



## Assessing the 20th Century Performance of Global Climate Models and Application to Climate Change Adaptation Planning

Item type	text; Electronic Dissertation
Authors	Geil, Kerrie L.
Publisher	The University of Arizona.
Rights	Copyright © is held by the author. Digital access to this material is made possible by the University Libraries, University of Arizona. Further transmission, reproduction or presentation (such as public display or performance) of protected items is prohibited except with permission of the author.
Downloaded	15-Apr-2017 07:40:39
Link to item	<a href="http://hdl.handle.net/10150/623015">http://hdl.handle.net/10150/623015</a>

ASSESSING THE 20<sup>TH</sup> CENTURY PERFORMANCE OF GLOBAL CLIMATE MODELS  
AND APPLICATION TO CLIMATE CHANGE ADAPTATION PLANNING

by

Kerrie Geil

---

Copyright © Kerrie Geil 2017

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF HYDROLOGY AND ATMOSPHERIC SCIENCES

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

WITH A MAJOR IN ATMOSPHERIC SCIENCES

In the Graduate College

THE UNIVERSITY OF ARIZONA

2017

THE UNIVERSITY OF ARIZONA  
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Kerrie Geil, titled **Assessing the 20<sup>th</sup> Century Performance of Global Climate Models and Application to Climate Change Adaptation Planning** and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

\_\_\_\_\_ Date: 12/2/2016  
Xubin Zeng

\_\_\_\_\_ Date: 12/2/2016  
Daniel Ferguson

\_\_\_\_\_ Date: 12/2/2016  
Michael Crimmins

\_\_\_\_\_ Date: 12/2/2016  
Stuart Marsh

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

\_\_\_\_\_ Date: 12/2/2016  
Dissertation Director: Xubin Zeng

## STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of the requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Kerrie L. Geil

## ACKNOWLEDGEMENTS

Many people have contributed to this dissertation and to the successful completion of my doctoral degree. Sincerest thanks to my PhD advisor, Dr. Xubin Zeng, for recognizing my strengths and weaknesses, for always being available to provide excellent guidance, and for giving me an unusual amount of autonomy. I have very much enjoyed my time under his advisement. It is his expertise, patience, and support that made this dissertation possible. I would like to express gratitude to Dr. Dan Ferguson, who helped steer me toward the career path that I was always meant to find. Dan has been an amazing mentor whose knowledge and perspectives have been particularly impactful. I am so thankful for his encouragement, guidance, and support. I would also like to thank Dr. Yolande Serra for her advisement and support during the first few years of this research and for her enthusiasm and patience.

Special thanks to my PhD committee members Dr. Stuart Marsh, Dr. Mike Crimmins, Dr. Dan Ferguson, Dr. Francina Dominguez, and Dr. Yolande Serra for pushing me to achieve and for their interdisciplinary guidance.

Additional thanks to the Land-Atmosphere-Ocean Interaction (LAOI) research group, led by Dr. Zeng, for the comradery and lively debates over the past six years. I am truly appreciative of the open and honest space we have created for constructive criticism, which makes us all better scientists.

I am grateful to the staff, investigators, and affiliates of the Climate Assessment for the Southwest who have shown me what successful transdisciplinary climate research looks like. I am especially appreciative of Ben McMahan for guiding my participation in research projects and of Ardeth Barnhart for being a role model in cross-boundary research facilitation.

This research was supported by the National Oceanic and Atmospheric Administration MAPP grant GC10-398, National Science Foundation grant AGS-0944101, National Aeronautics and Space Administration grant NNX14AM02G, and by the National Oceanic and Atmospheric Administration's Climate Program Office through grant NA12OAR4310124 with the Climate Assessment for the Southwest program at the University of Arizona.

Finally, I am thankful for the support of my parents and close friends who have always had confidence that I will achieve whatever I set out to. They have been encouraging, understanding, and emotionally supportive through this and all of my life's pursuits. Thank you.

## TABLE OF CONTENTS

<b>LIST OF TABLES .....</b>	<b>7</b>
<b>LIST OF FIGURES .....</b>	<b>8</b>
<b>ABSTRACT .....</b>	<b>12</b>
<b>CHAPTER 1: INTRODUCTION .....</b>	<b>14</b>
1.1 USING GLOBAL CLIMATE MODELS TO INFORM CLIMATE ADAPTATION PLANNING: MODELING ISSUES.....	14
1.1.a <i>Model Uncertainty, Natural Climate Variability, Scenario Uncertainty</i> .....	14
1.1.b <i>Model Evaluation</i> .....	16
1.1.c <i>Multimodel Ensemble Averaging</i> .....	18
1.1.d <i>Model Independence</i> .....	20
1.2 USING GLOBAL CLIMATE MODELS TO INFORM CLIMATE ADAPTATION PLANNING: APPLICATION ISSUES.....	21
1.2.a <i>Saliency, Credibility, Legitimacy</i> .....	22
1.2.b <i>Knowledge Exchange</i> .....	23
1.2.c <i>Ethics</i> .....	25
1.3 OBJECTIVES .....	27
<b>CHAPTER 2: PRESENT STUDY .....</b>	<b>29</b>
2.1 ASSESSMENT OF CMIP5 MODEL SIMULATIONS OF THE NORTH AMERICAN MONSOON SYSTEM .	29
2.2 QUANTITATIVE CHARACTERIZATION OF SPURIOUS NUMERICAL OSCILLATIONS IN 48 CMIP5 MODELS .....	31
2.3 EVALUATION OF THE MEAN STATE, TRENDS, AND VARIABILITY IN CMIP5 MONTHLY SURFACE AIR TEMPERATURE AT REGIONAL SCALES OVER THE U.S. FOR APPLICATION TO CLIMATE ADAPTATION PLANNING.....	33
2.4 FUTURE RESEARCH DIRECTIONS .....	34
<b>REFERENCES .....</b>	<b>38</b>
<b>APPENDIX A. ASSESSMENT OF CMIP5 MODEL SIMULATIONS OF THE NORTH AMERICAN MONSOON SYSTEM .....</b>	<b>42</b>
ABSTRACT.....	43
1. INTRODUCTION .....	43
2. OBSERVATIONAL DATA AND MODEL SIMULATIONS .....	45
a. <i>Observational data and reanalysis</i> .....	45
b. <i>Coupled general circulation model simulations</i> .....	45
3. ANNUAL CYCLE OF PRECIPITATION.....	45
4. SEASONAL SPATIAL CORRELATION OF MONSOON VARIABLES.....	47
5. MONSOON ONSET AND RETREAT.....	50
6. COMPOSITES.....	52
7. SUMMARY .....	54

ACKNOWLEDGMENTS .....	56
REFERENCES .....	56
<b>APPENDIX B. QUANTITATIVE CHARACTERIZATION OF SPURIOUS NUMERICAL OSCILLATIONS IN 48 CMIP5 MODELS .....</b>	<b>58</b>
ABSTRACT.....	59
1. INTRODUCTION .....	59
2. MODEL SIMULATIONS, OBSERVATIONS, AND METHODS .....	60
3. RESULTS.....	62
4. SUMMARY .....	65
ACKNOWLEDGMENTS.....	65
REFERENCES .....	65
SUPPORTING INFORMATION.....	67
<b>APPENDIX C. EVALUATION OF THE MEAN STATE, TRENDS, AND VARIABILITY IN CMIP5 MONTHLY SURFACE AIR TEMPERATURE AT REGIONAL SCALES OVER THE U.S. FOR APPLICATION TO CLIMATE ADAPTATION AND RESILIENCE EFFORTS ....</b>	<b>76</b>
ABSTRACT.....	77
1. INTRODUCTION .....	78
2. MODEL SIMULATIONS, OBSERVATIONS, AND METHODS .....	80
3. RESULTS .....	82
<i>a. Long-term mean state</i> .....	82
<i>b. Trends</i> .....	84
<i>c. Variability</i> .....	86
4. DISCUSSION.....	88
<i>a. Correlation between historical biases</i> .....	88
<i>b. Correlation between historical bias and projected change in the southwest</i> .....	89
<i>c. Choosing better performing multimodel ensembles for the southwest</i> .....	90
<i>d. 20<sup>th</sup> century performance and 21<sup>st</sup> century projections for the southwest US using three     different multimodel ensembles</i> .....	92
5. CONCLUSIONS .....	93
ACKNOWLEDGMENTS.....	96
APPENDIX: MODEL BIAS RANKINGS .....	96
REFERENCES .....	101
TABLES AND FIGURES.....	103

## LIST OF TABLES

TABLE A.1. CGCM model acronym, model expansion, three-letter model code, modeling center name, origin country, CMIP5 ensemble member identifier, and atmospheric (AGCM) and oceanic (OGCM) model grid resolution. Grid resolution format is (number of longitudinal grids) x (number of latitudinal grids) L (number of vertical levels). . . . .	46
TABLE A.2. Top and bottom five model rankings of spatial correlation $r$ for precipitation, 850-hPa index (geopotential height and winds), and 500-hPa index (geopotential height and winds), as compared to TRMM and ERA-Interim data, for the entire monsoon season, the development stage, the mature stage, and the decay stage over the extended NAMS domain. . . . .	50
TABLE A.3. Observed dates of monsoon onset and retreat for the core NAMS domain using TRMM data. . . . .	51
TABLE A.4. Model median dates of monsoon onset and retreat with corresponding lag from observations and standard deviation (in number of days) using the observation-defined absolute threshold criteria for the core NAMS domain. Observational information is listed in the last row for reference. . . . .	52
TABLE A.5. As in Table A.4, but using the model-relative threshold criteria. . . . .	54
TABLE B.S1. Model Information. . . . .	73
TABLE C.1. CMIP5 models used for this study. . . . .	103
TABLE C.2. Tmin and Tmax RMBAD MME models and total points for each model. . . . .	105
TABLE C.3. Tmin and Tmax TOP5 MME models and total points for each model. . . . .	105
TABLE C.A1. Model ranking by the average of the absolute value of seasonal mean bias for the period JAN 1900 – DEC 2005. . . . .	97
TABLE C.A2. Model ranking by absolute value of the JAN 1900 – DEC 2005 trend bias. Models with biases significantly different from observations at the 90% confidence level are colored brown. . . . .	98
TABLE C.A3. Model ranking by absolute value of the 1900 – 2005 max-month trend bias. Models with biases significantly different from observations at the 90% confidence level are colored brown. . . . .	99
TABLE C.A4. Model ranking by absolute value of the JAN 1900 – DEC 2005 monthly time series standard deviation bias. Here, the MME is the mean of individual model standard deviation values.	

Models with biases significantly different from observations at the 95% confidence level are colored brown. .... 100

**LIST OF FIGURES**

FIGURE A.1. Core NAMS domain (23.8758–28.8758N, 108.8758–104.8758W) and extended NAMS domain (15.1258–34.8758N, 119.8758–90.1258W), shown over regional topography...44

FIGURE A.2. Annual precipitation cycle (a) rms error (mm/day), (b) bias (%), and (c) phase lag yielding maximum correlation (months) for all models with respect to monthly P-NOAA observations over the core NAMS domain. Multimodel median and mean values are also shown in each panel. All metrics are calculated using 12 monthly climatological values for the period 1979–2005. Phase lag of +1 means the highest correlation of all 12 months is between monthly observations ordered January through December and monthly model precipitation ordered February through January. .... 48

FIGURE A.3. Annual precipitation cycle over the core NAMS domain for all models grouped by (a) small phase lag (lag = 0 months), (b) moderate phase lag (lag = 1 month), and (c) large phase lag (lag > 1 month). The multimodel mean (gray dashed line) for each category is shown in (a)–(c). (d) The all-model mean (dashed line) and spread (shading). Colors represent different modeling centers and solid vs dashed lines of the same color differentiate models from a common center. .... 49

FIGURE A.4. Climatological (1998–2010) area-averaged daily (thin line) TRMM precipitation index and 5-day running mean (heavy line) in millimeters per day for the core NAMS domain....51

FIGURE A.5. Daily precipitation climatology over the core NAMS domain for (top left) observations (1998–2010), (top right) multimodel mean (1979–2005), and 14 models (1979–2005). Short dashed lines depict the absolute threshold (1.3 mm/day), and long dashed lines depict model-relative thresholds for defining yearly monsoon onset and retreat dates. .... 53

FIGURE A.6. Monthly climatological precipitation (color shading), 850-mb geopotential height (contours), and 850-mb winds (vectors) for (top) the data, (middle) the best model composite, and (bottom) the worst model composite for (left) the development stage month of June, (center) the mature stage month of August, and (right) the decay stage month of October. .... 55

FIGURE B.1. Surface pressure (mb) for nine spectral models (spectral resolution increasing from top left to bottom right) shows the large range in wavelength and amplitude of the spurious numerical oscillations (aka Gibbs oscillations in spectral models). The middle plot displays the location of the transect (horizontal black line) used to quantify numerical oscillations in subsequent figures. .... 61

FIGURE B.2. Transects for a spectral model with (top) large, (middle) moderate, and (bottom) small amplitude surface pressure oscillations. The left column shows 12-month average transects during the 27 year study period, where the mean of each transect has been removed. The right column shows seasonal average transects for five separate years (mean removed). ..... 62

FIGURE B.3. Climatological transects for eight satellite observational data sets (see section 2). From top left to bottom right: Cross-Calibrated Multi-Platform (CCMP) zonal (U) and meridional wind (V), Tropical Rainfall Measuring Mission (TRMM) precipitation (pr), Atmospheric Infrared Sounder (AIRS) specific humidity (Q) and air temperature (T) at 1000 mb, Clouds and the Earth’s Radiant Energy System (CERES) downwelling longwave (rlds) and shortwave (rsds) radiation at the surface, and Moderate Resolution Imaging Spectroradiometer (MODIS) total cloud amount (clt). ..... 63

FIGURE B.4. RMSD percentile values of spectral and nonspectral models (refer to text for explanation of RMSD computation) for 13 variables, which include surface pressure (ps), near-surface u-wind (uas), and v-wind (vas), vertical velocity at 925 mb (wap 925) and 500 mb (wap 500), precipitation (pr), surface specific humidity (huss), surface air temperature (tas), surface incoming solar radiation (rsds) and incoming longwave radiation (rlds), geopotential height at 925 mb (zg 925) and 500 mb (zg 500), and total cloud amount (clt). RMSD values for observational data are shown on the bottom row. The red text indicates values greater than or equal to the observed value multiplied by a factor of 5, whereas the blue text indicates values below this threshold. Model results are shown in black for variables when there are no observations for comparison. .... 63

FIGURE B.5. Same as Figure B.4 except for the RMSD:IAV ratio. The red and orange text highlight values greater than unity and one half, respectively. Values less than one half are shown in black. .... 64

FIGURE B.S1. Climatological precipitation transects (black) and the corresponding running mean curves (red) used to compute RMSD values; shown for a spectral model (CMCC-CM; top), a finite volume model (GFDL-ESM2M; middle), and TRMM data (bottom). The number of points used to compute each running mean is labeled as ‘npts’. Note that the transect approaches the South American coast as one moves from west to east (left to right). ..... 68

FIGURE B.S2. The region of model transects shown over climatological (1979-2004) sea level pressure from the HadSLP2 dataset. The transects bisect the South Pacific subtropical surface high. Transects are chosen as the model latitude closest to 29° S, which ranges between 27.8° S and 30.3° S. The westernmost transect longitude is the closest model grid point approaching 112° W from the west and the easternmost longitude point is one to two model grid points westward of the first continental model grid point (dependent on the number of grid points used to compute the running mean). ..... 69

FIGURE B.S3. Examples of model transects that result in large values of RMSD and RMSD:IAV ratio, even when large Gibbs oscillations are not present along the entire transect. Note that all

transects include only ocean points. The culprits include (a) a premature decrease in surface pressure approaching land, (b) one large oscillation in surface pressure near the ocean-land transition, and (c-d) unrealistic steep gradients in meridional surface wind near the ocean-land transition. The CCMP meridional surface wind observations are shown in (e) for comparison to (c-d). . . . . 70

FIGURE B.S4. RMSD values for (a) spectral models and (b) non-spectral models for 13 variables: surface pressure (ps), near-surface u-wind (uas) and v-wind (vas), vertical velocity at 925 mb (wap925) and 500 mb (wap500), precipitation (pr), surface specific humidity (huss), surface air temperature (tas), surface incoming solar radiation (rsds) and incoming longwave radiation (rlds), geopotential height at 925 mb (zg925) and 500 mb (zg500), and total cloud amount (clt). Observational values are shown in black on the bottom row of each panel. Black circles indicate fields that are not available. Vertical red dashed lines represent the observational data value multiplied by a factor of 5. Red labels show values beyond each abscissa maximum and green labels show values for bars that are not easily visible. Vertical grid lines are shown with gray dashes. . . . . 71

FIGURE B.S5. Same as Figure B.S4, except for the RMSD:IAV ratio. . . . . 72

FIGURE C.1. The eight study areas shown over ASTER global DEM topography on our 1°x1° common grid. AK=Alaska, NW=northwest, SW=southwest, GP=great plains, MW=midwest, SE=southeast, NE=northeast, and US=United States. . . . . 106

FIGURE C.2. Bias in seasonal and annual mean Tmin and Tmax for the period 1900-2005 (where W, S, S, F, A on the x-axis stands for winter, spring, summer, fall, and annual, respectively). The MME-average is shown with a square marker and individual models are shown with dots along grey vertical model spread bars. Colored markers indicate bias significance at the 90% confidence level after adjusting for serial lag-1 autocorrelation. . . . . 107

FIGURE C.3. May time series of Tmin for the southwest region, where the solid line shows the observations, the blue dash shows the MME-average, and the gray shade shows the model spread. The green dash shows an example of one individual model. . . . . 108

FIGURE C.4. Annual average (a) and monthly average trends (b) in Tmin and Tmax. Only the month with the greatest MME-average trend bias is shown for each region in (b). Large circles are observations, large squares are the MME-average values, and individual models are shown with dots along grey vertical model spread bars. Filled large markers indicate MME-average and observed trends that are significant at the 90% confidence level, after adjusting for serial lag-1 autocorrelation. Color indicates that the modeled and observed trends are significantly different from each other at the 90% confidence level, after adjusting for serial lag-1 autocorrelation. . . . . 109

FIGURE C.5. Alaska max-month (October) trend in Tmax for (a) observations, (b) the 42-model MME-average, (c) the model with the smallest trend bias (MRIC3), and (d) the model with the largest trend bias (CESMF). Filled markers indicate that the trend is significant at the 90%

confidence level, after adjusting for serial lag-1 autocorrelation ( $p < 0.1$ ), while open markers are used for insignificant trends ( $p \geq 0.1$ ). The Pearson linear correlation coefficient ( $r$ ) of each trend is shown for reference. .... 110

FIGURE C.6. The time series standard deviation (a) where circles are the observations, dots along grey vertical model spread bars are individual models, and squares represent the average of all individual models. Color indicates models with variability that is significantly different than observations at the 95% confidence level, adjusted for serial lag-1 autocorrelation. (b) The detrended Tmax AK time series with mean removed for the 1960's, where the black line is the observed time series, the blue dashed line is the model with the smallest variability bias (ACC13) for the 106 year study period, and the green dashed line is the model with the largest variability bias (CMCCE). Gray shading shows the model spread. .... 111

FIGURE C.7. Linear regression between 20<sup>th</sup> century biases for 42 models. (a) Regression between annual trend bias and max-month trend bias for the GP region. (b) Regression between max-month trend bias and variability bias for the US region. (c) Regression between seasonal average mean bias and variability bias for the SW region. The regression line is shown in solid black with the corresponding correlation coefficient,  $r$ , located at the top center of each plot. A second regression that excludes the 10% of models (4) with the largest magnitude x-axis variable bias is shown with a dashed line and the corresponding  $r$  value is in parentheses at the top right of each plot. Colored  $r$  values represent significance at the 95% confidence level, adjusted for reduced model independence. .... 112

FIGURE C.8. Linear regression of 20<sup>th</sup> century bias to 21<sup>st</sup> century RCP 8.5 (a,b,c) and RCP 4.5 (d,e,f) projected change for the southwestern US using 33 models. Regression mean bias versus variability change (a,d), spring mean bias versus variability change (b,e), and annual trend bias vs annual trend change (c,f). The regression line is shown in solid black with the corresponding correlation coefficient,  $r$ , located at the top center of each plot. A second regression that excludes the 10% of models (3) with the largest magnitude x-axis variable bias is shown with a dashed line and the corresponding  $r$  value is in parentheses at the top right of each plot. Colored  $r$  values represent significance at the 95% confidence level, adjusted for reduced model independence...113

FIGURE C.9. Comparison of historical biases between the TOP5, RMBAD and ALL-model MMEs, for Tmin (top) and Tmax (bottom) using annual mean, March-April-May seasonal mean, annual trend, May trend, and standard deviation bias metrics. Large markers indicate the MME-average and individual models are shown with dots along grey model spread bars. .... 114

FIGURE C.10. Same as Figure 9, except for RCP 8.5 projected change. .... 115

## ABSTRACT

Rapid environmental changes linked to human-induced increases in atmospheric greenhouse gas concentrations have been observed on a global scale over recent decades. Given the relative certainty of continued change across many earth systems, the information output from climate models is an essential resource for adaptation planning. But in the face of many known modeling deficiencies, how confident can we be in model projections of future climate? It stands to reason that a realistic simulation of the present climate is at least a necessary (but likely not sufficient) requirement for a model's ability to realistically simulate the climate of the future. Here, I present the results of three studies that evaluate the 20<sup>th</sup> century performance of global climate models from phase 5 of the Coupled Model Intercomparison Project (CMIP5).

The first study examines precipitation, geopotential height, and wind fields from 21 CMIP5 models to determine how well the North American monsoon system (NAMS) is simulated. Models that best capture large-scale circulation patterns at low levels usually have realistic representations of the NAMS, but even the best models poorly represent monsoon retreat. Difficulty in reproducing monsoon retreat results from an inaccurate representation of gradients in low-level geopotential height across the larger region, which causes an unrealistic flux of low-level moisture from the tropics into the NAMS region that extends well into the post-monsoon season.

The second study examines the presence and severity of spurious Gibbs-type numerical oscillations across the CMIP5 suite of climate models. The oscillations can appear as unrealistic spatial waves near discontinuities or sharp gradients in global model fields (e.g., orography) and have been a known problem for decades. Multiple methods of oscillation reduction exist; consequently, the oscillations are presumed small in modern climate models and hence are rarely

addressed in recent literature. Here we quantify the oscillations in 13 variables from 48 global climate models along a Pacific ocean transect near the Andes. Results show that 48% of nonspectral models and 95% of spectral models have at least one variable with oscillation amplitude as large as, or greater than, atmospheric interannual variability.

The third study is an in-depth assessment model simulations of 20th century monthly minimum and maximum surface air temperature over eight US regions, using mean state, trend, and variability bias metrics. Transparent model performance information is provided in the form of model rankings for each bias type. A wide range in model skill is at the regional scale, but no strong relationships are seen between any of the three bias types or between 20<sup>th</sup> century bias and 21<sup>st</sup> century projected change. Using our model rankings, two smaller ensembles of models with better performance over the southwestern U.S. are selected, but they result in negligible differences from the all-model ensemble in the average 21<sup>st</sup> century projected temperature change and model spread. In other words, models of varied quality (and complexity) are projecting very similar changes in temperature, implying that the models are simulating warming for different physical reasons. Despite this result, we suggest that models with smaller 20<sup>th</sup> century biases have a greater likelihood of being more physically realistic and therefore, more confidence can be placed in their 21<sup>st</sup> century projections as compared to projections from models that have demonstrably poor skill over the observational period. This type of analysis is essential for responsibly informing climate resilience efforts.

## CHAPTER 1: INTRODUCTION

### 1.1 USING GLOBAL CLIMATE MODELS TO INFORM CLIMATE ADAPTATION PLANNING: MODELING ISSUES

Increasingly, science and technology are called upon to inform resilience efforts with respect to climate change and other social or environmental issues (President Barack Obama 2016, Kintisch 2006, UNESCO 2000). In the context of using climate model information responsibly to inform adaptation planning decisions, scientists must consider a number of remaining modeling issues and ambiguities.

#### *1.1.a Model Uncertainty, Natural Climate Variability, Scenario Uncertainty*

Models are only approximations of reality and therefore, multiple types of uncertainty will always exist. For climate simulations, uncertainty arises not only from the modeling process itself, but also from natural climate variability and scenario uncertainty (projections of future greenhouse gas (GHG) emissions).

In global climate models, uncertainty occurs for a variety of reasons, including limited theoretical and observational understanding of some earth system processes. For example, scientists still aren't sure of what exactly triggers El Nino-Southern Oscillation (ENSO) events, so while ENSO variability appears in some models, others still struggle with realistic ENSO simulation (Guilyardi 2015). This limitation also appears in the varied response across models to identical scenario forcing (Dessai et al. 2005). Structural uncertainty, also referred to as model inadequacy, is caused by difficulty in mathematically describing known processes accurately or

because certain processes (e.g. dynamic vegetation) may be missing or approximated (Knutti et al. 2010). Parametric uncertainty, arises because sub-grid scale processes such as cloud microphysics, convection, turbulence, and vegetation processes must be estimated or empirically derived and parameterized to the model grid scale (Knutti et al 2010). Additionally, some modeling centers over the past decade have transformed their already complex coupled global climate models (CGCMs) into even more complex earth system models (ESMs). The new capabilities of CMIP5 ESMs include parameterizations for aerosol chemistry and biogeochemical processes such as carbon and nitrogen cycling (Taylor et al. 2012, IPCC 2014). These model uncertainties and differing levels of model complexity contribute to the spread in CMIP5 model ability to simulate present-day and future climate.

Natural climate variability can be a large source of uncertainty, especially at regional spatial scales and on multi-decadal and shorter timescales (Sillmann et al. 2014, Northrop et al. 2014). Long-term model integrations from the CMIP5 suite are free-running, meaning that they are not initialized and forced with observed sea surface temperature and other observed conditions. Instead, they are spun up for a few hundred years to a quasi-equilibrium state using a plausible pre-industrial initialization. Then, this state becomes the new initialization for running the integrations forward through the present, forced only with observed time-varying atmospheric and land surface conditions (Stouffer et al. 2004, Taylor et al. 2012, Taylor et al. 2009, IPCC 2014). Because of this free-running nature, there is no reason for model natural variability to align with observed natural variability. For example, we can't expect free-running models to mirror the timing and intensity of observed ENSO events, which can result in short-term model climatologies, trends, and variability that differ from observations (Fyfe et al. 2013).

Model projections of future climate are driven by multiple scenarios of GHG emissions that prescribe future time-varying atmospheric and land surface conditions (Taylor et al. 2012). Scenario uncertainty arises from the fact that there is no way to know which of these future scenarios, if any, will align most closely with reality. Future global emissions depend on many unpredictable policy choices, technological developments, and economic considerations.

Considering these many uncertainties, how close to reality should we expect a global climate model to be? And, what level of complexity is required for realistic simulation of present-day and future climate conditions? Simple models of other complex systems have been shown to be reliable for decision making (Knutti 2010) and a more complex or higher resolution climate model may not be necessary to inform certain resilience efforts (Dunn et al. 2015). As climate models are increasingly used to inform adaptation decisions, these considerations should be included in discussions of uncertainty with decision makers.

### *1.1.b Model Performance Evaluation*

Unlike weather model forecasts that can be verified within a matter of days, the decadal-to century- long lead times of climate model projections make verification impossible on decision making time scales. Therefore, the credibility of climate models must be established by evaluating how well they simulate past and present-day climate conditions. It stands to reason that a realistic simulation of present climate conditions is at least a necessary (but likely not sufficient) requirement for a model's ability to realistically simulate (for the right reasons) the climate of the future. Clearly, understanding model performance as compared to observations on a variety of spatial and temporal scales during the 20th century is crucial to understanding model capability

and responsibly informing climate resilience efforts. Nonetheless, it is still unclear exactly how to judge model quality and reliability, and major questions remain including: 1) How does model performance in simulating present climate relate to future climate projections? 2) What metrics should be used to separate good models from bad models? and 3) How much skill in simulating present climate is due to calibration, tuning, or compensating errors?

Most straightforward metrics of assessing model performance in simulating present-day climate, such as root mean square error of the model climatology, don't correlate with future climate projections on a large scale (Knutti et al 2010b). Still, using projections from models that have demonstrably poor skill over the observational period to inform adaptation efforts is unwarranted and therefore, many methods of model ranking have been pursued (Giorgi et al. 2002, Schmittner et al. 2005, Dessai et al. 2005, Maxino et al. 2008, Perkins et al. 2007). Model performance similarities during the historical period have been shown to correlate to model projection similarities for certain variables on regional and global scales (Whetton et al. 2007), but defining historical performance metrics that relate to predictive skill is a largely unsolved problem (Knutti et al 2010b).

Separating good models from bad models depends on the question at hand. Researchers should evaluate climate models using metrics that are relevant to their specific purpose over their region of interest (Maxino et al. 2008, Knutti 2010a). Model quality should be assessed using multiple variables and techniques that go beyond simply examining mean statistics (Jun et al. 2008, Sun et al. 2015, Maxino et al. 2008, Knutti 2010). Even after developing application-relevant methods to assess model skill, the researcher must decide on where to place the threshold that distinguishes good models from bad ones for their particular purpose. This process is quite

ambiguous, with some researchers using natural breaks in performance (Maxino et al. 2008), whereas others use arbitrary top and bottom percentiles of ranked models (Geil et al. 2013, Geil and Zeng 2015). Knutti (2010a) suggests that it may be less controversial to eliminate the models that clearly perform the worst in any particular assessment than to agree on the best models.

Researchers should also consider that the extent of model calibration, tuning, and compensating errors is not transparent, which can lead to overconfidence in model capability. It is difficult to determine whether excellent agreement between model simulations and observations is the result of calibration and tuning or if the realistic simulations are actually correct for the right physical reasons (Knutti et al. 2008). Santer et al. (2009) suggest that using temporal and spatial variability evaluation metrics offers a more stringent test of model capability, since model developers are able to tune models to capture mean climate characteristics, whereas realistic representation of variability is difficult to achieve through tuning alone. Also, due to the limited number of earth system observations, model evaluation is probably often conducted with the same observational datasets that were used to develop and tune the model, which can lead to a warped view of model capability (Knutti et al. 2010b).

### *1.1.c Multimodel Ensemble Averaging*

It is also unclear as to how to aggregate model information to obtain future climate projections. Researchers must grapple with how to interpret a combination of models with differing levels of complexity, whether averaging models together makes physical sense, which models to include if multimodel ensemble (MME) averaging is appropriate, or whether

probabilistic methods of aggregation are more robust than MME averaging for a particular application.

Although CMIP model experiments are standardized, the suite of different models used to run those experiments represents a diverse range of model formulation, grid resolution, and complexity. Given these differences, individual models are simulating future warming for different reasons (Knutti et al 2008). In spite of this, any modeling center is allowed to contribute to CMIP archives regardless of model complexity or quality, which results in ambiguity as to how to interpret sets of CMIP models (Knutti 2010, Taylor et al. 2012).

The traditional method for aggregating model information is to create an MME average from all available model simulations. For mean climate simulation of multiple combined variables at global scales, an all-model MME average has been shown to outperform individual models, probably due to the cancelation of random modeling error (Reichler et al. 2008). However, an all-model MME may not be better than the single best model for any particular mean climate variable and selection of a few good models (up to about 5) for averaging has been shown to substantially decrease mean climate bias as compared to an all-model MME, at least for mean seasonal surface air temperature (Knutti et al. 2010). Depending on the context, averaging model information may not be physically meaningful. It may lead to unrealistic effects like the smoothing of spatially heterogeneous patterns (Knutti 2010) or yield physically implausible results. For example, in situations where there is a tipping point between multiple stable solutions, an average state may not exist (Knutti et al. 2010). If MME averaging is appropriate, the researcher is faced with the questions of which and how many models to include and whether to weight models by skill, which if improperly implemented could result in overconfidence that can be more damaging for

adaptation decision making than using equal weighting or not aggregating models at all (Knutti et al. 2010).

All-model MME averages are considered by some scientists to be a naïve approach to understanding model performance and projected climate changes (Jun et al 2008, Maxino et al 2008). Some argue that if including very poorly performing models in ensemble averages improves the average as compared to observations, then the improvement is for the wrong reasons (Maxino et al. 2008). Dessai et al. (2005) argue that using frequency distributions to aggregate model projections as opposed to MME averages is a better fit for identifying appropriate adaptation responses. Even when employing probabilistic methods though, the researcher must confront all the same questions regarding model quality, evaluation, and selection.

#### *1.1.d Model Independence*

Further complicating the interpretation of climate model information is the issue of model independence. Significant overlap exists between models due to the sharing of training data, human expertise, and model code (Pennell et al. 2011, Knutti 2010). Many models have highly correlated biases (Jun et al. 2008, Pennell et al. 2011, Knutti 2010) and considerable bias commonality exists beyond just models developed at the same center (Pennell et al. 2011). Model overlap results in an effective number of climate models that is much smaller than the total. While the effective number of climate models varies widely for individual model fields, Pennell et al. (2011) estimate that on average, the effective number of CMIP3 models for the northern hemisphere extratropics is between 7.5 and 9 from a total of 24. This raises issues for probabilistic

and statistical methods that assume model independence and leads to overconfidence in model projections, for example, due to confidence intervals that are too narrow (Pennell et al. 2011).

