Fall 2013

Evaluation of the CMIP5 Decadal Hindcasts in the State of California

Colin McKellar
San Jose State University

EVALUATION OF THE CMIP5 DECADAL HINDCASTS IN THE STATE OF
CALIFORNIA

A Thesis

Presented to

The Faculty of the Department of Meteorology and Climate Science

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Masters of Science

by

Colin J. McKellar

December 2013

The Designated Thesis Committee Approves the Thesis Titled

EVALUATION OF THE CMIP5 DECADAL HINDCASTS IN THE STATE OF CALIFORNIA

by

Colin J. McKellar

APPROVED FOR THE DEPARTMENT OF METEOROLOGY

AND CLIMATE SCIENCE

December 2013

Dr. Eugene C. Cordero	Department of Meteorology and Climate Science
Dr. Alison F. C. Bridger	Department of Meteorology and Climate Science
Dr. Bridget Thrasher	Climate Analytics Group and Climate Central Inc.

ABSTRACT

EVALUATION OF THE CMIP5 DECADAL HINDCASTS IN THE STATE OF

CALIFORNIA

by Colin J. McKellar

This study investigated the ability of the new Coupled Model Inter-comparison

Project phase 5 (CMIP5) decadal hindcasts to predict the observed decadal variability for

maximum temperature (Tmax) and minimum temperature (Tmin) in California over two

historical periods in the 20th century and one future period in the 21st century. Annual

and seasonal California temperature trends were computed by averaging 54 United States

Historical Climate Network version 2 temperature observations from 1960-1990 and

1980-2010. Modeled California temperatures were reconstructed with bi-linear

interpolation from the CMIP5 decadal hindcasts and 20th century experiments. The

individual model ensemble averages (MEA) and mean model ensemble averages

(MMEA) were then compared to the observations during 1960-1990 and 1980-2010. The

decadal hindcasts displayed a similar overall skill as the 20th century experiments in

predicting the observed annual and seasonal temperature trends during both historical

periods. However, the predictive skill for individual models showed that the decadal

hindcasts systematically improved the MEA predictions and that certain models, such as

the MRI-CGCM3, outperformed the MMEA in each experiment. Also the higher

performing models, such as the MRI-CGCM3, provided better future Tmax and Tmin

trend predictions. Future predictions show increasing annual and seasonal temperature

trends that indicate a longer growing season by the year 2035.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction and Background

California's agriculture industry is the fifth largest in the world and generates more than $40 billion per year in direct revenue from goods and services (California Agricultural Statistics Review 2012-2013). It is also responsible for an additional $100 billion each year from related economic activities (Cooley et al., 2009). The vitality of the agriculture industry in California is primarily dependent on the availability of freshwater resources (i.e., rivers, rain and snowfall, ground-water) for irrigation. However, climate change has directly altered the precipitation patterns that influence the availability, timing, and distribution of freshwater (Trenberth et al., 2007). Human-induced temperature changes have indirectly influenced the supply of freshwater by affecting the severity of droughts, the length of the growing season, and the types of crops that can grow. The supply of freshwater may vary significantly on a regional level, not only due to differences in climate, but also due to the agricultural practices, types of crops, and fresh water resources for the population.

Decadal climate prediction evaluates the 10- to 30-year evolution of climate, and provides forecasts that aim to improve understanding of the social, economic, and environmental implications of decadal climate variability (Meehl et al., 2009; Mehta et al., 2011). The first characteristic of decadal prediction is that variability is largest at the regional level, and secondly, the contribution of well-mixed anthropogenic greenhouse gases (GHGs) to the overall variability is low (Meehl et al., 2009; Taylor et al., 2012). The predictive skill of decadal forecasts are affected by the uncertainties in the natural short-term [i.e., El Nino-Southern Oscillation (ENSO)] and long-term (e.g. Pacific

Decadal Oscillation, Atlantic Meridional Overturning Circulation) atmosphere-ocean variability and the relative strength of the anthropogenic forced response (Mantua et al., 2002; Meehl et al., 2009; Mehta et al., 2011; Goddard et al., 2012). When compared to the global and continental scales, regional climate can vary significantly over smaller geographic areas depending upon the time, location, and type of land cover (Meehl et al., 2009; Abatzoglou et al., 2010; Cordero et al., 2010).

Although the effect of natural forcings, such as ENSO, on regional climate variability is greater than that at global scales, human-induced changes to the Earth's surface and oceans do alter local and regional climate patterns (LaDochy et al., 2007; Lebassi et al., 2011; Jin 2012). Variability is caused by changes to the types of land cover, for example, as a result of urbanization or irrigation, which in turn alter the albedo, surface emissivity, and surface energy balance. Changes in the surface energy balance affect the fluxes of heat, moisture, and momentum into the atmosphere. Additional changes to regional climate are due to atmospheric aerosols (Jin et al., 2010), ocean temperature and circulation (Latiff et al., 1996; Srokosz et al., 2012), variations in land and sea ice (Cavalieri et al., 2003), and changes to the carbon cycle (Taylor et al., 2012). External factors that can modify regional to global temperatures on a decadal scale include forcings, such as the 11-year solar cycle (Van Loon and Shea 1999; Van Loon et al., 2007) and volcanic eruptions (Robock 2000; Zanchettin et al., 2012). Over decadal to multi-decadal time periods, regional-scale human contributions can enhance or reduce the influence of natural variability (Lebassi et al., 2011; Meehl et al., 2012).

Previous research into natural variability can also provide additional predictive skill since decadal prediction is both an initial value problem and boundary condition problem (Meehl et al., 2009; Goddard et al., 2012). Prior advancements in seasonal prediction are transferrable to prediction on decadal time periods, including understanding and predicting short-term natural variability such as seasonal ENSO events (Barnston et al., 2012; Goddard et al., 2012). The contribution and interactions between the ocean and atmosphere also impact decadal to multi-decadal predictions due to the strong influence of decadal variability on regional climate patterns. Examples include Pacific Decadal Variability (PDV) and Atlantic Multi-decadal Variability (AMV) (Latif et al., 1994; Goddard et al., 2012; Srokosz et al., 2012). A better quantification of the natural variability will improve the predictive skill of 10-30 year forecasts (Branstator and Teng 2010; Msadek et al., 2010).

In earlier climate model inter-comparisons, decadal variability and prediction were examined through the use of Coupled Global Climate Model (CGCM) data from simulations that were initialized with uniform land and sea surface temperature (SST) distributions from the observed climate record. For example, the Coupled Model Intercomparison Project phase 3 (CMIP3) simulations of the 20th century are 50 to 150-yearlong projections initialized with pre-industrial fields. Future decadal predictions relied on various GHG emission scenario pathways and were typically initialized using data from 20th century climate model simulations (Meehl et al., 2007). Although there is some predictive skill that can be derived from previous decadal forecasts, variations at

regional scales due to natural variability are more difficult to ascertain (Meehl et al., 2009).

The nonlinear nature of the climate system ensures that no CGCM simulation can perfectly recreate observations because of internal variability that may not depend on the initial conditions (Hawkins and Sutton 2009; Brown et al., 2012; Goddard et al., 2012). Thus for any forecast, it is impossible to account for all time-evolving interactions between the natural and anthropogenic climate forcings. An approach to the larger uncertainty is to use a large ensemble of forecasts within a multi-model framework (Taylor et al., 2012). The CMIP3 dataset used initial states that were chosen at random times from the preindustrial control runs. Thus there was no correspondence in the ocean-induced internal variability between the observations and models. With the latest Coupled Model Inter-comparison Project phase 5 (CMIP5), new short-term decadal forecasts that start with initial land and ocean conditions are modeled within a multi-model experiment framework (Taylor et al., 2012).

Previous studies have shown that the initial conditions can increase or decrease the predictive skill over larger geographical areas in the 10-year decadal predictions (Kim et al., 2012). Additional skill is obtained in the 30-year predictions at larger scales due to the anthropogenic forcing caused by GHGs (Meehl et al., 2009, Taylor et al., 2012). The focus of the present study is to determine if the initial conditions prescribed by observations will improve the regional climate forecasts. The CMIP5 decadal hindcasts and 20th century experiments are evaluated according to their ability to predict the observed maximum (Tmax) and minimum (Tmin)

temperature trends in California over the complex regional climate zones across the state (Abatzoglou et al., 2010; Cordero et. al 2010). The performance of the decadal hindcasts and models is also used to determine future near-term predictions.

2. Data and Experimental Design

a. Observations

The Tmax and Tmin observations from the US Historical Climate Network version 2 (USHCNv2) were obtained from the National Climatic Data Center (NCDC). The observations used in this study are the monthly averaged temperatures from 54 USHCNv2 stations in California (Figure 1) spanning the periods 1960-1990 and 1980-2010.

1 berkeley
2 blythe
3 brawley_2sw
4 cedarville
5 chico_univ
6 chula_vista
7 colfax
8 cuyamaca
9 davis
10 death_valley
11 electra_ph
12 eureka_wso
13 fairmont
14 fort_bragg
15 fresno_wso
16 hanford
17 happy_camp
18 healdsburg
19 independence
20 indio_fire
21 lake_spaulding
22 lemon_cove
23 livermore
24 lodi
25 marysville
26 merced
27 mount_shasta
28 napa
29 needless
30 newport_beach
31 ojai
32 orland
33 orleans
34 pasadena
35 paso_robles
36 petaluma
37 quincy
38 redding
39 redlands
40 san_luis_obispo
41 santa_barbara
42 santa_cruz
43 santa_rosa
44 susanville
45 tahoe_city

46 tejon_rancho
47 tustin_irvine
48 ukiah
49 vacaville
50 wasco
51 weaverville
52 wil6_willows
53 yosemite
54 yreka
55 brookings
56 searchlight
57 parker
58 yuma_citrus

Figure 1. Map of the United Stated Historical Climate Network version 2 stations used for this study. Adapted from Cordero et al., 2010.

