. We first normalize the spectral densities by dividing their overall variances. We then compute the posterior distributions for the TVD values compared to a white-noise reference spectrum, which corresponds to a constant. The calculated TVD provides a measure of the difference between the variability in a given time series and white noise, used as a reference. By comparing such measures for a number of time series we have an assessment of their relative differences.

Chapter 3

3.1 Results

We apply the previously described methods to the 30-year time series of monthly-mean spatially averaged temperature data from the three sets of CanCM4, HadCM3, and two sets of GFDL simulations. We then compare the results from the different models as well as to those from the two observational products.

3.1.1 CanCM4 Results

Figure 3.1 shows the 95% posterior intervals for the baseline components of CanCM4 that we obtained from the DLM. We can see that the control runs are relatively flat compared to the other simulation types. This should be consistent over each model since the control runs lack changes in external forcings. Within each region we see overall increases in temperature. We can see 1-2 years of cooling trends in the global and tropical regions for the decadal and historical runs around 1982 and 1991 when the volcanoes El Chichón and Pinatubo erupted. We do not see this trend as much in the Northern and Southern hemisphere because the discount factors are so close to 1. The decadal predictions have around the same temperature as the reanalysis products in 1981 which we would expect because they were initialized from observed data. We do see some post-initialization secular drift in the decadal runs, most easily seen in the Northern Hemisphere. The model-generated baseline temperatures are more or less consistent with the observational product, however in the tropics the models appear to slightly overestimate the trend.

Figure 3.2 shows the 95% intervals for the posterior seasonal amplitudes for $k = 1, 2$ (annual and semi-annual cycles, respectively). We estimated the amplitudes with the DLM previously described. We do not include the plots of harmonics $k = 3$ and $k = 4$ because they appear to be insignificant. We see significant differences between the annual amplitudes of the observations and the model. None of the annual amplitudes from the simulations appear to be consistent with the observations except some slight overlap in the Southern Hemisphere decadal run. We also see significant differences between

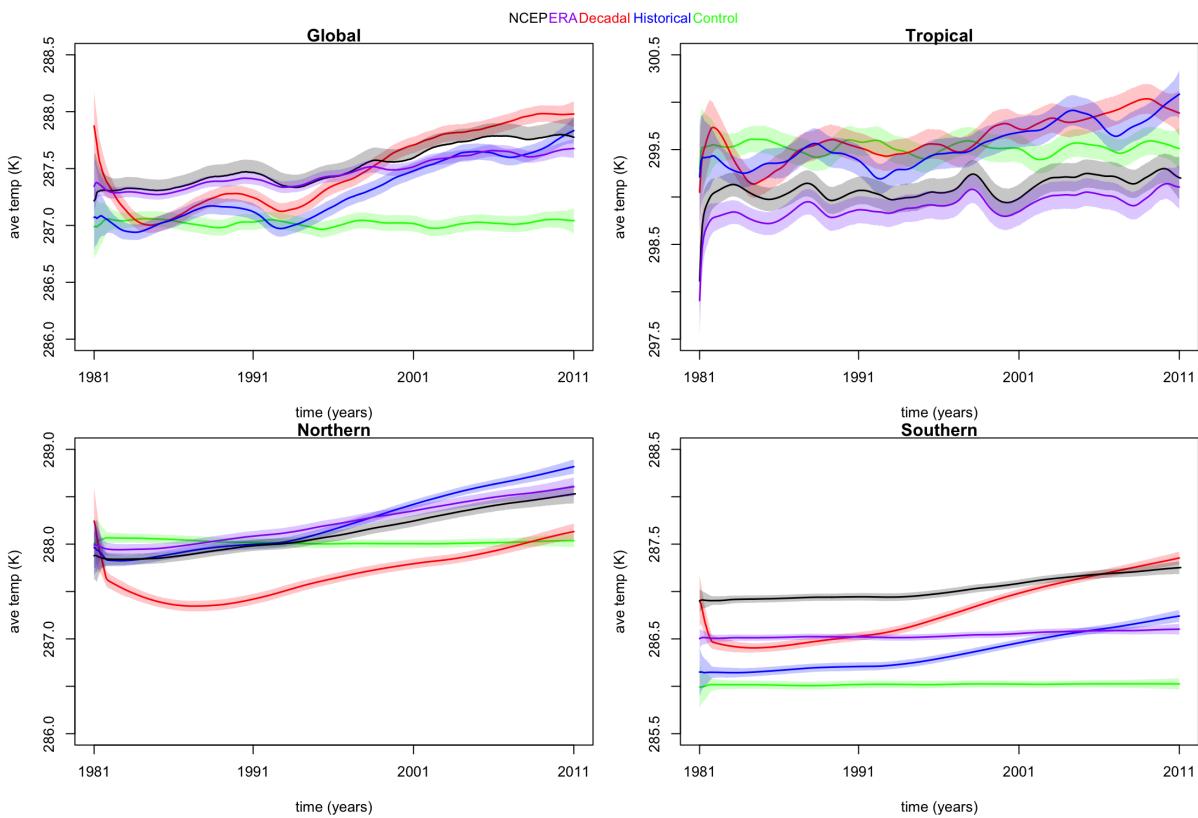


Figure 3.1: NCEP. ERA40. Decadal. Historical. Control. Baseline temperature estimates for CanCM4. Colors represent simulation types or observational products. Shaded regions represent 95% posterior intervals. Left column: Global and Northern Hemisphere. Right column: Tropical and Southern Hemisphere.