Model projections are likely to be biased toward an artificial consensus that is caused in part by model overlap (Pennell et al. 2011). It is reasonable to suspect that model similarities translate into a reduced range of climate change projections that don't sample the full range of uncertainty (Pennell et al. 2011, Knutti et al. 2008). In light of these facts, it is unclear which and how many models should be used to generate climate change projections and how much confidence can be placed on projections from a set of climate models that have so many commonalities.

## 1.2 USING GLOBAL CLIMATE MODELS FOR CLIMATE ADAPTATION PLANNING: APPLICATION ISSUES

Controlling and adapting to climate change are issues that are not easily defined and are complexly interwoven across scientific, technological, environmental, social, economic, and political boundaries. These types of wicked problems don't have "right" solutions, but their negative consequences can be mitigated through interdisciplinary collaboration, linking knowledge to action, and perseverance. Climate scientists can successfully contribute to mitigation and adaptation efforts by using collaborative knowledge exchange processes to deliver salient, credible, and legitimate information at the interface of science and decision making with serious consideration of the ethical issues at play.

### *1.2.a Salience, Credibility, Legitimacy*

Linking knowledge to action through the production of usable science requires that the scientific information produced be perceived by all stakeholders as salient, credible, and legitimate (Cash et al. 2003). Salience refers to the relevance of the information in the context of user needs, credibility refers to whether the information is perceived to be scientifically plausible and technically adequate, and legitimacy refers to whether the information itself, the producers of the information, and the process of creating the information are perceived to be unbiased and fair (Cash et al. 2003). These three elements generate trust across boundaries, shape knowledge exchange processes, and increase the likelihood that scientific information will be used (Lacey et al. 2015).

The salience, credibility, and legitimacy of climate information can be maximized when the producer is dedicated to understanding the realm of the user, including their organizational function, how information flows within their agency, how decisions are made, and their previous experience with climate information (Meadow et al. 2015). It is important for the producer to understand user priorities, decision making timelines, and the context in which the climate science will be applied (Brugger and Crimmins 2015). This type of knowledge will help the producer translate scientific information in understandable and salient ways, and also reveal the best ways to make the information available and accessible to the user. While keeping the focus on user information needs, the producer should use robust disciplinary and interdisciplinary scientific methods and address user and stakeholder concerns about bias and fairness throughout the process of scientific knowledge creation. For the user, usable science is understandable, accessible, salient, easy to integrate with existing knowledge, and fits into the user's decision framework (Meadow et

al. 2015). Jacobs et al. (2005) suggest asking a series of questions to assess the usefulness of scientific services and products: 1) Are producers asking and answering relevant questions at spatial and temporal scales relevant to the user? 2) Can the delivery of the scientific information be timely such that it is useful for decision making? 3) Are the scientific findings considered accurate and trustworthy by all stakeholders? 4) Is the scientific information provided in a format and translated in a way that is understandable to the user? 5) Is the scientific information useful given the constraints in the decision making process?

Generating usable science and working successfully at the boundary of science and decision making requires humility to recognize the limitations of one's own knowledge, and openness and respect for other systems of thought. A scientist functioning in this space will need to augment their disciplinary knowledge with expertise in effective communication across boundaries, facilitation, and policy development, and understand the processes of knowledge exchange (Preston et al. 2015).

### *1.2.b Knowledge Exchange*

Knowledge exchange is the process by which the interchange of knowledge occurs between scientific information producers and users or decision makers. It encompasses knowledge production, sharing, storage, mobilization, translation, and use (Cvitanovic et al. 2015). Historically, the transfer of knowledge has often followed the uni-directional knowledge-deficit model, where scientists as producers of knowledge make research available to potential users. Here, knowledge producers and users are two independent groups. The knowledge-deficit model considers the publication of scientific journal articles to adequately bring science knowledge into

the public domain and leaves the user or decision maker responsible for locating, understanding, and using the scientific information (Cvitanovic et al. 2015). This model may produce highly credible scientific information, but it is problematic for producing usable science that is also perceived as salient and legitimate by users. Success in achieving these three elements to increase the usability of scientific information is more likely to occur through collaborative and participatory knowledge exchange and research processes (Cash et al. 2003).

Contemporary approaches to improving knowledge exchange between scientists and decision makers include embedding, knowledge brokers, boundary organizations, and coproduction. Embedding refers to short-term professional development or permanent advisory-type positions for research scientists within organizations dominated by decision makers or for decision makers within scientific organizations. These types of positions facilitate the spread of knowledge across boundaries and the narrowing of priority knowledge gaps (Cvitanovic et al. 2015). The role of a knowledge broker is to facilitate the exchange of information among various stakeholders (e.g. researchers, practitioners, and policy makers). Knowledge brokers are typically based in science research teams or institutions, acting as intermediaries that develop relationships with science producers and users, and facilitating knowledge exchange across their networks (Cvitanovic et al. 2015). Boundary organizations facilitate knowledge exchange among diverse networks of stakeholders much like knowledge brokers, but they are established as a separate entity and are not typically embedded in institutional research teams. For this reason, boundary organizations can more effectively represent both sides of the science and decision making interface, while maintaining credibility through independence (Cvitanovic et al. 2015). Knowledge coproduction, a widely advocated form of knowledge exchange, refers to a process where all

relevant stakeholders participate in all aspects of the research program from onset to implementation and analysis (Cvitanovic et al. 2015).

The principles of coproduction are: 1) establishing long-term ongoing relationships between researchers and decision makers, 2) ensuring two-way communication, and 3) keeping the focus on the production of usable science (Meadow et al 2015). Successful coproduction is heavily reliant on iterativity (Meadow et al. 2015, Ferguson 2015), meaning that ongoing two-way interactions are essential and that the research and communication process itself should be malleable. Iterativity promotes evaluation and adjustment of research strategies and flexibility in research direction and methods (Brugger and Crimmins 2015). Up-front recognition of the need and importance of iterativity is crucial for all participants to successfully coproduce knowledge and generate usable science (Ferguson 2015). The process of coproduction promotes the salience, credibility, and legitimacy of the information produced by crossing communication divides between researchers and decision makers, translating knowledge to action, and active mediation of any conflicts that may arise (Cash et al. 2003).

### *1.2.c Ethics*

The most broadly discussed ethical issue pertaining to climate change is whether human beings have a responsibility to mitigate anthropogenic effects on the earth system to ensure sustainable use of natural resources and a livable environment for future generations. While this is an important issue, there are also many ethical issues pertaining to the interface of climate science research and decision making that should be considered. The development of a recognized system of professional ethics is being called for, in part, due to conflicting climate science research results

that can affect the direction of adaptation efforts. A system of professional ethics could enhance the quality control of climate science research that is intended to inform adaptation efforts and protect the interests of adaptation practitioners (Lacey et al. 2015).

For researchers working at the boundary of climate science and adaptation efforts, ethical considerations should come into play with the choice of research methods, presentation of uncertainty, interaction at the interface of science and decision making, and treatment of ambiguity. With respect to research methods, the researcher should be aware of modeling issues such as those discussed in Section 1.1 and create defensible research methods that navigate those issues, bearing in mind the intended use of the research. A contentious topic that often arises around adaptation efforts is whether downscaled climate information is needed in order to make sound adaptation decisions. Downscaled climate information can be no more reliable than the global climate model simulations on which the downscaling is based and does not automatically imply better information (Taylor et al. 2012). The assumptions and limitations of downscaling processes are often not well understood or explained, which has led to contradictions in climate projections (Hewitson et al. 2014). It may be the case, as in Australian viticulture, that the spatial scales relevant to end-user decision making can be captured with the grid resolution of current global climate models, which is information that can be elucidated by collaborative knowledge exchange processes and the understanding of user needs (Dunn et al. 2015). The uninformed or inappropriate use of downscaling also affects the transparency of uncertainty. Researchers may be unintentionally (or intentionally) presenting their findings as uncontroversial inputs to the user's decision making process, without proper understanding of their own methodological assumptions and uncertainties (Lacey et al 2015, Hewitson et al. 2014). Researchers interacting at the interface

of science and decision making should assume the role of an honest broker, providing factual information that does not reflect their own personal preferences, and never push specific methods or results because of benefit to themselves. Climate adaptation researchers, specifically, should help users understand the full range of adaptation options that may be available, instead of only the options that are relevant to themselves (Lacey et al 2015). Ambiguity in climate adaptation efforts is an opportunity to implement collaborative knowledge exchange processes that align with a broad set of values for exploring multiple adaptation pathways forward. Although, there is potential for asymmetric power relationships to develop in the face of ambiguity, where persuasive behaviors are used to exploit ambiguity for the gain of a single party over a broader benefit to all involved stakeholders. This risk should be understood by stakeholders working at the science and decision making interface and managed through ethical guidelines (Fleming et al. 2016).

### 1.3 OBJECTIVES

The credibility of climate models is established by evaluating how well they simulate past and present-day climate conditions. Science and technology are increasingly called upon to inform climate change adaptation and resilience efforts, and using projections from models that have demonstrably poor skill over the observational period to inform these efforts is unwarranted. This dissertation focuses on evaluating the skill of CMIP5 global climate models as compared to observations over the 20<sup>th</sup> century.

The work presented in Appendix A, published in the Journal of Climate (Geil et al. 2013), examines how well climate models simulate the North American monsoon system (NAMS) and

the causes of deficiencies in poorly performing models. Appendix B, published in *Geophysical Research Letters* (Geil and Zeng 2016), examines the presence and severity of unphysical numerical oscillations in global climate models that may affect the credibility of regional scale climate projections. Appendix C, which will be submitted to the *Journal of Applied Meteorology and Climatology*, is an in-depth assessment of model simulation of 20th century monthly minimum and maximum surface air temperature over the US on a regional basis.

In addition to these works, I have co-authored four other model evaluation and climate change projection studies. In Sheffield et al. (2013), I contributed information on historical model simulation of the NAMS, similar to the work in Appendix A. In Maloney et al. (2014), I contributed information on projected changes to the NAMS using the findings presented in Appendix A. I also assisted in the preparation of data and manuscript writing for a study examining the ability of CMIP5 models to simulate tropical depression wave activity and associated environmental factors in Serra and Geil (2016). Finally, in Zeng and Geil (2016), I assisted with the analysis and manuscript writing of a study developing decadal and long-term global warming projections based on an observational data-driven model.

## CHAPTER 2: PRESENT STUDY

### 2.1 ASSESSMENT OF CMIP5 MODEL SIMULATIONS OF THE NORTH AMERICAN MONSOON SYSTEM

Global and limited-area model simulations have been conducted in the past to evaluate the representation of the North American Monsoon System (NAMS) and the results show a wide range of model ability, but limited information has been published on this topic using the latest set of global climate models from phase 5 of the Coupled Model Intercomparison Project (CMIP5). The present study is comprised of a series of analyses aimed at assessing how well the CMIP5 suite of coupled general circulation models (CGCMs) is able to represent the NAMS.

Two analysis regions include a small  $4^{\circ}\times 5^{\circ}$  core domain in northwestern Mexico and an extended domain covering the larger NAMS region that encompasses most of Mexico and some of the US south and western states. Analyses include 1) comparison of the annual cycle of area-averaged monthly precipitation to observations and previously published CMIP3 results over the core monsoon domain, 2) a spatial correlation of monthly model precipitation, geopotential height, and wind to observations, 3) an assessment of monsoon onset and retreat dates as determined from daily precipitation, and 4) a model composite analysis of the best versus the worst representations of the NAMS.

There has been no improvement in the magnitude of the mean annual cycle of monthly precipitation over the core NAMS region since CMIP3, but the timing of seasonal changes in precipitation has improved with 27% more CMIP5 than CMIP3 models having zero phase lag. Despite this, a few models do not have a recognizable monsoon signal at all. Also, the multi-

model mean annual cycle is biased wet and exhibits the common problem of late monsoon termination.

Monsoon season correlations of monthly model output to observational data establish that most models have the highest correlation at the 500 hPa level and the lowest correlations for precipitation, however, relatively good or bad performance at the 500 hPa level is not predictive of 850 hPa level or precipitation performance.

The multi-model mean onset and retreat dates are 23 days early and 9 days late, respectively, using an absolute criteria for defining monsoon onset and retreat. Yearly model onset variability is comparable to that of the observational data, but yearly model retreat variability is much greater than what is seen in the observations. On average, model-relative criteria for determining onset and retreat dates result in less model bias compared the absolute criteria due to the prevailing wet bias in model precipitation.

An 850 hPa composite of best models reproduces the development and mature stages of the NAMS, but the composite of worst models fails to adequately illustrate most of the precipitation and circulation features seen in the observations. The large-scale circulation pattern bias seen in the best model composite is spatially consistent over the larger region influencing monsoon development, and thus still allows for a successful representation of the NAMS during the development and mature stages. In contrast, the spatial inconsistency of large-scale circulation pattern bias in the worst models prevents a realistic representation of the NAMS during the same period. Neither the composite of best or worst models realistically captures the retreat of the NAMS due to an extended connection to tropical moisture that causes excessive fall and winter precipitation. Models that best capture the relevant large-scale circulation patterns at low levels

usually have a realistic representation of the NAMS, while performance at mid-levels does not appear to be a major factor.

The importance of large-scale features to the representation of the NAMS in CMIP5 models is clear and for many models there is room for improvement in the representation of the NAMS by way of more accurate representation of low-level large-scale circulation features. Improvement in the representation of the NAMS in the best models is likely limited until increased model resolution allows for the capture of small-scale NAMS processes.

## 2.2 QUANTITATIVE CHARACTERIZATION OF SPURIOUS NUMERICAL OSCILLATIONS IN 48 CMIP5 MODELS

The presence of spurious numerical oscillations (SNOs) in global climate models has been known for decades and has been previously shown to cause poor representation of precipitation, wind, sea surface temperature, clouds, and more. The SNOs (in the form of Gibbs oscillations) are most prevalent in models that use spectral numerics and could compromise the results of regional climate analyses. This study provides a quantitative characterization of the SNOs in 48 CMIP5 models to draw awareness to the large SNOs present in these models.

An ocean transect at approximately 29° S that bisects the South Pacific High near the Andes is used to examine the SNOs where they are most easily visible: over the ocean and near a steep topographic gradient. We use 27-year climatological transects of monthly model variables over ocean points only and compute smoothed versions of the transects by applying a running mean. Observations are treated in the same way. Two metrics are used to identify and quantitatively

characterize SNOs along the transect. The first is the root mean square difference (RMSD) between the climatological transect of a variable and its running mean, representing an absolute measure of the oscillation amplitude. A relative measure is computed as the ratio of the RMSD to interannual variability (IAV). At each transect point, the standard deviation in time is first calculated using annual average values. IAV is then obtained as the average along the entire transect.

For variables that have observations for comparison, 40% of models on average have RMSD values greater than the RMSD value for observations along the transect multiplied by a factor of 5. Furthermore, 69% of the models have an RMSD:IAV ratio that is as large as, or larger than, interannual variability along the study transect for at least one variable. This translates to 95% of spectral models and 48% of non-spectral models having at least one RMSD:IAV ratio greater than unity. The largest SNOs by absolute and relative measures are seen in spectral models and in the surface pressure field, although smaller SNOs are visible in many of the variables examined. For eight of the thirteen variables, at least one model (or as many as half for surface pressure) has SNOs with amplitude as large as, or much larger than, the interannual variability of those variables along the transect. These variables include surface pressure, surface meridional winds, vertical velocity, surface air temperature, incoming surface radiation, and total cloud amount. Also, regardless of the numerical method employed, model resolution does not predict oscillation amplitude or prevalence.

The presence of large stationary numerical oscillations with amplitudes on the scale of atmospheric interannual variability suggests that these oscillations are spurious and should not be ignored. Despite this, SNOs are rarely mentioned in CMIP analysis literature probably because they are perceived as being small in modern climate models. Given past research by others and

our present findings, there is no reason to believe that the spurious oscillations are benign to climate simulations and they could very well have harmful impacts on the representation of variables at local, regional, and global scales.

### 2.3 EVALUATION OF THE MEAN STATE, TRENDS, AND VARIABILITY IN CMIP5 MONTHLY SURFACE AIR TEMPERATURE AT REGIONAL SCALES OVER THE U.S. FOR APPLICATION TO CLIMATE ADAPTATION PLANNING

Given the relative certainty of continued rapid change across many earth systems, local and regional decision makers are increasingly interested in climate change planning and adaptation methods. These decision makers need climate model projections on relevant temporal and spatial scales, as well as assessments of model reliability in order to make confident planning decisions. Here, we present an analysis of individual model simulations of monthly average minimum and maximum surface air temperature (Tmin, Tmax) to provide a clearer picture of model capability for adaptation and resilience planning efforts.

Area-averaged model and observed 20<sup>th</sup> century time series for eight regions in the United States are used to assess biases in model long-term mean state, trend, and variability. Model mean climate is examined using long-term annual and seasonal averages, linear trends are assessed using annual and monthly average temperature, and the standard deviation of the detrended monthly temperature anomaly time series is used to assess variability. Transparent model performance information is provided in the form of model rankings for each bias type.

A wide range in model skill is seen even for long term mean climate simulation where the highest skill is expected, and much of the ensemble cannot reproduce significant observed long term trends at monthly resolution. No strong relationships are seen between any of the three bias types or between 20<sup>th</sup> century bias and 21<sup>st</sup> century projected change. Using our model rankings, two smaller ensembles of models with better performance over the southwestern U.S. are selected and their 21<sup>st</sup> century projections are compared to those of the all-model ensemble. For the southwest, constraining temperature projections with multi-model ensembles that have small 20<sup>th</sup> century bias results in negligible differences in the multi-model ensemble average 21<sup>st</sup> century projected temperature change and model spread. In other words, models of varied quality (and complexity) are projecting very similar changes in temperature, implying that the models are simulating warming for different physical reasons.

Despite these results, we suggest that models with smaller 20<sup>th</sup> century biases have a greater likelihood of being more physically realistic with respect to both historical and future simulations, and therefore, more confidence can be placed in their 21<sup>st</sup> century projections as compared to projections from models that have demonstrably poor skill over the observational period. This type of analysis is essential for responsibly informing climate resilience efforts.

## 2.4 FUTURE RESEARCH DIRECTIONS

Much research is needed on how to responsibly utilize climate model information with respect to climate change vulnerability assessment and resilience planning efforts. This type of work can be broken down into many research sub-categories, of which the following list is by no

means comprehensive. 1) DEMONSTRATING MODEL CAPABILITY: transparent and in-depth multi-metric planning-relevant performance assessments of model skill compared to 20<sup>th</sup> century observations on a range of spatial and temporal scales. 2) DEFINING CONFIDENCE IN MODEL PROJECTIONS: how to interpret or adjust confidence in model projections for any given vulnerability or resilience effort based on 20<sup>th</sup> century model performance information. 3) APPLYING MODEL ENSEMBLES: what methods to use for choosing or combining climate model projections for vulnerability and resilience efforts, how to account for the issue of model overlap, how to describe and understand projection uncertainty in a planning context. 4) KNOWLEDGE EXCHANGE BEST PRACTICES: expansion of specific best practices for the facilitation of climate model knowledge exchange among scientists and practitioners when the end-goal is to use climate model projections for decision making. 5) KNOWLEDGE EXCHANGE ASSESSMENT: defining metrics to measure knowledge exchange success and the usability of any scientific information created.

The three manuscripts introduced in Sections 2.1-2.3 and presented in the Appendices of this dissertation mainly fall under the first research sub-category defined above “DEMONSTRATING MODEL CAPABILITY”, although the most recent manuscript (Appendix C) also includes aspects of sub-categories two “DEFINING CONFIDENCE IN MODEL PROJECTIONS” and three “APPLYING MODEL ENSEMBLES”. All three manuscripts spur follow-on research questions that also fall within the above-defined applied climate modeling research sub-categories.

In Appendix A we show a range of model skill in simulating the NAMS, although most models have large wet biases and do not properly simulate the monsoon season retreat due to large-scale low-level circulation issues that result in a prolonged connection to tropical moisture. We look at future projections for the core NAMS region in Maloney et al. (2014) and find a projected

MME-average annual mean drying of 22.2% by the end of the 21<sup>st</sup> century when using a 16-model ensemble that includes top performers and poor performers (based on the research in Appendix A). When using a 9-model ensemble of only top performers the MME-average annual mean projected drying is reduced to 15.4%, while the drying projected from the model that best simulates the NAMS in the 20<sup>th</sup> century is only 5.3%. Follow-on research questions could include: During which months is most of the projected drying occurring? What does large-scale circulation in the larger monsoon region look like in the future? What low-level circulation changes are causing the predicted drying in the good vs. poor historical performing models? Is future NAMS region circulation change related to historical model performance in a way that can inform our confidence in model projections of change? These questions are aimed at using model capability information to adjust our confidence in projections of future change, which falls under the second sub-category of climate model application research “DEFINING CONFIDENCE IN MODEL PROJECTIONS”. Answers to these questions could also help define appropriate methods for choosing or combining model projections in the NAMS region for resilience planning purposes, which is research sub-category three “APPLICATION OF MODEL ENSEMBLES”.

In Appendix B, we show that large spurious numerical oscillations are indeed present in most state-of-the-art climate models on the scale of atmospheric interannual variability near steep terrain. Spurious oscillations of this scale are likely not benign to climate simulations. Follow-on research could include the demonstration of future projection differences between ensembles of best and worst performing models with respect to spurious oscillations. An area that may yield impactful results is located off of the South American coast near the Andes, where some of the largest spurious oscillations are seen. How much do the oscillations affect the simulation of

regional upwelling and do they affect the simulation of El Niño-Southern Oscillation (ENSO) events? Do models with small or no spurious oscillations simulate ENSO frequency, intensity, and duration more realistically than models with large spurious oscillations? This type of research falls in large part under the first sub-category “DEMONSTRATING MODEL CAPABILITY”.

In Appendix C, we show a large range in model skill in simulating the 20<sup>th</sup> century mean, trend, and variability in monthly surface air temperature at regional scales in the US. Despite this fact, constraining temperature projections with multi-model ensembles that have small 20<sup>th</sup> century bias results in negligible differences in the multi-model ensemble average 21<sup>st</sup> century projected temperature change and model spread, at least for the southwest US region. This work focused only on regionally averaged temperature, therefore, the physical explanations for the large differences in model skill remain unknown. Follow-on research could use additional model variables to explain the physical reasons for large model biases, which may help shed light on why 20<sup>th</sup> century model bias is inconsequential to 21<sup>st</sup> century temperature projections. Also, extending the analysis at least to precipitation would be very informative for practitioners. These research ideas fall under sub-categories one “DEMONSTRATING MODEL CAPABILITY” and two “DEFINING CONFIDENCE IN MODEL PROJECTIONS”.

Another essential future research direction is the assessment of the limitations of downscaling techniques and how downscaled output relates to the original global model input. Downscaling techniques are currently being used for vulnerability and resilience efforts, and in regional and national climate assessments, despite the fact that the assumptions and limitations of downscaled climate products are not well understood. Responsible use of climate model information for decision making requires transparent and user-relevant assessments of downscaled

climate products. This research would be most closely associated with the first research sub-category “DEMONSTRATING MODEL CAPABILITY”, although it could easily include aspects of sub-categories two “DEFINING CONFIDENCE IN MODEL PROJECTIONS” and three “APPLYING MODEL ENSEMBLES”.

Finally, case studies in the realms of climate vulnerability assessment, resilience planning, and the knowledge exchange processes involved are critically needed in peer-reviewed publications. More documentation of both successful and unsuccessful vulnerability and resilience efforts are needed in order to improve knowledge exchange processes, clearly define best practices, and develop assessment metrics to measure the success of knowledge exchange processes and usable science products. Well-documented case studies could advance understanding in all of the five research sub-categories defined in this section.

## REFERENCES

- Brugger, J. and M. Crimmins, 2015: Designing institutions to support local-level climate change adaptation: insights from a case study of the US Cooperative Extension System. *Weather Climate and Society*, 7, 18-38.
- Cash, D.W., W.C. Clark, F. Alcock, N.M. Dickson, N. Eckley, D.H. Guston, J. Jager, and R.B. Mitchell, 2003: Knowledge systems for sustainable development. *Proc Natl Acad Sci*, 100, 8086-8091.
- Cvitanovic, C., A. Hobday, L. van Kerkhoff, S. Wilson, K. Dobbs, and N. Marshall, 2015: Improving knowledge exchange among scientists and decision-makers to facilitate the adaptive governance of marine resources: a review of knowledge and research needs. *Ocean & Coastal Management*, 112, 25–35.
- Dessai, S., X. Lu, and M. Hulme, 2005: Limited sensitivity analysis of regional climate change probabilities for the 21<sup>st</sup> century. *J Geophys Res*, 100, D19108.

- Dunn, M.R., J.A. Lindesay, and M. Howden, 2015: Spatial and temporal scales of future climate information for climate change adaptation in viticulture: a case study of user needs in the Australian winegrape sector. *Australian Journal of Grape and Wine Research*, 21, 226-239.
- Ferguson, D.B., 2015: Transdisciplinary climate research to support decision making. PhD dissertation, School of Geography and Development, The University of Arizona, 200 pp, <http://hdl.handle.net/10150/556812>.
- Fleming, A., and S.M. Howden, 2016: Ambiguity: A new way of thinking about responses to climate change. *Science of the Total Environment*, 571, 1271-1274.
- Fyfe, J.C, N.P. Gillett, and F.W. Zwiers, 2013: Overestimated global warming over the past 20 years. *Nature Climate Change*, 3, 767-769.
- Geil, K.L., Y. Serra, and X. Zeng, 2013: Assessment of CMIP5 Model Simulations of the North American Monsoon System. *J Climate*, 26, 8787-8801.
- Geil, K.L. and X. Zeng, 2015: Quantitative characterization of spurious numerical oscillations in 48 CMIP5 models. *Geophys Res Lett*, 42, 5066-5073.
- Giorgi, F. and L. Mearns, 2002: Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the “Reliability Ensemble Averaging” (REA) method. *J Climate*, 15, 1141-1158.
- Guilyardi, E., 2015: Challenges with ENSO in today’s climate models. National Oceanic and Atmospheric Administration, Accessed November 2015. [Available online at <https://www.climate.gov/news-features/blogs/enso/challenges-enso-today%E2%80%99s-climate-models>]
- Hewitson, B.C., J. Daron, R.G. Crane, M.F. Zermoglio, and C. Jack, 2014: Interrogating empirical-statistical downscaling. *Climatic Change*, 122, 539–554.
- IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.
- Jacobs, K., G. Garfin, and M. Lenart, 2005: More than just talk: connecting science and decision making. *Environ Sci Policy Sustain Dev*, 47, 6–21.
- Jun, M., R. Knutti, and D.W. Nychka, 2008: Spatial analysis to quantify numerical bias and dependence. *Journal of the American Statistical Association*, 103:483, 934-946.

- Kintisch, E., 2006: Ray Orbach Asks Science to Serve Society. *Science*, 313 (5795), 1874.
- Knutti, R., 2008: Why are climate models reproducing the observed global surface warming so well? *Geophys Res Lett*, 35, L18704.
- Knutti, R., 2010: The end of model democracy? *Climatic Change*, 102, 395-404.
- Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G.A. Meehl, 2010: Challenges in combining projections from multiple climate models. *J Climate*, 23, 2739-2758.
- Lacey J., S.M. Howden, C. Cvitanovic, A-M. Dowd, 2015: Informed adaptation: ethical considerations for adaptation researchers and decision-makers. *Glob Environ Change*, 32, 200-210.
- Maloney, E. and co-authors 2014: North American Climate in CMIP5 Experiments. Part III: Assessment of 21st Century Projections. *J Climate*, 27(6), 2230-2270.
- Maxino, C.C., B.J. McAvaney, A.J. Pitman, and S.E. Perkins, 2008: Ranking the AR4 climate models over the Murray-Darling Basin using simulated maximum temperature, minimum temperature and precipitation. *Int J Climatol*, 28, 1097-1112.
- Meadow, A.M., D.B. Ferguson, Z. Guido, A. Horangic, G. Owen, and T. Wall, 2015: Moving toward the deliberate coproduction of climate science knowledge. *Weather Climate and Society*, 7, 179-191.
- Northrop, P.J. and R.E. Chandler, 2014: Quantifying sources of uncertainty in projections of future climate. *J Climate*, 27, 8793-8808.
- Pennell, C. and T. Reichler, 2011: On the effective number of climate models. *J Climate*, 24, 2358-2367.
- Perkins, S.E., A.J. Pitman, N.J. Holbrook, and J. McAneney, 2007: Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *J Climate*, 20, 4356-4376.
- President Barack Obama to The Heads of Executive Departments and Agencies, Septembe 21 2016, The White House Office of the Press Secretary, Presidential Memorandum- Climate Change and National Security, <https://www.whitehouse.gov/the-press-office/2016/09/21/presidential-memorandum-climate-change-and-national-security>.
- Preston, B.L., L. Rickards, H. Funfgeld, and R.J. Keenan, 2015: Toward reflexive climate adaptation research. *Current Opinion in Environmental Sustainability*, 14, 127-135.

- Reichler, T., and J. Kim, 2008: How well do coupled models simulation today's climate? *Bull Amer Meteorol Soc*, 89 (3), 303-311.
- Santer, B.D., K.E. Taylor, P.J. Gleckler, C. Bonfils, T.P. Barnett, D.W. Pierce, T.M.L. Wigley, C. Mears, F.J. Wentz, W. Bruggemann, N.P. Gillett, S.A. Klein, S. Solomon, P.A. Stott, and M.F. Wehner, 2009: Incorporating model quality information in climate change detection and attribution studies. *Proc Natl Acad Sci*, 106, 14778–14783.
- Schmittner, A., M. Latif, and B. Schneider, 2005: Model projections of the North Atlantic thermohaline circulation for the 21<sup>st</sup> century assessed by observations. *Geophys Res Lett*, 32, L23710.
- Serra, Y.L. and K.L. Geil, 2016: Historical and projected eastern Pacific and Intra-Americas Sea TD-wave activity in a selection of IPCC AR5 models. *J Climate*, *in press*.
- Sheffield, J. and co-authors 2013: North American Climate in CMIP5 Experiments. Part I: Evaluation of 20th Century Continental and Regional Climatology. *J Climate*, 26 (23), 9209-9245.
- Sillmann, J., M.G. Donat, J.C. Fyfe, and F.W. Zwiers, 2014: Observed and simulated temperature extremes during the recent warming hiatus. *Environ Res Lett*, 9, 064023, doi: 10.1088/1748-9326/9/6/064023.
- Stouffer, R.J., A.J. Weaver, and M. Eby, 2004: A method for obtaining pre-twentieth century initial conditions for use in climate change studies. *Climate Dynamics*, 23 (3), 327-339.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2009: A summary of the CMIP5 experiment design. PCMDI Rep., 33 pp. [Available online at [http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor\\_CMIP5\\_design.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf).]
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull Amer Meteorol Soc*, 93, 485–498.
- United Nations Educational, Scientific and Cultural Organization (UNESCO), 2000. *Proceedings of the World Conference on Science*. London, UK: Banson.
- Whetton, P., I. Macadam, J. Bathols, and J. O'Grady, 2007: Assessment of the use of current climate patterns to evaluate regional enhanced greenhouse response patterns of climate models. *Geophys Res Lett*, 34, L14701.
- Zeng, X. and K.L. Geil, 2016: Global warming projection in the 21<sup>st</sup> century based on an observational data-driven model. *Geophys Res Lett*, 43, doi:10.1002/2016GL071035.

## APPENDIX A

### ASSESSMENT OF CMIP5 MODEL SIMULATIONS OF THE NORTH AMERICAN MONSOON SYSTEM

(Published in the Journal of Climate)

KERRIE L. GEIL<sup>1</sup>, YOLAND L. SERRA<sup>1</sup>, AND XUBIN ZENG<sup>1</sup>

<sup>1</sup>Department of Atmospheric Sciences, University of Arizona, Tucson, Arizona

Geil, K.L., Y. Serra, and X. Zeng, 2013: Assessment of CMIP5 Model Simulations of the North American Monsoon System. *J Climate*, 26, 8787-8801, doi: 10.1175/JCLI-D-13-00044.1.

© American Meteorological Society. Used with permission.



## Assessment of CMIP5 Model Simulations of the North American Monsoon System

KERRIE L. GEIL, YOLANDE L. SERRA, AND XUBIN ZENG

*Department of Atmospheric Sciences, University of Arizona, Tucson, Arizona*

(Manuscript received 18 January 2013, in final form 15 May 2013)

### ABSTRACT

Precipitation, geopotential height, and wind fields from 21 models from phase 5 of the Coupled Model Intercomparison Project (CMIP5) are examined to determine how well this generation of general circulation models represents the North American monsoon system (NAMS). Results show no improvement since CMIP3 in the magnitude (root-mean-square error and bias) of the mean annual cycle of monthly precipitation over a core monsoon domain, but improvement in the phasing of the seasonal cycle in precipitation is notable. Monsoon onset is early for most models but is clearly visible in daily climatological precipitation, whereas monsoon retreat is highly variable and unclear in daily climatological precipitation. Models that best capture large-scale circulation patterns at a low level usually have realistic representations of the NAMS, but even the best models poorly represent monsoon retreat. Difficulty in reproducing monsoon retreat results from an inaccurate representation of gradients in low-level geopotential height across the larger region, which causes an unrealistic flux of low-level moisture from the tropics into the NAMS region that extends well into the post-monsoon season. Composites of the models with the best and worst representations of the NAMS indicate that adequate representation of the monsoon during the early to midseason can be achieved even with a large-scale circulation pattern bias, as long as the bias is spatially consistent over the larger region influencing monsoon development; in other words, as with monsoon retreat, it is the inaccuracy of the spatial gradients in geopotential height across the larger region that prevents some models from realistic representation of the early and mid-season monsoon system.