The USHCNv2 is a high quality dataset that accounts for station biases by using a pairwise comparison between the individual station observations and the surrounding regional observations (Meene and Williams 2009). This includes documented and undocumented adjustments for station location and missing values. However, unlike

the first version of the United States Historical Climate Network, the data have not been

adjusted to account for the effects of urbanization (Meene et al., 2009).

b. Model Simulations

The multi-model ensemble simulations used in this study were obtained from the

CMIP5 archive and are listed in Table 1. The results of this study use CMIP5 data that

were available as of March 2012.

Table 1: The CMIP5 AOGCMs and ESMs examined for this study and the origins of each model (left). The decadal experiments (center) for each time period and the respective number of simulations for each model. The 20th century and RCP4.5 (right) experiments for each time period and the number of simulations for each model. a) Models simulate Tmax only. b) Annual ensemble size is one. c) Annual Tmax only.

Model	Originating Group(s)	Number of Decadal Ensemble Members for 30-year Hindcasts			Number of 20th Century Historical and RCP4.5 Ensemble Members		
		1960	1980	2005	1960	1980	2005
BCC-CSM1	Beijing Climate Center, China Meteorological Administration		2	2			
CANCM4 CANESM2[a]	Canadian Centre for Climate Modelling and Analysis	10	10	10	10	10[b]	10
CNRMCM5	Centre National de Recherches Meteorologiques	10	10	10	10	10	10
CSIRO-MK3-6-0[a]	CSIRO Atmospheric Research					10	10
HADCM3 HADGEM2-CC	Met Office Hadley Centre	10	10	10	10	10 1	10
INMCM4	Institute for Numerical Mathematics				1	1	1
ISPL-CM5-MR	Institute Pierre-Simon Laplace				1	1[c]	1
MIROC4h MIROC5 MIROC-ESM MIROC-ESM-CHEM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute	3	3	3	1 3	1 1 1	3 1
MPI-ESM	Max Planck Institute for Meteorology	3	3	3	3	3	3
MRI-CGCM3	Meterological Research Institute	3	3	3	3[b]	1	1
NORESM1-M	Norwegian Climate Center				3[b]	1	1

The CMIP5 dataset is derived from simulations by several coupled global atmosphere-

ocean general circulation models (AOGCMs) and earth-system models (ESMs) (Taylor et

al., 2012). These models are used to examine and predict the long-term climate forcing due to atmospheric GHGs. Examples include the 20th century simulations that attempt to re-create the observed variability and the 21st century experiments that test the model response to a relative concentration pathway (RCP) of well-mixed anthropogenic GHGs. The emission pathway RCP4.5 used in this study represents a scenario in which global atmospheric GHG concentrations first stabilize mid-century, producing a 4.5 W m^{-2} increase in radiative forcing, and then decrease in the second half of the 21st century. The second use of the AOGCMs and ESMs in the CMIP5 is to address the near-term decadal (10-year) to multi- decadal (30-year) variability and prediction using the decadal hindcasts. The 30-year decadal experiments used for our study helped test the impact of observed initial land and ocean temperatures as well as the impact of observed GHGs on decadal variability and prediction.

c. Experimental Design

In order to reproduce the regional annual and seasonal temperature observations for California, the 54 USCHNv2 station data were spatially averaged following a similar technique to that used by Cordero et al. (2010). The CMIP5 model temperatures for California were reconstructed by bi-linear interpolating from the GCM's grid resolution to the geographic position of each USHCN station.   Next, the 54 interpolated GCM station points were spatially averaged in a manner similar to the observations. Annual Tmax and Tmin of the observations and model data were calculated by taking the temporal average of the 12 monthly temperatures, while the seasonal datasets were constructed by taking the 3-month average using the standard definitions for winter,

9

December-January-February (DJF), spring, March-April-May (MAM), summer, June-July-August (JJA), and fall, September-October-November (SON). To assure data completeness in the observations, no three-month season was used if a single month was missing and no year was used if a single month or more was missing (Cordero et al., 2010).

For the CMIP5 model ensemble simulations, the individual model ensemble averages (MEA) were calculated by taking the average of all the simulated Tmax and Tmin trends. The number of simulations for each model ensemble is shown in Table 1. The experimental mean model ensemble average (MMEA) was calculated by taking the average of each model's individual MEA. The observations, MMEA, and individual MEA trends were then tested for statistical significance using the standard error of the trend estimate (Santer et al., 2000). The 30-year temperature trends for the observations and all CMIP5 simulations were computed using a linear least-square regression between the historical periods of 1960-1990, 1980-2010, and a future period of 2005-2035. Since several of the 20th century simulations ended in 2005, the first five years of the RCP4.5 experiment were used for the last five years of the historical period from 1980-2010 if the climate run corresponded to the same ensemble member and simulation.

Box and whisker plots were constructed for each individual model with an ensemble size of five or greater by using each simulated model trend. The box and whisker plots for the CMIP5 experiments were calculated by using the MEA trends. The box plots first allow for a comparison between the model or experimental trends and

observations. Secondly, the boxplot and whisker allow for the uncertainties to be compared between the different models and experiments when the initial distributions are unknown (Chang 2009). The main components of a boxplot are the minimum, median, maximum, 25th and 75th quartile, and outlier values. The "box" portion is constructed from the median and the 25th and 75th quartiles, or lower and upper portions of the box. This "box" is also known as the inter-quartile range (IQR) and quantifies the uncertainty and internal variability within an individual model(s) or between the different CMIP5 experiments. The IQR also represents the central tendency of the temperature trends since it indicates where 50% of the data are distributed relative to the mean. The "whiskers" are calculated by connecting the 25th quartile to the minimum value and 75th quartile to the maximum value. Each whisker represents 25% of the data and the length of each helps determine if the overall trends are positively distributed (i.e., a larger whisker between the 75th quartile and the maximum), or negatively distributed (i.e., a larger whisker between the 25th quartile and the minimum). The last portions of a box and whisker plot are the outliers, which are defined as $1.5\times (IQR) > 75^{th}$ quartile or $1.5\times (IQR) < 25^{th}$ quartile. If an outlier is suspected, a new minimum value is then calculated by subtracting $1.5\times (IQR)$ from the 25th quartile, while a new maximum is calculated by adding $1.5\times (IQR)$ to the 75th quartile. Any data values outside the new minimum and maximum values are classified as outliers. Lastly, comparing the width of the boxes for both experiments allows for an

initial comparison into the uncertainties without the need for extensive statistical analysis.

The skill of the experimental MMEAs and individual MEAs for the 1960-1990 and 1980-2010 historical periods was calculated by taking the absolute error between the modeled and observed trends. A total of ten simulated trends, comprised of the annual and seasonal Tmax and Tmin, were ranked between one and five. Each ranking corresponds to a step increase in the model error of $0.05\,^{\circ}\text{C dec}^{-1}$. For example, a ranking of one means there is an absolute error of $0\text{-}0.05\,^{\circ}\text{C dec}^{-1}$ between the model and observations. The final skill is determined by taking the average of the ten rankings, and a lower skill indicates that the experimental MMEA or MEA has a lower error when comparing the simulated trends to the observations.

3. Mean Model Ensemble Average Temperature Trends Using the Observations and Model Data

a. Annual and Seasonal Analysis: 1960-1990 and 1980-2010

Figure 2 shows the annual California Tmax (top) and Tmin (bottom) from the USHCNv2 and the experimental CMIP5 20th century historical and decadal mean model ensemble average (MMEA) for the time periods 1960-1990 (left) and 1980-2010 (right). The seasonal graphs are shown in Appendix A.

Annual Tmax

a

Annual Tmin

b
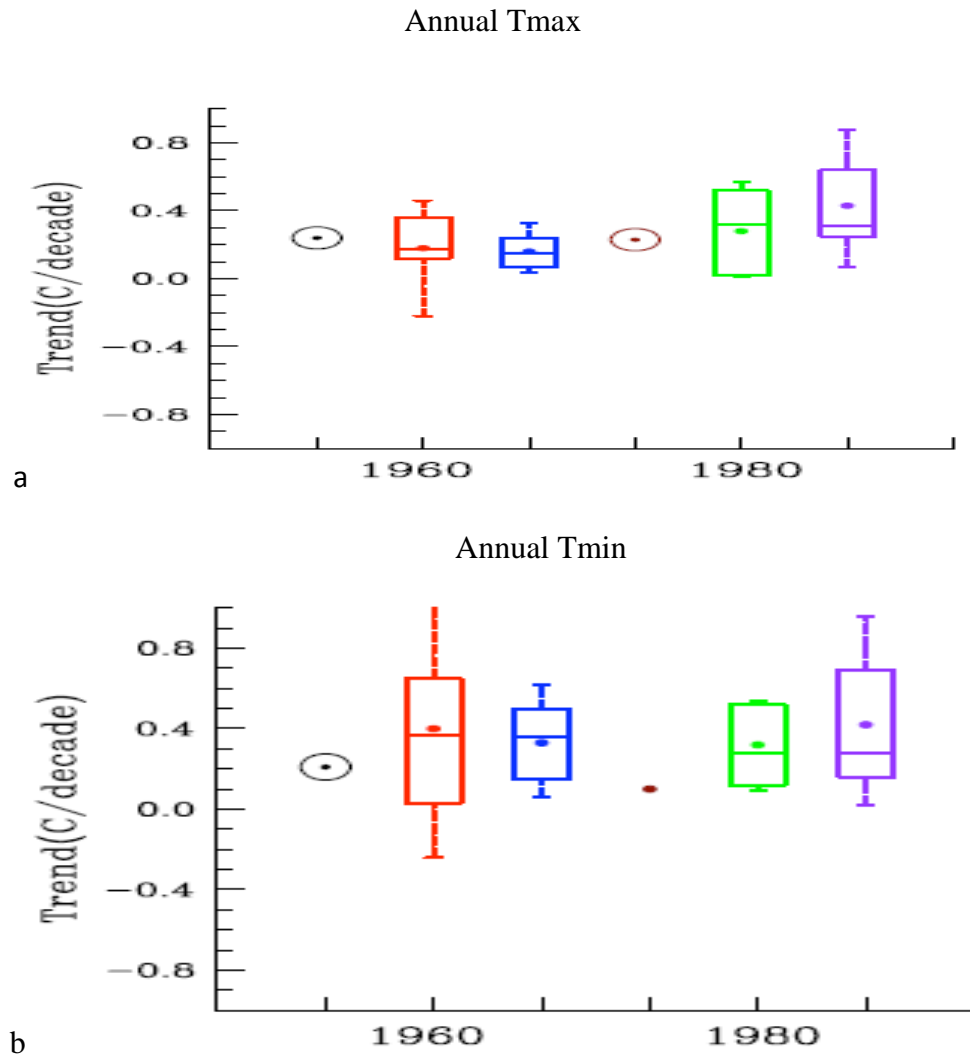
Figure 2. The mean model ensemble average (MMEA) and boxplots for annual (a) Tmax and (b) Tmin. The black and brown dots represent the 1960-1990 and 1980-2010 trend observations; circled dots display trends that are statistically significant at the 95% confidence interval. The red (1960-1990) and green (1980-2010) boxplots and dots represent the decadal trends and MMEA, while the blue (1960-1990) and purple (1980-2010) boxplots and dots represent the 20th century trends and MMEA. X represents outlying trends.