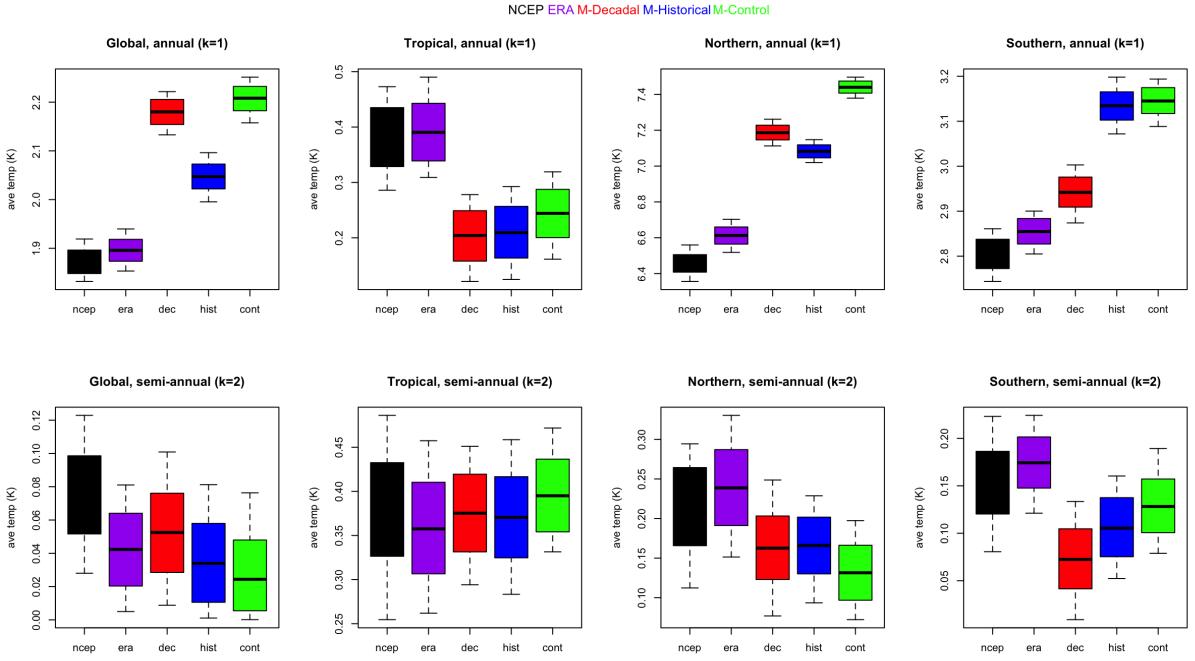


Figure 3.2: NCEP. ERA40. Decadal. Historical. Control. Annual and semi-annual posterior amplitude samples for CanCM4. Colors represent simulation type or observational product. Whiskers indicate maximum and minimum values while the boxes indicate the 95% posterior intervals.

the simulation types. Over the global domain, the historical run deviates from the rest while in the northern region, the control run differs, and in the south, the decadal run differs. The differences are not statistically significant in the tropics.

Figure 3.3 shows the 95% probability intervals on the posterior spectra, normalized with respect to white noise on the log scale. The spectral densities look smoothest in the domain with lower order q (i.e. global where $q=4$). The tropics has an order of 7 and is thus the least-smooth. We expect this because the tropics time series has the most noise. Within each region, the spectra from CanCM4 look similar but are slightly higher than the observation at the lowest frequencies.

The results from the TVD are shown in figure 3.4, which allows us to compare the spectra. The figure shows results for the comparison against a white-noise reference spectrum. Since none of the plots cover zero, we conclude that all observation and model time series are distinguishable from white-noise. For all spatial domains, the observations have lower TVD than the models suggesting that the models are farther from the white noise spectrum. The largest difference between observation and model is in the tropics.

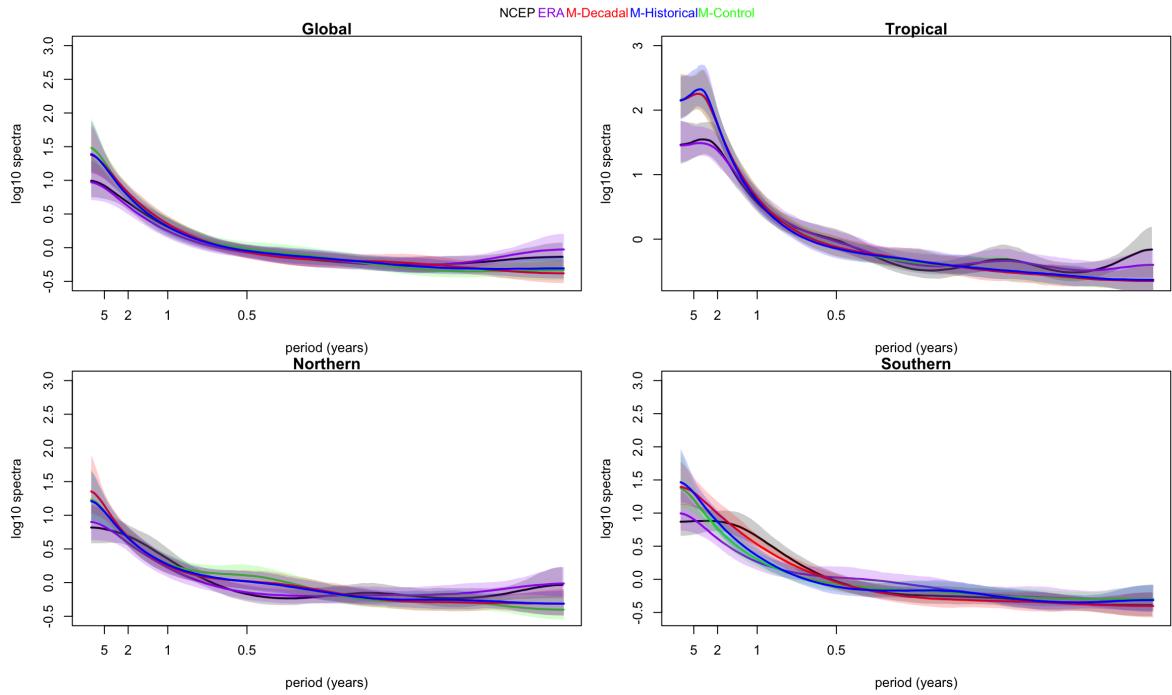


Figure 3.3: NCEP, ERA40, Decadal, Historical, Control, CanCM4 AR \log_{10} spectra normalized with respect to white-noise with 95% posterior intervals. The x-axis is labeled at select years ($2\pi/12\omega$). Colors represent simulation type or observational product.

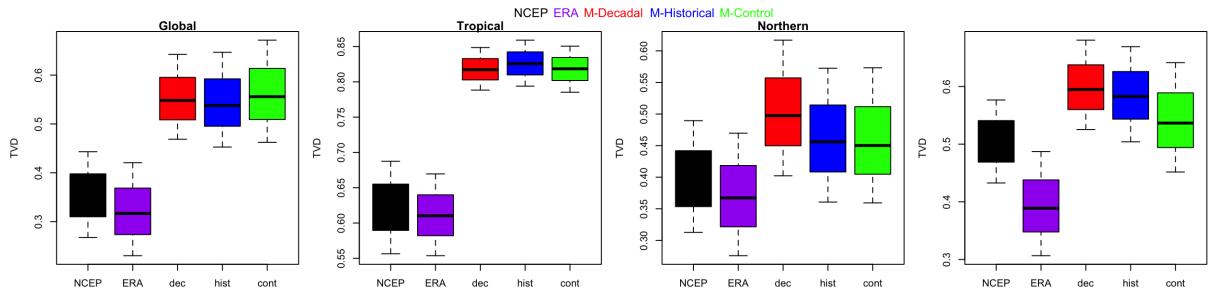


Figure 3.4: NCEP, ERA40, Decadal, Historical, Control, TVD with white-noise as the reference. Colors represent simulation type or observational product. Whiskers indicate the maximum and minimum values while boxes indicate 95% posterior intervals.