### 1. Introduction

The evolution of the North American monsoon system (NAMS) can be described as having development, mature, and decay stages similar to but less intense than its larger Asian counterpart. During the development stage (May–June), the extratropical jet weakens and migrates to the north resulting in decreased frequency of synoptic-scale transient activity from the midlatitudes over northern Mexico and the southwestern United States (Higgins et al. 1997). A thermal surface low forms in the desert regions (Rowson and Colucci 1992) and a pronounced anticyclone at jet stream level develops over northwestern Mexico (Barlow et al. 1998), analogous to the Tibetan high over Asia (Tang and Reiter 1984). Mid- to upper-level flow shifts from westerly in May–June

to easterly and southeasterly around the west side of the anticyclone by July (Douglas et al. 1993; Higgins et al. 1997). Low-level flow into the monsoon region is strongly influenced by the evolution of the North Atlantic subtropical high (NASH) and North Pacific subtropical high (NPSH). As the subtropical highs build and move northward, northwesterly flow from the NPSH is reduced over the northern Gulf of California and the westward extension of the NASH brings southerly flow into eastern Mexico and the U.S. Great Plains (Schmitz and Mullen 1996; Higgins et al. 1997; Barlow et al. 1998). Southerly winds flow over the Gulf of California (Badan-Dangon 1991; Douglas et al. 1993) and convective precipitation quickly spreads to the northwest along the western slopes and foothills of the Sierra Madre Occidental (SMO; Douglas et al. 1993). The mature stage (July–August) brings the precipitation maximum over the SMO and increased precipitation coincides with increased vertical moisture transport (Douglas et al. 1993; Schmitz and Mullen 1996). The decay stage (September–October) is conceptually the reverse of the development stage, but is

*Corresponding author address:* Kerrie L. Geil, Department of Atmospheric Sciences, University of Arizona, 1118 East Fourth Street, Tucson, AZ 85721-0081.  
E-mail: geil@atmo.arizona.edu

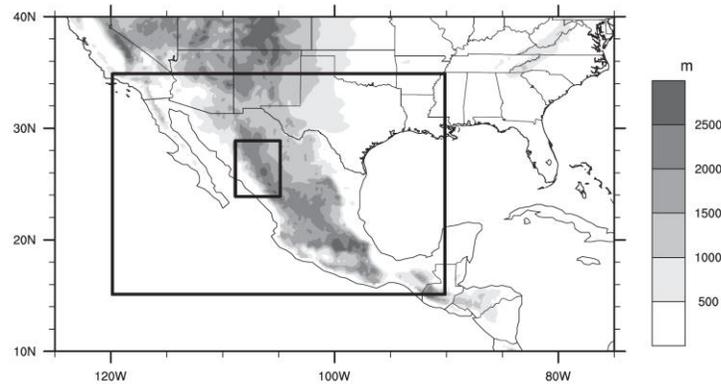


FIG. 1. Core NAMS domain (23.875°–28.875°N, 108.875°–104.875°W) and extended NAMS domain (15.125°–34.875°N, 119.875°–90.125°W), shown over regional topography.

more gradual (Higgins et al. 1997; Barlow et al. 1998). The main source of monsoon moisture has been debated over recent decades, but it is most likely that low-level moisture is advected mainly from the Gulf of California and the eastern Pacific, while the Gulf of Mexico contributes to upper-level moisture (Schmitz and Mullen 1996; Berbery 2001; Higgins et al. 2003). NAMS region continental moisture sources are also important, as precipitation recycling contributes to monsoon season rainfall (Bosilovich et al. 2003; Dominguez et al. 2008).

Global and limited-area model simulations have been conducted in the past to evaluate the representation of the NAMS and the results show a wide range of model ability. Arritt et al. (2000) demonstrated that the Met Office (UKMO) HadCM2 global model could simulate generally realistic NAMS circulation and precipitation, whereas Yang et al. (2001) showed that the National Center for Atmospheric Research (NCAR) CCM3 global model was unable to simulate these NAMS features. Liang et al. (2008) found that only 1 of 17 CMIP3 global models was able to realistically reproduce the NAMS precipitation annual cycle, interannual variability in precipitation, and key circulation patterns such as the monsoon high and the westward extension of the NASH with the associated low-level southerly flow. Stensrud et al. (1995) reproduced monsoon mesoscale circulation and the general features of deep convection with the Fourth-generation Pennsylvania State University–NCAR Mesoscale Model (MM4) limited-area model, and Berbery (2001) showed that NCEP's Eta limited-area model could additionally reproduce the diurnal cycle of moisture flux. Gao et al. (2007) used the MM5 limited-area model to demonstrate improvement over global models in representing spatial and temporal

precipitation patterns but found model deficiencies in representing the evolution of the diurnal cycle. Castro et al. (2007a) used the Regional Atmospheric Modeling System (RAMS) limited-area model driven with global reanalysis data and found that the model's enhanced representation of the surface boundary produced an acceptable diurnal cycle of summer precipitation in the monsoon region that was not captured by the driving reanalysis. A recent study by the same group using the Weather Research and Forecasting Model (WRF; Castro et al. 2012) showed the potential for limited-area models to improve seasonal NAMS forecasts. The use of higher resolution limited-area models that are able to capture the diurnal cycle of convection, as opposed to coarser general circulation models that do not have this capability, is stressed by Castro et al. (2007a,b, 2012) for drawing conclusions with respect to regional climate variability and prediction.

The present study is comprised of a series of analyses aimed at assessing how well the phase 5 of the Coupled Model Intercomparison Project (CMIP5) suite of coupled general circulation models (CGCMs) is able to represent the NAMS. The two analysis regions include a smaller core domain and a larger extended domain (Fig. 1). Our core domain is smaller but similar to that used by the North American Monsoon Experiment (NAME; Higgins et al. 2006) and related studies (e.g., Higgins and Gochis 2007; Gutzler et al. 2009), while our extended domain includes the larger NAMS region. The uniformity of the annual cycle of precipitation across all grid points within our core domain has been visually verified, as in Higgins et al. (1999), to ensure we have selected an area with a consistent monsoon precipitation signal.

The observational data and CMIP5 model simulations used in this study are described in section 2. The annual precipitation cycle and comparison to CMIP3 is presented in section 3, while section 4 evaluates the spatial progression of the monsoon using spatial gridpoint correlation. Daily precipitation is utilized to calculate monsoon onset and retreat dates in section 5, composites of the best and worst performing models are discussed in section 6, and a summary is presented in section 7.

## 2. Observational data and model simulations

### a. Observational data and reanalysis

Monthly precipitation observations are obtained from the recently developed National Oceanic and Atmospheric Administration (NOAA)  $0.5^\circ \times 0.5^\circ$  gridded precipitation dataset (P-NOAA) provided by Drs. Russ Vose and Ed Cook and is described by Castro et al. (2012). This dataset was created from station data and considers the dependence of precipitation on elevation, similar to the Parameter-Elevation Regressions on Independent Slopes Model (PRISM) dataset that covers only the United States (Daly et al. 1994). For our daily time resolution analysis, we use the Tropical Rainfall Measuring Mission (TRMM) 3B42v6 daily precipitation estimates (Huffman et al. 2007), which are provided on a  $0.25^\circ \times 0.25^\circ$  spatial grid. The dataset is created using several types of satellite measurements and also incorporates monthly station observations to improve accuracy. We have chosen to use the daily TRMM satellite dataset as opposed to a lower spatial resolution in situ daily dataset based on the importance of higher spatial resolution over variable terrain (Gochis et al. 2004). Also, in a study comparing different satellite-based precipitation estimates to 2004 NAME gauge data, Gochis et al. (2009) showed that the TRMM 3B42v6 product performs well over the monsoon region.

Monthly geopotential height and wind are provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERA-Interim; Dee et al. 2011) and are obtained from the Research Data Archive (RDA; <http://rda.ucar.edu: dataset ds627.0>). ERA-Interim is produced at spectral T255 horizontal resolution with 60 vertical layers and is provided at 6-hourly intervals on a  $0.75^\circ \times 0.75^\circ$  grid with 37 vertical pressure levels.

### b. Coupled general circulation model simulations

The source of CGCM climate simulations is the CMIP5 multimodel ensemble archive, made available online by the Program for Climate Model Diagnosis and

Intercomparison (PCMDI; <http://pcmdi3.llnl.gov/esgcat>). The historical experiment is chosen for this analysis, which imposes changing atmospheric and land surface conditions consistent with past observations, including changes in atmospheric composition due to anthropogenic and volcanic influences, solar forcing, concentrations of short-lived species and aerosols from both natural and anthropogenic sources, and land use (Taylor et al. 2009). For details regarding CMIP5 experimental design, the reader is referred to Taylor et al. (2009, 2012). Table 1 provides information on the 21 CGCMs used for this study, which have atmospheric components ranging in horizontal grid resolution from  $0.56^\circ \times 0.56^\circ$  in longitude by latitude to  $3.75^\circ \times 2.47^\circ$  and oceanic horizontal grids ranging from  $0.28^\circ \times 0.2^\circ$  to  $1.98^\circ \times 1.2^\circ$  resolution (Gent et al. 2011; Volodin et al. 2010; <http://data.giss.nasa.gov/modelE/ar5>). We recognize that this range of resolution is still relatively coarse for the representation of detailed topography and the resultant small-scale atmospheric (e.g., convective) processes. Model composite statistics of high versus low horizontal and vertical resolutions (not shown) for each of the analyses in this study did not reveal major differences in model performance, implying that even the highest resolution model examined is still too coarse to capture small-scale topographically influenced processes.

All observations, reanalysis, and simulations are regridded to the TRMM  $0.25^\circ \times 0.25^\circ$  master grid using bilinear spatial interpolation, which facilitates direct comparison. The reference period for this study is 1979–2005 for all model simulations, reanalysis, and observations, except for TRMM daily precipitation, which only includes the years 1998–2010. Testing using the ERA-Interim precipitation indicates that the precipitation metrics presented in this study are insensitive to the difference in reference periods (between 1979–2005 and 1998–2010).

## 3. Annual cycle of precipitation

The annual cycle of precipitation within the core NAMS domain is characterized by relatively dry winter months followed by an early spring minimum, a sharp rise during late spring leading to a summertime peak, and a return to a secondary minimum in the fall (see Fig. 3). The wettest months are July, August, and September, when the bulk of the annual precipitation occurs, whereas the driest months are March, April, and May. Following the methods of Liang et al. (2008), Fig. 2 compares the modeled and observed precipitation annual cycle using three metrics. For each model, the root-mean-square (rms) error of monthly mean rainfall is shown in Fig. 2a, the percent bias in annual rainfall

TABLE 1. CGCM model acronym, model expansion, three-letter model code, modeling center name, origin country, CMIP5 ensemble member identifier, and atmospheric (AGCM) and oceanic (OGCM) model grid resolution. Grid resolution format is (number of longitudinal grids)  $\times$  (number of latitudinal grids)  $\times$  (number of vertical levels).

Model	Model expansion	Code	Modeling group	Country	Ensemble member	AGCM grid resolution	OGCM grid resolution
BCC-CSM1.1	Beijing Climate Center, Climate System Model, version 1.1	BCC	Beijing Climate Center, China Meteorological Administration	China	r1i1p1	128 $\times$ 64L26	360 $\times$ 232L40
CanESM2	Second Generation Canadian Earth System Model	CAN	Canadian Centre for Climate Modelling and Analysis	Canada	r1i1p1	128 $\times$ 64L35	256 $\times$ 192L40
CCSM4	Community Climate System Model, version 4	CCS	National Center for Atmospheric Research	United States	r1i1p1	288 $\times$ 200L26	384 $\times$ 320L60
CNRM-CM5	Centre National de Recherches Météorologiques Coupled Global Climate Model, version 5	CNR	Centre National de Recherches Météorologiques/Centre Européen de Recherche et Formation Avancées en Calcul Scientifique	France	r1i1p1	256 $\times$ 128L31	362 $\times$ 292L42
CSIRO Mk3.6.0	Commonwealth Scientific and Industrial Research Organisation Mark, version 3.6.0	CSI	Commonwealth Scientific and Industrial Research Organization, Queensland Climate Change Centre of Excellence	Australia	r1i1p1	192 $\times$ 96L18	192 $\times$ 189L31
GFDL-CM3	Geophysical Fluid Dynamics Laboratory Climate Model, version 3	GF3	NOAA/Geophysical Fluid Dynamics Laboratory	United States	r1i1p1	144 $\times$ 90L48	360 $\times$ 200L50
GFDL-ESM2G	Geophysical Fluid Dynamics Laboratory Earth System Model with Generalized Ocean Layer Dynamics (GOLD) component	GFG			r1i1p1	144 $\times$ 90L24	360 $\times$ 210L63
GFDL-ESM2M	Geophysical Fluid Dynamics Laboratory Earth System Model with Modular Ocean Model 4 (MOM4) component	GFM			r1i1p1	144 $\times$ 90L24	360 $\times$ 200L50
GISS-E2-R	Goddard Institute for Space Studies Model E, coupled with the Russell ocean model	GIS	National Aeronautics and Space Administration (NASA) Goddard Institute for Space Studies	United States	r6i1p1	144 $\times$ 90L40	288 $\times$ 180L32
HadCM3	Hadley Centre Coupled Model, version 3	HA3	Met Office Hadley Centre	United Kingdom	r1i1p1	96 $\times$ 73L19	288 $\times$ 144L20
HadGEM2-CC	Hadley Centre Global Environment Model, version 2—Carbon Cycle	HGC			r1i1p1	192 $\times$ 145L60	360 $\times$ 216L40
HadGEM2-ES	Hadley Centre Global Environment Model, version 2—Earth System	HGE			r1i1p1	192 $\times$ 145L38	360 $\times$ 216L40
INM-CM4.0	Institute of Numerical Mathematics Coupled Model, version 4.0	INM	Institute of Numerical Mathematics	Russia	r1i1p1	180 $\times$ 120L21	360 $\times$ 340L40
IPSL-CM5A-LR	L'Institut Pierre-Simon Laplace Coupled Model, version 5A, coupled with Nucleus for European Modelling of the Ocean (NEMO), low resolution	IPL	L'Institut Pierre-Simon Laplace	France	r1i1p1	96 $\times$ 96L39	182 $\times$ 149L31
IPSL-CM5A-MR	L'Institut Pierre-Simon Laplace Coupled Model, version 5A, coupled with NEMO, mid resolution	IPM			r1i1p1	143 $\times$ 144L39	182 $\times$ 149L31

TABLE 1. (Continued)

Model	Model expansion	Code	Modeling group	Country	Ensemble member	AGCM grid resolution	OGCM grid resolution
MIROC-ESM	Model for Interdisciplinary Research on Climate, Earth System Model	MIE	Japan Agency for Marine-Earth Science and Technology.	Japan	r11p1	128 × 64L80	256 × 192L44
MIROC4h	Model for Interdisciplinary Research on Climate, version 4 (high resolution)	M14	Research Institute (The University of Tokyo), and National Institute for Environmental Studies		r11p1	640 × 320L56	1280 × 912L48
MIROC5	Model for Interdisciplinary Research on Climate, version 5	M15			r11p1	256 × 128L40	256 × 224L50
MPI-ESM-LR	Max Planck Institute Earth System Model, low resolution	MPI	Max Planck Institute for Meteorology	Germany	r11p1	192 × 96L47	256 × 220L40
MRI-CGCM3	Meteorological Research Institute Coupled Atmosphere–Ocean General Circulation Model, version 3	MRI	Meteorological Research Institute	Japan	r11p1	320 × 160L48	360 × 368L51
NorESM1-M	Norwegian Earth System Model, version 1 (intermediate resolution)	NOR	Norwegian Climate Centre	Norway	r11p1	144 × 96L26	320 × 384L26

based on the mean monthly climatological observed rainfall value ( $1.66 \text{ mm day}^{-1}$ ) is shown in Fig. 2b, and the phase lag in number of months is shown in Fig. 2c. All three metrics are calculated for the core NAMS domain using 12 monthly climatological precipitation values. A correlation is computed between the model annual cycle and the observed cycle at each (monthly interval) time lag and the phase lag is defined as the time lag with the highest correlation. The range of rms error in annual rainfall totals is  $0.76\text{--}2.74 \text{ mm day}^{-1}$  and the average model error is  $1.47 \text{ mm day}^{-1}$ . The large majority of models are biased wet, with an average bias of 51.3% and a range from  $-42\%$  to  $136\%$ . Although there is no bias calculation in Liang et al. (2008) for comparison, similarity between the range of rms error ( $0.46\text{--}2.23 \text{ mm day}^{-1}$ ) in their study of CMIP3 models and that of the CMIP5 models in this analysis indicates that there has been no improvement in the magnitude of the simulated annual cycle of monthly precipitation and, in fact, the lowest and highest rms values have increased slightly since the previous generation of CGCMs. On the other hand, there does seem to be improvement in the timing of seasonal precipitation shifts, with 13 out of 21 (62%) CMIP5 models having a phase lag of zero months as compared to 6 out of 17 (35%) CMIP3 models in Liang et al. (2008). Figure 3 shows the precipitation annual cycle for all models separated into three groups by phase lag value. Small, moderate, and large phase lag models are defined as those with zero, 1-month, and greater than 1-month phase lags, respectively. The only model that captures all characteristic features of the annual cycle with the proper timing is the Met Office Hadley Centre (MOHC) HGE model; however, this model is too wet for 11 out of 12 months. Of the small and moderate phase lag models, a common problem is the difficulty in ending the monsoon season, as reflected by the insufficient fall minimum and high precipitation in the fall and winter seasons. This problem is also seen in the multimodel mean.

#### 4. Seasonal spatial correlation of monsoon variables

To verify the appropriate spatial progression of NAMS onset and retreat, we assess the spatial pattern of monsoon variables within the extended NAMS analysis region. This is accomplished using a simple point-to-point spatial correlation of climatological monthly model values of precipitation, geopotential height at 500 and 850 hPa, and meridional and zonal wind at 500 and 850 hPa, to the corresponding TRMM satellite data or ERA-Interim data. Pressure level correlation indices at 500 and 850 hPa are created by averaging the

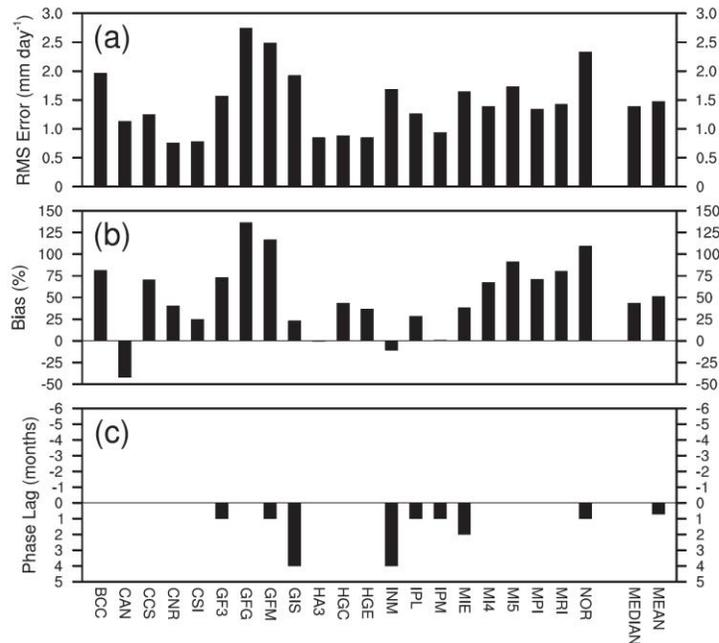


FIG. 2. Annual precipitation cycle (a) rms error ( $\text{mm day}^{-1}$ ), (b) bias (%), and (c) phase lag yielding maximum correlation (months) for all models with respect to monthly P-NOAA observations over the core NAMS domain. Multimodel median and mean values are also shown in each panel. All metrics are calculated using 12 monthly climatological values for the period 1979–2005. Phase lag of +1 means the highest correlation of all 12 months is between monthly observations ordered January through December and monthly model precipitation ordered February through January.

corresponding correlations of geopotential height and winds. It is important to note that Poisson grid filling (using standard routines included with the NCAR Command Language V6.0.0; NCAR 2012) is used to interpolate the 850-hPa model output below ground to obtain continuous fields at this pressure level. Correlations are calculated for each month during the May through October monsoon season and the average correlations of the May–June (MJ) early season, the July–August (JA) midseason, the September–October (SO) late season, and the full May–October (MJJASO) season are used to rank the models.

Table 2 shows the results of the correlation analysis for the five highest and lowest ranked models. The range of highest to lowest ranked correlation is usually smallest for the 500-hPa index (e.g., seasonal range of  $r = 0.64$ – $0.91$ ) and largest for precipitation (e.g., seasonal range of  $r = 0.23$ – $0.79$ ). The IPL model even has a negative precipitation correlation ( $r = -0.05$ ) during the decay stage. The highest correlations occur earlier in

the monsoon season, while the lowest correlations occur later in the season. Seasonally, 16 out of 21 models perform better at 500 hPa than at 850 hPa (not explicitly shown in Table 2), partly because of less small-scale variability and hence fewer spatial degrees of freedom at 500 hPa. It is interesting that three of the five models that do not perform better at 500 hPa than at 850 hPa (CNR, HGC, and HGE) consistently have the highest precipitation correlations. Relatively good or bad correlation at the 500-hPa level is not predictive of 850-hPa level or precipitation correlation; however, better correlation at the 850-hPa level usually corresponds to better correlation of precipitation during the monsoon season.

These results and the results of the annual cycle of precipitation analysis are used to choose the models appropriate for the daily time resolution onset and retreat analysis that follows. Models with a large phase lag ( $>1$  month) and models in the bottom quintile of precipitation or 850-hPa index seasonal correlations have the poorest representations of the NAMS and

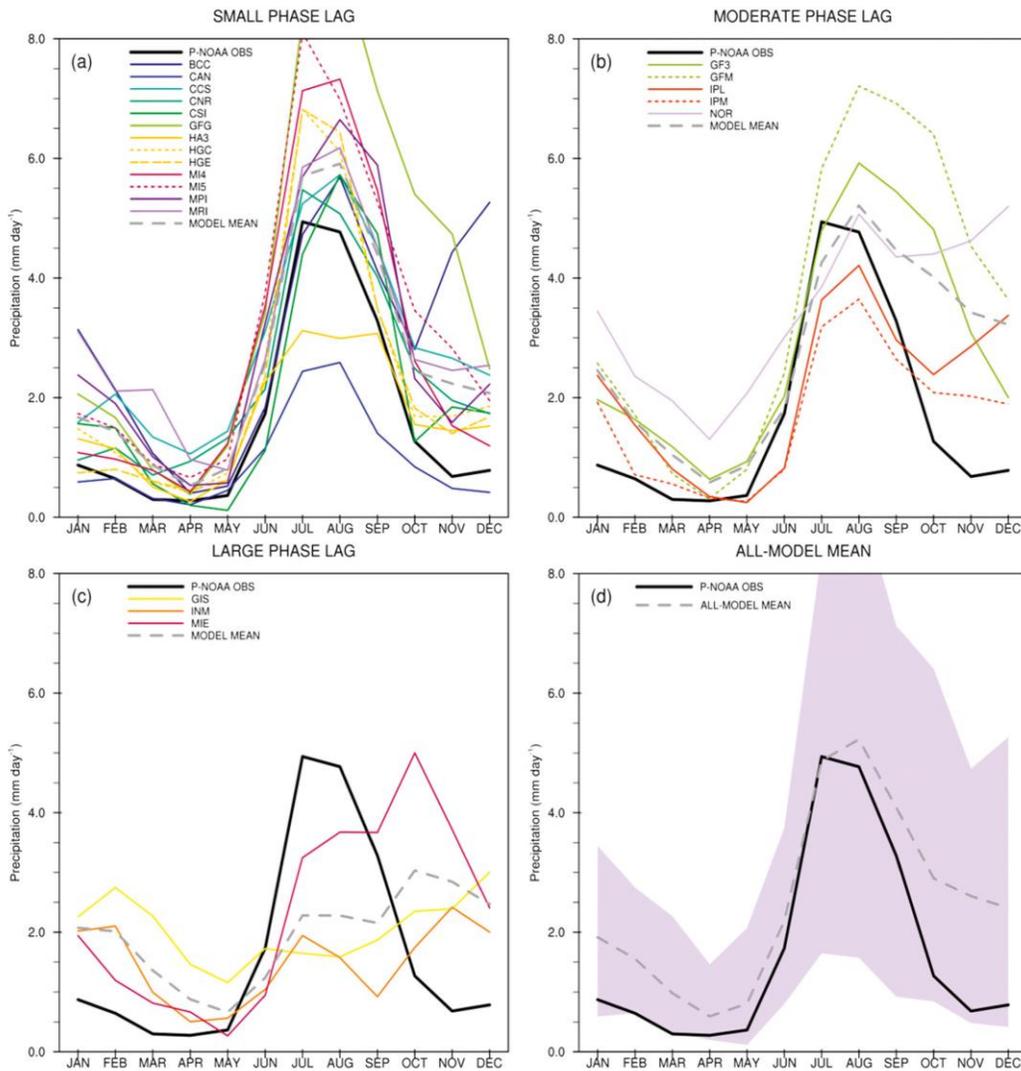


FIG. 3. Annual precipitation cycle over the core NAMS domain for all models grouped by (a) small phase lag (lag = 0 months), (b) moderate phase lag (lag = 1 month), and (c) large phase lag (lag > 1 month). The multimodel mean (gray dashed line) for each category is shown in (a)–(c). (d) The all-model mean (dashed line) and spread (shading). Colors represent different modeling centers and solid vs dashed lines of the same color differentiate models from a common center.

are therefore eliminated from the daily analysis. This eliminates seven models: namely, the GIS, INM, IPL, IPM, MIE, MI4, and NOR models.

Note that neither the correlation nor the correlation difference between models needs to be statistically significant for the ranking in Table 2. Actually, it is not easy to address the statistical significance because of spatial

autocorrelation in both the meridional and zonal directions. To reduce the spatial autocorrelation, we have thinned the data at 10, 20, 30, and 40 gridbox increments (i.e., with 2.5°, 5°, 7.5°, and 10° distances between adjacent data points). All correlations over these coarse grids for the top five ranked models in each category and for all other models are found to be statistically

TABLE 2. Top and bottom five model rankings of spatial correlation  $r$  for precipitation, 850-hPa index (geopotential height and winds), and 500-hPa index (geopotential height and winds), as compared to TRMM and ERA-Interim data, for the entire monsoon season, the development stage, the mature stage, and the decay stage over the extended NAMS domain.

	Seasonal (MJJASO)			Development (MJ)			Mature (JA)			Decay (SO)		
	Rank		$r$	Rank		$r$	Rank		$r$	Rank		$r$
Precipitation	1	CNR	0.79	1	HA3	0.88	1	HA3	0.77	1	CSI	0.84
	2	HGE	0.79	2	HE	0.85	2	HGE	0.76	2	CNR	0.80
	3	HGC	0.79	3	HGC	0.84	3	CNR	0.75	3	HGC	0.79
	4	CSI	0.77	4	CNR	0.82	4	HGC	0.75	4	HGE	0.77
	5	HA3	0.77	5	INM	0.81	5	MRI	0.74	5	MRI	0.68
	17	GF3	0.55	17	GFG	0.55	17	GIS	0.60	17	GFG	0.46
	18	NOR	0.55	18	MPI	0.54	18	MPI	0.59	18	MIE	0.43
	19	MI4	0.53	19	IPM	0.53	19	MI4	0.39	19	GF3	0.25
	20	IPM	0.34	20	NOR	0.50	20	IPM	0.37	20	IPM	0.12
	21	IPL	0.23	21	IPL	0.49	21	IPL	0.25	21	IPL	-0.05
	850-hPa index	1	CNR	0.90	1	CNR	0.92	1	CNR	0.89	1	CNR
2		HGE	0.85	2	HA3	0.89	2	HGE	0.89	2	HA3	0.81
3		HA3	0.85	3	CAN	0.87	3	HGC	0.89	3	HGE	0.79
4		HGC	0.84	4	HGC	0.87	4	MRI	0.88	4	CAN	0.78
5		CAN	0.83	5	HE	0.87	5	MI5	0.85	5	MPI	0.78
17		INM	0.68	17	INM	0.75	17	INM	0.73	17	GFG	0.55
18		IPM	0.66	18	IPM	0.72	18	NOR	0.71	18	IPM	0.53
19		MIE	0.57	19	GIS	0.55	19	MIE	0.71	19	GIS	0.49
20		GIS	0.54	20	IPL	0.54	20	IPL	0.63	20	MIE	0.45
21		IPL	0.53	21	MIE	0.54	21	GIS	0.59	21	IPL	0.42
500-hPa index		1	GFG	0.91	1	GFG	0.95	1	CCS	0.92	1	MI5
	2	MPI	0.91	2	CCS	0.94	2	GFG	0.92	2	GFM	0.89
	3	HA3	0.90	3	MPI	0.94	3	IPM	0.91	3	NOR	0.89
	4	CCS	0.90	4	IPM	0.94	4	MPI	0.91	4	MPI	0.89
	5	GFM	0.89	5	BCC	0.93	5	NOR	0.90	5	CNR	0.88
	17	HGC	0.77	17	NOR	0.84	17	BCC	0.73	17	IPL	0.76
	18	HGE	0.77	18	HE	0.84	18	MI4	0.72	18	IPM	0.74
	19	MRI	0.75	19	HGC	0.83	19	HGE	0.63	19	GIS	0.73
	20	MI4	0.75	20	GIS	0.80	20	HGC	0.60	20	MRI	0.66
	21	GIS	0.64	21	MRI	0.78	21	GIS	0.40	21	MI4	0.66

significant at the 0.001 confidence level based on a Student's  $t$  test. Other methods of computing statistical significance with a more robust technique for accounting for spatial autocorrelation were not pursued because of the apparent small value added for the type of analyses presented in this study.

### 5. Monsoon onset and retreat

Interannual variability of the NAMS impacts local ecosystems, agriculture, and the general public; therefore, there is considerable interest in determining the NAMS annual onset and retreat dates (e.g., Higgins et al. 1997; Zeng and Lu 2004; Ellis et al. 2004; Arias et al. 2012). Identification of model annual onset and retreat dates using daily precipitation provides a higher time resolution evaluation of monsoon season phasing than the monthly phase lag analysis in section 3. We employ the method of Higgins et al. (1997) for identifying the onset of monsoon rains. A precipitation index

(PI) is created by time averaging TRMM daily precipitation observations at each grid point in the core domain and subsequently averaging all grid points together. The PI time series, shown in Fig. 4, is used to set the precipitation magnitude and duration threshold criteria for defining monsoon onset at  $\geq 1.3 \text{ mm day}^{-1}$  and 3 days, respectively. These criteria are tested using individual years of daily TRMM observations to ensure they yield a set of reasonable onset dates. For each year, monsoon onset occurs the first time the threshold criteria are satisfied after 1 May. Since precipitation is much more variable following than preceding the monsoon season, the daily precipitation time series for each year of data are examined to set the monsoon retreat criteria at  $< 1.3 \text{ mm day}^{-1}$  and 7 days. For each year, monsoon retreat occurs when the threshold criteria are first satisfied after 1 September. Table 3 shows the observed monsoon onset and retreat dates for each year based on these criteria. The average calendar date of onset for the core domain is 18 June, which agrees with

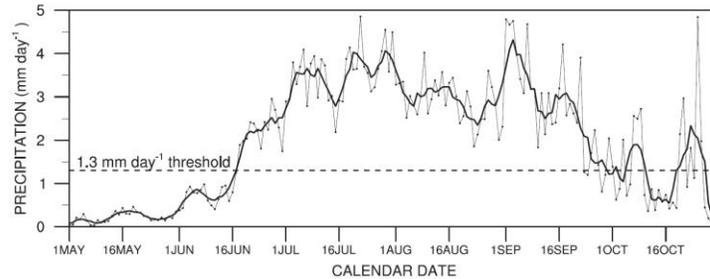


FIG. 4. Climatological (1998–2010) area-averaged daily (thin line) TRMM precipitation index and 5-day running mean (heavy line) in millimeters per day for the core NAMS domain.

the results of Higgins et al. (1999) (see their Fig. 12). The median date of onset is 19 June and the median date of retreat is 28 September.

Table 4 shows model median onset and retreat dates together with the corresponding lag from observations (in number of days) using the observation-defined threshold criteria, herein referred to as the absolute criteria. The CSI model onset is closest to the observed median onset date by a wide margin with a +3-day onset lag (Table 4) and also has a reasonable onset standard deviation (Table 4) and visual clarity of the onset in the daily climatology (Fig. 5). Most models have a standard deviation of onset dates that is within  $\pm 3$  days of that of the observations (15.2 days) and daily climatological precipitation that displays a clear monsoon onset, although generally onset is early. These properties are reflected in the multimodel average standard deviation of onset (14.9 days; Table 4); daily climatological precipitation (Fig. 5); and early 28 May onset date (Table 4), which is 22 days earlier than the observed date. The HGE model retreat is closest to the observed median retreat date with a +4-day retreat lag, but the standard deviation of retreat dates almost doubles that of the observational data and retreat is not as visually clear in the daily climatology as it is in the observations. The standard deviation of retreat dates for most models is amplified by at least a factor of 2 using the absolute criteria, as compared to the 8.3-day standard deviation of the observations. Most models also extend the monsoon past the observed retreat date. Additionally, most models lack a visually clear retreat in the daily climatological precipitation, with the end of the monsoon season blending into excessive fall and winter precipitation. These properties are also seen in the multimodel average standard deviation of retreat (18.1 days; Table 4); daily climatological precipitation (Fig. 5); and late 6 October retreat date (Table 4), which is 8 days later than the observed date.