The 1960-1990 USHCNv2 observations indicate statistically significant warming (95% confidence level) for Tmax +0.24°C dec^{-1} and Tmin +0.21°C dec^{-1}. The 1960-1990 decadal and 20th century MMEA (indicated by colored dots) for Tmax are close to the observations, with warming trends of +0.18°C dec^{-1} and +0.16°C dec^{-1}, while for Tmin, the decadal and historical experiments are above the observations with positive trends of +0.40°C dec^{-1} and +0.32°C dec^{-1}. The annual Tmax and Tmin trends between the decadal and 20th century experiments during 1960-1990 are not statistically different at the 95% confidence interval.

The annual observed trends for California Tmax and Tmin during 1980-2010 show statistically significant warming in Tmax +0.23°C dec^{-1}, while Tmin displays a slight positive trend of +0.10°C dec^{-1}. The 1980-2010 decadal hindcast MMEA for Tmax also indicates warming of +0.28°C dec^{-1} and is closer to the observations compared to the historical trend of +0.43°C dec^{-1}. For Tmin, both the decadal trend +0.32°C dec^{-1} and the historical trend +0.42°C dec^{-1} are well above the observations. Lastly, annual trends show more variation between the decadal and historical experiments during this time frame compared to 1960-1990. However, the differences in the MMEA during 1980-2010 between the decadal and 20th century temperature trends are not statistically different.

Seasonal temperature trends (Appendix A) from 1960-1990 indicate that the USHCNv2 observations exhibit similar significant warming in MAM Tmax

+0.66°C dec^{-1} and Tmin +0.63°C dec^{-1}. The remaining seasonal trends are lower by

at least 0.20°C dec^{-1} and did not show significant state-wide warming or cooling, with

similar trends between Tmax and Tmin. Previous studies have shown significant

warming during MAM that is larger than that in other seasons (Cayan et al., 2008;

Cordero et al., 2010). However, neither the decadal nor the 20th century experiments

simulate this large warming during MAM, instead displaying lower trends by roughly

0.50°C dec^{-1}. Lastly, the difference in the MMEA Tmax and Tmin trends between

both experiments and for all seasons during 1960-1990 is not statistically significant.

Seasonal trends from the USHCNv2 during 1980-2010 (Appendix A) display the

largest significant warming during JJA in both Tmax +0.36°C dec^{-1} and Tmin

+0.26°C dec^{-1}. The decadal hindcasts most closely predict the observed warming for

Tmax +0.37°C dec^{-1}and Tmin +0.29°C dec^{-1}. The 20th century experiments also show

warming similar to the decadal hindcasts for Tmax +0.35°C dec^{-1} and slightly lower

trends in Tmin +0.20°C dec^{-1}. All the remaining seasonal USHCNv2 temperature

trends do not show any significant warming or cooling in Tmax and Tmin. In general,

the magnitude of the MMEA for the 1980-2010 experiments closely matches the

observed seasonal trends. However, none of the seasonal temperature trends during

1980-2010 are statistically different between the decadal and 20th century experiments.

b. Comparison of Uncertainty in the CMIP5 Experiments

The boxplots help quantify the annual (Figure 2) and seasonal (Appendix A) uncertainty between the decadal and 20th century experiments. For example, annual Tmax from 1960-1990 shows a similar IQR, or internal variance, for the decadal and 20th century experiments, while for Tmin, the IQR is larger in the decadal experiment by 0.30°C dec^{-1}. The annual IQR for Tmax and Tmin from 1980-2010 between the decadal and 20th century experiments was similar in magnitude. On the contrary, the seasonal variability in both the decadal and 20th century experiments was noticeably lower during 1960-1990 for both temperatures when compared to 1980-2010. However, the seasonal variability in the decadal experiment during 1980-2010 is larger for Tmax compared to Tmin.

Another measure of uncertainty is shown in the boxplots by the length of the whiskers. The relative length of each whisker, represented by the length of the line from the quartile to the extreme value, helps determine if the temperature variations display a normal distribution. For example, the annual Tmax trends for the 1960-1990 historical experiment and both 1980-2010 experiments are positively distributed while the 1960-1990 decadal experiment shows a negative distribution. Lastly, the extreme maximum and minimum temperature trends and the difference in extremes identify large outliers that address the model and experimental climate sensitivities. For example, from 1980-2010, the seasonal historical forecasts for Tmax and Tmin show a large negative outlying trend of -1.5°C dec^{-1}.

c. Skill of the CMIP5 Experiments

The absolute error in the annual and seasonal MMEA between all of the decadal and 20th century experiments and the observations is used to determine the overall experimental skill, which is shown in Table 2.

Table 2: The overall skill of each experiment.

Overall Decadal Skill	3.1
Overall Historical Skill	3.3
1960-1990 Decadal Skill	3.2
1960-1990 Historical Skill	3.3
1980-2010 Decadal Skill	2.9
1980-2010 Historical Skill	3.3

The overall decadal skill was better at 3.1 when compared to the 20th century skill of 3.3. When looking at individual time periods, the results show that the skill is similar between the experiments during 1960-1990, while from 1980-2010 the decadal experiment again shows a higher skill. Also, the decadal experiments show a higher skill for 1980-2010, while the 20th century experiments display a similar skill during both historical time periods. But, when comparing the skill rankings to the experimental MMEA in the Figure 2, certain seasons display more error when compared to other seasons during both

time periods and for both experiments. One example includes the MMEA temperature

trends for MAM in which the predictions were well below the observations.

4. Model Ensemble Average Temperature Trends for the CMIP5 Experiments

The structure of the CMIP5 allows for multiple AOGCMs and ESMs to be examined in their response to a climate-forcing scenario. The core CMIP5 experiments provide information that helps identify reasons why different climate models give different solutions given similar initial conditions (Taylor et al., 2012), and they also allow for inter-model comparisons for each forecast. Beyond the required core experiments, individual modeling agencies (Table1) have the option to investigate several critical model attributes in further detail with the so- called tier 1 and tier 2 simulations. Due to the complex nature of the climate system and the uncertainties introduced into the model simulations and experiments, large ensembles must be constructed to examine and quantify initial state and model errors (Mehta et al., 2011; Goddard et al., 2012). Furthermore, more hindcasts allow model skill to be assessed, different periods of variability to be sampled, and sources of improved prediction to be determined (Goddard et al., 2012).