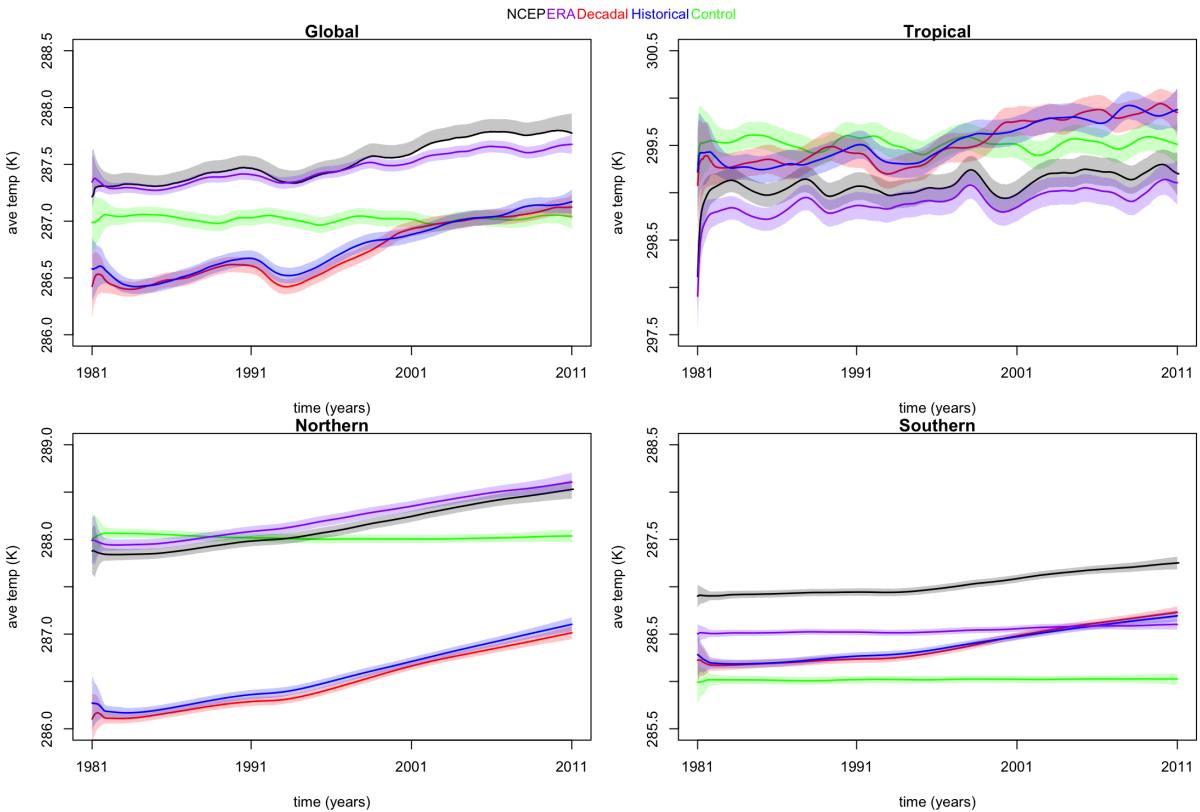


Figure 3.5: NCEP. ERA40. Decadal. Historical. Control. Baseline temperature estimates for HadCM3. Colors represent simulation types or observational products. Shaded regions represent 95% posterior intervals. Left column: Global and Northern Hemisphere. Right column: Tropical and Southern Hemisphere.

3.1.2 HadCM3 Results

We see the baseline components for HadCM3 in Figure 3.5. As we stated previously, we would expect the decadal predictions to have the same value as the observations in 1981. In every region except the tropics, the decadal value is significantly different than the observations. In the Northern Hemisphere, we can see significant model-versus-observations differences throughout the 30-year baselines. The post-initialization drift is not as extreme in this model and it appears that the model underestimates the temperature in every domain.

Figure 3.6 shows the posterior amplitudes for HadCM3. For the annual harmonic, the control run has significantly larger amplitudes in the global and Northern regions. There are no differences between model amplitudes in the tropics and Southern domains but we do see significant differences in model-versus-observation comparisons in every domain for $k = 1$. The semi-annual amplitudes have lower distributions for the control run in the global, Northern, and Southern domains.

Figure 3.7 illustrates the posterior spectra of HadCM3 normalized with respect to white-noise. We again see that the models have higher spectral values than the observations at low frequencies. It

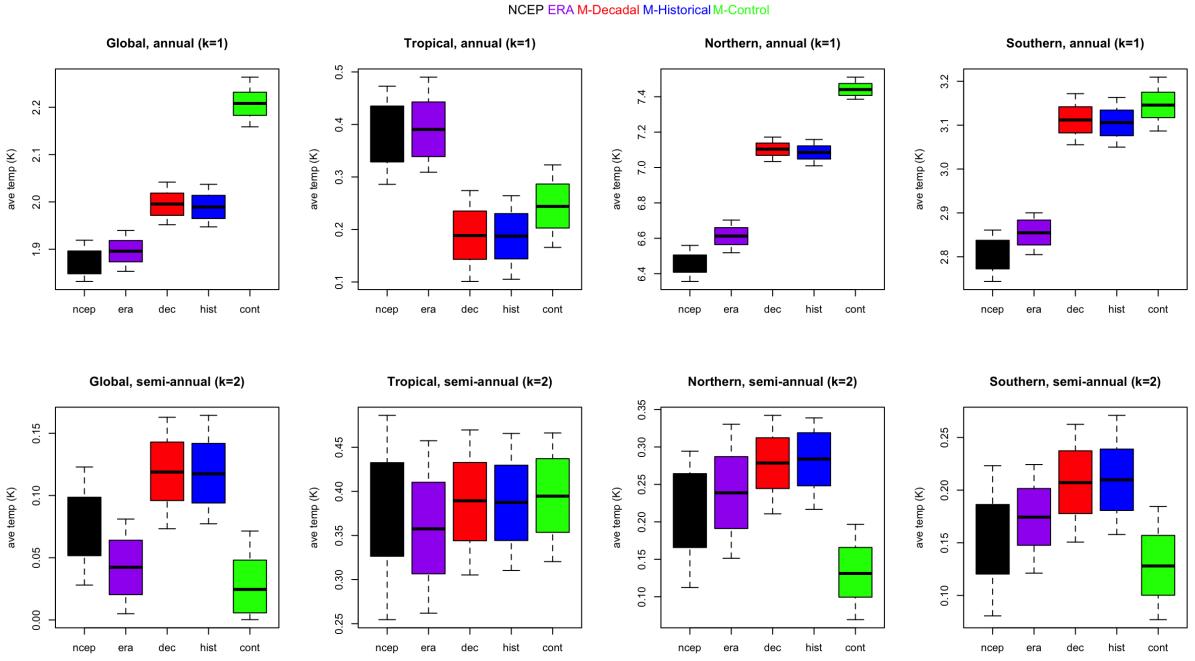


Figure 3.6: NCEP, ERA40, Decadal, Historical, Control. Annual and semi-annual posterior amplitude samples for HadCM3. Colors represent simulation type or observational product. Whiskers indicate maximum and minimum values while the boxes indicate the 95% posterior intervals.

appears that the spectral densities of each simulation type are consistent within the region.