The use of absolute criteria may not be suitable for many models due to errors in annual precipitation cycle amplitude; therefore, we have additionally computed model-relative threshold criteria as an alternate method of defining monsoon onset and retreat. Our method is constructed around that of Zeng and Lu (2004), who used a normalized precipitable water index to define globally unified summer monsoon onset and retreat dates. We use precipitation instead of precipitable water to define a normalized precipitation index (NPI) for the TRMM observations,

$$NPI = \frac{P_{\text{threshold}} - P_{\text{min}}}{P_{\text{max}} - P_{\text{min}}},$$

where  $P_{\text{threshold}}$  is the daily precipitation magnitude threshold of  $1.3 \text{ mm day}^{-1}$  and  $P_{\text{min}}$  and  $P_{\text{max}}$  are the 13-yr climatological values of the area-averaged minimum and maximum monthly TRMM precipitation in

TABLE 3. Observed dates of monsoon onset and retreat for the core NAMS domain using TRMM data.

Onset			Retreat		
20	Jul	1998	9	Sep	1998
17	Jun	1999	20	Sep	1999
31	May	2000	28	Sep	2000
3	Jul	2001	10	Oct	2001
3	July	2002	19	Sep	2002
17	Jun	2003	27	Sep	2003
4	Jun	2004	27	Sep	2004
10	Jul	2005	21	Sep	2005
19	Jun	2006	2	Oct	2006
10	Jun	2007	7	Oct	2007
28	Jun	2008	28	Sep	2008
18	May	2009	5	Oct	2009
2	Jul	2010	28	Sep	2010
Std dev: 15.2 days			Std dev: 8.3 days		
Avg: 18 Jun			Avg: 27 Sep		
Median: 19 Jun			Median: 28 Sep		

TABLE 4. Model median dates of monsoon onset and retreat with corresponding lag from observations and standard deviation (in number of days) using the observation-defined absolute threshold criteria for the core NAMS domain. Observational information is listed in the last row for reference.

Model	Median onset	Lag	Std dev	Median retreat	Lag	Std dev
BCC	2 Jun	-17	17.3	17 Sep	-11	18.4
CAN	8 Jun	-11	17.0	14 Sep	-14	10.1
CCS	9 May	-41	10.5	16 Oct	18	22.8
CNR	11 May	-39	14.4	3 Oct	5	17.9
CSI	22 Jun	3	13.1	6 Oct	8	11.9
GF3	21 May	-29	12.9	26 Oct	28	18.9
GFG	15 May	-35	15.3	18 Oct	20	22.6
GFM	27 May	-23	18.7	18 Oct	20	27.9
HA3	28 May	-22	18.3	22 Sep	-6	16.6
HGC	25 May	-25	14.1	7 Oct	9	15.7
HGE	6 Jun	-13	17.3	2 Oct	4	14.9
MI5	21 May	-29	10.8	12 Oct	14	23.8
MPI	5 Jun	-14	14.2	8 Oct	10	13.9
MRI	4 Jun	-15	14.6	6 Oct	8	18.2
Average	28 May	-22	14.9	6 Oct	8	18.1
Median	28 May	-22	14.5	7 Oct	9	18.0
Obs	19 Jun	0	15.2	28 Sep	0	8.3

the core NAMS domain. The computed NPI value (0.345) and the 27-yr climatological values of area-averaged minimum and maximum monthly precipitation from each model are then used to solve the NPI equation for the model-relative precipitation thresholds ( $P_{\text{threshold}}$ ; mm day<sup>-1</sup>), while the duration criterion for all models is held constant at 3 days for onset and 7 days for retreat. Onset and retreat are again defined as when the criteria are first satisfied after 1 May and 1 September, respectively. Model median onset and retreat dates with the corresponding lag from observations (in number of days) using the model-relative threshold criteria are shown in Table 5, and Fig. 5 depicts the difference between absolute and model-relative thresholds, shown over the daily climatological precipitation for each model and for the multimodel mean. The HGE and BCC models have onset dates that are closest to the observed median onset with a -2-day lag (Table 5), reasonable onset standard deviations (Table 5), and visual clarity of the onset in the daily climatologies (Fig. 5). The CSI model-relative retreat is the closest to the observed median retreat date with a +2-day lag, but the standard deviation of the retreat dates is about 1.5 times that of the observational data. Furthermore, although a retreat is visible in the daily climatology of CSI, the fall minimum is weak and quickly leads to exaggerated fall and winter precipitation. The multimodel average median onset is 12 days early on 7 June and the standard deviation of onset (Table 5) is similar to that of the observational data, while the average median

retreat is 4 days early on 24 September and the standard deviation of retreat (Table 5) is much larger than that of the observational data. The model-relative criteria act to adjust the multimodel onset and retreat dates 10 and 4 days closer to the observed dates, respectively.

Interestingly, high monthly and seasonal precipitation correlations do not necessarily imply monsoon onset and retreat dates that are close to observations. The best example of this is demonstrated with the results from the CNR model, which has the highest 850-hPa correlations and ranks in the top four models for precipitation correlations throughout the entire monsoon season (Table 2) but has a median absolute onset date that is 39 days earlier (Table 4) and a model-relative onset that is 29 days earlier (Table 5) than the observed date. A visual inspection of the daily precipitation field for each year (not shown) reveals consistent small precipitation events in early May that satisfy the definition of monsoon onset even though a clearer onset signal is visible later in the daily climatological precipitation. Adjustment of the monsoon onset definition was not able to remedy the problem. The disparity between the daily and monthly resolution analysis results for the CNR model is a good example of how important insight can be gained with higher time resolution model output and how it is possible for lower time resolution model output to be misleading. Still, the daily precipitation analysis could also be deceptive on its own, as an investigation of the six models deemed inappropriate for the daily analysis (not shown) reveals that it is possible for models to have relatively good onset and retreat dates over the small core NAMS region without the proper spatial progression of precipitation over time within the surrounding larger NAMS region.

## 6. Composites

Our final analysis visually elucidates the previously demonstrated wide range in model ability to reproduce key spatial and temporal features of the NAMS. Composites of the monthly fields of precipitation, 850-hPa geopotential height, and 850-hPa winds are constructed to illustrate the major differences between models with the best and worst representations of the NAMS based on the measures discussed below.

Models with large phase error at monthly time resolution (GIS, INM, and MIE) are not considered for compositing because of the unrecognizable or very poor representation of the annual cycle of NAMS region precipitation. Also, the monthly fields are examined for each individual model to rule out the presence of any spurious model output. During this process, anomalous large-scale waves were discovered in the geopotential height field from the CSI model and consequently this

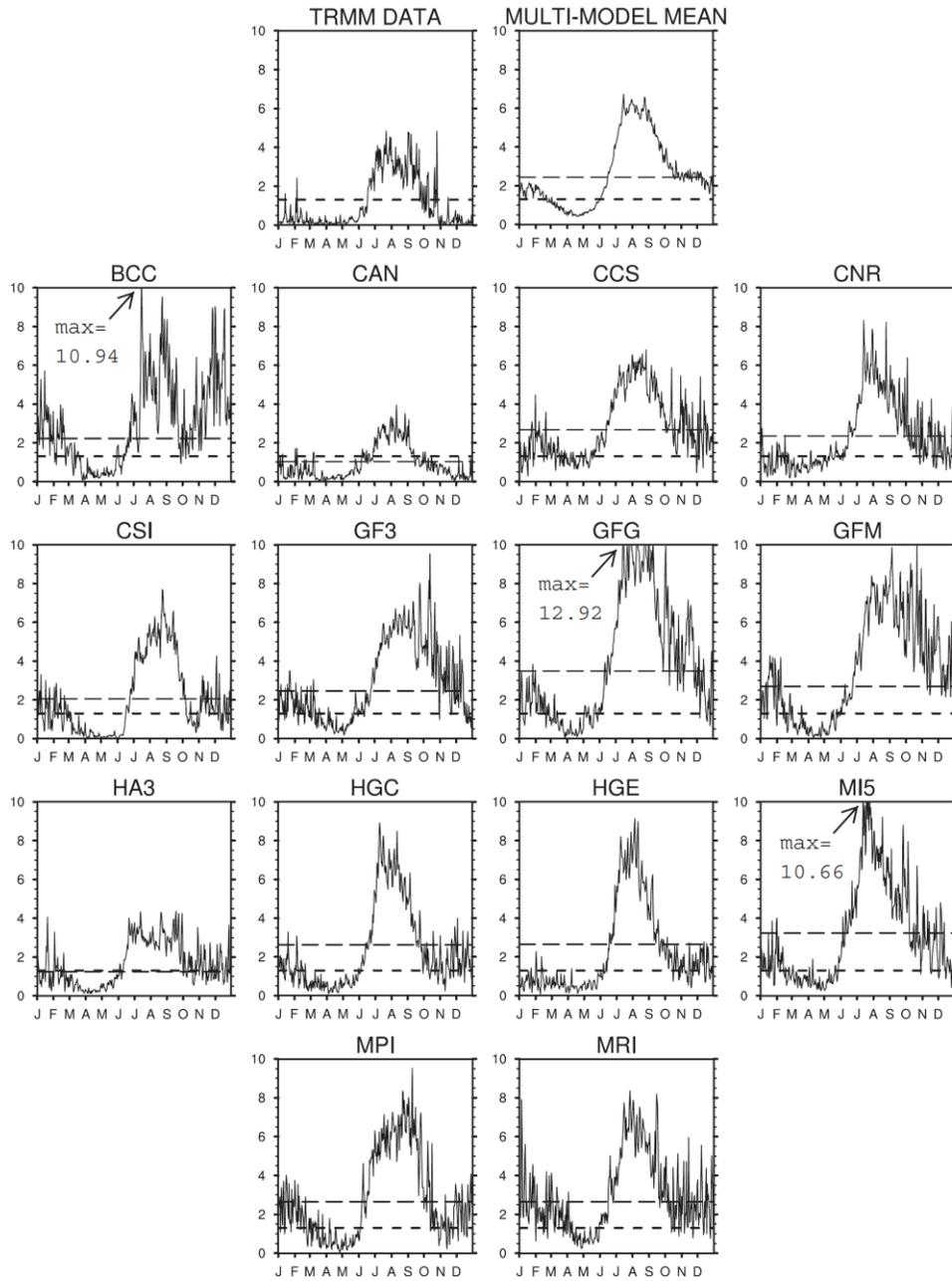


FIG. 5. Daily precipitation climatology over the core NAMS domain for (top left) observations (1998–2010), (top right) multimodel mean (1979–2005), and 14 models (1979–2005). Short dashed lines depict the absolute threshold ( $1.3 \text{ mm day}^{-1}$ ), and long dashed lines depict model-relative thresholds for defining yearly monsoon onset and retreat dates.

TABLE 5. As in Table 4, but using the model-relative threshold criteria.

Model	Median onset	Lag	Std dev	Median retreat	Lag	Std dev
BCC	17 Jun	-2	17.8	10 Sep	-18	10.8
CAN	3 Jun	-16	16.2	6 Sep	-22	7.8
CCS	20 May	-30	17.7	2 Oct	4	16.5
CNR	21 May	-29	15.9	21 Sep	-7	12.4
CSI	26 Jun	7	11.3	30 Sep	2	12.5
GF3	4 Jun	-15	18.2	6 Oct	8	18.9
GFG	10 Jun	-9	17.6	6 Oct	8	15.4
GFM	8 Jun	-11	20.7	5 Oct	7	28.1
HA3	28 May	-22	18.4	12 Sep	-16	10.8
HGC	7 Jun	-12	16.8	19 Sep	-9	10.5
HGE	17 Jun	-2	17.1	22 Sep	-6	12.4
M15	5 Jun	-14	10.9	25 Sep	-3	16.4
MPI	14 Jun	-5	14.3	1 Oct	3	13.3
MRI	9 Jun	-10	15.1	21 Sep	-7	15.5
Average	7 Jun	-12	16.3	24 Sep	-4	14.4
Median	8 Jun	-11	17.0	24 Sep	-4	12.9
Obs	19 Jun	0	15.2	28 Sep	0	8.3

model is not considered for compositing either. The best models must rank within the top five of the remaining models for seasonal precipitation and 850-hPa correlations and must have a daily absolute onset lag of less than 30 days, resulting in a composite of the HA3, HGE, and HGC models. The worst models must rank within the bottom five of the remaining models for seasonal precipitation and 850-hPa correlations, resulting in a composite of the IPL, IPM, and NOR models. We also note that all models satisfying the requirements for the best composite have been previously categorized as having small phase error at monthly time resolution, whereas all models satisfying the requirements for the worst composite have been previously categorized as having moderate phase error at monthly time resolution (see Figs. 3a,b).

Figure 6 shows precipitation, 850-hPa geopotential height, and 850-hPa winds for the observational data, the best composite, and the worst composite for the development stage month of June, the mature stage month of August, and the decay stage month of October. Circulation features important to the proper representation of the NAMS are clearly seen in the observations (Fig. 6, top). During the development and mature stages, the NASH, as seen in the geopotential height field, strengthens and extends toward the west. The westward extension of the NASH brings easterly flow and moisture into the eastern portion of the study region from the Gulf of Mexico and the northward movement of the NPSH allows for moisture transport into the western portion of the region from the eastern Pacific and the Gulf of California, enhancing convergence and precipitation over land. During this period, a clear northward

progression of precipitation along the Sierra Madre Occidental from southern Mexico is visible. The observations also show (Fig. 6, top) cross-equatorial flow, southwesterly winds in the ITCZ that create regions of enhanced convergence, and generally weak winds over the eastern North Pacific. Flow along the western coast of Mexico brings moisture from the ITCZ toward the Mexican mainland, whereas this flow weakens and turns offshore during the decay stage, which effectively severs the connection between the monsoon region and ITCZ moisture. Monsoon decay also brings the eastward retreat of the NASH; the weakening of the NASH, NPSH, and ITCZ; and the southeasterly retreat of precipitation along the Sierra Madre due to reduced convergence and moisture availability.

The composite of the three best models (Fig. 6, middle) illustrates most of the circulation features seen in the observations during the development and mature stages and demonstrates that current CGCM's are capable of realistically representing the NAMS, even at horizontal resolutions coarser than 1°. That being said, some significant issues are also apparent. For example, the 850-hPa geopotential heights are biased low, precipitation is biased high, and the connection to tropical moisture extends well into the fall, resulting in a poor representation of monsoon retreat.

The composite of the three worst models (Fig. 6, bottom) fails to adequately reproduce most of the circulation features seen in the observations. Although the composite does show strengthening and weakening of the NASH and NPSH, the NASH is overextended toward the west and the NPSH is far too weak, resulting in an anomalously strong north-south gradient of geopotential height over the eastern North Pacific. The gradient in the tropics produces strong zonal winds within and to the north of the ITCZ, preventing the proper development of the ITCZ and cutting off ITCZ moisture to the monsoon region. Reduced convergence over Latin America results in extremely poor representation of the tropical wet season over this area and prevents the observed progression of monsoon precipitation from the southeast toward the northwest along the mountainous west coast of Mexico. During the decay stage, the retreat of the subtropical highs shifts the unrealistic tropical gradient in geopotential height toward the north, which steers winds and moisture over the eastern North Pacific to the northwest along the Mexican coast, resulting in exaggerated precipitation over the core NAMS domain during the fall.

## 7. Summary

A total of 21 CMIP5 coupled general circulation models are examined to determine how well the models

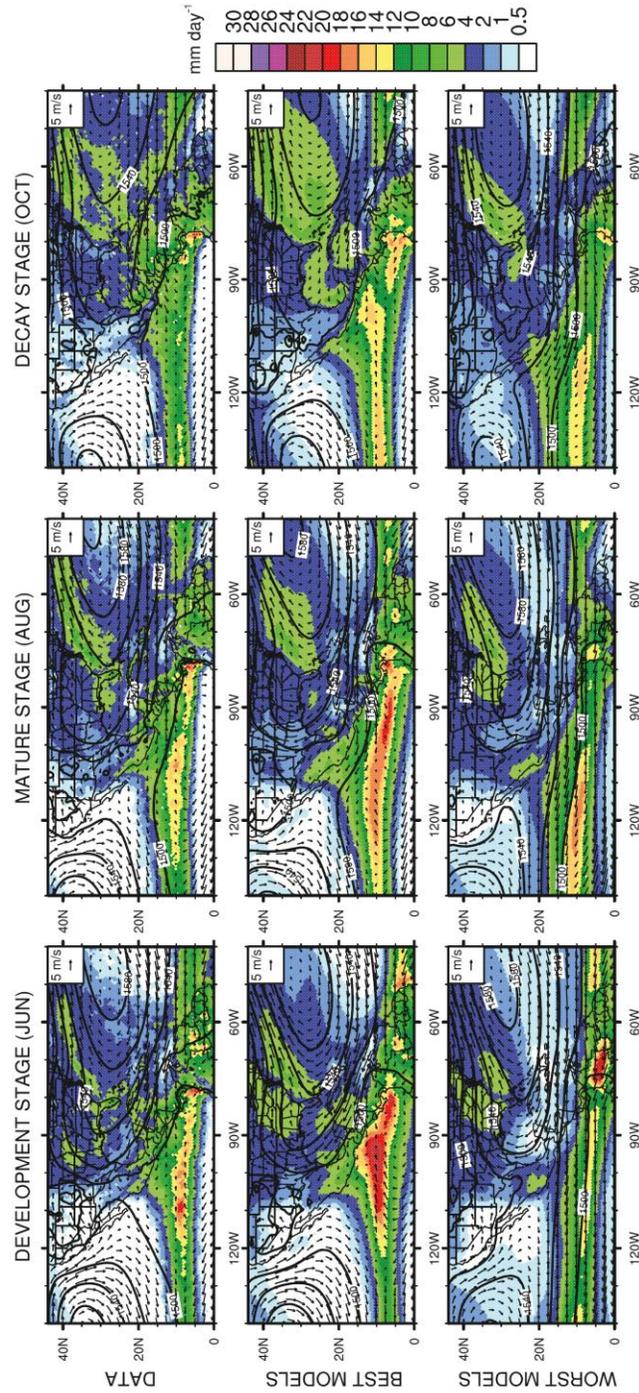


FIG. 6. Monthly climatological precipitation (color shading), 850-mb geopotential height (contours), and 850-mb winds (vectors) for (top) the data, (middle) the best model composite, and (bottom) the worst model composite for (left) the development stage month of June, (center) the mature stage month of August, and (right) the decay stage month of October.

represent the North American monsoon system. Analyses include a comparison of the annual cycle of area-averaged monthly precipitation over a core monsoon domain to CMIP3 results of Liang et al. (2008), a spatial correlation of monthly model precipitation and pressure level variables to data, an assessment of monsoon onset and retreat dates as determined from daily precipitation, and a model composite analysis of the best versus the worst representations of the NAMS.

There has been no improvement in the magnitude (rms error and bias) of the mean annual cycle of monthly precipitation over the core NAMS region since CMIP3, but the timing of seasonal changes in precipitation has improved with 27% more CMIP5 than CMIP3 models having zero phase lag. Despite this, a few models do not have a recognizable monsoon signal at all. Also, the multimodel mean annual cycle is biased wet and exhibits the common problem of late monsoon termination.

Monsoon season correlations of monthly model output to observational data establish that most models have the highest correlation at the 500-hPa level and the lowest correlations for precipitation; however, relatively good or bad performance at the 500-hPa level is not predictive of 850-hPa level or precipitation performance.

The multimodel mean onset and retreat dates are 23 days early and 9 days late, respectively, using the absolute criteria for defining monsoon onset and retreat. Yearly model onset variability is comparable to that of the observational data, but yearly model retreat variability is much greater than what is seen in the observations. On average, the model-relative onset and retreat dates are an improvement over the absolute dates because of the prevailing wet bias in model precipitation.

The 850-hPa composite of best models reproduces the development and mature stages of the NAMS, but the composite of worst models fails to adequately illustrate most of the precipitation and circulation features seen in the observations. The large-scale circulation pattern bias seen in the best model composite is spatially consistent over the larger region influencing monsoon development and thus still allows for a successful representation of the NAMS during the development and mature stages. In contrast, the spatial inconsistency of large-scale circulation pattern bias in the worst models prevents a realistic representation of the NAMS during the same period. Neither the composite of best models nor the composite of worst models realistically captures the retreat of the NAMS because of an extended connection to tropical moisture that causes excessive fall and winter precipitation. Models that best capture the relevant large-scale circulation patterns at low levels usually have a realistic representation of the NAMS,

while performance at midlevels does not appear to be a major factor.

We have shown the importance of large-scale features to the representation of the NAMS in a suite of CMIP5 models that are still relatively coarse for capturing the detailed regional topography and resultant small-scale NAMS processes. Model composites of high versus low horizontal and vertical resolutions (not shown) did not reveal major differences in model performance with respect to NAMS representation, implying that even the highest resolution model examined is still too coarse to capture small-scale topographically influenced processes. There is room for improvement in the representation of the NAMS for many models by way of more accurate representation of low-level large-scale circulation features, but improvement in the representation of the NAMS in the best models is likely limited until increased model resolution allows for the capture of small-scale NAMS processes. Finally, we encourage subsequent CMIP collaborations to output more daily model fields, which were not available for all models and variables examined in this study, thereby limiting most analyses to monthly time resolution.

*Acknowledgments.* This work was supported by the National Oceanic and Atmospheric Administration MAPP Grant GC10-398. We would like to thank Drs. Russ Vose and Ed Cook for providing the P-NOAA dataset. We also thank Dave Gochis and two anonymous reviewers for their constructive suggestions. We acknowledge the Computational and Information Systems Laboratory (CISL) at the National Center for Atmospheric Research (NCAR) for maintaining the RDA and the National Science Foundation for their sponsorship of NCAR. We additionally acknowledge the World Climate Research Programme's (WCRP) Working Group on Coupled Modelling (WGCM), which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their model output. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

#### REFERENCES

- Arias, P. A., R. Fu, and K. C. Mo, 2012: Decadal variation of rainfall seasonality in the North American monsoon region and its potential causes. *J. Climate*, **25**, 4258–4273.
- Arritt, R. W., D. C. Goering, and C. J. Anderson, 2000: The North American monsoon system in the Hadley Centre coupled ocean-atmosphere GCM. *Geophys. Res. Lett.*, **27**, 565–568.

- Badan-Dangon, A., C. E. Dorman, M. A. Merrifield, and C. D. Winant, 1991: The lower atmosphere over the Gulf of California. *J. Geophys. Res.*, **96** (C9), 877–896.
- Barlow, M., S. Nigam, and E. H. Berbery, 1998: Evolution of the North American monsoon system. *J. Climate*, **11**, 2238–2257.
- Berbery, E. H., 2001: Mesoscale moisture analysis of the North American monsoon. *J. Climate*, **14**, 121–137.
- Bosilovich, M. G., Y. C. Sud, S. D. Schubert, and G. K. Walker, 2003: Numerical simulation of the large-scale North American monsoon water sources. *J. Geophys. Res.*, **108**, 8614, doi:10.1029/2002JD003095.
- Castro, C. L., R. A. Pielke Sr., and J. O. Adegoke, 2007a: Investigation of the summer climate of the contiguous United States and Mexico using the Regional Atmospheric Modeling System (RAMS). Part I: Model climatology (1950–2002). *J. Climate*, **20**, 89–110.
- , —, —, S. D. Schubert, and P. J. Pegion, 2007b: Investigation of the summer climate of the contiguous United States and Mexico using the Regional Atmospheric Modeling System (RAMS). Part II: Model climate variability. *J. Climate*, **20**, 3866–3887.
- , H. Chang, F. Dominguez, C. Carrillo, J. Schemm, and H. H. Juang, 2012: Can a regional climate model improve the ability to forecast the North American monsoon? *J. Climate*, **25**, 8212–8237.
- Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.*, **33**, 140–158.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.
- Dominguez, F., P. Kumar, and E. R. Vivoni, 2008: Precipitation recycling variability and ecoclimatological stability—A study using NARR data. Part II: North American monsoon region. *J. Climate*, **21**, 5187–5203.
- Douglas, M. W., R. A. Maddox, and K. Howard, 1993: The Mexican monsoon. *J. Climate*, **6**, 1665–1677.
- Ellis, A. W., E. M. Saffell, and T. W. Hawkins, 2004: A method for defining monsoon onset and demise in the southwestern USA. *Int. J. Climatol.*, **24**, 247–265.
- Gao, X., J. Li, and S. Sorooshian, 2007: Modeling intraseasonal features of 2004 North American monsoon precipitation. *J. Climate*, **20**, 1882–1896.
- Gent, P. R., and Coauthors, 2011: The Community Climate System Model version 4. *J. Climate*, **24**, 4973–4991.
- Gochis, D. J., A. Jimenez, C. J. Watts, J. Garatuza-Payan, and W. J. Shuttleworth, 2004: Analysis of 2002 and 2003 warm-season precipitation from the North American Monsoon Experiment event rain gauge network. *Mon. Wea. Rev.*, **132**, 2938–2953.
- , S. W. Nesbitt, W. Yu, and S. F. Williams, 2009: Comparison of gauge-corrected versus non-gauge corrected satellite-based quantitative precipitation estimates during the 2004 NAME enhanced observing period. *Atmósfera*, **22**, 69–98.
- Gutzler, D. S., and Coauthors, 2009: Simulations of the 2004 North American monsoon: NAMAP2. *J. Climate*, **22**, 6716–6740.
- Higgins, R. W., and D. Gochis, 2007: Synthesis of results from the North American Monsoon Experiment (NAME) process study. *J. Climate*, **20**, 1601–1607.
- , Y. Yao, and L. Wang, 1997: Influence of the North American monsoon system on the U.S. summer precipitation regime. *J. Climate*, **10**, 2600–2622.
- , Y. Chen, and A. V. Douglas, 1999: Interannual variability of the North American warm season precipitation regime. *J. Climate*, **12**, 653–680.
- , and Coauthors, 2003: Progress in Pan American CLIVAR research: The North American monsoon system. *Atmósfera*, **16**, 29–65.
- , and Coauthors, 2006: The NAME 2004 field campaign and modeling strategy. *Bull. Amer. Meteor. Soc.*, **87**, 79–94.
- Huffman, G. J., and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.*, **8**, 38–55.
- Liang, X. Z., J. Zhu, K. E. Kunkel, M. Ting, and J. X. L. Wang, 2008: Do CGCMs simulate the North American monsoon precipitation seasonal-interannual variability? *J. Climate*, **21**, 4424–4448.
- NCAR, cited 2012: NCAR Command Language, version 6.0.0. [Available online at <http://www.ncl.ucar.edu/>]
- Rowson, R. W., and S. J. Colucci, 1992: Synoptic climatology of thermal low-pressure systems over south-western North America. *Int. J. Climatol.*, **12**, 529–545.
- Schmitz, J. T., and S. L. Mullen, 1996: Water vapor transport associated with the summertime North American monsoon as depicted by ECMWF analyses. *J. Climate*, **9**, 1621–1634.
- Stensrud, D. J., R. L. Gall, S. L. Mullen, and K. W. Howard, 1995: Model climatology of the Mexican monsoon. *J. Climate*, **8**, 1775–1794.
- Tang, M., and E. R. Reiter, 1984: Plateau monsoons of the Northern Hemisphere: A comparison between North America and Tibet. *Mon. Wea. Rev.*, **112**, 617–637.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2009: A summary of the CMIP5 experiment design. PCDMI Rep., 33 pp. [Available online at [http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor\\_CMIP5\\_design.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf)]
- , —, and —, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498.
- Volodin, E. M., N. A. Dianskii, and A. V. Gusev, 2010: Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations. *Izv. Atmos. Ocean. Phys.*, **46**, 414–431, doi:10.1134/S000143381004002X.
- Yang, Z. L., D. Gochis, and W. J. Shuttleworth, 2001: Evaluation of the simulations of the North American monsoon in the NCAR CCM3. *Geophys. Res. Lett.*, **28**, 1211–1214.
- Zeng, X., and E. Lu, 2004: Globally unified monsoon onset and retreat indexes. *J. Climate*, **17**, 2241–2248.

## APPENDIX B

### QUANTITATIVE CHARACTERIZATION OF SPURIOUS NUMERICAL OSCILLATIONS IN 48 CMIP5 MODELS

(Published in Geophysical Research Letters)

KERRIE L. GEIL<sup>1</sup> AND XUBIN ZENG<sup>1</sup>

<sup>1</sup>Department of Atmospheric Sciences, University of Arizona, Tucson, Arizona

Geil, K.L. and X. Zeng, 2015: Quantitative characterization of spurious numerical oscillations in 48 CMIP5 models. *Geophys Res Lett*, 42, 5066-5073, doi:10.1002/2015GL063931.

© American Geophysical Union. Used with permission.



## RESEARCH LETTER

10.1002/2015GL063931

## Key Points:

- Large spurious numerical oscillations (SNOs) still exist in many CMIP5 models
- The oscillations are often larger than atmospheric interannual variability
- SNOs could compromise climate simulations and must be substantially reduced

## Supporting Information:

- Table S1
- Figures S1–S5 and Table S1

## Correspondence to:

K. L. Geil,  
geil@atmo.arizona.edu

## Citation:

Geil, K. L., and X. Zeng (2015), Quantitative characterization of spurious numerical oscillations in 48 CMIP5 models, *Geophys. Res. Lett.*, *42*, 5066–5073, doi:10.1002/2015GL063931.

Received 19 MAR 2015  
Accepted 11 MAY 2015  
Published online 29 JUN 2015

©2015. American Geophysical Union.  
All Rights Reserved.

## Quantitative characterization of spurious numerical oscillations in 48 CMIP5 models

Kerrie L. Geil<sup>1</sup> and Xubin Zeng<sup>1</sup>

<sup>1</sup>Department of Atmospheric Sciences, University of Arizona, Tucson, Arizona, USA

**Abstract** Spurious numerical oscillations (SNOs) (e.g., Gibbs oscillations) can appear as unrealistic spatial waves near discontinuities or sharp gradients in global model fields (e.g., orography) and have been a known problem in global models for decades. Multiple methods of oscillation reduction exist; consequently, the oscillations are presumed small in modern climate models and hence are rarely addressed in recent literature. Here we use two metrics to quantify SNOs in 13 variables from 48 Coupled Model Intercomparison Project Phase 5 models along a Pacific ocean transect near the Andes. Results show that 48% of nonspectral models and 95% of spectral models have at least one variable with SNO amplitude as large as, or greater than, atmospheric interannual variability. The impact of SNOs on climate simulations should be thoroughly evaluated and further efforts to substantially reduce SNOs in climate models are urgently needed.

## 1. Introduction

Global climate models play a critical role in our understanding of climate processes and our ability to make climate projections. They are an invaluable tool in a climate scientist's toolbox, and projections from these models are increasingly used by nonscientists for climate planning and adaptation purposes. The ability to model climate has undergone vast improvement in recent decades, but model development is an ongoing process and it is no secret that model deficiencies still exist. Many deficiencies are complex and difficult to pinpoint, whereas other deficiencies are well known and can be reduced or eliminated using proven methods.

Spurious numerical oscillations (hereafter referred to as SNOs) are a well-known source of numerical noise in global climate models. In spectral models, SNOs (also known as (aka) Gibbs oscillations) are unrealistic spatial waves that appear in model fields, such as orography, that contain discontinuities or sharp gradients and are mainly associated with the transformation of the truncated spectral representation of a field to physical space. For spherical harmonics, used in spectral global climate models, the oscillations come from two sources, namely, the Fourier transform for longitude and the Legendre transform for latitude. Similar looking SNOs are present near sharp gradients in nonspectral models (e.g., models that use only finite difference, finite element, and finite volume methods), although these oscillations are usually more localized.

Spectral model results have been shown to be sensitive to the transformed orography and spectral resolution, where simulation of variables such as precipitation over mountainous terrain is more realistic across multiple scales when using smoothed orography and when the same model is run at higher resolution versus a lower resolution [Lindberg and Broccoli, 1996; Yorgun and Rood, 2014, 2015]. Local-, regional-, and global-scale precipitation patterns, among other model variables, can be affected by SNOs. Local issues can include grid-point storms near mountainous terrain caused by spurious vertical velocity associated with SNOs [Webster *et al.*, 2003] and a connection between SNOs and unrealistic bands of precipitation [Bouteloup, 1995]. Locally and regionally, poor representation of precipitation near mountainous terrain [Bala *et al.*, 2008; Yorgun and Rood, 2014, 2015] has been associated with spectral numerics. SNOs have been the cause of unrealistic “spotty” precipitation over the Sahel region of Africa [Navarra *et al.*, 1994] and have also been shown to be detrimental to global precipitation patterns [Lindberg and Broccoli, 1996]. Additionally, SNOs have been associated with the poor representation of low clouds, radiation, surface wind stress, and sea surface temperature near upwelling regions [Bala *et al.*, 2008], clouds, and low-level meridional wind [Navarra *et al.*, 1994] and near-surface winds [Bouteloup, 1995]. It is also important to note that SNO amplitude is variable across model quantities, which can result

in physical inconsistencies, nonlinear interactions between model physics and dynamics, and have serious impacts on model budgets [Jablonowski and Williamson, 2011]. This point becomes especially important when considering tracer transport in a model, where spurious oscillations can cause unphysical quantities such as negative mixing ratios.