a. Analysis of Ensemble Size on Model Trends

The annual AOGCM and ESM predictions during 1960-1990 (left) and 1980-2010 (right) for the decadal experiments are shown at the top of Figure 3 (Tmax) and Figure 4 (Tmin).
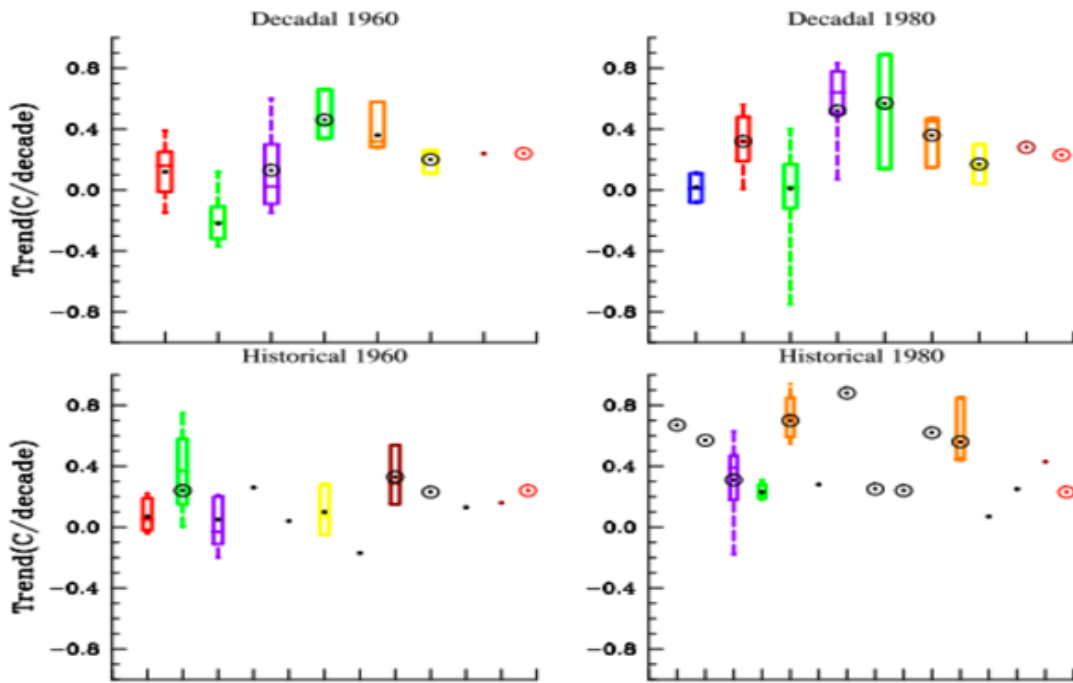
Figure 3. Annual model ensemble means (MEA) for Tmax. The decadal (top), and 20th century (bottom) experiments between the periods 1960-1990 (left), 1980-2010 (right). Boxplots represent each models individual trend distribution and the dots are the MEA. As a reference, brown dots indicate the MMEA and red dots display the observations. Circled dots show trends that are significant at the 95% confidence interval.
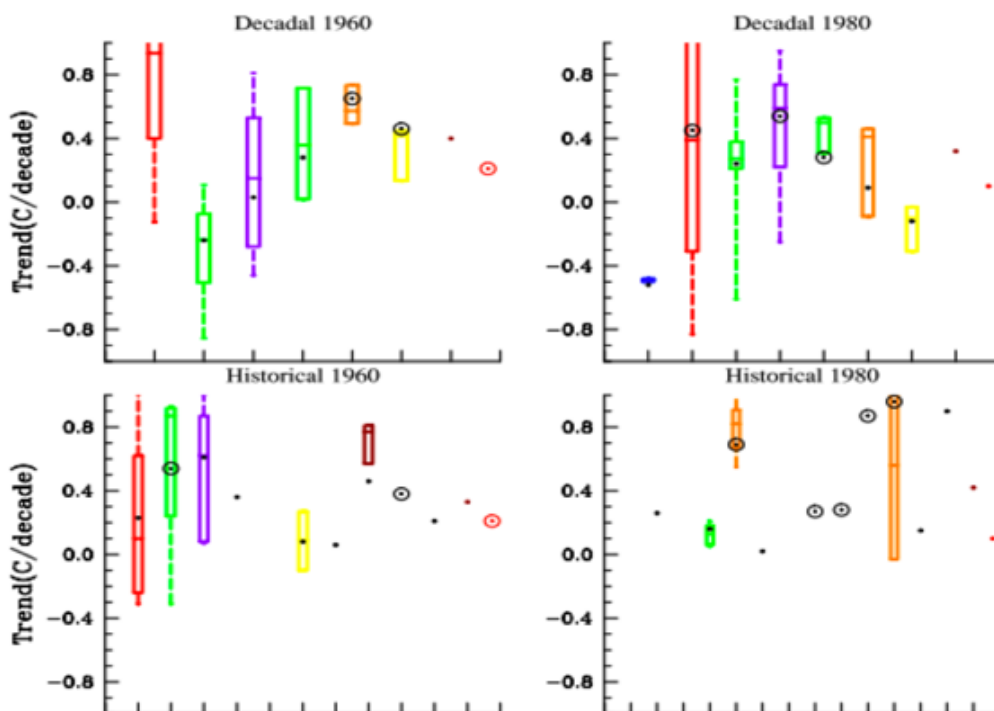
Figure 4. Annual model ensemble means (MEA) for Tmin. The decadal (top), and 20th century (bottom) experiments between the periods 1960-1990 (left), 1980-2010 (right). Boxplots represent each models individual trend distribution and the dots are the MEA. As a reference, brown dots indicate the MMEA and red dots display the observations. Circled dots show trends that are significant at the 95% confidence interval.

The box plot IQR represents each climate model's internal variability and temperature trend distribution, while the dots show the MEA. The decadal models with an ensemble size of ten include CanCM4, CNRM-CM5, and HadCM3. The models with a smaller ensemble size of three include MIROC4h, MPI-ESM-LR, and MRI-CGCM3, while the BCC-CSM1 has an ensemble size of two. In the 1960-1990 decadal hindcast, all three of the large ensemble size models demonstrated an increase in internal variability (larger box) in Tmin compared to Tmax, while two out of three of the smaller ensemble size models display a decrease in the range of trends for Tmin, and MPI-ESM has no difference between temperatures. The differences in internal variability between Tmax and Tmin during 1980-2010 indicate that two out of three of the larger ensemble size models display an increase in variability, and all models show a large increase in the range of trends for Tmin compared to Tmax. For the smaller ensemble size models, two show a decreased range of trends, and one model, MRI-CGCM3, displays no change.

When comparing the larger ensemble models, the internal variability (size of box) is noticeably lower for Tmax compared to Tmin during 1960-1990 and 1980-2010. However the total range in the trends, or the difference between the maximum and minimum value, is twice as large for Tmin compared to Tmax. When comparing the smaller ensemble size models, note that the box represents the total range and the median since it requires at least five trends to construct a box and whisker plot. The smaller ensemble size models typically have a smaller range in trends for both temperatures. When comparing the range in model trends between the different sized ensembles, the smaller ensemble size models have a significantly lower range in trends. In certain

instances, the range of trends for the smaller ensemble sizes is lower when compared to the internal variability or the larger ensemble sizes.

When comparing temperature trends during 1960-1990, the models with the largest ensemble size of ten (CANESM2, CNRMCM5, and HadCM3) display a larger increase in the internal variability and total range of Tmin trends compared to Tmax. The smaller ensemble size of three models (MIROC-ESM, MRI-CGCM3, and MPI-ESM) do not display any change in the range of trends between Tmax and Tmin. Several of the temperature trends in the 1980-2010 20th century historical experiment have only one model simulation, but the variability in the 20[th]century model temperatures for either time period within individual models cannot be accurately determined with only one climate simulation. An example of the error that can occur with only one simulation is displayed by the MIROC-ESM-CHEM during the seasonal 1980-2010 Tmax and Tmin trends, with large negative outlying trends of up to $1.5^{\circ}C$ dec^{-1}.

b. Influence of Initial Conditions on Model Temperature Trends

When comparing the skill between the decadal and historical experiments, there is not a large improvement in the MMEA compared to the observations. However, individual AOGCMs and ESMs do show improvement in their predictive skill when utilizing initial conditions. For example, when comparing the skill of the four models (CNRM-CM5, HadCM3, MPI-ESM-LR, and MRI-CGCM3) used in both the decadal and the historical experiments, there is a systematic improvement in the overall skill of each model when using initial conditions. The average overall skill of the four decadal models in the decadal experiment is 3.3, while the historical skill is 3.9.

Determining the skill of an individual model can provide further insight into how different model attributes contribute to different solutions. As previously documented by Mehta et al., (2011), decadal climate prediction is influenced by systematic model biases that are generated by well-known modeling problems (other than resolution). Examples include errors in the individual ocean and atmospheric models as well as interactions between these models and other climate components, such as sea and land ice and earth surface processes. The overall individual skill for each model is shown in Table 3, while the model skill for 1960-1990 is shown in Table 4, and the skill from 1980-2010 is shown in Table 5. The results show that one model, the ESM named MRI- CGCM3, outperformed all other models and the MMEA. In addition, the performance of the ESM in both experiments is generally better compared to the AOGCMs. Lastly, a large contribution to the overall error is the low level of basic understanding about the natural climate patterns in the decadal time frame.

Table 3. The overall MMEA and the individual MEA skill score for the decadal (top) and historical (bottom) experiments. The bold values indicate situations in which an individual model`s ensemble mean is better than the MMEA ranking.

Overall Decadal Skill

MMEA	3.1
CanCM4	3.4
CNRM-CM5	3.2
HadCM3	3.7
MIROC4h	3.6
MPI-ESM-LR	3.5
MRI-CGCM3	**2.7**

Overall Historical Skill

MMEA	3.3
CNRM-CM5	3.7
HadCM3	4.4
IPSL-CM5A-MR	4.3
MIROC-ESM	**2.8**
MIROC5	3.6
MPI-ESM-LR	4.2
MRI-CGCM3	**3.1**
NorESM1	**2.8**

Table 4. The overall MMEA and the individual MEA rankings for the decadal (top) and
historical (bottom) experiments from a) 1960-1990 and b) 1980-2010. The bold values
indicate situations in which an individual model`s ensemble mean is better than the
MMEA rankings.

Experiment Model Skill

Decadal 1960-1990

MMEA	3.2
CanCM4	3.7
CNRM-CM5	**3.1**
HadCM3	3.5
MIROC4h	3.6
MPI-ESM-LR	3.5
MRI-CGCM3	**3.0**

20th century 1960-1990

MMEA	3.3
CanESM2	**3.2**
CNRM-CM5	**3.3**
HadCM3	4.3
INMCM4	**3.2**
IPSL-CM5A-MR	3.5
MIROC-ESM	**2.7**
MIROC5	**2.6**
MPI-ESM-LR	3.7
MRI-CGCM3	**3.2**
NorESM1-M	**2.7**

Table 5. Same as Table 4 except the rankings are for 1980-2010.