We can see the TVD with white-noise as the reference in figure 3.8. In the global and tropical domain, we see significant model-versus-observation differences while in the other two regions the difference is less severe, but still apparent. There are no differences within model simulations and the models have a spectral density further from white-noise than the observations.

3.1.3 GFDL Results

Figure 3.9 illustrates the baseline components from the DLM for the GFDL model. As previously stated there is no data available for a 30-year decadal run, so this analysis was left out. We notice a slight pre-initialization drift in the historical runs. The control run appears to follow the observations well in the global and tropical domain while the control is over estimating the temperature in the Southern Hemisphere. The temperature drop after the eruption of Pinatubo is most easily seen in the historical run for the tropics, while it is almost unrecognizable globally.

Figure 3.10 represents the posterior amplitudes of the GFDL model as well as the observations. For the annual harmonics, we see significant differences in amplitudes between both simulations and model-versus-observation globally and in the Northern Hemisphere. The tropical domain shows no differences between amplitudes and the Southern Hemisphere shows differences between model-versus-

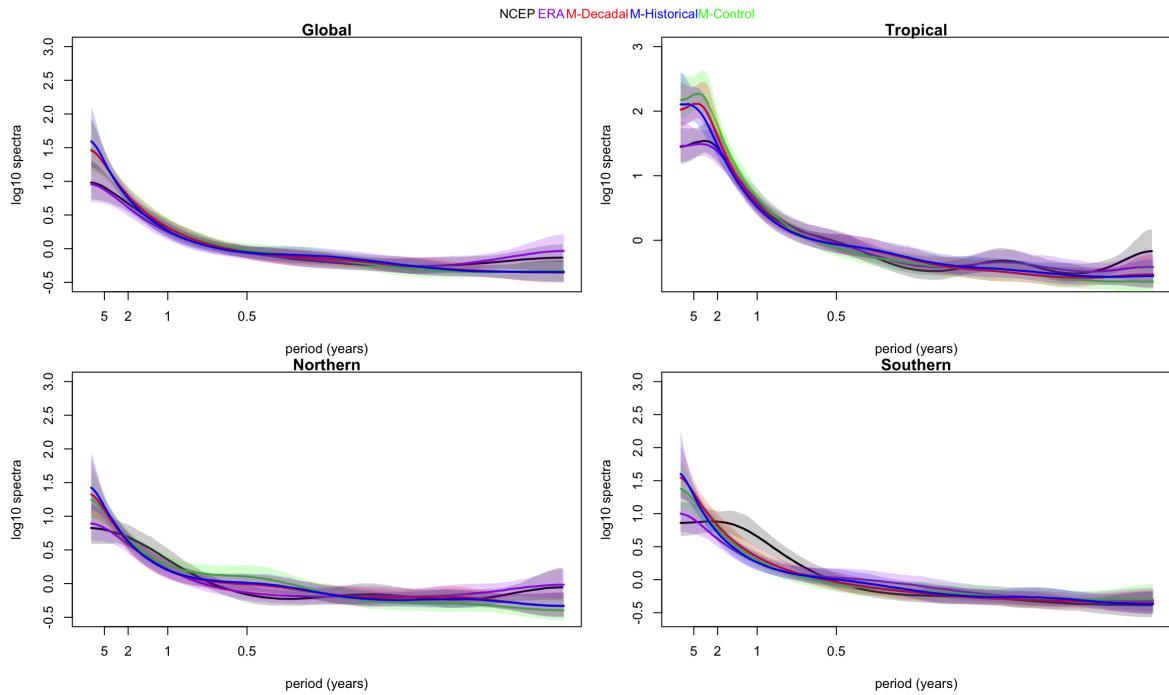


Figure 3.7: NCEP, ERA40, Decadal, Historical, Control. HadCM3 AR \log_{10} spectra normalized with respect to white-noise with 95% posterior intervals. Colors represent simulation type or observational product. The x-axis is labeled at select years($2\pi/12\omega$).

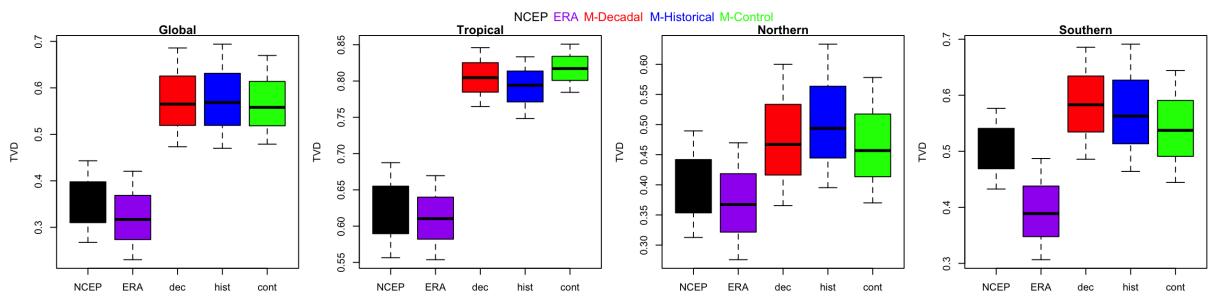


Figure 3.8: NCEP, ERA40, Decadal, Historical, Control. TVD for HadCM3 with white-noise as the reference. Colors represent simulation type or observational product. Whiskers indicate the maximum and minimum values while boxes indicate 95% posterior intervals.