A number of techniques have been developed to reduce the unwanted SNOs in order to improve model results. In spectral models, these techniques include filters that are applied to the harmonic coefficients of the spectral representation of a field [e.g., Hoskins, 1980; Sardeshmukh and Hoskins, 1984; Navarra *et al.*, 1994; Lindberg and Broccoli, 1996] and a variational method that minimizes the difference between an actual field and the grid point representation of the spectral field using a cost function that allows for geographic modulation [Bouteloup, 1995]. In nonspectral models, SNOs can be reduced by physical diffusion of subgrid-scale energy within model parameterizations [Pielke, 2002; Warner, 2011], with the use of spatial diffusion terms (filters) in the predictive equations to numerically diffuse shorter wavelengths [Pielke, 2002; Warner, 2011] and with the use of certain implicit numerical schemes [Navarra *et al.*, 1994; Warner, 2011]. For an additional source of detailed information on filters and diffusion in global spectral and nonspectral models, the reader is referred to Jablonowski and Williamson [2011].

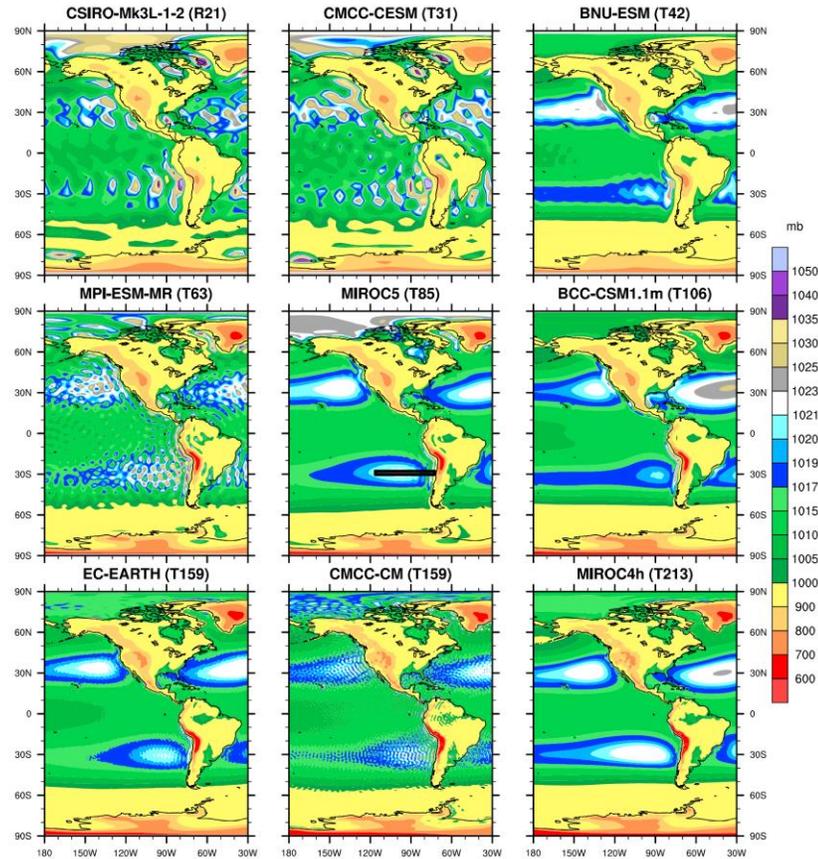
Given the detrimental nature of SNOs to model simulations, climate simulations based on models containing these oscillations could be compromised. The questions are then how pervasive are SNOs in Coupled Model Intercomparison Project Phase 5 (CMIP5) models and what are their quantitative characteristics? The purpose of this paper is to address these questions and more generally to draw greater awareness of SNOs in CMIP5 models, with the hope that modeling groups will act to substantially reduce oscillation biases.

## 2. Model Simulations, Observations, and Methods

The source of climate simulations is the CMIP5 multimodel ensemble archive (<http://pcmdi9.llnl.gov>). For this analysis, we use one ensemble member for 27 years (1979–2005) of the historical experiment, which imposes changing atmospheric and land surface conditions consistent with past observations. More detail regarding CMIP5 experimental design is provided in Taylor *et al.* [2009, 2012]. Table S1 in the supporting information provides information, including atmospheric model component resolution, on the models used for this study.

For model evaluations, we choose observational data sets with minimal or no use of global models during the data development process. For this reason, reanalyses are not used. Wind observations are from the Cross-Calibrated Multi-Platform (CCMP) ocean surface wind vector analyses [Atlas *et al.*, 2011], precipitation observations are from the Tropical Rainfall Measuring Mission (TRMM) 3B43v7 product [Huffman *et al.*, 2007], and air temperature and specific humidity observations at 1000 mb are from the Atmospheric Infrared Sounder (AIRS) [Aumann *et al.*, 2003]. Observations of downwelling shortwave and longwave radiation at the surface come from the Clouds and the Earth's Radiant Energy System (CERES) [Kato *et al.*, 2013], and total cloud amount is from the Moderate Resolution Imaging Spectroradiometer (MODIS) [Hubanks *et al.*, 2008]. All observations used are at monthly time resolution, and the number of consecutive years used is data set dependent, with the shortest observational span being 8 years for AIRS (2003–2010) and the longest being 24 years for CCMP (1988–2011).

An ocean transect at approximately 29°S that bisects the South Pacific High near the Andes is used to examine the SNOs where they are most easily visible based on Figure 1: over the ocean and near a steep topographic gradient. We use 27-year climatological transects of monthly model variables over ocean points only and compute smoothed versions of the transects by applying a running mean. Three points are used to compute the running mean for the large majority of models, but we use four or five points for a small number of models based on visual inspection of the best fit for each model (see Figure S1 in the supporting information), in an attempt to obtain the smoothest running mean possible without losing too many end points. This is necessary due to varying model resolutions and varying SNO wavelengths even within specific spectral models. A sensitivity test of our results to the number of points used to compute the running mean was performed and revealed low sensitivity. The transect latitude is chosen as the

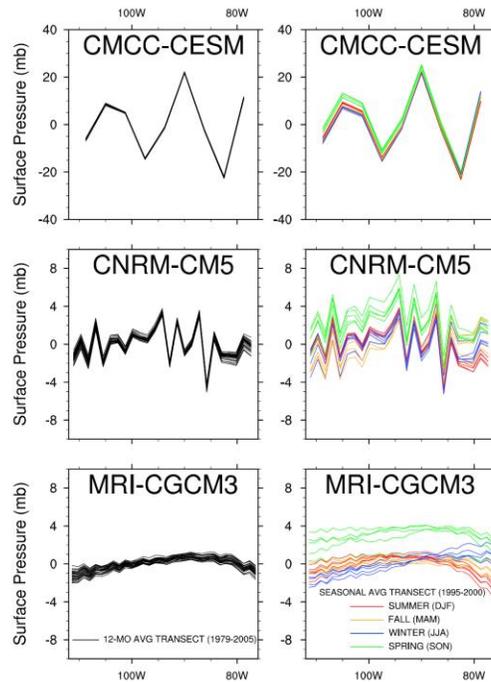


**Figure 1.** Surface pressure (mb) for nine spectral models (spectral resolution increasing from top left to bottom right) shows the large range in wavelength and amplitude of the spurious numerical oscillations (aka Gibbs oscillations in spectral models). The middle plot displays the location of the transect (horizontal black line) used to quantify numerical oscillations in subsequent figures.

closest model latitude to 29°S; therefore, it varies between 27.8°S and 30.3°S. Observations are treated in the same way except that the number of years used for the transect climatology is data set dependent and the number of points used for the running mean varies from 3 to 10.

Polynomial and spline curve fitting methods were also attempted, but the running mean is the superior smoothing method because it consistently aligns best with the unsmoothed transects. The small area that contains all the model and data transects is shown in the middle plot of Figure 1. A closer view of the transect area over climatological (1979–2004) sea level pressure from the HadSLP2 data set [Allan and Ansell, 2006] is shown in Figure S2 in the supporting information. See Figure S1 in the supporting information for examples of the variable transects and running mean curves.

We identify and quantitatively characterize SNOs along the transect using two metrics. The first is the root-mean-square difference (RMSD) between the climatological transect of a variable and its running mean, representing an absolute measure of the oscillation amplitude. A relative measure is computed as the ratio



**Figure 2.** Transects for a spectral model with (top) large, (middle) moderate, and (bottom) small amplitude surface pressure oscillations. The left column shows 12 month average transects during the 27 year study period, where the mean of each transect has been removed. The right column shows seasonal average transects for five separate years (mean removed).

details). The linear correlation between the average oscillation wavelength along the transect and the physical grid spacing in spectral models is high ( $\rho = 0.88$ ) and the average wavelength ranges from  $2.5\Delta x$  to  $4.5\Delta x$  (where  $\Delta x$  is grid size in the zonal direction). For nonspectral models, the correlation is slightly lower ( $\rho = 0.69$ ) and the oscillation wavelength is  $2\Delta x$ , with few exceptions.

Some models, such as the MIROC4h and BCC-CSM1.1 m shown in Figure 1, have very limited or small SNOs, and the biggest issue appears to be the spike in surface pressure just off the west coast of continents, especially near high terrain like the Andes (also see Figure S3 in the supporting information). Most other models also have this issue, but the problem exists in combination with larger-amplitude SNOs that spread across the oceans. Models with larger-amplitude SNOs, such as the CSIRO-Mk3L-1-2, CMCC-CESM, and MPI-ESM-MR models, do not even coherently capture the surface subtropical centers of high pressure. Note that SNOs are present not just over the South Pacific Ocean but can be seen globally over ocean regions and can also be seen over some land regions (see, for example, the Amazon region in the MPI-ESM-MR, MIROC5, and CMCC-CM plots in Figure 1).

Figure 2 shows the seasonal and interannual variations of SNOs for three spectral models with large-, moderate-, and small-amplitude oscillations in surface pressure. It is clear that the minima and maxima of the oscillations are stationary both interannually and interseasonally. The stationarity of the oscillations reinforces that they are spurious oscillations as opposed to physical waves resultant from other model processes.

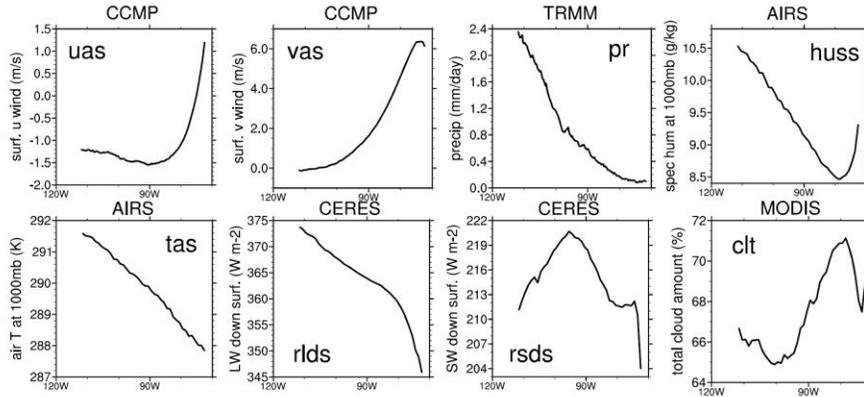
Before computing the model SNO metrics, we first determine if physical oscillations exist along the transect based on observations. Figure 3 shows climatological transects for eight observational data sets. While there

of the RMSD to interannual variability (IAV). At each transect point, the standard deviation in time is first calculated using annual average values. IAV is then obtained as the average along the entire transect.

### 3. Results

Ideally, we would begin the analysis by looking at orography, but many of the spectral models do not provide the transformed field, such that spurious numerical oscillations are not present in the model orography output available from the PCMDI archive. Therefore, a suitable starting point is the surface pressure field.

Figure 1 shows the large range in wavelength and amplitude of SNOs from nine spectral models. In spectral models, the oscillation wavelength decreases with increasing spectral resolution. For example, along the ocean transect shown in the middle plot (also see Figure S2 in the supporting information), the coarse resolution (R21) CSIRO-Mk3L-1-2 model has an average SNO wavelength of  $\sim 12^\circ$ . The wavelength decreases to  $\sim 9^\circ$  in the MPI-ESM-MR model (midresolution of T63) and to  $\sim 3^\circ$  in the finer resolution (T159) CMCC-CM model (see Table S1 for model



**Figure 3.** Climatological transects for eight satellite observational data sets (see section 2). From top left to bottom right: Cross-Calibrated Multi-Platform (CCMP) zonal ( $U$ ) and meridional wind ( $V$ ), Tropical Rainfall Measuring Mission (TRMM) precipitation ( $pr$ ), Atmospheric Infrared Sounder (AIRS) specific humidity ( $Q$ ) and air temperature ( $T$ ) at 1000 mb, Clouds and the Earth’s Radiant Energy System (CERES) downwelling longwave ( $rlds$ ) and shortwave ( $rsds$ ) radiation at the surface, and Moderate Resolution Imaging Spectroradiometer (MODIS) total cloud amount ( $ct$ ).

are no large oscillations such as those seen in Figures 1 and 2, some observations do show small bumps or ripples along the transect (see precipitation ( $pr$ ), surface downwelling shortwave radiation ( $rsds$ ), and total cloud amount ( $clt$ ) plots in Figure 3). For this reason, we compute the SNO metrics based on the observations as a reference to be compared to model results. Only when the model metric values are much greater than the data metric values can we claim the model oscillations are spurious. For RMSD, the threshold value is arbitrarily defined as the observed metric value multiplied by a large factor of 5. For the RMSD:IAV ratio, we use a threshold value of unity (meaning the amplitude of spurious oscillations is at least as large as atmospheric interannual variability along the transect) and a more restrictive threshold of one half, which is still at least 5 times the RMSD:IAV ratio value for any given observational variable.

RMSD values along the ocean transect for 13 variables (columns) are shown for the 90th, 50th, and 10th percentile spectral and nonspectral models in Figure 4, along with the RMSD value based on observations (bottom row) if available. RMSD values larger than the observational value times a factor of 5 can be seen even in the 10th percentile of spectral models (only for downwelling longwave radiation at the surface;  $rlds$ ). At the 90th percentile, seven out of the eight spectral model variables with observations for

	ps	uas	vas	wap	wap	pr	huss	tas	rsds	rlds	zg	zg	clt	
	(mb)	(m/s)	(m/s)	925 (Pa/s)	500 (Pa/s)	(mm/day)	(g/kg)	(K)	( $W/m^2$ )	( $W/m^2$ )	925 (m)	500 (m)	(%)	
spectral	90th	8.428	0.225	0.375	0.031	0.008	0.164	0.065	0.427	8.782	4.668	3.201	1.718	5.208
	50th	2.339	0.081	0.187	0.017	0.006	0.076	0.049	0.136	3.153	1.389	0.549	0.496	1.523
	10th	0.915	0.031	0.096	0.003	0.002	0.027	0.020	0.055	1.262	0.816	0.235	0.288	0.645
finite	90th	1.425	0.332	0.691	0.016	0.030	0.068	0.079	0.289	3.609	2.604	1.826	1.453	2.171
	50th	0.175	0.073	0.193	0.002	0.004	0.023	0.034	0.078	0.703	0.372	0.589	0.269	0.437
	10th	0.026	0.018	0.016	0.001	0.002	0.009	0.019	0.021	0.248	0.116	0.158	0.118	0.265
OBS	---	0.023	0.030	---	---	0.021	0.023	0.034	0.383	0.147	---	---	0.165	

**Figure 4.** RMSD percentile values of spectral and nonspectral models (refer to text for explanation of RMSD computation) for 13 variables, which include surface pressure ( $ps$ ), near-surface u-wind ( $uas$ ), and v-wind ( $vas$ ), vertical velocity at 925 mb ( $wap$  925) and 500 mb ( $wap$  500), precipitation ( $pr$ ), surface specific humidity ( $huss$ ), surface air temperature ( $tas$ ), surface incoming solar radiation ( $rsds$ ) and incoming longwave radiation ( $rlds$ ), geopotential height at 925 mb ( $zg$  925) and 500 mb ( $zg$  500), and total cloud amount ( $clt$ ). RMSD values for observational data are shown on the bottom row. The red text indicates values greater than or equal to the observed value multiplied by a factor of 5, whereas the blue text indicates values below this threshold. Model results are shown in black for variables when there are no observations for comparison.

	ps	uas	vas	wap 925	wap 500	pr	huss	tas	rsds	rlds	zg 925	zg 500	clt	
spectral	90th	10.735	0.523	0.958	6.611	0.976	0.539	0.305	1.208	1.923	1.846	0.474	0.182	1.630
	50th	3.361	0.223	0.580	2.758	0.565	0.236	0.195	0.469	0.664	0.474	0.115	0.056	0.610
	10th	1.181	0.084	0.237	0.805	0.209	0.081	0.068	0.172	0.244	0.314	0.041	0.024	0.254
finite	90th	1.826	0.705	1.718	3.548	3.098	0.177	0.319	0.893	0.691	0.840	0.260	0.148	0.643
	50th	0.240	0.164	0.431	0.493	0.475	0.086	0.137	0.208	0.137	0.127	0.085	0.023	0.138
	10th	0.029	0.036	0.039	0.320	0.189	0.025	0.076	0.066	0.048	0.048	0.022	0.011	0.079
OBS	---	0.042	0.063	---	---	0.076	0.067	0.073	0.086	0.073	---	---	0.060	

Figure 5. Same as Figure 4 except for the RMSD:IAV ratio. The red and orange text highlight values greater than unity and one half, respectively. Values less than one half are shown in black.

comparison have RMSD values larger than the threshold. For nonspectral models, large RMSD values are only found for one variable at the 50th percentile (surface meridional wind; vas) and for six out of the eight variables with observations for comparison at the 90th percentile, although many of these values are smaller than those seen for spectral models. Some models, mostly spectral, have RMSD values that are very large. For example, the largest RMSD values for surface pressure, downwelling shortwave radiation at the surface, and total cloud amount are 18.88 mb, 9.31 W m<sup>-2</sup>, and 5.23%, respectively (see Figure S4 in the supporting information for individual model results).

Figure 5 shows that the observed RMSD:IAV ratio (bottom row) is always less than 0.1 for each of the eight data sets. Regardless of spectral versus nonspectral numerics, many models have RMSD:IAV ratios that are less than the threshold values. Variables that are relatively oscillation-free include specific humidity at the surface (huss) and geopotential height at 925 and 500 mb (zg 925, zg 500). For most spectral models, the RMSD:IAV ratio is highest for surface pressure and vertical velocity at 925 mb (see Figure S5 in the supporting information for individual model results). Large oscillations are seen in these two variables even at the 10th percentile in spectral models. For one spectral model, the amplitude of SNOs in the surface pressure field is a staggering 29.32 times larger than the year-to-year variability along the transect (Figure S5 in the supporting information). In nonspectral models, the largest-amplitude oscillations are found in the vertical velocity field at 925 and 500 mb. Nonspectral models have no RMSD:IAV ratios over the threshold values at the 50th percentile, whereas spectral models show large RMSD:IAV ratios at the 50th percentile for surface pressure, vertical velocity at 925 and 500 mb, meridional surface wind, incoming shortwave and longwave radiation at the surface, and total cloud amount (Figure 5). At the 90th percentile, both spectral and nonspectral models have large RMSD:IAV ratio values for most variables, although most spectral model values are larger.

Of the 48 models in this study, 69% of the models have an RMSD:IAV ratio larger than unity for at least one variable, and this increases to 83% of all models if considering an RMSD:IAV ratio of one half. This statistic can be translated in terms of spectral versus nonspectral numerical methods as follows: 95% of spectral models and 48% of nonspectral models have at least one RMSD:IAV ratio greater than unity, and these percentages increase to 100% of spectral and 70% of nonspectral models if considering an RMSD:IAV ratio of one half.

While model resolution affects the wavelength of SNOs as discussed previously, it is not a predictor of oscillation amplitude or prevalence. The linear correlation between average oscillation amplitude along the transect and physical grid spacing is low for spectral ( $\rho = 0.37$ ) and nonspectral models ( $\rho = 0.34$ ). The models with large values of both metrics for multiple variables span the gamut of model resolution. These models include CMCC-CESM (low resolution of T31/3.75°), the MPI suite (moderate resolution of T63/1.875°), and the MRI suite (higher resolution of T159/1.125°). Additionally, the best performing spectral models with respect to both metrics (EC-EARTH, MIROC4h, and BNU-ESM) are a mixture of lower and higher-resolution models, with resolutions of T159/1.125°, T213/0.5625°, and T42/2.8°, respectively (see Figures S4 and S5 in the supporting information for individual model results and Table S1 for additional model details).

Note that sometimes large values of both metrics can be caused not by large SNOs along the entire transect, but instead by more localized issues near the ocean-land transition. This occurs mostly in nonspectral models. For example, the premature decrease of surface pressure over ocean points approaching land (as illustrated

for some models in Figure S3 in the supporting information) causes high values of both metrics for the FGOALS-g2, INM-CM4, and most of the GFDL suite models. There are also a few spectral models with generally small spurious oscillations in surface pressure that end up with large values for both metrics because of one very large oscillation near the ocean-land transition. These models include BCC-CSM1.1 m and the MIROC suite (see the BCC-CSM1.1 m, MIROC5, and MIROC4h plots in Figure 1 and Figure S3 in the supporting information). Steep and most likely spurious gradients are also seen in the meridional surface wind field over ocean points approaching land in the GISS suite and HadCM3 models (Figure S3 in the supporting information).

#### 4. Summary

The presence of spurious numerical oscillations (SNOs) in global climate models has been known for decades and has been previously shown to cause poor representation of precipitation, wind, sea surface temperature, clouds, and more. The SNOs (in the form of Gibbs oscillations) are most prevalent in models that use spectral numerics and could compromise the results of scientific climate analyses. This study provides a quantitative characterization of the SNOs in 48 CMIP5 models to draw awareness to the large SNOs present in these models.

For variables that have observations for comparison, 40% of models on average have RMSD values greater than the RMSD value for observations along the transect multiplied by a large factor of 5 (see Figure S4 in the supporting information). Furthermore, 69% of the models have an RMSD:IAV ratio that is as large as, or larger than, interannual variability along the study transect for at least one variable. This translates to 95% of spectral models and 48% of nonspectral models having at least one RMSD:IAV ratio greater than unity. The largest SNOs by absolute and relative measures are seen in spectral models and in the surface pressure field, although smaller SNOs are visible in many of the variables examined. For 8 of the 13 variables, at least one model (or as many as half for surface pressure) has SNOs with amplitude as large as, or much larger than, the interannual variability of those variables along the transect. These variables include surface pressure, surface meridional winds, vertical velocity, surface air temperature, incoming surface radiation, and total cloud amount. Also, regardless of the numerical method employed, model resolution does not predict oscillation amplitude or prevalence.

The presence of large stationary numerical oscillations with amplitudes on the scale of atmospheric interannual variability suggests that these oscillations are spurious and should not be ignored. Despite this, SNOs are rarely mentioned in CMIP analysis literature probably because they are perceived as being small in modern climate models. Given past research by others and our present findings, there is no reason to believe that the spurious oscillations are benign to climate simulations and they could very well have harmful impacts on the representation of variables at local, regional, and global scales. Future studies are needed to quantify how the SNOs affect model climate processes or the quality and robustness of the model simulations in general. The statistics presented in this paper could affect the design of future CMIP analyses, and we encourage the CMIP analysis community to address the potential impacts of these findings. The CMIP modeling groups are also urged to share information on the specific treatments of the oscillations or lack thereof so that SNOs can be substantially reduced in all climate models (e.g., for future CMIP activities).

#### References

- Allan, R., and T. Ansell (2006), A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004, *J. Clim.*, *19*, 5816–5842, doi:10.1175/JCLI3937.1.
- Atlas, R., R. N. Hoffman, J. Ardizzone, S. M. Leidner, J. C. Jusem, D. K. Smith, and D. Gombos (2011), A cross-calibrated, multiplatform ocean surface velocity product for meteorological and oceanographic applications, *Bull. Am. Meteorol. Soc.*, *92*, 157–174, doi:10.1175/2010BAMS2946.1.
- Aumann, H. H., et al. (2003), AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products, and processing systems, *IEEE Trans. Geosci. Remote Sens.*, *41*(2), 253–264, doi:10.1109/TGRS.2002.808356.
- Bala, G., R. B. Road, D. Bader, A. Mirin, D. Ivanova, and C. Drui (2008), Simulated climate near steep topography: Sensitivity to numerical methods for atmospheric transport, *Geophys. Res. Lett.*, *35*, L14807, doi:10.1029/2008GL033204.
- Bouteloup, Y. (1995), Improvement of the spectral representation of the Earth topography with a variational method, *Mon. Weather Rev.*, *123*, 1560–1573.
- Hoskins, B. J. (1980), Representation of the Earth topography using spherical harmonics, *Mon. Weather Rev.*, *108*, 111–115.
- Hubanks, P. A., M. D. King, S. Platnick, and R. Pincus (2008), MODIS Atmosphere L3 Gridded Product Algorithm Theoretical Basis Document No. ATBD-MOD30, *Tech. Rep.*, National Aeronautics and Space Administration.

#### Acknowledgments

The AIRS data set used in this work was obtained from the ESGF obs4MIPs project (<https://www.earthsystemcog.org/projects/obs4mips/>). The CCMP data set can be obtained through the NASA/JPL/PODAAC (<http://podaac.jpl.nasa.gov/>). TRMM data are from NASA/GES/DISC ([disc.sci.gsfc.nasa.gov/](http://disc.sci.gsfc.nasa.gov/)). CERES data are provided by the NASA LaRC (<http://ceres.larc.nasa.gov/>). The level 3 MODIS MCD08\_M3\_NC data product is obtained from NASA/LP/DAAC and USGS/EROS ([https://lpdaac.usgs.gov/data\\_access/](https://lpdaac.usgs.gov/data_access/)). We acknowledge the WCRP Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table S1 in the supporting information) for producing and making available their model output. For CMIP the DOE/PCMDI provides coordinating support and led development of software infrastructure. This work was supported by the NSF (AGS-0944101) and NASA (NNX14AM02G).

The Editor thanks two anonymous reviewers for their assistance in evaluating this paper.

- Huffman, G. J., R. F. Adler, D. T. Bolvin, G. Gu, E. J. Nelkin, K. P. Bowman, Y. Hong, E. F. Stocker, and D. B. Wolff (2007), The TRMM multi-satellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scale, *J. Hydrometeorol.*, *8*(1), 38–55, doi:10.1175/JHM560.1.
- Jablonowski, C. and D. L. Williamson (2011), The pros and cons of diffusion, filters and fixers in atmospheric general circulation models, in *Numerical Techniques for Global Atmospheric Models, Lecture Notes in Computational Science and Engineering*, vol. 80, edited by P. H. Lauritzen et al., pp. 381–494, Springer, Heidelberg, Germany.
- Kato, S., N. G. Loeb, F. G. Rose, D. R. Doelling, D. A. Rutan, T. E. Caldwell, L. Yu, and R. A. Weller (2013), Surface irradiances consistent with CERES-derived top-of-atmosphere shortwave and longwave irradiances, *J. Clim.*, *26*, 2719–2740, doi:10.1175/JCLI-D-12-00436.1.
- Lindberg, C., and A. J. Broccoli (1996), Representation of topography in spectral climate models and its effect on simulated precipitation, *J. Clim.*, *9*, 2641–2659.
- Navarra, A., W. F. Stern, and K. Miyakoda (1994), Reduction of the Gibbs oscillation in spectral model simulations, *J. Clim.*, *7*, 1169–1183.
- Pielke, R. A., Sr. (2002), *Mesoscale Meteorological Modeling, International Geophysics Series*, vol. 78, Academic Press, San Diego, Calif.
- Sardeshmukh, P. D., and B. J. Hoskins (1984), Spatial smoothing on the sphere, *Mon. Weather Rev.*, *112*, 2524–2529.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2009), A summary of the CMIP5 experiment design, PCDMI Rep., 33 pp. [Available online at [http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor\\_CMIP5\\_design.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf)]
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of CMIP5 and the experiment design, *Bull. Am. Meteorol. Soc.*, *93*, 485–498, doi:10.1175/BAMS-D-11-00094.1.
- Warner, T. T. (2011), *Numerical Weather and Climate Prediction*, Cambridge Univ. Press, Cambridge, U. K.
- Webster, S., A. R. Brown, D. R. Cameron, and C. P. Jones (2003), Improvements to the representation of orography in the Met Office Unified Model, *Q. J. R. Meteorol. Soc.*, *123*, 1989–2010, doi:10.1256/qj.02.133.
- Yorgun, M. S., and R. B. Rood (2014), An object-based approach for quantification of GCM biases of the simulation of orographic precipitation. Part 1: Idealized simulations, *J. Clim.*, *27*, 9139–9154, doi:10.1175/JCLI-D-14-00051.1.
- Yorgun, M. S., and R. B. Rood (2015), An object-based approach for quantification of GCM biases of the simulation of orographic precipitation. Part 2: Quantitative analysis, *J. Clim.*, *27*, 9139–9154, doi:10.1175/JCLI-D-14-00730.1, in press.



*Geophysical Research Letters*

Supporting Information for

**Quantitative Characterization of Spurious Numerical Oscillations in 48 CMIP5 Models**

Kerrie L. Geil and Xubin Zeng

Department of Atmospheric Sciences, University of Arizona, Tucson, Arizona

**Contents of this file**

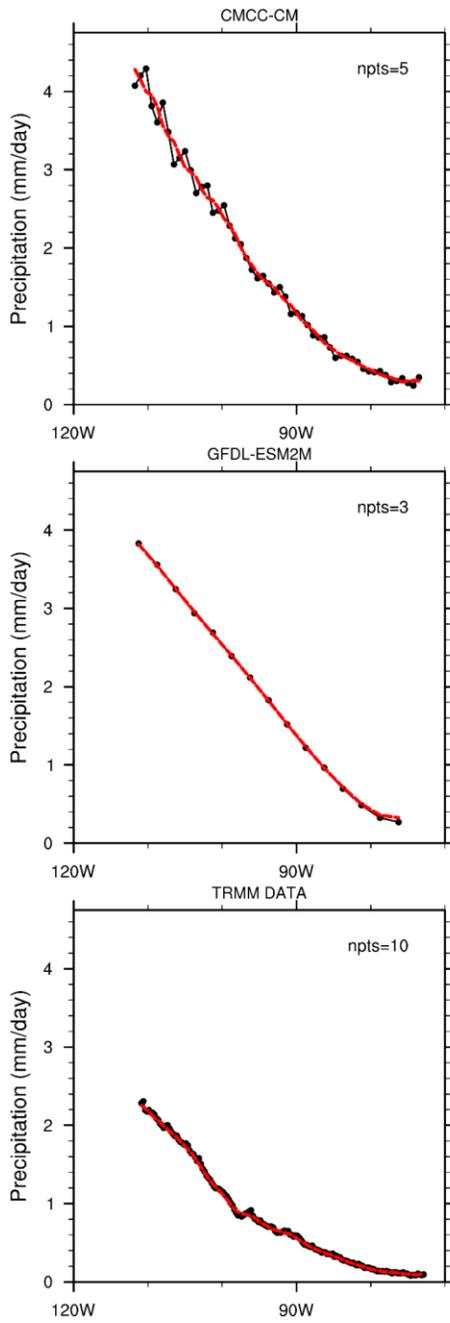
Figures S1 to S5

**Additional Supporting Information (Files uploaded separately)**

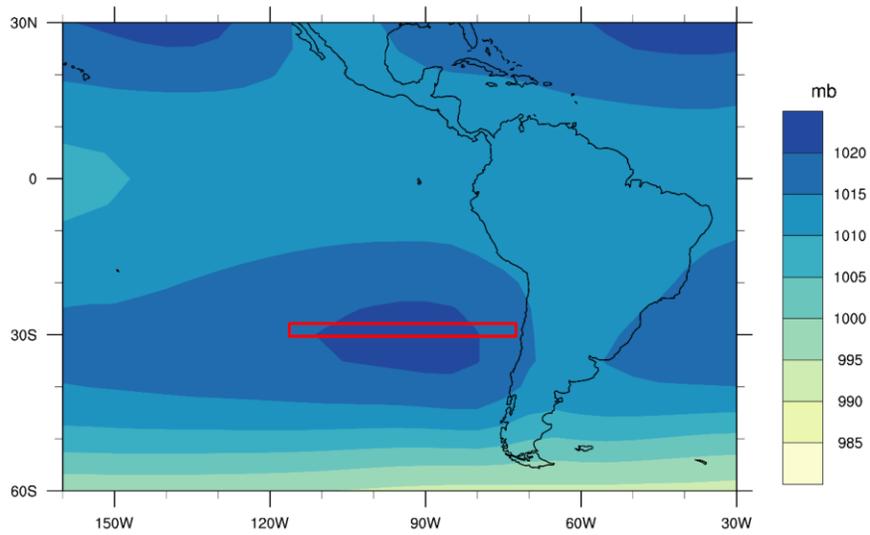
Table S1

**Introduction**

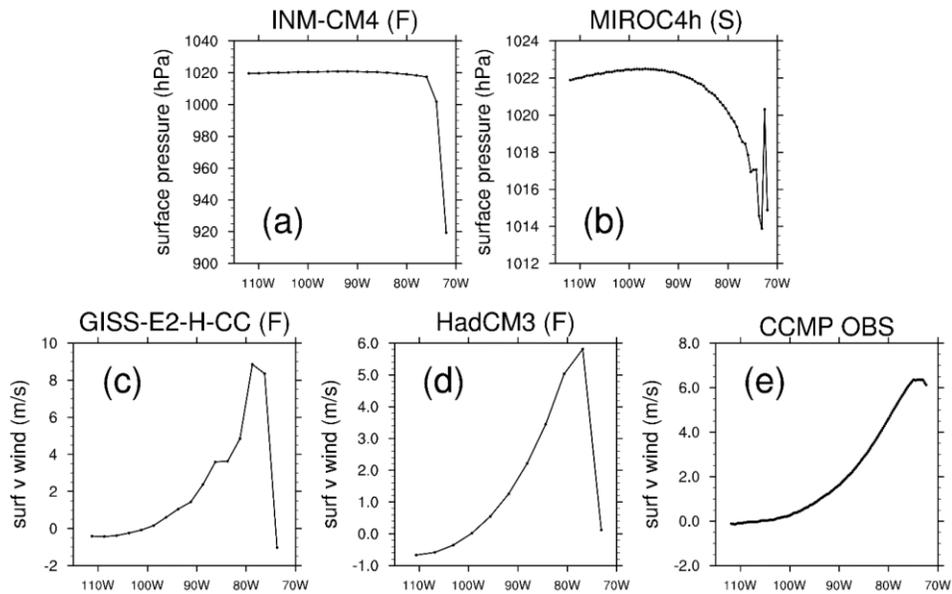
Figures S1 through S3 included in this supplementary information provide detailed illustration of the study methods, Figures S4 and S5 show detailed results for individual models, and Table S1 is a comprehensive list of models used.