Decadal 1980-2010

MMEA	2.9
BCC-CSM1	3.6
CanCM4	3.1
CNRM-CM5	3.3
HadCM3	3.8
MIROC4h	3.6
MPI-ESM-LR	3.4
MRI-CGCM3	**2.3**

20th century 1980-2010

MMEA	3.3
CanESM2	
CNRM-CM5	4
CSIRO-Mk3-6-0	
HadCM3	4.4
HadGEM2-CC	**2.2**
IPSL-CM5A-MR	5.0
MIROC-ESM	**2.8**
MIROC-ESM-CHEM	4.5
MIROC5	4.5
MPI-ESM-LR	4.7
MRI-CGCM3	**2.9**
NorESM1-M	**2.9**

5. Future CMIP5 Decadal and 21st Century Predictions

As noted by Mehta et al. (2011), predictable climate signals during a decadal time period at the regional level are complicated by local anthropogenic influences, such as urbanization, irrigation, and aerosols. These local forcings impact decadal climate, depending on the region and time. In addition, the future forecasts need to address the magnitude of the natural regional climate variability as well as the interactions and feedbacks within a changing climate system. The decadal hindcasts can improve predictions by eliminating the error associated with initial conditions. The annual decadal (left) and 21st century RCP4.5 (right) results for Tmax (top) and Tmin (bottom) are shown in Figure 5.
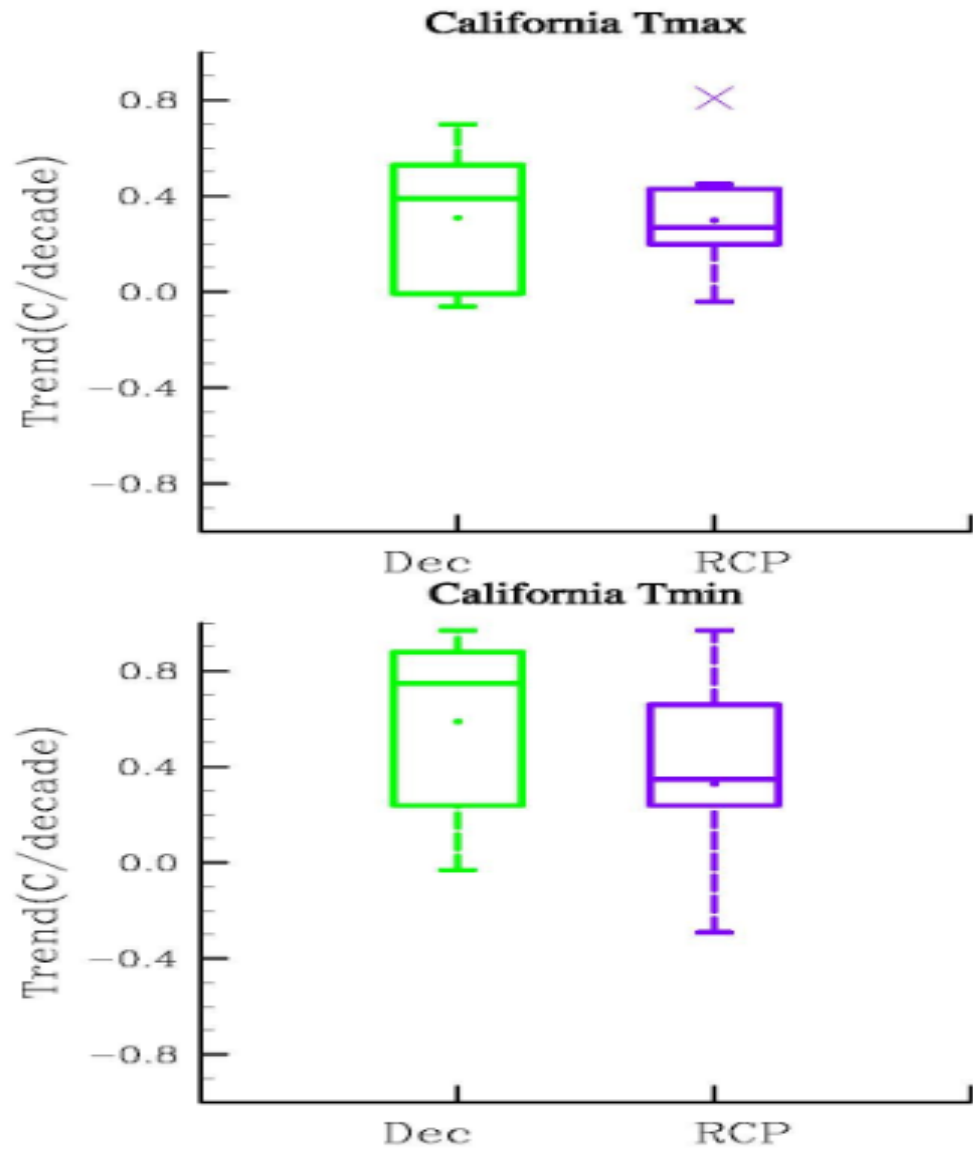
Figure 5. Future Tmax and Tmin temperature trends for the annual Tmax (top) and Tmin (bottom). The green boxplot and dot represent the average median trend and MMEA for the decadal 2005 experiment. The purple boxplot and dot shows the average median trend and MMEA for the RCP4.5 experiment. The X shows outlying trends.

The contribution from anthropogenic GHGs is more influential on decadal prediction by the end of the 30-year forecasts. As a result, a direct comparison of the natural variability between the decadal and RCP4.5 experiments from 2005-2035 is possible. For example, the natural variability for the future annual decadal hindcast was greater by $0.1\,^{\circ}$C dec^{-1} for both Tmax and Tmin. Another characteristic was that the annual 2005-2035 decadal hindcasts displayed a wider range of natural variability for Tmax and Tmin of $0.5\,^{\circ}$C dec^{-1} compared to $0.15\,^{\circ}$C dec^{-1} in the RCP4.5 experiment. One future MMEA trend that is noticeably different is the annual decadal Tmin, which is twice as large as Tmax and the other seasonal Tmin trends. However, this trend pattern in future Tmin also occurred during the decadal 1960-1990 hindcast.

The future seasonal results for Tmax and Tmin are shown in Appendix B. The difference in the natural variability in Tmax between both future experiments is greatest in MAM and SON by $0.10\,^{\circ}$C dec^{-1}, while all other seasons displayed little to no difference. For Tmin, the greatest difference in the natural variability between experiments occurred in MAM. Also, the overall range of Tmax trends was greater in the RCP4.5 experiments by nearly $0.3\,^{\circ}$C dec^{-1} during JJA and SON, and during DJF and SON for Tmin. Lastly, the total range of the temperature trends was greater in all seasonal RCP4.5 experiments by nearly $0.30\,^{\circ}$C dec^{-1}.

When looking at future forecasts for California from 2005-2035, the MMEA between the decadal and RCP4.5 experiments is less reliable for future predictions. One reason is that when looking at the performance of the decadal hindcasts in the historical

experiments, their skill is similar to the 20th century forecasts. In fact, future results

display roughly similar MMEAs between the decadal and 21st century RCP4.5

experiments.   However individual models, such as the MRI-CGCM3, can possibly

provide better future predictions since it accounted for more processes and out-

performed the experimental MMEA in the historical forecasts. The future trends for the

MRI-CGCM3 are displayed in Table 6 (Tmax) and Table 7 (Tmin).

Table 6. The 2005-2035 MEA Tmax trends for the MRI-CGCM3 in the decadal and
RCP4.5 experiments. The decadal and RCP4.5 MMEA are for reference. All trends are
in $^{\circ}$C dec^{-1}.

	Decadal MRI-CGCM3	Decadal MMEA	RCP4.5 MRI-CGCM3	RCP4.5 MMEA
Annual Tmax	0.39	0.31	0.20	0.3
DJF Tmax	0.12	0.34	0.11	0.19
MAM Tmax	0.36	0.27	0.19	0.26
JJA Tmax	0.36	0.26	0.2	0.40
SON Tmax	0.56	0.27	0.36	0.38

Table 7. The 2005-2035 MEA Tmin trends for the MRI-CGCM3 in the decadal and RCP4.5 experiments. The decadal and RCP4.5 MMEA are for reference. All trends are in $^{\circ}$C dec^{-1}.

	Decadal MRI-CGCM3	Decadal MMEA	RCP4.5 MRI-CGCM3	RCP4.5 MMEA
Annual Tmin	0.63	0.60	0.10	0.33
DJF Tmin	0.20	0.35	0.03	0.12
MAM Tmin	0.37	0.27	0.22	0.23
JJA Tmin	0.39	0.28	0.31	0.41
SON Tmin	0.53	0.33	0.34	0.36

The best performing climate model, the MRI-CGCM3, displayed greater Tmax and Tmin trends for all of the decadal hindcasts when compared to the RCP4.5 forecasts. Also, other than DJF, the decadal MEA for the MRI-CGCM3 showed greater Tmax and Tmin trends when compared to the decadal MMEA. The implications of these results are that California will have a longer growing season by 2035. The first reason is the large increase in annual Tmin and two, the large warming temperature trends in SON.

6. Discussion

This section compares the current results to that in the literature and offers hypotheses to explain the results. This first includes explaining why the decadal hindcast experiments are only slightly improved compared to the historical experiments. Secondly, there is an examination into the role of ocean circulations in the decadal hindcasts. Third, there is a discussion of the uncertainties involved in decadal prediction, and finally, there are hypotheses explaining the future results.

The results show that there is a modest increase in the overall predictive skill with the decadal hindcast experiments. However, individual models display a systematic improvement in skill with the decadal hindcasts compared to the 20th century historical experiments. Although using initial conditions did translate into better predictions, the uncertainty that remains is that sampling the decadal time-frame is less than ideal since the natural variability leads to circumstances where the initial dates occur in different anomalous periods in the climate record (Argues et al., 2012), as was demonstrated by the MAM decadal hindcast Tmax and Tmin trends. However, unlike prior studies (i.e., Meehl et al., 2009; Goddard et al., 2012), initial conditions increase the predictive skill of decadal hindcasts at a regional scale. Lastly, as suggested by Goddard et al. (2012), initial conditions in the decadal hindcasts show an increase in forecast skill, as is demonstrated by the increase in the 1980-2010 decadal experiments.