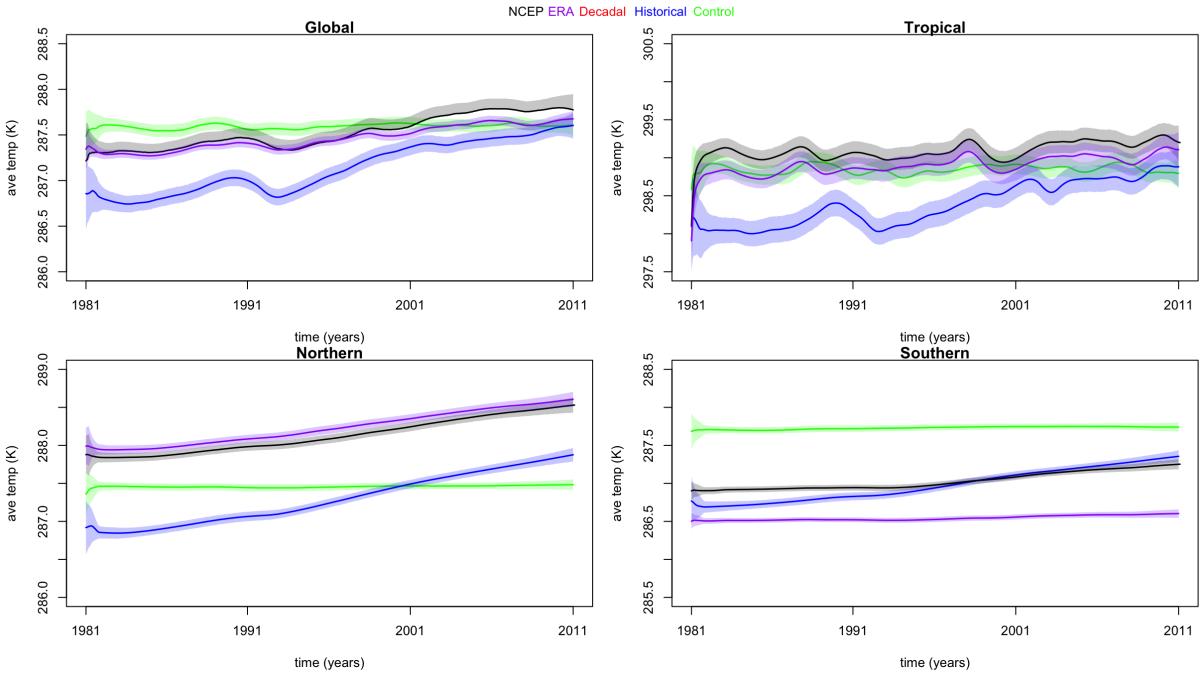


Figure 3.9: NCEP. ERA40. Decadal. Historical. Control. Baseline temperature estimates for GFDL. Colors represent simulation types or observational products. Shaded regions represent 95% posterior intervals. Left column: Global and Northern Hemisphere. Right column: Tropical and Southern Hemisphere.

observation, but not between simulations.

We can see the spectral densities for GFDL in figure 3.11. We see the same smoothness properties as with the other two models. The densities for the models are higher than they are for the observations at low frequencies. Again, the spectra for the three simulation types are very similar.

Figure 3.12 shows the TVD distributions for GFDL that we use to compare the spectral densities. There are model-versus-observation differences in TVD in every domain, but the differences are not as extreme in the Northern Hemisphere.

3.1.4 Comparing Models

Below, we summarize the previously stated results in a way that makes model-versus-model comparisons straightforward. Theoretically, there should not be significant differences between the models because they all represent the same climate system. Nevertheless, they represent the climate system with different physics, different numerical approaches to the solution of the equations that describe the system, different parameterizations for phenomena that happen at small scales, use different initializations and different hardware to run the models.

Figure 3.13 shows the baseline components of all model simulations (including MIROC5) as well as the observations in one plot. All control runs are relatively flat because they are not influenced by

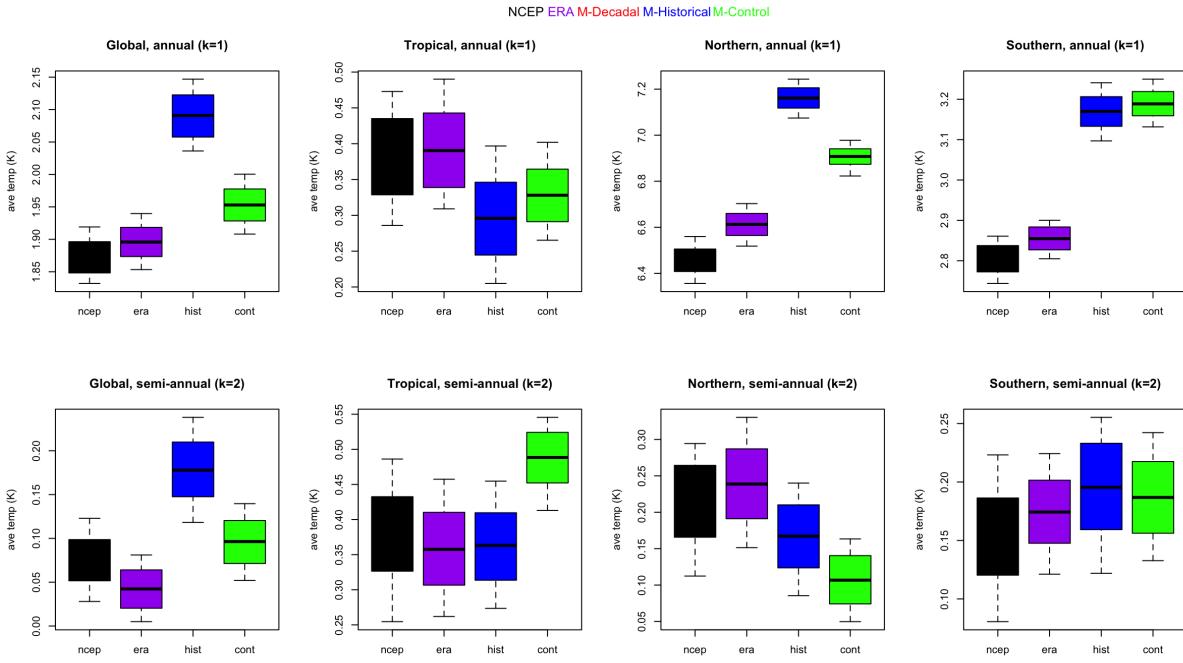


Figure 3.10: NCEP. ERA40. Decadal. Historical. Control. Annual and semi-annual posterior amplitude samples for GFDL. Colors represent simulation type or observational product. Whiskers indicate maximum and minimum values while the boxes indicate the 95% posterior intervals.