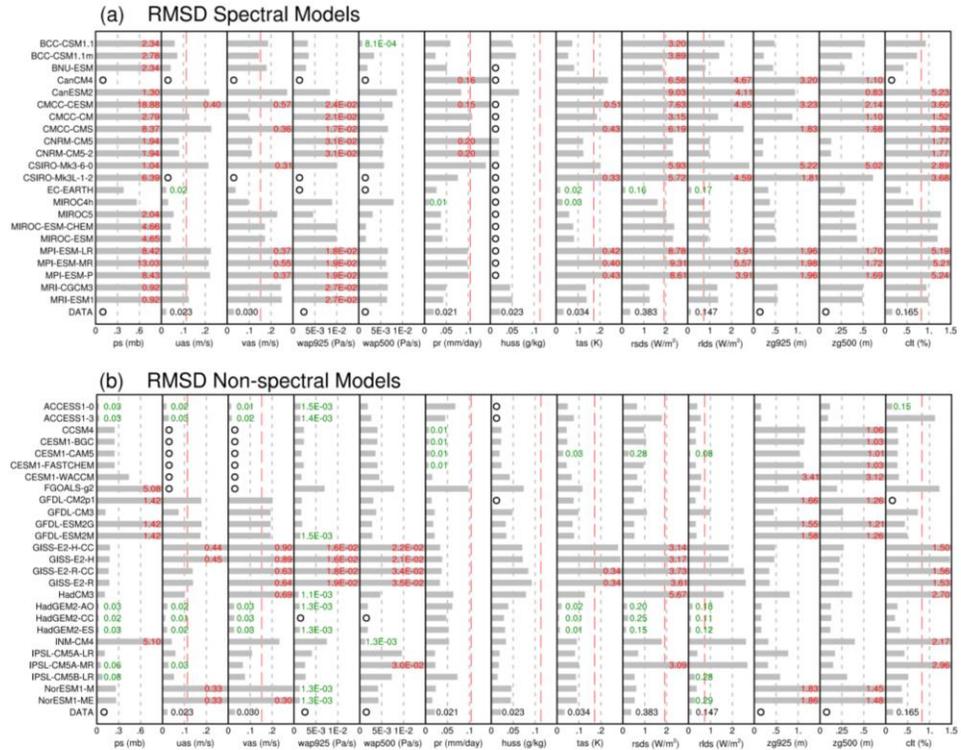
**Figure S1.** Climatological precipitation transects (black) and the corresponding running mean curves (red) used to compute RMSD values; shown for a spectral model (CMCC-CM; top), a finite volume model (GFDL-ESM2M; middle), and TRMM data (bottom). The number of points used to compute each running mean is labeled as 'npts'. Note that the transect approaches the South American coast as one moves from west to east (left to right).



**Figure S2.** The region of model transects shown over climatological (1979-2004) sea level pressure from the HadSLP2 dataset. The transects bisect the South Pacific subtropical surface high. Transects are chosen as the model latitude closest to 29° S, which ranges between 27.8° S and 30.3° S. The westernmost transect longitude is the closest model grid point approaching 112° W from the west and the easternmost longitude point is one to two model grid points westward of the first continental model grid point (dependent on the number of grid points used to compute the running mean).



**Figure S3.** Examples of model transects that result in large values of RMSD and RMSD:IAV ratio, even when large Gibbs oscillations are not present along the entire transect. Note that all transects include only ocean points. The culprits include (a) a premature decrease in surface pressure approaching land, (b) one large oscillation in surface pressure near the ocean-land transition, and (c-d) unrealistic steep gradients in meridional surface wind near the ocean-land transition. The CCMP meridional surface wind observations are shown in (e) for comparison to (c-d).



**Figure S4.** RMSD values for (a) spectral models and (b) non-spectral models for 13 variables: surface pressure (ps), near-surface u-wind (uas) and v-wind (vas), vertical velocity at 925 mb (wap925) and 500 mb (wap500), precipitation (pr), surface specific humidity (huss), surface air temperature (tas), surface incoming solar radiation (rds) and incoming longwave radiation (rlds), geopotential height at 925 mb (zg925) and 500 mb (zg500), and total cloud amount (cft). Observational values are shown in black on the bottom row of each panel. Black circles indicate fields that are not available. Vertical red dashed lines represent the observational data value multiplied by a factor of 5. Red labels show values beyond each abscissa maximum and green labels show values for bars that are not easily visible. Vertical grid lines are shown with gray dashes.

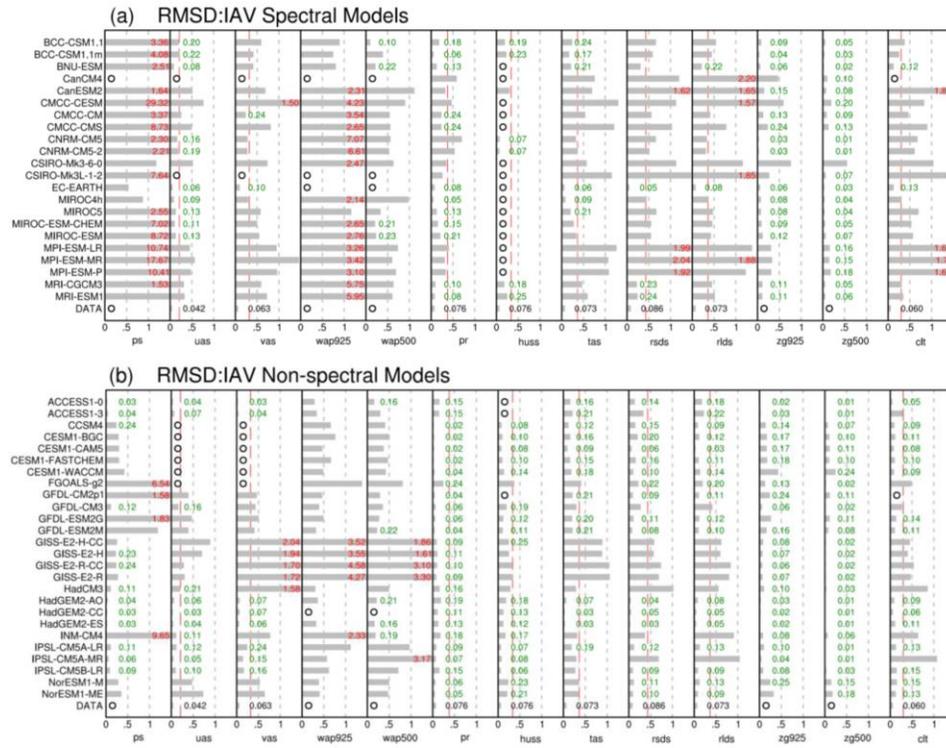


Figure S5. Same as Figure S4, except for the RMSD:IAV ratio.

**Table S1.** Model Information

Model	Modeling Group	Country	AGCM Resolution <sup>a</sup>	Spectral or Finite
ACCESS1-0	Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM)	Australia	192x144L38 (1.875x1.25)	F
ACCESS1-3			192x144L38 (1.875x1.25)	F
BCC-CSM1.1	Beijing Climate Center, China Meteorological Administration	China	128x64L26 (2.813x2.813) T42	S
BCC-CSM1.1m			320x160L26 (1.125x1.125) T106	S
BNU-ESM	Beijing Normal University	China	128x64L26 (2.813x2.813) T42	S
CanCM4	Canadian Centre for Climate Modelling and Analysis	Canada	128x64L35 (2.813x2.813) T63	S
CanESM2			128x64L35 (2.813x2.813) T63	S
CCSM4	National Center for Atmospheric Research	USA	288x200L26 (1.25x0.9)	F
CESM1-BGC			288x192L30 (1.25x0.938)	F
CESM1-CAM5			288x192L30 (1.25x0.938)	F
CESM1-FASTCHEM			288x192L30 (1.25x0.938)	F
CESM1-WACCM			144x96L66 (2.5x1.875)	F
CMCC-CESM	Centro Euro-Mediterraneo sui Cambiamenti Climatici	Italy	96x48L39 (3.75x3.75) T31	S
CMCC-CM			480x240L31 (0.75x0.75) T159	S
CMCC-CMS			192x96L95 (1.875x1.875) T63	S
CNRM-CM5	Centre National de Recherches Meteorologiques/Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique	France	256x128L31 (1.4x1.4) TL127	S
CNRM-CM5-2			256x128L31 (1.4x1.4) TL127	S

CSIRO-Mk3.6.0	Commonwealth Scientific and Industrial Research Organization, Queensland Climate Change Centre of Excellence	Australia	192x96L18 (1.875x.1875) T63	S
CSIRO-Mk3L-1-2	University of New South Wales	Australia	64x56L18 (5.625x3.2) R21	S
EC-EARTH	EC-EARTH Consortium	multiple (Europe)	320x160L62 (1.125x1.125) T159	S
FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and CESS, Tsinghua University	China	128x60L26 (2.813x3)	F
GFDL-CM2.1	NOAA Geophysical Fluid Dynamics Laboratory	USA	144x90L24 (2.5x2)	F
GFDL-CM3			144x90L48 (2.5x2)	F
GFDL-ESM2G			144x90L24 (2.5x2)	F
GFDL-ESM2M			144x90L24 (2.5x2)	F
GISS-E2-H	NASA Goddard Institute for Space Studies	USA	144x90L40 (2.5x2)	F
GISS-E2-H-CC			144x90L40 (2.5x2)	F
GISS-E2-R			144x90L40 (2.5x2)	F
GISS-E2-R-CC			144x90L40 (2.5x2)	F
HadCM3	Met Office Hadley Centre	UK	96x73L19 (3.75x2.5)	F
HadGEM-AO	National Institute of Meteorological Research/Korea Meteorological Administration	Korea	192x145L38 (1.875x1.25)	F
HadGEM2-CC	Met Office Hadley Centre	UK	192x145L38 (1.875x1.25)	F
HadGEM2-ES			192x145L38 (1.875x1.25)	F
INM-CM4	Institute for Numerical Mathematics	Russia	180x120L21 (2x1.5)	F
IPSL-CM5A-LR	Institut Pierre-Simon Laplace	France	96x95L39 (3.75x1.875)	F
IPSL-CM5A-MR			144x143L39 (2.5x1.25)	F
IPSL-CM5B-LR			96x95L39 (3.75x1.875)	F
MIROC4h	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	Japan	640x320L56 (0.563x0.563) T213	S
MIROC5			256x128L40 (1.4x1.4) T85	S

MIROC-ESM			128x64L80 (2.813x2.813) T42	S
MIROC-ESM- CHEM			128x64L80 (2.813x2.813) T42	S
MPI-ESM-LR			192x96L47 (1.875x1.875) T63	S
MPI-ESM-MR			192x96L95 (1.875x1.875) T63	S
MPI-ESM-P		Germany	192x96L47 (1.875x1.875) T63	S
MRI-CGCM3			320x160L48 (1.125x1.125) TL159	S
MRI-ESM1		Japan	320x160L48 (1.125x1.125) TL159	S
NorESM1-M			144x96L26 (2.5x1.9)	F
NorESM1-ME		Norwegian Climate Centre	144x96L26 (2.5x1.9)	F

<sup>a</sup> Atmospheric (AGCM) model resolution is given for the physical grid with a format of (# of longitudinal grids) x (# of latitudinal grids)

L (# of vertical levels) followed by the approximate grid spacing in degrees longitude by degrees latitude. Spectral resolution, if applicable, appears below the physical grid resolution where the 'T' and 'TL' stand for triangular truncations at full Gaussian and linear grids (on which the physical parameters are computed), respectively and 'R' stands for rhomboidal truncation.

## **APPENDIX C**

### **EVALUATION OF THE MEAN STATE, TRENDS, AND VARIABILITY IN CMIP5 MONTHLY SURFACE AIR TEMPERATURE AT REGIONAL SCALES OVER THE US FOR APPLICATION TO CLIMATE ADAPTATION AND RESILIENCE EFFORTS**

(To be submitted to the Journal of Applied Meteorology and Climatology)

**KERRIE L. GEIL<sup>1</sup> AND XUBIN ZENG<sup>1</sup>**

<sup>1</sup>Department of Atmospheric Sciences, University of Arizona, Tucson, Arizona

## **Abstract**

Model simulations of 20th century monthly minimum and maximum surface air temperature over eight US regions are assessed using mean state, trend, and variability bias metrics. Transparent model performance information is provided in the form of model rankings for each bias type. A wide range in model skill is seen even for long term mean climate simulation where the highest skill is expected, and much of the ensemble cannot reproduce significant observed long term trends at monthly resolution. No strong relationships are seen between any of the three bias types or between 20<sup>th</sup> century bias and 21<sup>st</sup> century projected change. Using our model rankings, two smaller ensembles of models with better performance over the southwestern U.S. are selected, but they result in negligible differences from the all-model ensemble in the average 21<sup>st</sup> century projected temperature change and model spread. In other words, models of varied quality (and complexity) are projecting very similar changes in temperature, implying that the models are simulating warming for different physical reasons. Despite this result, we suggest that models with smaller 20<sup>th</sup> century biases have a greater likelihood of being more physically realistic and therefore, more confidence can be placed in their 21<sup>st</sup> century projections as compared to projections from models that have demonstrably poor skill over the observational period. This type of analysis is essential for responsibly informing climate resilience efforts.

## 1. Introduction

Rapid environmental changes linked to human-induced increases in atmospheric greenhouse gas concentrations have been observed on local to global scales over recent decades. Given the relative certainty of continued rapid change across many earth systems, local and regional decision makers are increasingly interested in planning for projected climate changes and employing mitigation strategies. These decision makers need climate model projections on relevant time and spatial scales, as well as assessments of model reliability in order to make confident planning decisions.

Preliminary resources for this type of planning in the US include national and regional climate assessment reports such as the National Climate Assessment (NCA; Melillo et. al 2014) and the Assessment of Climate Change in the Southwest United States (Garfin et. al 2013). Aimed at decision makers, these reports use plain language to condense the most up-to-date scientific knowledge on national- and regional-scale climate observations and future change projections. In the context of regional climate resilience efforts, one of the main problems with reports like the NCA is the lack of information regarding model reliability. Climate projections are almost exclusively presented using a multi-model ensemble (MME) average of all available models, regardless of individual model performance. This method is likely inappropriate on smaller regional planning scales due to the large spread in model ability at these scales (Maxino et al. 2008, Perkins et al. 2007). It stands to reason that a realistic simulation of the present climate is at least a necessary (but likely not sufficient) requirement for a model's ability to realistically simulate the climate of the future for the right reasons. As such, understanding model performance at regional

scales over the 20th century is crucial to discerning model reliability, avoiding the use of the worst performing models, and informing our confidence in climate projections.

Three main issues can be found from previous regional performance assessments. First, there is a dearth of climate model information available at local to regional planning-relevant scales. Second, performance information for specific models is frequently neglected in favor of the MME mean and model spread. Third, the measures used to evaluate model performance are often not comprehensive or quantitative enough to provide sufficient evidence of model reliability for planning decisions. For example, similar to the national and regional climate assessment reports, some studies simply evaluate model performance based on long term mean climate (Reichler et al. 2008, Macadam et al. 2010). Other studies focus on the simulation of trends (Sillmann et al. 2014) or climate variability (Yao et al. 2013, Maxino et al. 2008, Perkins et al. 2007). Many studies examine both mean climate and variability, but don't provide clear quantitative information regarding individual model performance at decision relevant scales (Sillmann et al. 2013a, Wuebbles et al. 2014, Santer et al. 2009, Cheng et al. 2015, Sun et al. 2015). Overall, there are a lack of studies that use all of these relevant evaluation measures (long term means, trends, and variability) to quantify individual model performance at scales applicable to adaptation planning.

Here, we present a planning-relevant analysis of individual model simulations of monthly average minimum and maximum surface air temperature (Tmin, Tmax). We evaluate the ability of all available CMIP5 climate models to simulate 20<sup>th</sup> century long term mean climate, trends, and variability in order to provide a clearer picture of model capability for resilience efforts. Our analysis was designed based on the interest of an energy utility in southern Arizona in assessing

their heat-related vulnerabilities and increasing their resilience to future temperature change. The model performance information presented here is intended for similar application by researchers and practitioners working at the intersection of climate science and decision making.

## 2. Model Simulations, Observations, and Methods

The source of model simulations is the CMIP5 multimodel ensemble archive (<http://pcmdi.llnl.gov>). We use a 106-year period (1900-2005) of the historical experiment, which imposes changing atmospheric and land surface conditions consistent with past observations. Details regarding CMIP5 experimental design can be found in Taylor et al (2009, 2012). One ensemble member is chosen for each model with monthly minimum and maximum surface air temperature and topography available in the archive (42 models in total). Table 1 lists the models used and provides the modeling group, country origin, and abbreviation code we have assigned to each model.

Observations of Tmin and Tmax are from the Berkeley Earth 1°x1° gridded monthly land data product (Rhode et al. 2013), which provides temperature over land only. Although this study is limited to model performance at monthly time resolution, we choose the Berkeley Earth monthly observations because of the corresponding daily data product, which will provide consistency for further model evaluation at daily resolution.

Preprocessing steps include regridding all models to a common grid, applying an elevation-based model temperature correction, creating an MME-average, and area-averaging over eight study areas. The 1°x1° grid of the Berkeley Earth observations is used as the common grid and model information is adjusted to this grid using bilinear spatial interpolation. The temperature

correction is applied to each model using a lapse rate of 6.5 °C/km and the elevation difference between modeled and observed topography. Since the Berkeley Earth gridded observations do not include elevation information, we use the ASTER Global Digital Elevation Model (DEM) dataset (NASA JPL, 2009) scaled down to our 1-degree common grid by area averaging. At this point, MME-average Tmin and Tmax are created by averaging together the regridded, elevation-corrected spatial fields of all 42 models. Then, a monthly time series is obtained for all models by area-averaging over eight study areas. A land mask is applied to each model so that only grids with greater than 50% land area are included in each area-average. The seven regional study areas used are based on the regions delineated by the NCA and are shown in Figure 1 (SW=southwest, NW=northwest, GP=great plains, MW=midwest, NE=northeast, SE=southeast, AK=Alaska). We also include an additional continental US study area that comprises most of the lower 48 states.

The area-averaged model and observed 20<sup>th</sup> century time series for each region are used to assess biases in model long-term mean state, trend, and variability. First, we examine model mean climate using long-term annual and seasonal averages. Here, significant biases are identified at the 90% confidence level using a two-tailed t-test for difference in means and adjusting for lag-1 autocorrelation. Next, we look at linear trends in annual average temperature as well as in monthly average temperature for the month with the largest difference between observed and MME-average trends for each region. Significant trends at the 90% confidence level are again identified using a two-tailed t-test, adjusted for lag-1 autocorrelation. Significance is computed in the same way for the trend of the time series difference between models and observations to identify modeled and observed trends that are significantly different from each other. Finally, we use the standard deviation of the detrended monthly temperature anomaly time series to assess variability,

where the anomalies are computed by removing monthly climatological means. Significant differences between model and observed variability at the 95% confidence level are identified using a two-tailed f-test, adjusted for lag-1 autocorrelation.

Results are then used to rank model performance with respect to each type of bias. Each type of bias ranking is divided into five categories to aid in the selection of better performing multimodel ensembles, so that the reader can quickly identify the bias magnitude for each model. The bounds of the categories were arbitrarily chosen to separate out very good and very poor model performance, and to distribute average performing models relatively evenly over the remaining categories. In the discussion section of this paper, we use the southwest US region to demonstrate how the rankings can be used to select better performing multimodel ensembles and then compare future projections from these ensembles to projections using all available models.

### 3. Results

#### *a. Long-term mean state*

Starting with model mean state bias, we see a range in regional performance. Figure 2 shows the MME-average bias and model bias spread for the seasonal and annual mean climate states. For both T<sub>min</sub> and T<sub>max</sub>, seasonal mean state MME-average bias is much smaller in some regions (average absolute seasonal bias is 0.52 °C and 0.80 °C for GP, 0.38 °C and 0.76 °C for SE T<sub>min</sub> and T<sub>max</sub>, respectively) than in others (average absolute seasonal bias is 2.40 °C and 1.35 °C for SW, 1.68 °C and 2.08 °C for AK T<sub>min</sub> and T<sub>max</sub>). The same is true for the annual mean MME-average bias. It is clear that for most regions, the majority of models are biased warm for T<sub>min</sub>, although the largest model biases are cold biases that act to compensate for most of the

overwhelming model warm biases. The same phenomenon is seen in Tmax, where the majority of models are biased cold in most regions, but this bias is mostly compensated by a lesser number of models with larger warm biases. Across all eight regions on average, Tmax MME-average absolute bias is slightly larger than Tmin bias (by 0.23 °C seasonally and 0.21 °C annually), but the model spread in Tmax bias is smaller than the spread in Tmin bias (by 1.06 °C seasonally and 1.56 °C annually). This indicates that Tmax model biases of opposite signs are less compensating than they are for Tmin. Also of note is that the average model spread in annual mean bias for both Tmin and Tmax (12.87 °C and 11.31 °C) is almost as large as that of seasonal mean bias (13.60 °C and 12.53 °C), showing a large range in model ability even at the coarser timescale. On a seasonal and annual basis, the MME-average bias is usually significant for both Tmin and Tmax across all regions. For confident decision making in the context of climate change adaptation and resilience efforts, the MME-average performance and spread of model biases in simulating even long-term mean climate at the regional scale are inadequate.

Additionally, Figure 2 clearly shows that low annual mean bias can be mistaken for good performance when seasonal biases are of opposite signs. For example, without looking beyond the annual mean bias, there is no way to know that the annual mean Tmin bias in the NW region is the result of compensating seasonal biases, whereas this is not the case for the low annual mean bias in the SE region. Use of this metric alone is not comprehensive enough to accurately assess model skill and can cause over-confidence in model ability.

Assessment of vulnerability to climate change and the resulting planning decisions may focus on one particular season or even one month of the year. Often, the MME-average bias and model spread increase when looking at the long-term mean state for a single month. Figure 3 shows

the May time series for the SW region, where the MME-average bias is  $-3.83$  °C and the model bias spread is  $13.27$  °C on average, as compared to the summer and annual mean bias and model spread in Figure 2 (summer MME-average bias  $-3.14$ , spread  $12.57$ ; annual MME-average bias  $-2.40$ , spread  $12.87$ ).

All models are ranked on the basis of Tmin and Tmax seasonal bias in the appendix (Table A1). For most regions, the majority of models are biased  $1-3$  °C. There are two instances where the MME-average outperforms all individual models (in the GP and US regions for Tmin), and this is caused by compensating model biases. In regions where bias is stronger in one direction, many models outperform the MME-average. For example, in the SW region where most models show Tmin cold biases, 67% of individual models have a smaller average seasonal bias than the MME-average. Similarly, in the NE region where most models are biased warm in Tmin and cold in Tmax, 19% of individual models outperform the MME-average for Tmin and 26% for Tmax.

#### *b. Trends*

Next, we look at model ability to simulate 20<sup>th</sup> century trends. Figure 4 shows the trends in annual average temperature as well as in monthly average temperature for the month with the largest difference between observed and MME-average trends (herein referred to as max-month) for each region. For the annual average trends shown in Figure 4a, there is a significant difference between the MME-average and observed Tmin values for half of the regions and regardless of significance, the MME-average value is always an underestimation (by  $0.24$  °C/century on average). In the case of Tmax, there is no significant difference between the MME-average and observed values for any region, although the MME-average value is almost always a slight

overestimation (by 0.14 °C/century on average). The magnitude of the model spread in annual mean trend is similar between Tmin and Tmax for each region (average difference in model spread is 0.27 °C/century), but while most models agree on the sign of the Tmin trends, the models are more often split on the sign of Tmax trends. Also, the AK model spread for both Tmin and Tmax is much larger than any other regions, indicating the difficulty that models have in simulating realistic 20<sup>th</sup> century climate in this region.

Keeping in mind that only the month with the largest MME-average trend bias is shown in Figure 4b, most MME-average trends in max-month temperature are significantly different from observations. For each region, the greatest MME trend bias is found almost exclusively in fall or winter. The model spread in max-month trends (4.39 °C/century for Tmin, 3.82 °C/century for Tmax on average) is much larger than for annual average trends (2.06 °C/century Tmin, 2.21 °C/century for Tmax on average) and there is no model consensus on the sign of max-month trends. It is also interesting that when the observed max-month trend in Tmax is significant (shown with large filled circles in Fig 4b, e.g. for the GP, MW, NE, and US regions), the MME-average trend is always an underestimation. This is also true for Tmin max-month and annual average trends.

All MME-average trends in annual and max-month temperature are significant (filled large squares in Figure 4a,b), whereas there is one region for Tmin and four regions for Tmax where the observed max-month temperature trend is not significant (open circles in Figure 4b). A time series example of this is shown in Figure 5, which compares the observed, MME-average, and single model max-month (October) trends in Tmax for the AK region. The observed trend (Figure 5a) is -0.75 °C/century and is not significant ( $p=0.31$ ), whereas the MME-average trend (Figure 5b) is significant and of the opposite sign (+1.04 °C/century,  $p<0.01$ ). Many models, such as MRIC3

(Figure 5c), reproduce small insignificant trends ( $-0.60$  °C/century,  $p=0.41$ ), but the models that produce larger significant trends, such as CESMF (Figure 5d;  $+3.76$  °C/century,  $p<0.01$ ) overwhelm the MME-average toward significant bias.

All models are ranked on the basis of Tmin and Tmax annual and max-month trend bias in the appendix. For annual average trend bias (Table A2), the MME-average usually falls within the 0.2-0.5 °C/century category for Tmin and within the 0-0.2 °C/century category for Tmax. Most models have biases of less than 1 °C/century and depending on the region, 12-45% of individual models have smaller trend biases than the MME-average. For max-month trend bias (Table A3) the MME-average usually falls within the 1-2 °C/century category for both Tmin and Tmax. Most models have biases greater than 1 °C/century and 40-52% of individual models have smaller biases than the MME-average.

### *c. Variability*

Last, we examine 20<sup>th</sup> century variability. Figure 6a shows the standard deviation of the modeled and observed detrended monthly temperature anomaly time series. Here, the MME-average is computed as the average of individual model standard deviation values, as opposed to computing variability from the average MME temperature field because averaging model fields together smooths out most variability. Without visualization of individual model results, it may appear that many models have much greater than observed variability in Tmin and Tmax for most regions, but the large model spread is usually caused by only a few models. This is reflected in the fact that, with the exception of the AK region, the MME-average variability is very close to that of observations. The MME-average variability bias for the AK region is  $0.72$  °C for Tmin and  $0.48$

°C for Tmax, whereas across all other regions on average (excluding AK), it is 0.23 °C for Tmin and 0.08 °C for Tmax. The magnitude of the model spread varies across the regions, ranging from 0.43 °C (US Tmax) to 2.46 °C (NW Tmin), and the spread is often smaller for Tmax than Tmin (0.53 °C smaller on average).

An illustration of model ability to simulate temperature variability is shown in Figure 6b, where we zoom in to the modeled and observed Tmax anomaly time series for the AK region during the randomly chosen decade of the 1960's (other decades examined appeared similar). The trend and climatological monthly mean of each series has been removed. The model spread in gray shading shows the tendency to overestimate variability in the region. The models with the smallest (ACC13, 0.04 °C) and largest (CMCCE, 1.17 °C) standard deviation bias (for the entire 1900-2005 study period) are shown as an example of individual model performance. While the model with small bias (blue) performs well, the model with the largest bias (green) clearly shows more extreme warm and cold temperatures as compared to observations, especially during winter months. Future change in climate variability, including extremes (which are not assessed here), is a very important consideration in assessing vulnerability to climate change, as changes in variability can affect the frequency of heat waves, freezing temperatures, flooding and drought events, etc. For this reason, determining the models that simulate realistic variability and using that information to constrain vulnerability assessment is essential.

All models are ranked on the basis of Tmin and Tmax standard deviation bias in the appendix (Table A4). Most models have biases of less than 0.5 °C for Tmin and Tmax, except in the AK region, where most models are biased greater than 0.5 °C. For any given resilience application, climate model variability performance information at monthly time resolution may

not be sufficient for decision making. We suggest using the results presented here in combination with a decision-relevant variability assessment of daily resolution model information.

#### 4. Discussion

##### *a. Correlation between historical biases*

To understand how the historical model biases relate to each other, we perform a linear regression analysis between each type of bias (seasonal mean, annual trend, max-month trend, and variability) for each region. For each regression, the linear correlation coefficient ( $r$ ) is computed and tested for significance at the 95% level using a two-tail t-test. Substantial overlap is known to exist between models due to the sharing of training data, human expertise, and model code, which results in an effective number of climate models that is much smaller than the total (Pennell et al. 2011, Knutti 2010). While the effective number of models varies widely for individual model fields, Pennell et al. (2011) estimated that on average, the effective number of CMIP3 models for the northern hemisphere extratropics is between 7.5 and 9 from a total of 24. To account for model overlap, we roughly estimate the effective number of independent models as one third of the total (14 effective models from 42 total). The effective number of models is used to compute the degrees of freedom for significance testing. To ensure robust tests of significance, we also compute the linear correlation coefficient a second time excluding the 10% of models (4) with the largest magnitude variable bias.

No strongly significant relationships are found, although significance (using both measures) between the annual trend bias and max-month trend bias is seen in four regions (SW, NW, GP, US) for  $T_{min}$  and three regions (SW, GP, US) for  $T_{max}$ . Figure 7a shows this

relationship for GP Tmin where the correlation values are 0.66 and 0.63. The only other relationship that is significant using both measures is between max-month trend bias and variability bias for the US region, but the correlation is not particularly strong, as shown in Figure 7b, and the same relationship is not significant in any of the smaller regions. All other linear correlations between regression variables are either insignificant by both measures or significant only if computed using all 42 models, such as the relationship shown in Figure 7c.

*b. Correlation between historical bias and projected change in the southwest*

A total of 33 models (marked with asterisks in Table 1), have output available for the historical, RCP 4.5, and RCP 8.5 experiments. For the discussion of future projections over the southwest US we adjust our metrics slightly, according to the interest expressed by an electric company in southern Arizona in spring time temperature changes. We compare the 20<sup>th</sup> century (1900-2005) bias to the 21<sup>st</sup> century (2006-2099) projected change of five metrics: annual mean, spring mean (March, April, May average), annual trend, May trend, and monthly time series standard deviation for Tmin and Tmax. Linear correlations and statistical significance are computed using the same methods as in section 4a and model overlap is again accounted for by estimating the effective number of independent models as one third of the total (11 effective models from 33 total). Three statistically significant correlations between 20<sup>th</sup> century bias and 21<sup>st</sup> century projected change are found, but they all prove fairly weak visually.

Figure 8 shows these significant correlations, which include Tmax mean bias to RCP 8.5 projected variability change (Figure 8a), Tmax spring mean bias to RCP 8.5 projected variability change (Figure 8b), and Tmax annual trend bias to RCP 4.5 projected annual trend change (Figure

8f). None of these significant relationships hold true for both RCPs (Figure 8d,e,c). The reason that the significant and insignificant correlations shown in Figure 8 are all negative is unclear, while the many other insignificant correlations (not shown) are a mixture of positive and negative relationships. No other significant correlations were found between any of the biases and projections for Tmin or Tmax.

*c. Choosing better performing multimodel ensembles for the southwest*

Regardless of the lack of significant relationship between 20<sup>th</sup> century biases or between 20<sup>th</sup> century bias and 21<sup>st</sup> century projections, it is still worthwhile to at least eliminate demonstrably poor performing models from the MME when the intended use of the MME is to inform climate resilience efforts. Using the regional mean, trend, and variability bias rankings in the Appendix, we choose two better performing MMEs to examine southwest temperature projections. Many methods exist to accomplish this task, but the process of choosing models should be relevant to the intended use of projection information. We are not using the max-month trend bias rankings (Table A3) in the model selection process because it is not as relevant as the annual trend bias to our intended use of the projections.