The decadal results show that using the initial SSTs does provide extra forecast skill by accounting for the initial ocean heat content (Taylor et al., 2012). However, it is important to note that the CMIP5 decadal hindcasts spin up the sub-surface ocean

currents. This introduces large uncertainty at the decadal scale because a deeper portion of the ocean circulation is important for prediction. Thus, in-situ ocean observations will provide more information into critical sub-surface ocean properties, such as temperature and salinity (Latiff and Barnett 1994; Meehl et al., 2009; Mehta et al., 2011). This would help account for an important feedback system between the atmosphere and ocean in which, over decadal periods, larger-scale pressure and wind patterns in the atmosphere affect ocean currents by altering the wind stress on the ocean surface. In turn, this can influence the transport of warm or cold water and act to alter SST gradients. The changes in SST gradients, in turn, influence atmospheric circulations due to the thermal wind balance (Latiff and Barnett 1994; Meehl et al., 2009).

It is well understood that climate model differences arise from differences in the dynamical cores, the amount of sub-grid parameterizations of the atmosphere and ocean, and the interactions with other components of the climate system (Mehta et al., 2012). This issue is important for prediction since models will produce different solutions to the same initial conditions. For example, calculating the surface energy balance with an AOGCM will produce a different solution than an ESM. This is because an ESM uses a predictive model for more components of the climate system that are more realistic compared to an AOGCM that typically use parameterizations. This will reduce errors caused by the radiative forcing calculation due to such variables as aerosols, land surface changes, sea and land ice, etc. The results also show that most ESMs display a higher skill and outperform other models. One such model was the

MRI-CGCM3, which is a fully coupled ESM with more modeled components. This is compared to a model with less skill, such as the HadCM3, which is not an ESM and has more parameterizations.

The decadal hindcasts are still in the experimental stage, which makes it difficult to attribute the changes in temperature trends due only to individual model characteristics. Other equally important factors, such as ensemble size and assimilation techniques, still need to be tested for their uncertainties in the experiments (Taylor et al., 2012; Goddard et al., 2012). The tier 1 experiments in the CMIP5 look at ensemble size as a possible factor for increased decadal prediction. The larger ensemble size of ten has, on average, a larger overall range in trends compared to an ensemble size of three. However, the larger ensemble sizes allows for more characteristics of the distribution to be sampled with a box and whisker plot. This shows that an ensemble size of three or lower may be too small to accurately sample the unforced variability due to the greater uncertainties at both a regional and decadal scale. This also shows that unforced variability is the largest factor contributing to decadal variability and varies significantly with time, location, and initial conditions.

Based on the performance of the decadal hindcasts during the two historical periods, the future decadal hindcast MMEA temperature trends will only be slightly better than the RCP4.5 experiment. However, individual models, such as the MRI-CGCM3 and ESMs, might provide a more accurate assessment of future trends. The first reason is that there is a high uncertainty in determining future trends with only one hindcast, as shown by the results from the two historical hindcasts that display a large

variation in modeled trends. The second reason is that the individual models may

provide a better forecast since each model has different model components and

parameterizations. Lastly, the ESMs may give better predictions because every

component of the climate system influences decadal prediction.

7. Conclusion

This study evaluated the predictive skill of the CMIP5 near-term forecasts by comparing the mean model ensemble average (MMEA) from the decadal hindcasts and 20th century experiments with the observed Tmax and Tmin trends for California between 1960-1990 and 1980-2010. The study also compared the individual MEA and predictive skill for each AOGCM or ESM in order to evaluate various modeling aspects, such as ensemble size and assimilation methods. The first goal was to examine the new decadal hindcasts in order to evaluate the limitations and uncertainties of predicting natural decadal variability. The second goal was to produce a modeled decadal hindcast at a regional scale in order to provide valuable and versatile climate information for local resource managers and governments.

The observed annual and seasonal Tmax and Tmin trends for California were reconstructed using monthly data from 54 USHCNv2 stations in a method similar to that of Cordero et al. (2010). The CMIP5 model temperatures for California were reconstructed using bi-linear interpolation between the GCM`s grid resolution and the geographic position of each USHCN station. The individual MEAs were calculated by averaging all model ensemble simulations, and the experimental MMEA was calculated by taking the average of each individual MEA. The overall skill of each experiment and model was calculated by taking the absolute error between either the MEA or the MMEA and the observations for both the annual and seasonal Tmax and Tmin trends. The results showed that the MMEA for the decadal experiments displayed a slightly higher skill in predicting both the observed annual and seasonal temperature trends.

The box and whisker plots represent the median, quartiles, and outliers which were calculated from each experiment's MEA. The annual Tmax trends (Figure 2a) demonstrated a lower uncertainty compared to Tmin (Figure 2b.), and MAM showed the highest error while JJA had the lowest error between the MMEA and observations. Lastly, the overall variance in trends was larger in the 1960-1990 annual and seasonal decadal experiments as well as the 1980-2010 20th century annual and seasonal experiments. However, the internal variability was generally similar between the decadal and 20th century forecasts and both historical experiments during 1960-1990 and 1980-2010 were not statistically different at the 95% confidence interval.

Each AOGCM and ESM was analyzed to examine the tier 1 experiments of the CMIP5. First, the overall skill and variance was similar between the AOGCMs and ESM in both experiments. Secondly, a larger factor in decadal prediction is the model's initial ensemble size. An example of the error generated by a small ensemble size occurred with the 1980-2010 20th century experiments in which a model with one simulation (MIROC-ESM-CHEM) contributed to a very large experimental error. Lastly, the natural variability at the regional scale and decadal time frame is too large to sample with three or fewer simulations, while larger ensembles provide a more accurate prediction of the uncertainties and experimental model biases at a regional scale.

By using initial conditions, the decadal hindcasts have shown an improvement over prior datasets used to examine decadal variability and prediction. Prior studies have suggested that initial conditions would add to predictability by accounting for early model errors and observed ocean heat content (Meehl et al., 2009; Mehta et al., 2011).

More importantly, the CMIP5 decadal hindcast framework helped eliminate some of the inter-model experiment errors and biases (Taylor et al., 2012; Goddard et al., 2012). In California, previous studies have suggested that Tmax is controlled by local scale anthropogenic forcings while Tmin is controlled generally by natural large scale forcings (LaDochy et al., 2007; Cordero et al., 2010).

Although the hindcast experiments did not drastically improve the predictive skill in the decadal time-frame, the best performing model can still provide additional insight into future temperature trends for California. For this study, the MRI-CGCM3 out-performed all other models and the experimental MMEA during the historical time-frame and therefore can possibly provide better future predictions. The MRI-CGCM3 displayed drastically different trends between the future decadal and RCP4.5 experiments, especially for the future annual Tmin trends. The MRI-CGCM3 trends showed a large increase in the annual Tmin and both temperature trends in SON which suggests a longer growing season by 2035. Lastly, this conclusion demonstrates that using the initial conditions can lead to drastically different solutions, even when using a higher performing climate model.

Remaining studies into decadal prediction would include analyzing the Argo float model data since this includes in-situ observations of sub-surface temperature, currents, and salinity. It has long been determined that sub-surface ocean conditions play a very important role in decadal prediction due to the persistence of predictable ocean SST anomalies (Latiff and Barnett 1994; Mehta et al., 2011) and include ocean-atmospheric patterns such as the ENSO, PDO, and AMOC. Although the natural climate

variability is greater, shorter-term anthropogenic forcings from land use changes and aerosols will also play a role on decadal temperature trends. This is because these forcings are highly spatially heterogeneous and changing in time and will act to change the surface energy balance and alter the natural variability.

REFERENCES

Abatzoglou, J. T., K. T. Redmond, and L. M. Edwards, 2009: Classification of regional climate variability in the state of California. *J. Climate App. Meteor.*, **48,** 1527-1541.

Abatzoglou, J. T., and K. T. Redmond, 2007: Asymmetry between trends in spring and autumn temperature and circulation regimes over western North America. *Geophys. Res. Lett.,* **34,** *doi:10.1029/2007GL030891.*

Alfaro, E., A. Gershunov, and D. R. Cayan, 2006: Prediction of summer maximum and minimum temperature over the Central and Western United States: The roles of soil moisture and sea surface temperature. *Journal of Climate,* **19,** 1407-1421.

Arguez, A., I. Durre, S. Applequist, R. S. Vose, M. F. Squires, X. Yin, R. R. Heim Jr., and T. W. Owen, 2012: NOAA`S 1981-2010 U.S. climate normal. *Bull. Amer. Meteor. Soc.,* **93,** 1687-1697.

Branstator, G., and H. Teng, 2010: Two limits of initial-value decadal predictability in a CGCM. *Journal of Climate,* **23,** 6292-6311.

Barnston, A. G., M. K. Tippett, M. L. L`Heureux, S. Li, and D. G. DeWitt, 2012: Skill of real-time seasonal ENSO model predictions during 2002-11. Is our capability increasing? *Bull. Amer. Meteor. Soc.,* **93,** 631-651.

Bonfils, C., P. B. Duffy, B. D. Santer, T. M. L. Wigley, D. B. Lobell, T. J. Phillips, and C. Doutriaux, 2008: Identification of external influences on temperatures in California. *Climatic Change,* **87,** 43-55.