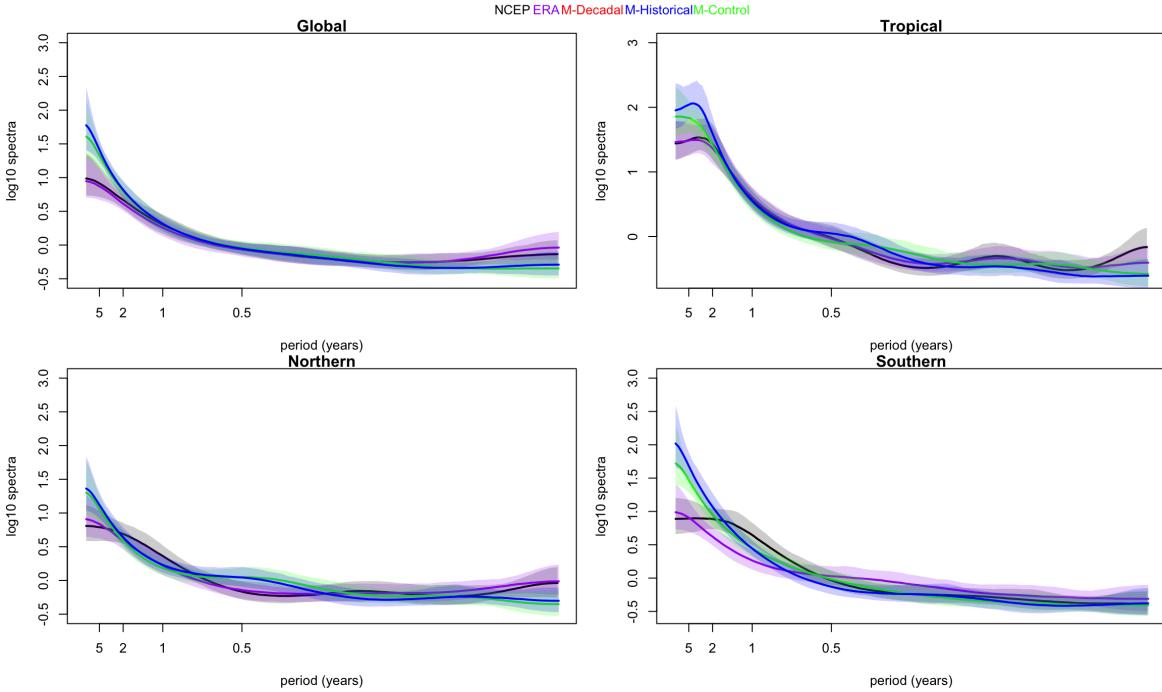


Figure 3.11: NCEP. ERA40. Decadal. Historical. Control. GFDL AR \log_{10} spectra normalized with respect to white-noise with 95% posterior intervals. The x-axis is labeled at select years($2\pi/12\omega$). Colors represent simulation types or observational products.

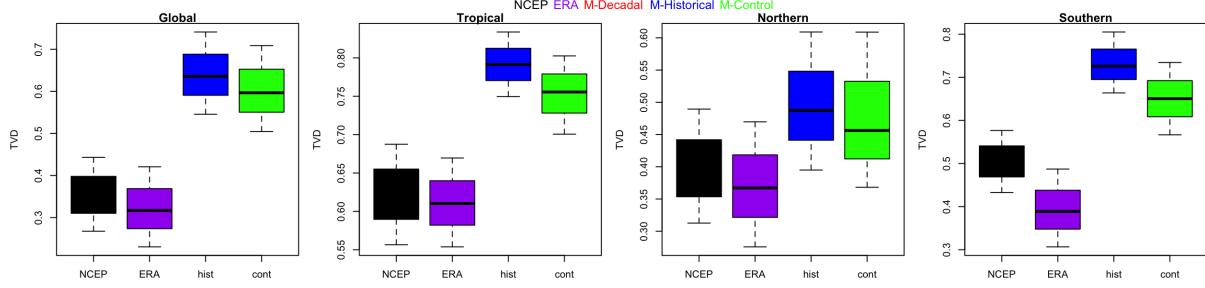


Figure 3.12: NCEP. ERA40. Decadal. Historical. Control. TVD for GFDL with white-noise as the reference. Whiskers indicate the maximum and minimum values while boxes indicate 95% posterior intervals. Colors represent simulation types or observational products.

external forcings. CanCM4 is the only model with decadal predictions that have the same values as the observations in 1981. We see stronger evidence of post-initialization drift in the decadal run of CanCM4 than in the other models. CanCM4 and HadCM3 have completely overlapping control runs. We can also see many differences between models. Globally, MIROC5 systematically overestimates the temperature while HadCM3 underestimates. The tropics appears to have the best fit to the observations. In the Northern Hemisphere, HadCM3 systematically underestimates the temperatures and in the Southern Hemisphere, the best fitting model appears to be GFDL.

We see interesting results in the amplitudes. Figure 3.14 shows all amplitudes plotted next to one another to make comparisons apparent. We can see disagreements in the amplitudes between models and observations in almost every model type and spatial domain. There are also several model-versus-model differences. For example the Southern Hemisphere decadal runs all have significantly different amplitudes. In the tropics there appear to be less differences between amplitudes of all model types.

Figure 3.15 shows the log spectra for all observations and models. It is hard to see a difference between each model's spectral density as they are very similar. Since we use the same AR order, the densities have similar smoothness across domains. We can clearly see that at the lowest frequencies, all of the models have higher spectra than the observations.

Finally, we show the TVD for each model and observation. We do not find significant model-versus-model differences in any of the spatial domains and simulation types. There are however, model-versus-observation differences in nearly every plot. The observations are clearly closer to white-noise than any of the models.