We develop a simple point-based method using the bias categories in Tables A1, A2, and A4, which were chosen fairly arbitrarily in order to separate excellent and poor performance, while distributing average performing models across the remaining categories relatively evenly. For Tmin and Tmax separately, each model is given points based on the bias category it falls in for each of the three bias types. Zero points are given for the smallest bias category, one point for the second smallest bias category, and so on, up to four points for the largest bias category. Summed

across the three bias types, the total points possible for each model ranges between zero and twelve, with fewer points indicating better performance. This process is completed twice, once for Tmin bias and again for Tmax bias.

A fairly straight forward way of choosing a better performing MME is to eliminate models that clearly do not simulate realistic 20<sup>th</sup> century climate. We call this process “RMBAD” and select MMEs for Tmin and Tmax separately. All models that fall within the largest and second to largest bias categories for any bias type (i.e. 3 or 4 points for any bias) are eliminated first. Any remaining models with total points of 5 or higher are eliminated second. The RMBAD selection process results in an MME of 24 models for Tmin and 19 models for Tmax (Table 2) from a total of 33 available models. In a planning context, we suggest using this type of MME for probabilistic studies and for generation of future potential climate scenarios.

Another way of choosing a better performing MME is to select a number of the best performing models. Knutti et al. (2010) show that for mean seasonal surface air temperature, selection of a few good models for averaging substantially decreases mean climate bias as compared to an all-model MME average, and the greatest bias reduction is seen when about five good models are chosen. Based on these results, we use our points system to choose five models with good 20<sup>th</sup> century performance in a process we call “TOP5”. First, we decided to discount the importance of annual trend bias (as compared to mean and variability bias) based on the theoretical ability of models to capture small 20<sup>th</sup> century observed linear trends. The observed 20<sup>th</sup> century annual trend is less than 1 °C/century for Tmin and Tmax over the southwest US. If modeled and observed natural variability near the beginning and end of the century is not similar, then for small observed trends, it is possible that the modeled trend will be significantly different, therefore

imposing equal weight on trend bias during the model selection process may be unreasonably restrictive. Second, while it is clear that considerable bias commonality exists beyond just models developed at the same center (Pennell et al. 2011), commonalities between models from different modeling centers are not transparent and specific model overlap information is not readily available. In an attempt to account for at least a portion of model overlap, we require the models chosen using the TOP5 process to be from separate modeling centers. The models resulting from the TOP5 process for Tmin and Tmax are listed in Table 3. Because this type of MME is so restricted in number of models, we suggest using this method for multimodel averaging and planning purposes where average projections may be sufficient.

*d. 20<sup>th</sup> century performance and 21<sup>st</sup> century projections over the southwest US using three different multimodel ensembles*

Figure 9 compares the performance of the RMBAD, TOP5, and all-model MMEs in simulating 20<sup>th</sup> century climate characteristics. For the mean and variability metrics, the RMBAD- and TOP5- average biases are an improvement over the all-model average bias. For example, the RMBAD-average and TOP5-average reduce the magnitude of the all-model-average spring time bias in Tmin of -3.23 °C by 37% (1.19 °C) and 57% (1.83 °C), respectively. Similarly, the RMBAD-average and TOP5-average reduce the all-model-average variability bias in Tmax of 0.112 °C by 27% (0.030 °C) and 94% (0.105 °C). As intended, the RMBAD and TOP5 MME selection processes greatly reduce model spread for the mean and variability bias metrics. These result are not seen for the trend metrics due to the down-weighting of trend performance in the

TOP5 selection process and because for the SW region, relatively few models fall in the worst trend bias categories chosen for Table A2.

Figure 10 shows the 21<sup>st</sup> century projected temperature changes for the SW region for all three MMEs. The TOP5-, RMBAD-, and all-model- average projected changes in Tmin and Tmax are nearly identical for all five metrics. There is very small or no reduction in projection uncertainty (model spread) between the all-model MME and the RMBAD MME, and the small model spread of the TOP5 MME is more than likely caused by the reduction in number of models rather than by any real reduction of uncertainty. Results for RCP 4.5 (not shown) are very similar.

## 5. Conclusions

We have examined the capacity of all available CMIP5 global climate models to simulate 20<sup>th</sup> century long term mean climate, trends, and variability in monthly average minimum and maximum surface air temperature on planning-relevant spatial scales in the US. Transparent information regarding individual model performance over the continental US and seven smaller US regions (based on the National Climate Assessment) is provided in the form of bias rankings in order to create a clearer picture of model ability for resilience efforts.

At regional scales in the US, the model spread in annual mean bias is between 11 and 14 °C, revealing a large range in model ability even at coarse timescales. Although for most regions, the majority of models have mean seasonal and annual biases of less than 3 °C in magnitude. Annual trend bias is generally less than 1 °C/century and max-month trend bias is 1-3 °C/century for most regions. While MME-average annual trends may be fairly close to observations, many individual models have annual trend biases nearly as large, or larger than, the trend itself. For the

monthly trends examined, the MME-average underestimates observed significant trends and many models do not agree on the sign of regional trends at monthly time resolution. Models tend to overestimate variability, especially during winter months, although most models have variability bias of less than 0.5 °C for most regions. Models perform worst over the AK region for every bias metric.

No strong correlations exist between any of the 20<sup>th</sup> century biases for any region, nor do the 20<sup>th</sup> century biases correlate strongly with 21<sup>st</sup> century projected mean, trend or variability change. Using the bias rankings to select two better performing MMEs, we examine 21<sup>st</sup> century temperature projections for the southwest US as compared to using an all-model MME. We find that constraining temperature projections with MMEs that have small 20<sup>th</sup> century biases results in negligible differences in the MME-average 21<sup>st</sup> century projected climate change and model spread. In other words, models of varied quality (and complexity) are projecting very similar changes in temperature, implying that models are simulating warming for different physical reasons. More research is required to elucidate the physical reasons why CMIP5 model projections of regional warming are so similar, despite the large range in simulation capability.

Further study is also needed to assess if our findings remain true in other regions of the world or for other variables such as precipitation. It is important to note that the models shown here to perform well for temperature may not perform well for other variables. For example, in Geil et al. (2013) we show that for a region in northwest Mexico located a few degrees south of the southwest region in the present study, one of the worst representations of seasonal precipitation is seen in the MIRCE model, which is a TOP5 Tmax model in the present study. Similarly, our previous work shows the three models from the UK Met Office have by far the best representations

of seasonal precipitation in northwest Mexico, whereas we show here that they are not top performers for temperature in the southwest.

Regardless of the current disconnect between 20<sup>th</sup> century model performance and 21<sup>st</sup> century projected change, relying on information from an all-model MME where each model is given equal value is often considered to be a naïve approach to understanding model performance and projected climate changes (Jun et al 2008, Maxino et al 2008). Also, it is unknown if the relationships between 20th century bias and projected change will remain static as model development continues. We suggest that models with smaller 20th century biases have a greater likelihood of being more physically realistic with respect to both historical and future climate simulations, and therefore, more confidence can be placed in their 21st century projections as compared to projections from models that have demonstrably poor skill over the observational period. This type of analysis is essential to clarifying our confidence in climate change projections and responsibly informing climate resilience efforts.

Lastly, this study is based solely on monthly resolution model information. It is unknown if the non-relationship between 20<sup>th</sup> century bias and projected change remains when examining higher time resolution model information, for instance, whether an assessment of bias in the frequency, intensity, and duration of heat waves affects the projected change in heat wave characteristics. For resilience efforts that are based on model projections, we stress the importance of carrying this work one step further, through assessment of daily resolution model information using specific metrics relevant to the decisions at hand, and constraining vulnerability assessment to those models showing reasonable 20<sup>th</sup> century performance.

## Acknowledgements

The Berkeley Earth temperature dataset was obtained from <http://berkeleyearth.org>. ASTER GDEM is a product of NASA and METI. We acknowledge the WCRP Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1) for producing and making available their model output. For CMIP the DOE/PCMDI provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. This work was supported by NOAA's Climate Program Office through grant NA12OAR4310124 with the Climate Assessment for the Southwest program at the University of Arizona, NASA (NNX14AM02G), and the Agnese Nelms Haury Program in Environment and Social Justice.

## **Appendix: Model Bias Rankings**

The following tables provide model bias rankings for each of the eight US regions. For all appendix tables, categories are inclusive on the lower bound and exclusive on the upper bound, and the 42-model MME is shown in bold.

**TABLE A1.** Model ranking by the average of the absolute value of seasonal mean bias for the period JAN 1900 – DEC 2005.

TMIN								
ISEASONAL BIAS\	SW	NW	GP	MW	NE	SE	AK	US
0 - 1 °C	CCSM4, CESMC, CESMB, CESMF*, ACC13, MIRCS, MPIEP*	HADEE, HADEC, GFDEG	<b>MME42</b> , NOR1M, CNRMC, CMCCC, CNRM2*	CMCCC, NOR1M, HADC3*, BNUEM	CMCCC, CMCCS, CNRM2*, MIRC3, HADC3*, MR1E1*, CNRMC, CESMC, <b>MME42</b> , MPIEL	HADEC, CESMB, CCSM4, CESMF*, CNRM2*, CNRMC, <b>MME42</b> , CMCCC, GFDEG, GIE2R, GFDEM, HADEE, NOR1M, GIERC, GFDC3, MIRC3, CESMC, MR1E1*, BNUEM	CESMC, CMCCC, FGOG2*	<b>MME42</b> , CESMC, NOR1M, CMCCC, CCSM4, MR1E1*, CMCCS, MIRC3, BCC1, HADEE, GFDEG, GIERC, GIE2R, BCCCM, CESMF*, CESMB, CMCC*, CSI36, MPIEL, CSI3L*, BNUEM, MPIEM, GFDEM, HADEC
1 - 2 °C	MIRCE, MIRCC, NOR1M, MPIEL, CMCC*, GFDEM, MPIEM, GIEHC, GIE2H, BNUEM, CMCCS, BCCCM, ACC10, GFDEG	<b>MME42</b> , CESMC, CMCCC, MPIEM, CANE2, GFDEM, MPIEP*, CMCC*, FGOG2*, MPIEL, CMCCS, BNUEM, ACC10, CCSM4, NOR1M, BCC1, CESMW*, BCCCM, CESMB, CESMF*, CSI3L*, GIERC, GIE2R, CNRMC, CSI36	BNUEM, HADEC, HADC3*, MIRC3, CMCCS, BCC1, MR1E1*, GIERC, GFDEM, GFDEG, CESMC, GIE2R, HADEE, CMCC*, CCSM4, CESMF*, GIE2H, CESMB, GIEHC, GFDC3	GFDEG, CMCCS, GIE2R, <b>MME42</b> , BCC1, GIERC, CNRM2*, CESMC, CNRMC, HADEC, CMCC*, GIE2H, MIRC3, GFDC3, MR1E1*, HADEE, MPIEL, MPIEM, GIEHC	GFDEG, CMCC*, NOR1M, MPIEM, CESMF*, INMC4, HADEC, CESMB, BNUEM, CCSM4, HADEE, MPIEP*, CSI36, GIEHC, BCC1, GIE2R, CSI3L*, GIERC, BCCCM	GIE2H, HADC3*, CMCCS, BCC1, GIEHC, BCCCM, CESMW*, CSI36	CMCCS, MPIEP*, HADC3*, MPIEM, CESMW*, <b>MME42</b> , MPIEL, CMCC*	CNRMC, GFDC3, ACC10, MPIEP*, GIE2H, CNRM2*, GIEHC, HADC3*
2 - 3 °C	GIE2R, GIERC, GFDC3, CMCCC, CANE2, MR1E1*, MIRC3, <b>MME42</b> , HADEE, CESMW*, CSI3L*, BCC1, CSI36	CNRM2*, GFDC3, MIRC3, HADC3*, MR1E1*, GIEHC, GIE2H, MIRCC, MIRCE	CSI36, BCCCM, MPIEL, ACC10, FGOG2*, CSI3L*, MPIEM, MPIEP*, MIRC5	GFDEM, CSI36, CCSM4, MPIEP*, CSI3L*, CESMF*, BCCCM, INMC4, CESMB, FGOG2*, ACC10	GIE2H, GFDC3, FGOG2*, GFDEM, ACC10, CESMW*	MPIEL, INMC4, FGOG2*, CMCC*, CSI3L*, MPIEM, MPIEP*, MIRC5, ACC10, ACC13, CANE2	BCC1, CCSM4, CESMB, CNRMC, CESMF*, BNUEM, CSI36, CNRM2*, CANE2, ACC10, MIRC3, GFDEG, HADEC, HADEE, NOR1M, INMC4	CANE2, CESMW*, MIRCC, MIRCE, FGOG2*, MIRC5, ACC13, INMC4
3 - 5 °C	HADEC, FGOG2*	ACC13, MIRC5, INMC4	MIRCC, INMC4, MIRCE, CESMW*, CANE2, ACC13	CESMW*, MIRC5, MIRCC, ACC13, MIRCE	MIRC5, ACC13, IPCAM, MIRCC, MIRCE, CANE2	MIRCC, MIRCE	MR1E1*, BCCCM, GFDC3, ACC13, IPCAM, GFDEM, MIRC5, GIE2R, GIERC, MIRCC	---
≥ 5 °C	HADC3*, CNRMC, CNRM2*, IPCAM, INMC4, IPCAL, IPCBL	IPCBL, IPCAM, IPCAL	IPCAM, IPCAL, IPCBL	CANE2, IPCAM, IPCAL, IPCBL	IPCAL, IPCBL	IPCAM, IPCAL, IPCBL	MIRCE, GIEHC, GIE2H, IPCAL, CSI3L*, IPCBL	IPCAM, IPCAL, IPCBL

TMAX								
ISEASONAL BIAS\	SW	NW	GP	MW	NE	SE	AK	US
0 - 1 °C	CMCCS, MPIEP*, CESMC	MIRC5	CMCCC, <b>MME42</b> , BNUEM, CESMC, NOR1M, INMC4	INMC4, <b>MME42</b> , CESMC	CESMW*, MIRCC, MIRCE, CESMB	CESMB, CCSM4, ACC10, CESMF*, MIRCE, MIRCC, ACC13, MIRC5, <b>MME42</b> , CESMW*, CESMC	GIEHC, GIE2R, GIERC, MIRC5, GIE2H	CESMC, INMC4, CESMB, CESMF*
1 - 2 °C	MPIEL, MPIEM, ACC13, IPCAL, INMC4, IPCBL, CMCCC, <b>MME42</b> , MIRC5, CESMF*, CESMB, CCSM4, CSI3L*, CMCC*, MIRCC, ACC10, HADEC, BCCCM, CNRMC	CESMC, INMC4, GIE2H, CESMB, <b>MME42</b> , CESMF*, GIEHC, MIRCE, CSI3L*, CCSM4, MIRCC, CNRMC, ACC13, NOR1M	CMCCS, MPIEL, MPIEM, MPIEP*, BCC1, CMCC*, ACC10, CCSM4, CESMF*, GIE2H, GIERC, CESMB, GIE2R, GIEHC, CESMW*, CNRM2*, ACC13, CSI3L*, MIRC5, GFDEM, MR1E1*, MIRC3, GFDC3, CNRMC	MPIEP*, CESMW*, CNRMC, ACC10, CNRM2*, ACC13, MPIEM, CCSM4, NOR1M, MIRC5, CESMB, CESMF*, GIEHC, MPIEL, GIE2H	ACC13, CCSM4, CESMC, CESMF*, NOR1M, BCCCM, MIRC5, <b>MME42</b> , CANE2, INMC4, ACC10	BNUEM, MPIEP*, MPIEM, MPIEL, CNRMC, CMCCS, CNRM2*, CMCC*, NOR1M, BCC1, GIEHC, HADEE, CSI3L*, INMC4	MIRC3, MR1E1*, MIRCE, MIRCC, CESMC, INMC4	ACC13, CCSM4, CESMW*, MIRC5, CNRMC, <b>MME42</b> , MIRCC, ACC10, MIRCC, CNRM2*, MPIEP*, NOR1M, CMCC*, HADEE, GIEHC, GIE2H, CSI3L*
2 - 3 °C	GIEHC, CSI36, HADEE, GIE2H, CNRM2*, CANE2, CESMW*, NOR1M, GIE2R, GIERC, HADC3*, BNUEM, IPCAM	CNRM2*, ACC10, CSI36, CANE2, IPCAL, BCCCM, GFDC3, GIE2R, MR1E1*, MPIEP*, HADEC, CMCCS, CESMW*, GIERC, HADC3*, MIRC3, HADEE	GFDEG, HADEC, HADC3*, MIRCC, HADEE, MIRCE, CSI36, BCCCM	BCC1, BNUEM, CMCCS, MR1E1*, MIRC3, HADEE, GIERC, MIRCE, MIRCC, CMCCC, GIE2R, CMCC*, GFDC3, CSI3L*, HADEC, GFDEM	HADEE, BNUEM, GIE2H, GIERC, CNRMC, CNRM2*, GIE2R, BCC1, MPIEP*, CMCC*, CSI36, GIEHC, HADEC, MPIEM	CMCCC, HADC3*, GIE2H, CANE2, HADEC, CSI36, MR1E1*, GIERC, MIRC3, GIE2R, GFDC3	<b>MME42</b> , ACC13, CANE2	MPIEM, CMCCS, BCCCM, MPIEL, CMCCC, HADC3*, HADEC, BCC1, CANE2, BNUEM, CSI36, GIE2R, GIERC, MR1E1*, MIRC3
3 - 5 °C	BCC1, MR1E1*, MIRC3, GFDC3, GFDEM, GFDEG	CMCCC, CMCC*, BCC1, BNUEM, MPIEL, MPIEM, GFDEM, IPCBL, FGOG2*, IPCAM	FGOG2*	BCCCM, HADC3*, CSI36, GFDEG, FGOG2*, CANE2	HADC3*, GFDC3, CMCCC, MPIEL, MR1E1*, MIRC3, IPCBL, CMCCS, GFDEM, CSI3L*, IPCAL, GFDEG	BCCCM, FGOG2*, GFDEM, IPCAL, IPCBL, GFDEG	GFDEM, CSI36, CESMW*, CNRMC, CESMB, CESMF*, CSI3L*, CCSM4, HADC3*, GFDC3, GFDEG, CNRM2*, CMCCC, FGOG2*, ACC10, MPIEP*, BCCCM, HADEE, CMCCS, NOR1M, HADEC	GFDC3, IPCAL, GFDEM, IPCBL, GFDEG, FGOG2*
≥ 5 °C	FGOG2*	GFDEG	CANE2, IPCAL, IPCBL, IPCAM	IPCAL, IPCBL, IPCAM	FGOG2*, IPCAM	IPCAM	MPIEM, BCC1, MPIEL, BNUEM, CMCC*, IPCAL, IPCAM, IPCBL	IPCAM

\* Models for which RCP 4.5 and/or RCP8.5 simulations are not available.

**TABLE A2.** Model ranking by absolute value of the JAN 1900 – DEC 2005 trend bias. Models with biases significantly different from observations at the 90% confidence level are colored brown.

TMIN								
[BIAS]	SW	NW	GP	MW	NE	SE	AK	US
0 - 0.2 °C / century	CESMB, INMC4, CNRM2*, GFDEM, IPCBL, BCC1, IPCAM, <b>MME42</b> , HADC3*, NOR1M, HADEE, CESMF*, BCCCM, CESMW*, MRIC3, MRIE1*, CESMC, CNRMC, CSI3L*, CMCCC, CMCC*, GFDEG	CESMW*, GFDEM, IPCBL, MPIEL, CESMF*, CESMC, MRIE1*, CNRM2*	INMC4, CESMF*, FGOG2*, IPCBL, CSI3L*, GFDEG, IPCAM, CCSM4, BCC1, MRIE1*	BCC1, CESMF*, CESMW*, MPIEP*, CESMB, FGOG2*, INMC4	CESMW*, MIRCE, FGOG2*, CESMF*, GIE2R, CCSM4, BCC1, IPCAM, MPIEP*	MRIE1*, BNUEM, CESMW*, MPIEP*, MRIC3, MIRCE, IPCBL, FGOG2*, IPCAM, CSI3L*, CESMB, BCCCM, GIEHC, GFDEG, BCC1	HADEE, CANE2, FGOG2*, GFDEM, IPCAL, <b>MME42</b> , CESMW*, MIRCE, IPCAM, IPCBL	FGOG2*, INMC4, IPCBL, CESMW*, IPCAM, CESMB, BCCCM, BCC1, CESMF*, MRIE1*
0.2 - 0.5 °C / century	FGOG2*, HADEC, CMCCS, MIRCE, CANE2, GIEHC, <b>MIRCC</b> , ACC10, MPIEM, CCSM4	CESMB, INMC4, CANE2, FGOG2*, IPCAM, <b>MME42</b> , MPIEM, CCSM4, BCCCM, CMCCC, MPIEP*, BCC1, CSI3L*, MRIC3	BCCCM, CNRMC, CNRM2*, <b>MME42</b> , CESMB, MPIEP*, MIRCE, CMCCC, CESMW*, GFDEM, GIE2H, ACC10, NOR1M, CMCCS, MRIC3, BNUEM	MPIEL, MRIE1*, GFDEG, CNRMC, IPCBL, CCSM4, CSI3L*, MRIC3, CMCCC, GIEHC, <b>MME42</b> , GIE2R, GIE2H, CNRM2*	MRIE1*, GIEHC, ACC10, INMC4, IPCBL, MRIC3, <b>MME42</b> , CESMB, MRIC3, GIE2H, CMCC*, BNUEM, HADEE, GFDEM, BCCCM, CMCCS, GIERC, MPIEL, GFDC3, IPCAL	INMC4, CESMF*, <b>MME42</b> , CCSM4, NOR1M, ACC10, CESMC, CNRMC, HADEE, <b>CSI36</b> , CMCCC, MPIEL, MRIC3, MPIEM, CANE2, CMCC*, CNRM2*, GIERC, GIE2R, GIE2H	GIE2R, CESMC, INMC4, GIERC, HADEC, BCC1, MRIE1*, GIE2H, MPIEL, MPIEM	CCSM4, <b>MME42</b> , CNRM2*, GFDEM, MPIEP*, CSI3L*, CNRMC, MRIC3, CMCCC, GFDEG, ACC10, GIEHC, NOR1M, MPIEL, MRIC3, GIE2H, MIRCE, CANE2, CMCCS, CMCC*
0.5 - 1 °C / century	GIE2H, ACC13, MPIEL, GIERC, MRIC3, BNUEM, MPIEP*, GIE2R, GFDC3, CSI36, IPCAL	BNUEM, ACC10, GIEHC, CNRMC, GFDEG, GFDC3, NOR1M, HADC3*, GIE2H, MIRCE, CMCCS, GIERC, MRIC3, MIRCC, CMCC*	CMCC*, MPIEM, HADC3*, CANE2, MIRCC, GIEHC, MPIEL, MRIC3, IPCAL, CSI36, GIE2R, GIERC, GFDC3, CESMC, HADEC, ACC13	BNUEM, IPCAM, CMCC*, ACC10, CMCCS, NOR1M, CANE2, IPCAL, BCCCM, MRIC3, GFDEM, GFDC3, CSI36, MIRCE, MPIEM, HADC3*, GIERC, CESMC, HADEC	CNRMC, NOR1M, CANE2, CMCCC, HADEC, ACC13, CNRM2*, MIRCC, HADC3*, CSI36, CSI3L*, GFDEG, MPIEM, CESMC	MIRCC, HADC3*, IPCAL, HADEC, GFDEM, ACC13, GFDC3, CMCCS	MIRCC, ACC10, MPIEP*, CMCCC, GFDC3, CMCCS, BCC1, NOR1M, HADC3*, MRIC3, CNRMC, CSI36, CESMB	CESMC, BNUEM, GIE2R, HADC3*, MPIEM, GFDC3, GIERC, IPCAL, HADEC, ACC13, CSI36, HADEC, MIRCC
1 - 2 °C / century	---	ACC13, CSI36, IPCAL, HADEE, HADEC	HADEC	ACC13, MIRCC, HADEC	---	---	CMCC*, GIEHC, MRIC3, BCCCM, CSI3L*, GFDEG, CCSM4, ACC13, CESMF*	---
≥ 2 °C / century	---	---	---	---	---	---	BNUEM	---

TMAX								
[BIAS]	SW	NW	GP	MW	NE	SE	AK	US
0 - 0.2 °C / century	MIRCE, NOR1M, CSI3L*, FGOG2*, CNRM2*, IPCAM, IPCBL, HADC3*, <b>MME42</b> , CNRMC, GFDEG, BCC1, CESMC, CESMB, BCCCM, CESMW*	MRIE1*, HADC3*, ACC10, CMCCS, BCC1, CNRMC, FGOG2*, CSI3L*, <b>MME42</b> , MIRCE, IPCAM, CANE2, GFDEG	IPCAM, MIRCE, CESMC, CNRM2*, CESMB, GFDEM, HADC3*, IPCBL, CMCCS, MIRCC, CNRMC, MRIC3, GIE2H, ACC10, <b>MME42</b> , CSI36	BCCCM, GIE2R, GIEHC, CANE2, CMCCS, CESMC, CMCC*, IPCBL, MRIC3, GIE2H, INMC4, <b>MME42</b> , CCSM4	GFDEM, CESMC, INMC4, CMCC*, CMCCS, MRIC3, BCCCM, HADEC, CNRMC, GIERC, MIRCC, MRIE1*, GFDC3, CMCCC, CANE2, FGOG2*, CSI3L*, GIE2H, NOR1M, <b>MME42</b>	GIEHC, MIRCC, GFDEM, CMCC*, CNRM2*, GIE2H, HADEC, GIE2R, ACC13, CANE2, ACC10, MRIE1*	MIRCC, MRIE1*, GIERC, IPCBL, <b>MME42</b>	MRIE1*, IPCAM, CESMC, ACC10, CNRMC, GFDEM, CNRM2*, <b>MME42</b> , MIRCE, CMCCS, MRIC3, CSI3L*, CANE2, MRIC3, FGOG2*, IPCBL
0.2 - 0.5 °C / century	MIRCC, CMCCS, GFDEM, CCSM4, MIRCC, CESMF*, MRIE1*, ACC10, CANE2, IPCAL, CMCC*	MIRCC, NOR1M, MRIC3, CNRM2*, GIERC, MRIC3, GFDC3, GIEHC, GIE2H, CMCC*, IPCBL, CESMC, ACC13	GFDC3, CMCC*, CANE2, BCCCM, ACC13, MRIC3, MRIE1*, CESMF*, CESMW*, FGOG2*, GIERC	IPCAM, CNRM2*, GFDEM, ACC10, MRIE1*, ACC13, GFDC3, HADC3*, HADEE, CESMB, CSI3L*, GIERC, CSI36, NOR1M, GFDEG, FGOG2*	ACC10, ACC13, GIEHC, GFDEG, IPCAM, GIE2R, CNRM2*, HADC3*, HADEE, MIRCE, MRIC3, BCC1, MPIEP*	HADC3*, GIERC, MPIEP*, MRIC3, <b>MME42</b> , GFDC3, HADEE, CESMB, CSI36, BCC1, CNRMC, IPCBL, CESMC, INMC4, <b>IPCAM</b> , BCCCM, CMCCS	MIRCE, CMCCC, CESMC, HADEC, IPCAL, FGOG2*, GFDEM, GFDC3, CANE2, ACC10, HADC3*, IPCAM, INMC4, GIE2R, CESMW*, HADEE	NOR1M, GIE2H, HADC3*, GFDEG, CMCC*, GIE2R, BCCCM, GIEHC, GFDC3, ACC13, GIERC, BCC1, HADEE, CESMB, INMC4, MIRCC, CSI36
0.5 - 1 °C / century	MRIC3, MPIEM, MPIEP*, GFDC3, CSI36, ACC13, HADEE, INMC4, CMCCC, BNUEM, GIERC, MPIEL, GIE2H, GIE2R	GFDEM, GIE2R, CESMW*, CESMB, MPIEM, INMC4, MPIEL, CESMF*, CSI36, HADEE, CCSM4, HADEC, BCCCM, IPCAL	NOR1M, GIE2R, GIEHC, CCSM4, CSI3L*, BCC1, GFDEG, HADEE, INMC4, IPCAL, MPIEP*, BNUEM, HADEC, MPIEM	MIRCC, CNRMC, MIRCE, CESMF*, MPIEP*, HADEC, CMCCC, BCC1, CESMW*, MPIEL, MIRCC	CSI36, CCSM4, BNUEM, CESMF*, IPCAL, IPCBL, CESMW*, MPIEL, CESMB	CSI3L*, CMCCC, NOR1M, MRIC3, FGOG2*, MIRCE, CCSM4, CESMF*, CESMW*, GFDEG, BNUEM, IPCAL, MPIEM	CMCCS, MPIEL, MPIEP*, NOR1M, MRIC3, MPIEM, CMCC*, GIE2H, CNRM2*, MRIC3, CNRMC, BCC1, CSI36	CCSM4, CESMW*, CESMF*, MPIEP*, CMCCC, HADEC, IPCAL, MPIEL, BNUEM, MPIEM
1 - 2 °C / century	GIEHC, HADEC	MPIEP*, BNUEM, CMCCC	MPIEL, CMCCC	BNUEM, IPCAL, MPIEM	MPIEM	MPIEL	BCCCM, GFDEG, CSI3L*, GIEHC, CESMB, ACC13, CCSM4	---
≥ 2 °C / century	---	---	---	---	---	---	CESMF*, BNUEM	---

\* Models for which RCP 4.5 and/or RCP8.5 simulations are not available.

**TABLE A3.** Model ranking by absolute value of the 1900 – 2005 max-month trend bias. Models with biases significantly different from observations at the 90% confidence level are colored brown.