Brown, P. T., E. C. Cordero, and S. A. Mauget, 2012: Reproduction of 20th Century inter- to multi-decadal surface temperature variability in radiatively forced coupled climate models. *J. Geophys. Res.,* doi:10.1029/2011JD016864.

California Agricultural Statistics Review, cited 2013: Agricultural statistics overview [Available online at http://www.cdfa.ca.gov/Statistics/Agriculturalstatisticsoverview].

Cavalieri, D. J., C. L. Parkinson, and K. Y. Vinnikov, 2003: 30-year satellite record reveals contrasting Arctic and Antarctic decadal sea ice variability. *Geophys. Res. Lett.*, **30,** doi:10.1029/2003GL018031.

Cayan, D. R., E. P. Maurer, M. D. Dettinger, M. Tyree, and K. Hayhoe, 2008: Climate change scenarios for the California region. *Climatic Change,* **87,** S21-S42, doi:10.1007/s10584-007-9377-6.

Chang, G., Helium. Cited 2012. How to read and interpret a box plot. [Available online at http://www.helium.com/items/1309964-box-plot-basics].

Christy, J. R., W. B. Norris, K. Redmond, and K. P. Gallo, 2005: Methododolgy and results of calculating central California surface temperature trends: evidence of human-induced climate change? *J. Climate,* **19,** 548-563.

Cooley, H., J. Christian-Smith, and P. Gleick, 2009: Sustaining California agriculture in an uncertain future. *Pacific Institute,* IBSN: 1-893790-21-5.

Cordero, E. C., W. Kessomkiat, J. Abatzoglou, and S. A. Mauget, 2010: The identification of distinct patterns in California temperature trends. *Climatic Change,* doi.10.1007/s10584-011-0023-y.

Dettinger, M. D., and D. R. Cayan, 1994: Large-scale atmospheric forcing of recent trends toward early snowmelt runoff in California. *Journal of Climate,* **8,** 606-622.

Duffy, P. B., C. Bonfils, and D. Lobell, 2007: Interpreting recent temperature trends in California. *EOS Trans. AGU,* **88,** 409-410.

Ebbesmeyer, C. C., D. R. Cayan, D. R. McLain, F. H. Nichols, D. H. Peterson, and K. T. Redmond, 1991: 1976 step in the Pacific climate: Forty environmental changes between 1968-75 and 1977-1984. *In: Proc. 7th Ann. Pacific Climate Workshop.* Calif. Dept. of Water Resources, Interagency Ecol. Stud. Prog. Report 26.

Goddard, L., J. W. Hurrell, B. P. Kirtman, J. Murphy, T. Stockdale, and C. Vera, 2012: Two time scales for the price of one (almost). *Bull. Amer. Meteor. Soc.,* **93,** 621-629.

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

Jin, M. S., 2012: Developing an index to measure urban heat island effect using satellite land skin temperature and land cover observations. *Journal of Climate,* **25,** 6193-6200.

Jin, M., M. J. Shepherd, and W. Zheng, 2010: Urban surface temperature reduction via the urban aerosol direct effect: A remote sensing and WRF model sensitivity study, *Advances in Meteor.*, **2010,** doi:10.1155/2010/681587.

Kim, J. S., K. Y. Kim, and S. W. Yeh, 2012: Statistical evidence for the natural variation of the central Pacific El Nino. *J. Geophys. Res.,* **117,** C06014, doi:*10.1029/2012JC008003*.

LaDochy, S., R. Medina, and W. Patzert, 2007: Recent California climate variability: spatial and temporal patterns in temperature trends. *Clim. Res.,* **33,** 159-169.

Latif, M., and T. P. Barnett, 1994: Causes of decadal climate variability over the North Pacific and North America. *Science*, **226**, 634-637.

Latif, M., and T. P. Barnett, 1996: Decadal climate variability over the North Pacific and North America: Dynamics and predictability. *Journal of Climate,* **9,** 2407-2423.

Lebassi, B. J. Gonzalez, D. Fabris, E. Maurer, N. Miller, C. Milesi, and R. Bornstein, 2010: Observed 1970-2005 cooling of summer daytime temperatures in coastal California. *Journal of Climate,* doi: 10.1175/ 2008JCLI2111.1.

Mantua, N. J., 2002: Pacific decadal oscillation (PDO). *Encyclopedia of Global Environmental Change,* **1,** 592-594.

Meehl, G. A. and Coauthors, 2009: Decadal prediction. Can it be skillful? *Bull. Amer. Meteor. Soc.,* **90,** 1467-1481.

Meehl, G. A., and Coauthors, 2012: Climate system response to external forcings and climate change projections in CCSM4. *Journal of Climate*, **25,** 3661-3683.

Meehl, G. A., J. M. Arblaster, and C. Tebaldi, 2007: Contributions of natural and anthropogenic forcing to changes in temperature extremes over the United States. *Geophys. Res. Lett.,* **34,** doi: 10.1029/2007GL030948.

Meene, M. J., and C. N. Williams, Jr., 2009: Homogenization of temperature series via pairwise comparisons. *Journal of Climate,* **22,** 1700-1717.

Meene, M. J., C. N. Williams, and R. S. Vose, 2009: The United States historical climatology network monthly temperature data – version 2. *Bull. Amer. Meteor. Soc.,* **90,** 993-1107.

Mehta, V., G. Meehl, L. Goddard, J. Knight, A. Kumar, M. Latif, T. Lee, A. Rosati, and D. Stammer, 2011: Decadal climate predictability and prediction. Where are we? *Bull. Amer. Meteor. Soc.,* **92,** 637-640.

Miller, A. J., D. R. Cayan, T. P. Barnett, N. E. Graham, and J. M. Oberhuber, 1994: Interdecadal variability of the Pacific Ocean: model response to observed heat flux and wind stress anomalies. *Climate Dyn.,* **9,** 287-302.

Msadek, R., C. Frankignoul, and Z. X. Li, 2010: Mechanisms of the atmospheric response to North American multidecadal variability: a model study. *Climate Dyn.,* doi:10.1007/s00382-010-0958-0.

Robock, A., 2000: Volcanic eruptions and climate. *Review in Geophys.,* **38,** 191-219.

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

Srokosz, M., M. Baringer, H. Bryden, S. Cunningham, T. Delworth, S. Lozier, J. Marotzke, and R. Sutton, 2012: Past, present, and future changes in the Atlantic Meridional overturning circulation. *Bull. Amer. Meteor. Soc.,* **93,** 1663-1676.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experimental design. *Bull. Amer. Meteor. Soc.,* **93,** 485-497.

Trenberth, K. E., and Coauthors, 2007: Observations: Surface and atmospheric climate change. *Climate Change 2007: The Physical Science Basis.* Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (Eds.)]. Cambridge University Press, Cambridge, UK.

Van Loon, H., and D. J. Shea, 1999: A probable signal of the 11-year solar cycle in the troposphere of the northern hemisphere. *Geophys. Res. Lett.,* **26,** 2893-2896.

Van Loon, H., G. A. Meehl, and D. J. Shea, 2007: Coupled air-sea response to solar forcing in the Pacific region during northern winter. *J. Geophys. Res.,* **112,** doi: 10.1029/2006JD007378.

Zanchettin, D., C. Timmreck, H. F. Graf, A. Rubino, S. Lorenz, K. Lohmann, K. Kueger, and J. H. Jungclaus, 2012a: Bi-decadal variability excited in the coupled ocean-atmosphere system by strong tropical volcanic eruptions. *Climate Dyn.,* **39,** 419-444.
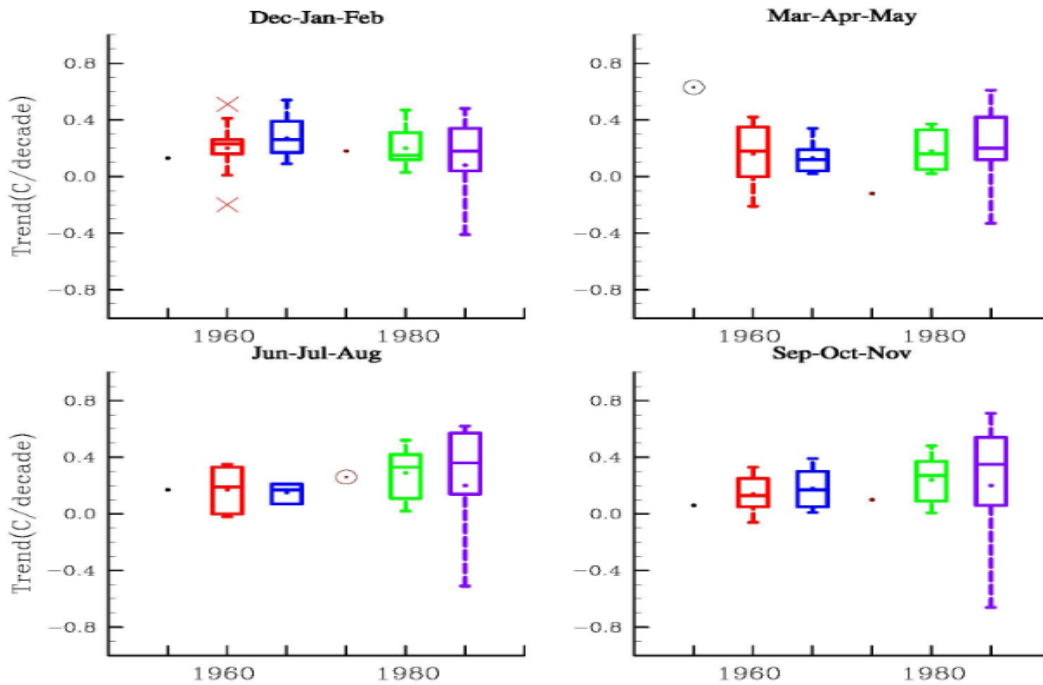
# APPENDIX A: Historical Seasonal Graphs