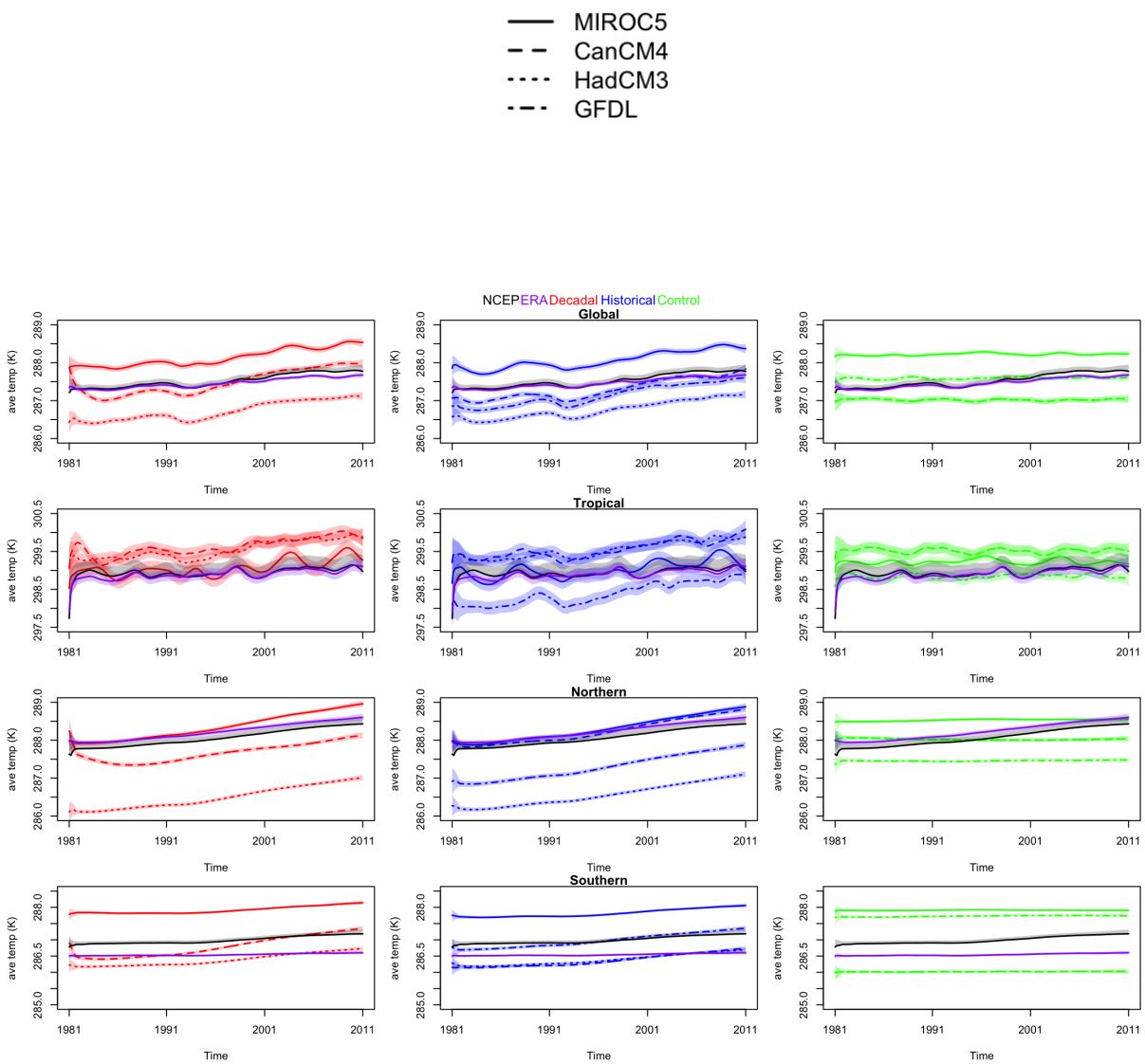


Figure 3.13: NCEP, ERA40, Decadal, Historical, Control. Baseline temperature estimates for all models and observations. Colors represent simulation types or observational products. Shaded regions represent 95% posterior intervals.

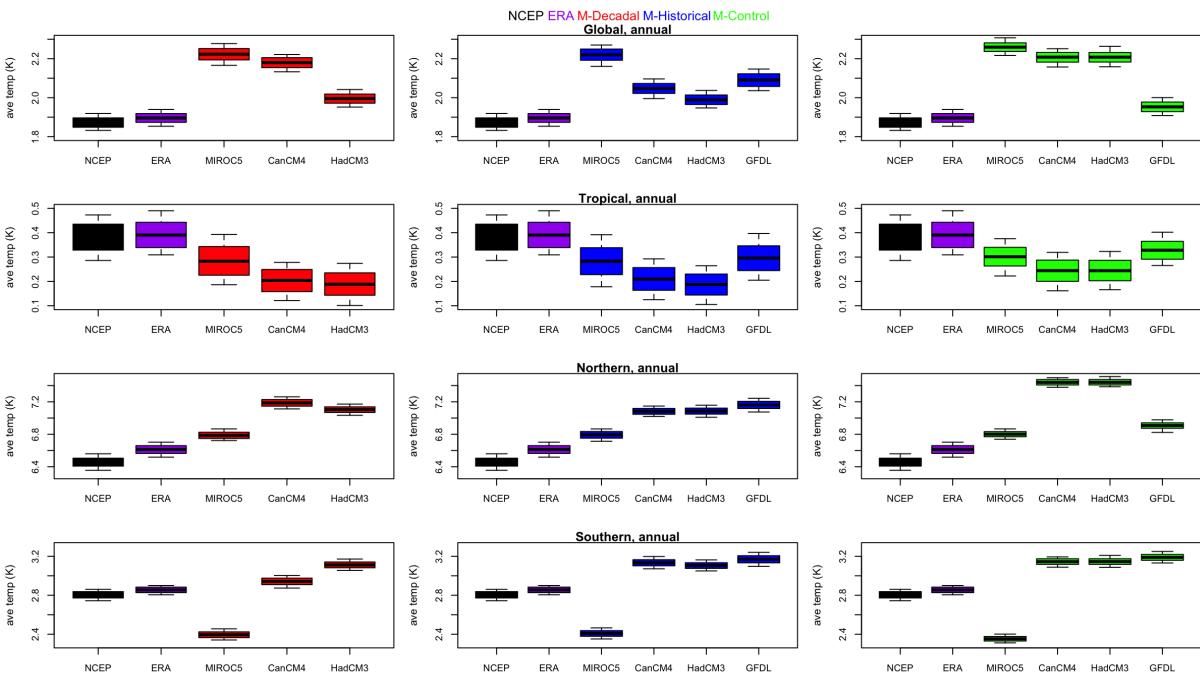


Figure 3.14: NCEP. ERA40. Decadal. Historical. Control. Annual and semi-annual posterior amplitude samples for all models and observations. Colors represent simulation or observational product. Whiskers indicate maximum and minimum values while the boxes indicate the 95% posterior intervals.

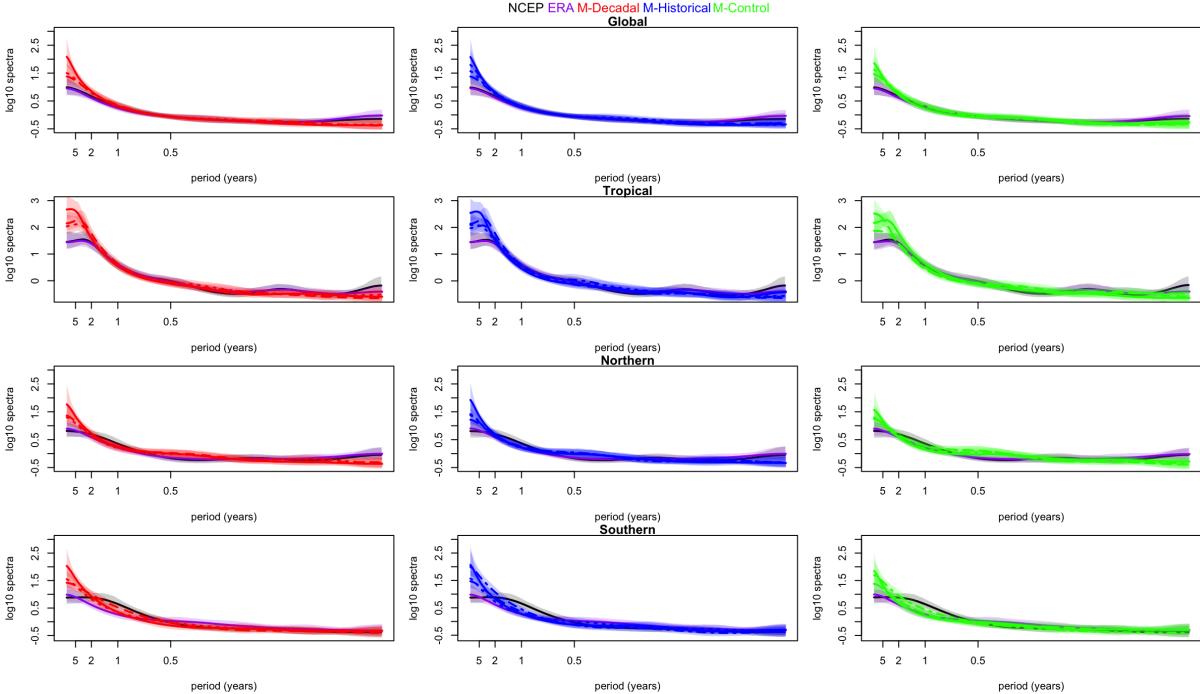


Figure 3.15: NCEP. ERA40. Decadal. Historical. Control. AR \log_{10} spectra normalized with respect to white-noise with 95% posterior intervals for all models and observations. The x-axis is labeled at select years($2\pi/12\omega$). Colors represent simulation type or observational product.

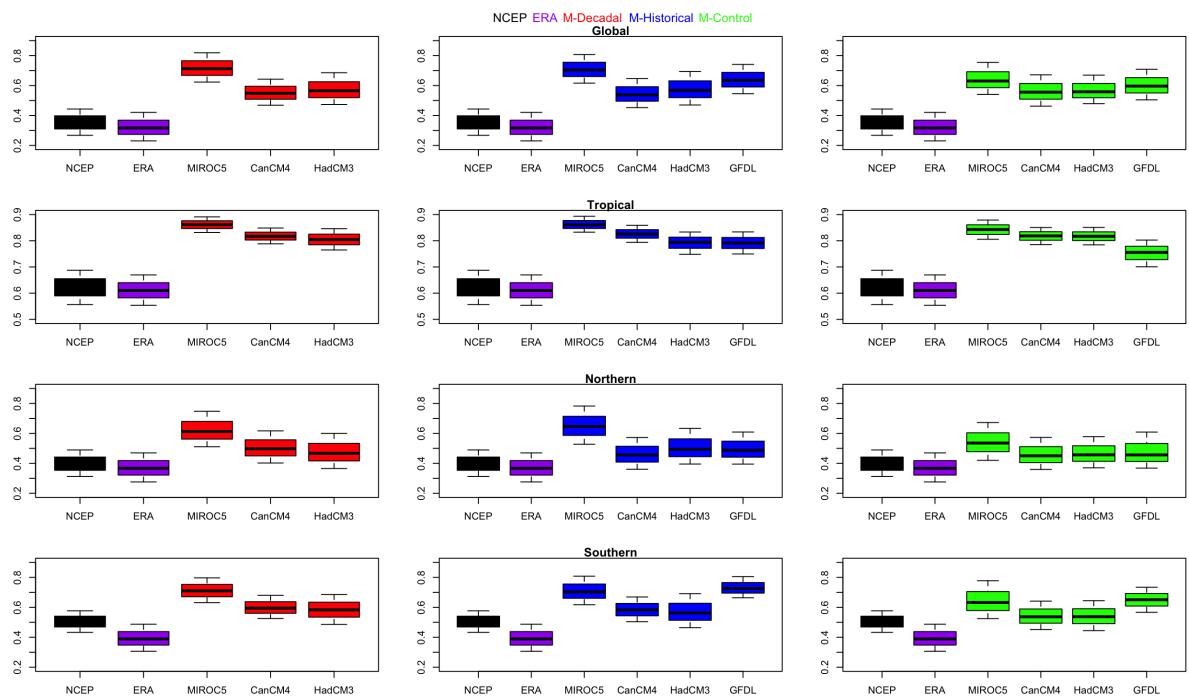


Figure 3.16: NCEP. ERA40. Decadal. Historical. Control. TVD for all models and observations with white-noise as the reference. Whiskers indicate the maximum and minimum values while boxes indicate 95% posterior intervals. Colors represent simulation type or observational product.

Chapter 4

4.1 Conclusions

We implemented a model diagnostic statistical methodology in order to compare four climate simulations. We implemented multivariate Dynamic Linear Models (DLM) to estimate temperature baseline and seasonality. We fit multivariate autoregressive (AR) models to the residuals of the DLM in order to characterize internal variability through their spectral density. We study these over four spatial domains (global, tropical, Northern and Southern Hemispheres) and for three types of model simulations (control, decadal, and historical).

We found some interesting results from our analysis. We see evidence of post-initialization drift in all of the decadal predictions, most prominently in CanCM4. None of the baseline or seasonal components of the climate models fully agreed with the observational products. This is seen most clearly when examining the posterior amplitudes and TVDs. Despite these significant differences each residual time series produced very similar spectral estimates of internal temperature variability within each region. Because we have these similar spectra, we know that our analysis methods are consistently extracting comparable components of internal variability for the simulation data within each region (Barata, R., Prado, R., & Sansò, B. 2019). Using the methodology presented in the paper stated above, we were able to assess and compare the behavior of climate model simulations and observational products.

Bibliography

- [1] Barata, R., Prado, R., & Sansò, B.: Comparison and assessment of large-scale surface temperature in climate model simulations, 2019.
- [2] Cheylek, P., Li, J., Dubey, M., Wang, M., and Lesins, G., "Observed and model simulated 20th century arctic temperature variability: Canadian earth system model canesm2," *Atmospheric Chemistry and Physics Discussions*, 2011.
- [3] Delworth, T.L., Rosati, A., and Anderson, W., "Simulated Climate and Climate Change in the GFDL CM2.5 High-Resolution Coupled Climate Model" (2012).
- [4] Prado, R. & West, M." Time series: modeling, computation, and inference, CRC Press, 2010.
- [5] Fujiwara, M., Wright, J.S., Manney, G.L., Gray, L.J., Anstey, J., Birner, S., Davis, S., Gerber, E.P., Harvey, V.L., Hegglin, M.i., "Introduction to the SPARC reanalysis intercomparisons project (S-RIP) and overview of the reanalysis systems," *Atmospheric Chemistry and Physics*, 2017.
- [6] Valdes, P.J., Armstrong, E., "The BRIDGE HadCM3 family of climate models: HadCM3@Bristol", 2017.
- [7] West, M. & Harrison, J. "Bayesian forecasting & dynamic models", vol. 1030, Springer New York City, 1999.