Seasonal Tmax
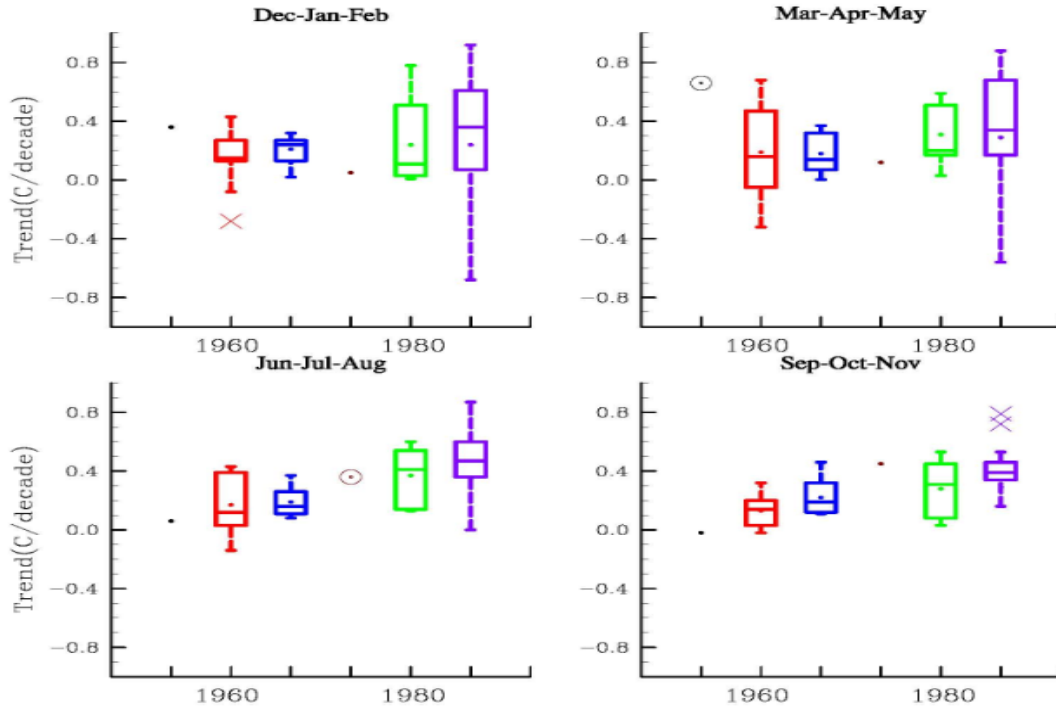
The seasonal historical Tmax trends for DJF (top-left), MAM (top-right), JJA (bottom-left), and SON (bottom-right). The black and brown dots represent the 1960-1990 and 1980-2010 trend observations; circled dots display trends that are statistically significant at the 95% confidence interval. The red (1960-1990) and green (1980-2010) boxplots and dots represent the decadal trends and MMEA, while the blue (1960-1990) and purple (1980-2010) boxplots and dots represent the 20th century trends and MMEA. X represents outlying trends
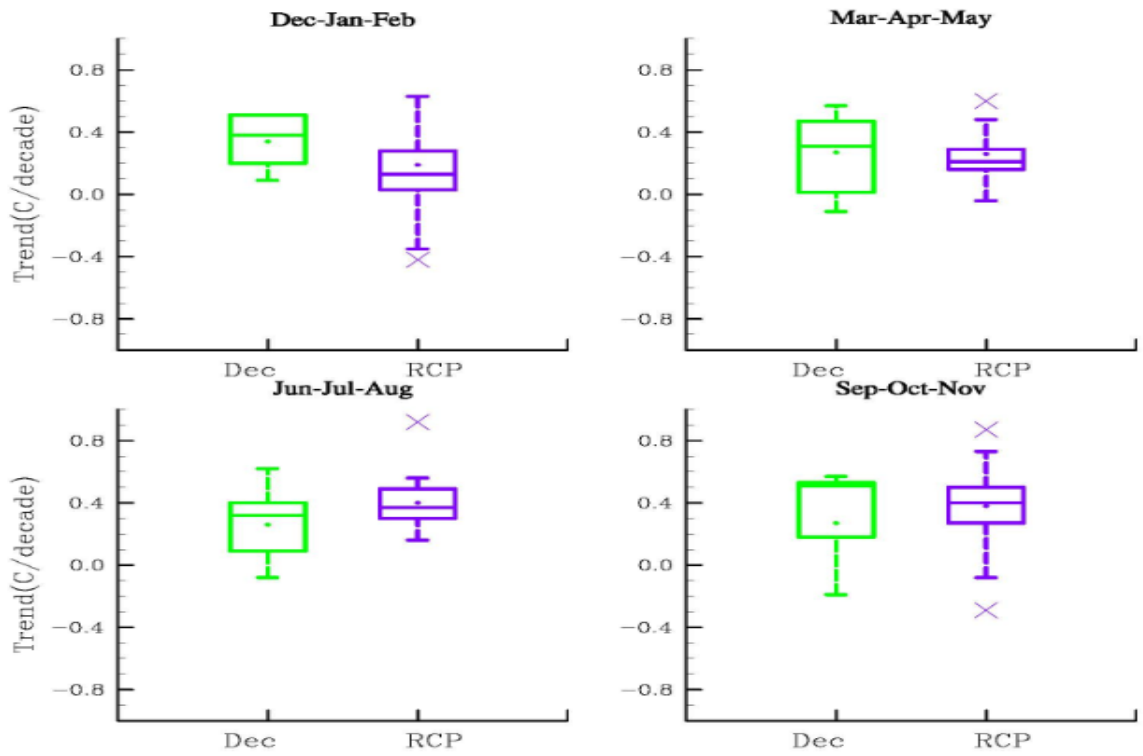
Seasonal Tmin

The seasonal historical Tmin trends for DJF (top-left), MAM (top-right), JJA (bottom-left), and SON (bottom-right). The black and brown dots represent the 1960-1990 and 1980-2010 trend observations; circled dots display trends that are statistically significant at the 95% confidence interval. The red (1960-1990) and green (1980-2010) boxplots and dots represent the decadal trends and MMEA, while the blue (1960-1990) and purple (1980-2010) boxplots and dots represent the 20th century trends and MMEA. X represents outlying trends.

APPENDIX B: Future Seasonal Graphs
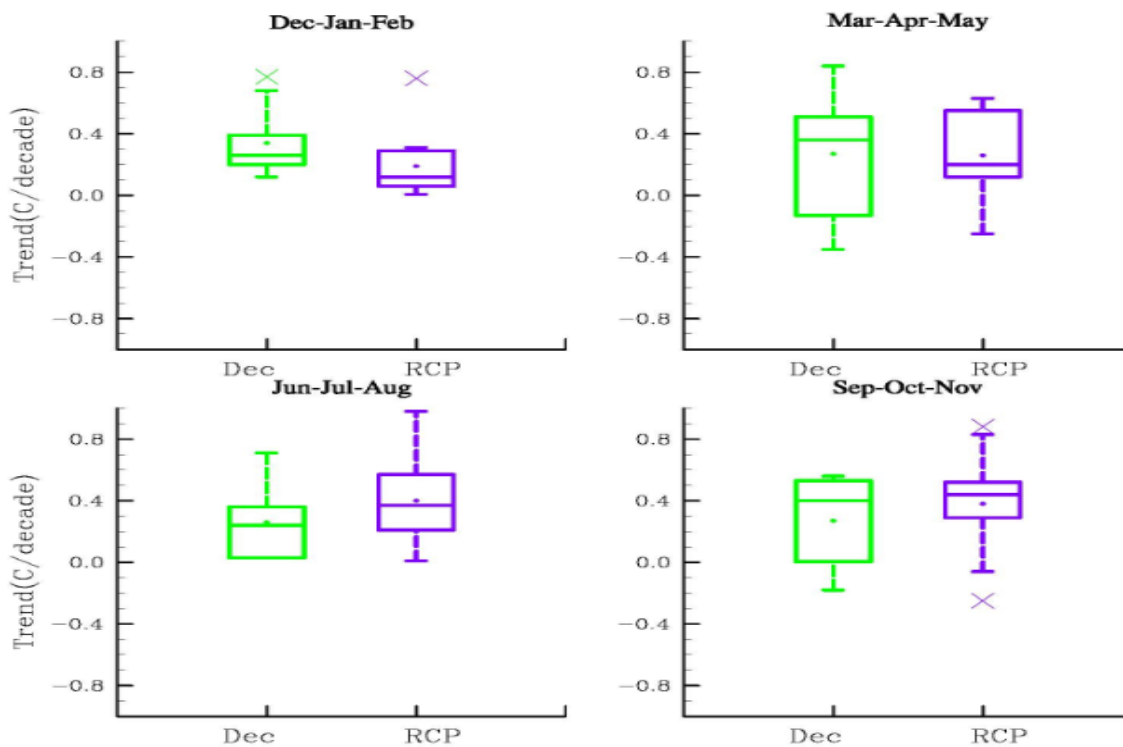
The seasonal future Tmax trends for DJF (top-left), MAM (top-right), JJA (bottom-left), and SON (bottom-right). The green boxplot and dot represent the average median trend and MMEA for the decadal 2005 experiment. The purple boxplot and dot shows the average median trend and MMEA for the RCP4.5 experiment. The X shows outlying trends

The seasonal future Tmin trends for DJF (top-left), MAM (top-right), JJA (bottom-left), and SON (bottom-right). The green boxplot and dot represent the average median trend and MMEA for the decadal 2005 experiment. The purple boxplot and dot shows the average median trend and MMEA for the RCP4.5 experiment. The X shows outlying trends