TASMIN								
[BIAS]	SW (MAY)	NW (JAN)	GP (FEB)	MW (FEB)	NE (FEB)	SE (NOV)	AK (OCT)	US (FEB)
0 - 0.2 °C / century	CESMB, MRIE1*, INMC4, FGOG2*, MPIEM, CMCCS, CMCC*, MPIEP*	CESMF*, CMCC, BCCCM	IPCAM	CMCC, INMC4, IPCAL, IPCBL	CESMW*	CCSM4, IPCAL	ACC13	IPCAM, INMC4
0.2 - 0.5 °C / century	CNRM, MPIEL, IPCAL, GFDEG, MIRCC, BCC1, CMCC, GFDEM, CESMW*	CNRM2*, INMC4, CESMB, MPIEP*, CESMC, BNUEM	MPIEP*, BCCCM	MPIEM, GFDEG	GFDC3, IPCAM, MPIEM	HADC3*, CESMB, BNUEM, GIE2R, MPIEL, GIEHC	MIRIC3, HADEC, HADC3*	MPIEP*, MPIEM, GFDEM, CMCC
0.5 - 1 °C / century	HADEC, CSI3L*, MRIC3, GIE2R, <b>MME42</b> , NOR1M, CANE2, BNUEM, CCSM4, CNRM2*, CESMF*, HADEC, MIRCE, HADC3*, IPCAM, GIEHC	GFDEM, GIERC, IPCAM, CCSM4, MRIE1*	INMC4, GFDEG, MPIEM, GFDEM	---	MIRCS, MPIEP*, MIRCE, IPCBL, CESMB, CCSM4, HADEC, INMC4	CESMF*, CMCC*, CSI36, GIE2H, MIRCE, BCCCM, HADEC, CESMW*	CMCCS, GFDEG, MIRCS, MRIE1*, CMCC	CESMC, CCSM4, CESMW*, GFDEG, IPCBL, BNUEM, GFDC3, MPIEL, CESMB, BCCCM, FGOG2*
1 - 2 °C / century	IPCBL, CSI36, GIERC, BCCCM, GIE2H, MIRCS, CESMC, ACC13, GFDC3, ACC10	MRIC3, CSI3L*, CANE2, MPIEM, GIE2R, GFDC3, <b>MME42</b> , BCC1, ACC10, MIRCE, IPCAL, CESMW*, CSI36, GIE2H, FGOG2*, GIEHC, MPIEL	CMCC, CCSM4, FGOG2*, IPCBL, ACC10, GFDC3, CSI3L*, MPIEL, CESMC, CANE2, <b>MME42</b> , MRIC3, HADC3*, BNUEM, BCC1, MRIE1*, CNRM2*, CESMF*	FGOG2*, BNUEM, GFDEM, IPCAM, MPIEP*, GFDC3, MIRCE, CESMW*, CESMF*, MIRCS, BCC1, ACC10	GFDEG, CMCC, BCC1, NOR1M, CESMC, <b>MME42</b> , IPCAL, HADEC, FGOG2*, MPIEL, ACC10, GIEHC, CMCC*, GIERC, BNUEM, ACC13, CMCCS, CANE2, CSI3L*	MRIE1*, GIERC, BCC1, MPIEM, <b>MME42</b> , CANE2, ACC10, MPIEP*, FGOG2*, MIRCS, CMCC, CESMC, GFDEG, CNRM, CSI3L*, ACC13, GFDC3, CNRM2*, MRIC3, HADEC, INMC4, GFDEM, CMCCS	MPIEL, GFDC3, CESMB, MPIEM, CNRM, ACC10, INMC4, MIRCC, CSI36, CNRM2*	<b>MME42</b> , BCC1, ACC10, MRIC3, GIEHC, CESMF*, MIRCE, CSI3L*, MIRCS, CANE2, CNRM2*, NOR1M, IPCAL, ACC13, MRIE1*, GIERC, HADEC, CMCCS, GIE2H, CNRM, CMCC*
≥ 2 °C / century	---	NOR1M, CNRM, HADC3*, MIRCS, HADEC, HADEC, IPCBL, ACC13, CMCCS, MIRCC, CMCC*, GFDEG	IPCAL, CESMW*, GIEHC, GIERC, GIE2H, ACC13, MIRCE, NOR1M, CNRM, MIRCS, HADEC, CMCC*, CESMB, CMCCS, MIRCC, GIE2R, CSI36, HADEC	BCCCM, <b>MME42</b> , CSI3L*, GIEHC, CNRM2*, HADEC, CMCCS, CESMC, CCSM4, NOR1M, CANE2, CESMB, CMCC*, GIERC, MRIC3, ACC13, CNRM, MPIEL, MIRCC, GIE2H, MRIE1*, GIE2R, HADEC, CSI36, HADC3*	CNRM, MRIC3, GIE2R, BCCCM, MIRCC, GIE2H, CSI36, MRIE1*, CNRM2*, HADC3*	NOR1M, IPCAM, MIRCC, IPCBL	CMCC*, <b>MME42</b> , CSI3L*, IPCAL, GFDEM, IPCAM, BCC1, HADEC, GIE2R, CESMW*, CANE2, NOR1M, IPCBL, MIRCE, GIE2H, GIERC, FGOG2*, CESMC, MPIEP*, CCSM4, GIEHC, BCCCM, BNUEM, CESMF*	GIE2R, MIRCC, HADC3*, HADEC, CSI36

TASMAX								
[BIAS]	SW (NOV)	NW (NOV)	GP (FEB)	MW (FEB)	NE (FEB)	SE (JAN)	AK (OCT)	US (FEB)
0 - 0.2 °C / century	CSI36, CANE2, ACC13	CSI36, GIE2R, CMCC*	GFDEG, MPIEP*, GFDEM	GFDEG	CESMW*, MPIEM	---	MIRIC3	MPIEM, IPCAL, CMCC, INMC4
0.2 - 0.5 °C / century	MIRCC, HADEC, GIEHC	CANE2, ACC10, ACC13	BCCCM, MPIEM, CSI3L*, IPCAL	MPIEM, INMC4, IPCAL	GFDC3, MIRCS	CSI3L*, CMCC*	CMCC, ACC13, CMCCS	GFDEM
0.5 - 1 °C / century	BCC1, GFDC3, ACC10, CESMC, CESMF*, CMCC*, GIE2R, HADC3*, IPCBL, MIRCE	GIEHC, IPCAM, BCC1, GFDEG, MRIC3, CESMC, HADEC, GFDC3	INMC4, CCSM4, CESMC, MPIEL, FGOG2*, MRIC3, CMCC	CESMW*, GFDEM	MIRCE, HADEC, IPCAM, MPIEP*, CMCC, CESMC, GFDEG	GIERC, NOR1M, CESMC, CNRM2*, CANE2, CMCCS, ACC13, MIRCC, BNUEM, INMC4, BCCCM, GIE2R	GFDEG, HADEC, MRIE1*, HADC3*, MIRCS, MPIEL, IPCAL	MPIEP*, GFDEG, CESMW*, IPCBL, CESMC, MPIEL, BCCCM, CCSM4, BNUEM, GFDC3
1 - 2 °C / century	MRIC3, IPCAM, BCCCM, CNRM, <b>MME42</b> , CMCC, NOR1M, MRIE1*, GFDEG, MIRCS, GIE2H, GFDEM, HADEC, CESMW*, CMCCS, GIERC, CSI3L*, CNRM2*, CESMB, MPIEP*, BNUEM	MIRCC, HADC3*, FGOG2*, MIRCE, <b>MME42</b> , HADEC, MPIEM, CMCC, INMC4, CNRM, GIERC, MIRCS, CESMF*, CESMW*, NOR1M, CNRM2*, CCSM4, BCCCM, BNUEM, CSI3L*	GFDC3, HADC3*, CESMW*, ACC10, MRIE1*, CANE2, ACC13, <b>MME42</b> , NOR1M, CESMF*, BNUEM, IPCBL, BCC1, CNRM2*, MIRCS, GIERC, GIE2H	CMCC, MPIEP*, IPCBL, MIRCE, GFDC3, FGOG2*, MIRCS, BNUEM, BCCCM, ACC10, CESMF*, CSI3L*, HADEC, <b>MME42</b> , CMCC*	CESMF*, CESMB, GFDEM, CMCC*, ACC10, IPCBL, NOR1M, GIERC, CCSM4, <b>MME42</b> , BCC1, CSI3L*, INMC4, GIEHC, MRIE1*, HADEC, IPCAL, FGOG2*, MPIEL, BNUEM, CMCCS, ACC13	GIE2H, FGOG2*, CSI36, GFDEG, GIEHC, HADEC, IPCAM, CESMW*, <b>MME42</b> , ACC10, CNRM, IPCAL, MIRCS, MRIE1*, MPIEM, MPIEL, CESMF*, GFDEM, HADC3*, GFDC3	INMC4, GFDC3, CNRM, MPIEM, IPCAM, ACC10, CESMB, MIRCC, CNRM2*, CMCC*, <b>MME42</b> , CSI36, BCC1, CESMW*	CSI3L*, CESMB, CESMF*, FGOG2*, MIRCS, BCC1, ACC13, MIRCE, MRIE1*, GIEHC, NOR1M, IPCAM, CMCC*, GIERC, CNRM2*, CANE2, MRIE1*, GIE2H, CNRM, HADEC, CMCCS, HADC3*
≥ 2 °C / century	MPIEM, CCSM4, FGOG2*, INMC4, IPCAL, MPIEL	CMCCS, IPCAL, MPIEL, GFDEM, MPIEP*, CESMB	MIRCE, GIEHC, CMCC*, CESMB, IPCAM, MIRCC, CNRM, HADEC, CMCCS, GIE2R, CSI36, HADEC	BCC1, CESMC, CCSM4, NOR1M, GIEHC, ACC13, MRIC3, MPIEL, CMCCS, GIERC, IPCAM, CESMB, CNRM2*, CANE2, CNRM, MRIE1*, MIRCC, GIE2H, HADC3*, HADEC, CSI36, GIE2R	GIE2H, GIE2R, BCCCM, CANE2, MRIE1*, MRIC3, MIRCC, CNRM, CSI36, CNRM2*, HADC3*	HADEC, MRIC3, MPIEP*, MIRCE, BCC1, CMCC, CESMB, CCSM4, IPCBL	NOR1M, GFDEM, GIE2R, CANE2, CSI3L*, MIRCE, FGOG2*, HADEC, CESMC, GIE2H, IPCBL, GIERC, CCSM4, BCCCM, MPIEP*, GIEHC, BNUEM, CESMF*	MIRCC, GIE2R, HADEC, CSI36

\* Models for which RCP 4.5 and/or RCP8.5 simulations are not available.

**TABLE A4.** Model ranking by absolute value of the JAN 1900 – DEC 2005 monthly time series standard deviation bias. Here, the MME is the mean of individual model standard deviation values. Models with biases significantly different from observations at the 95% confidence level are colored brown.

TASMIN								
(STDDEV BIAS)	SW	NW	GP	MW	NE	SE	AK	US
0 - 0.1 °C	MRIE1*, MPEL, HADEE, MIRCS, ACC10, CESMW*, GFDC3, GIEHC, MRIC3, CSI36, FGOG2*, GIEZH, MPIEP*, GIE2R, MIRCC, HADEE	ACC10, HADEC, MRIC3, MIRCC, CSI36, GIERC, HADEE, GIE2R, HADC3*, GFDC3, MIRCE, MRIE1*, CSI3L*, MIRCS	MRIE1*, CESMC, INMC4, CMCC*, GFDEM, CESMW*, BNUEM, GIERC, CSI36, CMCCC	GIEZH, GIERC, CANE2, CMCC*, GFDEG, CMCCC, ACC13	GIEHC, MIRCE, GIE2R, GIEZH, MIRCS, MIRCC, CNRMC, GFDEM	CNRMC, CESMC, MRIE1*, ACC10, MRIC3, GFDC3, GFDEM, BNUEM, GIERC, CNRM2*, FGOG2*	ACC13	MIRCS, ACC13, MIRCE, GIERC, MIRCC, GIEZH, GFDEM, CSI3L*, GFDC3, GIE2R, GIEHC, CSI36, CNRMC, MPEL, MPIEM, GFDEG
0.1 - 0.2 °C	CMCCS, GIERC, MPIEM, ACC13, CMCCC, MIRCE, CSI3L*, CESMC, GFDEG, CNRM2*, CNRMC, CMCC*	FGOG2*, MPEL, GIEZH, GFDEM, GIEHC, MPIEP*, CNRMC	GFDEG, ACC13, GIEZH, ACC10, NOR1M, CNRMC, CANE2, CNRM2*, GIEHC, MPEL, MPIEM, MRIC3, MME42, GIE2R, GFDC3	MIRCS, GIE2R, BNUEM, GIEHC, CNRM2*, GIEHC, MPIEM, MPEL	GIERC, CMCCC, ACC13, CSI3L*, CSI36, GFDEG	GIEZH, MME42, CMCCC, CANE2, CSI36, ACC13, GFDEG, GIE2R, CSI3L*, HADEC, CESMW*, BCCC1	GIE2R, CMCCC, GIERC	CMCCC, CANE2, CNRM2*, CMCC*, MPIEP*, CMCCS, ACC10, HADEE, MME42
0.2 - 0.5 °C	NOR1M, MME42, HADC3*, GFDEM, BCCC1, CCSM4, CESMB, CESM*, CANE2, BNUEM, BCCCM	CNRM2*, CESMC, CESMW*, ACC13, MPIEM, MME42, INMC4, GFDEG, CMCCS, CANE2, CMCC*, CMCCC, CESMF*	CMCCS, MIRCS, CSI3L*, HADC3*, MPIEP*, HADEE, BCCC1, FGOG2*, MIRCC, MIRCE, HADEC, CESMF*, CESMB, CCSM4	CNRMC, FGOG2*, MRIE1*, CSI3L*, CESMC, MIRCE, CMCCS, MME42, INMC4, MIRCC, NOR1M, CESMW*, MPIEP*, BCCC1, CSI36, CNRM2*, MRIC3, CESMB, HADC3*	CNRM2*, CANE2, BCCC1, CMCCS, MPEL, GFDC3, MPIEM, NOR1M, CMCC*, MME42, MPIEP*, FGOG2*, BNUEM, INMC4, IPCAM, CESMB, CESMC, ACC10	INMC4, MPEL, BCCCM, MPIEM, MPIEP*, GIEHC, HADEE, HADC3*, CMCCS, MIRCS, CNRM2*, GIEHC, ACC10, HADEE, CESMB, CCSM4, CESMF*	GIEZH, MRIE1*, HADC3*, GFDEM, GFDEG, GFDC3, CNRM2*, GIEHC, ACC10, HADEE, MPIEP*	FGOG2*, BCCC1, INMC4, CESMC, CESMW*, MRIE1*, HADC3*, BNUEM, NOR1M, HADEC, MRIC3, BCCCM, CESMB, CCSM4, CESMF*
0.5 - 1 °C	---	CCSM4, BCCC1, BNUEM, CESMB, NOR1M, BCCCM	BCCCM	CCSM4, ACC10, HADEE, CESMF*, BCCCM, HADEC, IPCAM, IPCAL	MRIE1*, CESMW*, BCCCM, CCSM4, CESMF*, HADEC, HADC3*, MRIC3, HADEC, IPCAL	MIRCC, MIRCE, IPCAM, IPCAL	MIRCE, CANE2, MRIC3, MPEL, CNRMC, INMC4, MPIEM, HADEC, CSI36, MIRCC, MRIC3, MME42, CMCCS, CESMC, CSI3L*, BCCCM	IPCAL, IPCAM, IPCBL
≥ 1 °C	INMC4, IPCAM, IPCAL, IPCBL	IPCAL, IPCBL, IPCAM	IPCAL, IPCAM, IPCBL	IPCAL	IPCAL	IPCAL	IPCAL, CESMB, BCCC1, CESMF*, CCSM4, CMCC*, FGOG2*, NOR1M, IPCBL, CESMW*, IPCAL, BNUEM	---

TASMAX								
(STDDEV BIAS)	SW	NW	GP	MW	NE	SE	AK	US
0 - 0.1 °C	CMCCS, GFDC3, CESMW*, MPEL, MPIEP*, NOR1M, CMCC*, IPCBL, MPIEM, MIRCC, ACC13, MIRCE, MRIC3, MME42	MPEL, CESMC, CESMW*, ACC10, MPIEP*, ACC13, HADEC, HADC3*, MME42, CNRMC, CNRM2*, HADEE	ACC13, CSI36, BNUEM, FGOG2*, CANE2, INMC4, MME42, CNRM2*	CESMC, NOR1M, MIRCC, CSI3L*, MIRCE, CANE2, INMC4, CMCCC, CMCCS, MPEL, CSI36	CSI36, INMC4, MIRCE, ACC13, CNRMC, CMCCC, GIERC, NOR1M, FGOG2*, MIRCS, MIRCC	CESMB, IPCBL, CCSM4, CESMF*, ACC13, MME42, CNRM2*, CANE2, CMCC*	ACC13	MPEL, MIRCC, CSI36, FGOG2*, GFDC3, ACC13, INMC4, MRIC3, MPEL, NOR1M, MRIE1*, CMCCC, MIRCE, IPCAL, MRIC3, CESMC, CMCCS, CNRMC, MME42, IPCAM, GIERC, CSI3L*, CMCC*, CNRM2*
0.1 - 0.2 °C	BCCC1, MRIE1*, GIEZH, CESMC, CMCCC, MIRCS, INMC4, GFDEG, FGOG2*, GIEHC, IPCAL, ACC10, CNRM2*, CESMF*, HADEE	CSI36, IPCBL, INMC4, MPIEM, GFDEM, GFDC3, MIRCE, MIRCS, CMCC*, MIRCC, CMCCS, CMCCC	CMCC*, MPIEP*, CNRMC, CMCCS, ACC10, GFDC3, CESMC, BCCC1, IPCAM	MIRCS, CMCC*, FGOG2*, MIRCS, GFDC3, MPIEM, GFDEG, GFDEM, MME42	GFDEM, CESMC, GIE2R, CNRM2*, CSI3L*, MRIC3, GIEHC, GFDEG, MRIE1*, CMCC*, GIEZH	CMCCC, ACC10, CESMW*, MRIC3, FGOG2*, MIRCS, GFDC3, NOR1M, GIEZH, MPEL, INMC4, IPCAM, GIEHC, CNRMC	MRIE1*, CSI36	MPIEP*, ACC10, IPCBL, CESMW*, CANE2, GFDEG, GIEZH, GIE2R, GFDEM, GIEHC, HADC3*, HADEE, BCCC1
0.2 - 0.5 °C	CSI36, BCCCM, GFDEM, CESMB, CSI3L*, CANE2, CNRMC, IPCAM, GIERC, CCSM4, HADEC, GIE2R, HADC3*, BNUEM	GFDEG, CSI3L*, FGOG2*, NOR1M, CESMF*, GIE2R, GIERC, CCSM4, BCCC1, MRIC3, IPCAL, MRIE1*, IPCAM, CESMB, GIEZH, BNUEM, GIEHC, CANE2	MPEL, CESMW*, CESMB, CESMF*, GFDEG, HADC3*, MPEL, CMCCC, IPCAL, CCSM4, GFDEM, NOR1M, HADEC, CSI3L*, BCCCM, GIERC, GIEHC, MRIC3, GIE2R, MIRCC	CESMW*, MRIE1*, CESMB, GIERC, BNUEM, GIEZH, ACC13, GIE2R, CNRMC, MPIEP*, CCSM4, CESMF*, IPCAM, HADC3*, CNRM2*, BCCCM, GIEHC, IPCAL	MME42, GFDC3, CMCCS, CANE2, HADC3*, CESMB, BCCC1, MPIEM, CESMW*, ACC10, MPEL, CESMF*, CCSM4, MPIEP*, BCCCM, BNUEM, HADEE	GIERC, MRIE1*, MPEL, CSI3L*, IPCAL, CSI36, CMCCS, MPIEP*, BNUEM, HADEC, GIE2R, HADEC, CESMC, BCCCM, MIRCC, MIRCE, GFDEG, BCCC1	GIE2R, GIERC, INMC4, CSI3L*, HADC3*, GFDEM, GIEZH, CMCCC, MRIC3, MIRCE, GFDC3, BCCCM, GFDEG, CANE2, GIEHC, MIRCC, MME42	HADEC, CESMB, BCCCM, BNUEM, CESMF*, CCSM4
0.5 - 1 °C	---	BCCCM	MIRCE, MRIE1*	BCCC1, ACC10, HADEC, HADEC, IPCBL	HADEC, IPCAM, IPCAL	GFDEM, HADC3*	MPIEP*, ACC10, IPCAL, HADEC, CNRM2*, MPEL, MPIEM, MRIC3, CNRMC, HADEC, CESMC, IPCAM, BCCC1, CMCCS, IPCBL	---
≥ 1 °C	---	---	---	---	IPCAL	---	FGOG2*, CESMB*, CESMF*, CCSM4, CESMW*, BNUEM, NOR1M, CMCC*	---

\* Models for which RCP 4.5 and/or RCP8.5 simulations are not available.

## References

- Cheng, L., T.J. Phillips, and A. AghaKouchak, 2015: Non-stationary return levels of CMIP5 multi-model temperature extremes. *Climate Dyn*, 44, 2947-2963, doi: 10.1007/s00382-015-2625-y.
- Garfin, G., A. Jardine, R. Merideth, M. Black, and S. LeRoy, 2013: *Assessment of Climate Change in the Southwest United States: A Report Prepared for the National Climate Assessment*. A report by the Southwest Climate Alliance. Washington, DC: Island Press.
- Geil, K.L., Y. Serra, and X. Zeng, 2013: Assessment of CMIP5 Model Simulations of the North American Monsoon System. *J Climate*, 26, 8787-8801.
- Jun, M., R. Knutti, and D.W. Nychka, 2008: Spatial analysis to quantify numerical bias and dependence. *J Amer Stat Assoc*, 103:483, 934-946, doi:10.1198/016214507000001265.
- Knutti, R., 2008: Why are climate models reproducing the observed global surface warming so well? *Geophys Res Lett*, 35(L18704), doi:10.1029/2008GL034932.
- Knutti., R., 2010: The end of model democracy? *Climatic Change*, 102, 395-404, doi:10.1007/s10584-010-9800-2.
- Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G.A. Meehl, 2010: Challenges in combining projections from multiple climate models. *J Climate*, 23, 2739-2758, doi:10.1175/2009JCLI3361.1
- Macadam, I., A.J. Pitman, P.H. Whetton, and G. Abramowitz, 2010: Ranking climate models by performance using actual values and anomalies: Implications for climate change impact assessments. *Geophys Res Lett*, 37, L16704, doi:10.1029/2010GL043877.
- Maxino, C.C., B.J. McAvaney, A.J. Pitman, and S.E. Perkins, 2008: Ranking the AR4 climate models over the Murray-Darling Basin using simulated maximum temperature, minimum temperature and precipitation. *Int J Climatol*, 28, 1097-1112, doi: 10.1002/joc.1612.
- Melillo, Jerry M., Terese (T.C.) Richmond, and Gary W. Yohe, 2014: *Climate Change Impacts in the United States: The Third National Climate Assessment*. U.S. Global Change Research Program, 841 pp. doi:10.7930/J0Z31WJ2.
- NASA JPL, 2009, ASTER Global Digital Elevation Model. Version 2. NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (<https://lpdaac.usgs.gov>), accessed Nov 2015, doi:10.5067/ASTER/ASTGTM.002.
- Pennell, C. and T. Reichler, 2011: On the effective number of climate models. *J Climate*, 24, 2358-2367, doi: 10.1175/2010JCLI3814.1

- Perkins, S.E., A.J. Pitman, N.J. Holbrook, and J. McAneney, 2007: Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *J Climate*, 20, 4356-4376, doi: 10.1175/JCLI4253.1
- Reichler, T., and J. Kim, 2008: How well do coupled models simulation today's climate? *Bull Amer Meteor Soc*, 89, 303–311, doi: 10.1175/BAMS-89-3-303.
- Rohde, R., R.A. Muller, R. Jacobsen, E. Muller, S. Perlmutter, A. Rosenfeld, J. Wurtele, D. Groom, and C. Wickham, 2013: A new estimate of average earth surface land temperature spanning 1753 to 2011. *Geoinfor Goestat: An Overview*, 1:1, doi:10.4172/2327-4581.1000101.
- Santer, B.D and co-authors, 2009: Incorporating model quality information in climate change detection and attribution studies. *Proc Natl Acad Sci*, 106(35), 14778-14783, doi: 10.1073/pnas.0901736106.
- Sillmann, J., V. Kharin, X. Zhang, F.W. Zwiers, and D. Bronaugh, 2013: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J Geophys Res Atmos*, 118, 1716-1733, doi:10.1002/jgrd.50203.
- Sillmann, J., M.G. Donat, J.C. Fyfe, and F.W. Zwiers, 2014: Observed and simulated temperature extremes during the recent warming hiatus. *Environ Res Lett*, 9, 064023, doi: 10.1088/1748-9326/9/6/064023.
- Sun, Q., C. Miao, and Q. Duan, 2015: Comparative analysis of CMIP3 and CMIP5 global climate models for simulating the daily mean, maximum, and minimum temperatures and daily precipitation over China. *J Geophys Res Atmos*, 120, 4806-4824, doi:10.1002/2014JD022994.
- Taylor, K.E., R.J. Stouffer, and G.A. Meehl, 2009: A summary of the CMIP5 experiment design. PCDMI Rep., 33 pp. [Available online at [http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor\\_CMIP5\\_design.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf).]
- Taylor, K.E., R.J. Stouffer, and G.A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull Amer Meteor Soc*, 93, 485-498, doi: 10.1175/BAMS-D-11-00094.1.
- Wuebbles, D. and co-authors, 2014: CMIP5 climate model analysis: climate extremes in the United States. *Bull Amer Meteor Soc*, 95, 571-583, doi:10.1175/BAMS-D-12-00172.1.
- Yao, Y., Y. Luo, J. Huang, and Z. Zhao, 2013: Comparison of monthly temperature extremes simulated by CMIP3 and CMIP5 models. *J Climate*, 26, 7692-7707, doi:10.1175/JCLI-D-12-00560.1.

## Tables and Figures

**Table 1.** CMIP5 models used for this study.

Model	Code	Modeling Center	Country
ACCESS1-0	ACC10*	Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM)	Australia
ACCESS1-3	ACC13*		
BCC-CSM1.1	BCCC1*	Beijing Climate Center, China Meteorological Administration	China
BCC-CSM1.1m	BCCCM*		
BNU-ESM	BNUEM*	Beijing Normal University	China
CanESM2	CANE2*	Canadian Centre for Climate Modelling and Analysis	Canada
CCSM4	CCSM4*	National Science Foundation, US Department of Energy, and National Center for Atmospheric Research	USA
CESM1-BGC	CESMB*		
CESM1-CAM5	CESMC*		
CESM1-FASTCHEM	CESMF		
CESM1-WACCM	CESMW		
CMCC-CESM	CMCCE	Centro Euro-Mediterraneo sui Cambiamenti Climatici	Italy
CMCC-CM	CMCCC*		
CMCC-CMS	CMCCS*		
CNRM-CM5	CNRMCM*	Centre National de Recherches Meteorologiques/Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique	France
CNRM-CM5-2	CNRM2		
CSIRO-Mk3.6.0	CSI36*	Commonwealth Scientific and Industrial Research Organization, Queensland Climate Change Centre of Excellence	Australia
CSIRO-Mk3L-1-2	CSI3L	University of New South Wales	Australia
FGOALS-g2	FGOG2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and CESS, Tsinghua University	China
GFDL-CM3	GFDC3*	NOAA Geophysical Fluid Dynamics Laboratory	USA
GFDL-ESM2G	GFDEG*		
GFDL-ESM2M	GFDEM*		
GISS-E2-H	GIE2H*	NASA Goddard Institute for Space Studies	USA
GISS-E2-H-CC	GIEHC*		
GISS-E2-R	GIE2R*		
GISS-E2-R-CC	GIERC*		
HadCM3	HADC3	Met Office Hadley Centre	UK
HadGEM2-CC	HADEC*		
HadGEM2-ES	HADEE*		
INM-CM4	INMC4*	Institute for Numerical Mathematics	Russia

IPSL-CM5A-LR	IPCAL*	Institut Pierre-Simon Laplace	France
IPSL-CM5A-MR	IPCAM*		
IPSL-CM5B-LR	IPCBL*		
MIROC5	MIRC5*	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	Japan
MIROC-ESM	MIRCE*		
MIROC-ESM-CHEM	MIRCC*		
MPI-ESM-LR	MPIEL*	Max Planck Institute for Meteorology	Germany
MPI-ESM-MR	MPIEM*		
MPI-ESM-P	MPIEP		
MRI-CGCM3	MRIC3*	Meteorological Research Institute	Japan
MRI-ESM1	MRIE1		
NorESM1-M	NOR1M*	Norwegian Climate Centre	Norway

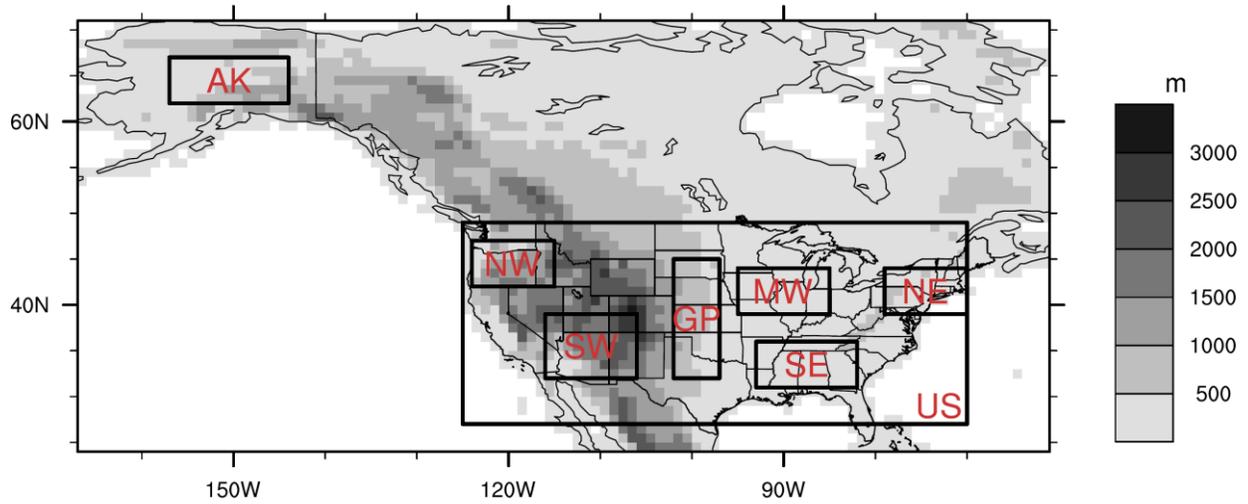
\* Models with output available for the historical experiment from 1900-2005, and for the RCP 4.5 and RCP 8.5 experiments from 2006-2100.

**Table 2.** Tmin and Tmax RMBAD MME models and total points for each model.

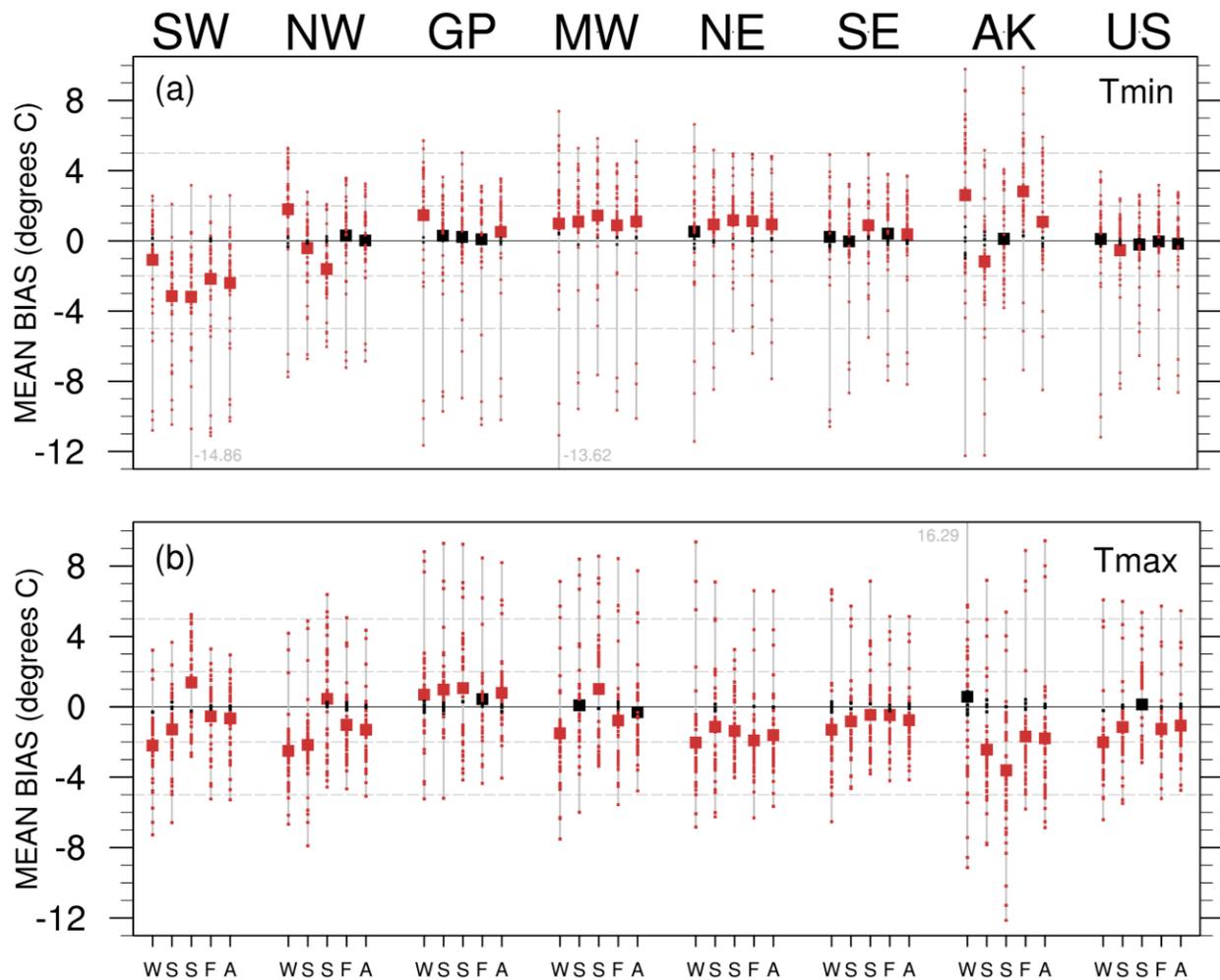
TMIN	POINTS	TMAX	POINTS
ACC10	2	ACC10	3
ACC13	3	ACC13	3
BCCC1	4	BCCCM	3
BCCCM	3	CCSM4	4
CCSM4	3	CESMB	3
CESMB	2	CESMC	1
CESMC	1	CMCCC	4
CMCCC	3	CMCCS	1
CMCCS	3	CNRMC	3
CSI36	4	INMC4	4
GFDC3	4	IPCAL	3
GFDEG	2	IPCAM	4
GFDEM	3	IPCBL	1
GIE2H	3	MIRC5	3
GIEHC	2	MIRCC	2
GIE2R	4	MIRCE	1
HADEE	2	MPIEL	3
MIRC5	2	MPIEM	3
MIRCC	2	NOR1M	2
MIRCE	3		
MPIEL	3		
MPIEM	3		
MRIC3	2		
NOR1M	3		

**Table 3.** Tmin and Tmax TOP5 MME models and total points for each model.

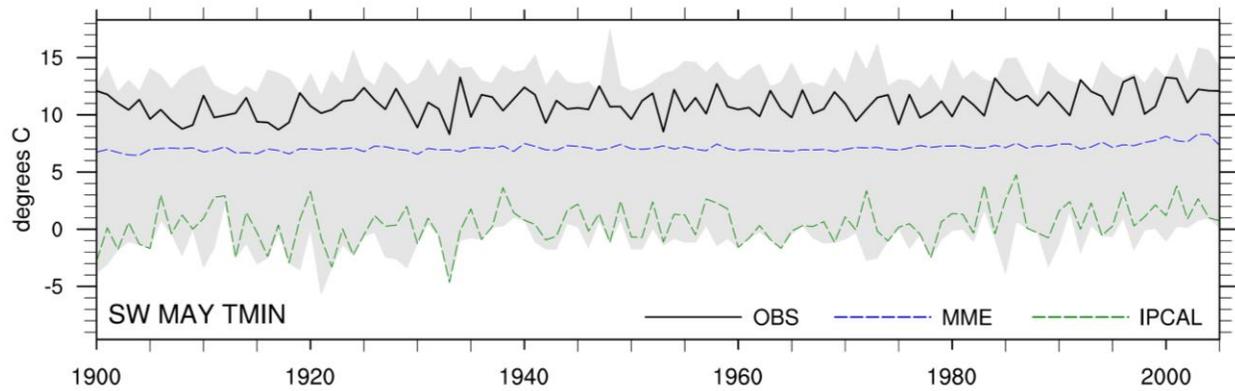
TMIN	POINTS	TMAX	POINTS
ACC10	2	CESMC	1
CESCM	1	CMCCS	1
GIEHC	2	IPCBL	1
MIRC5	2	MIRCE	1
MPIEL	3	MPIEL	3



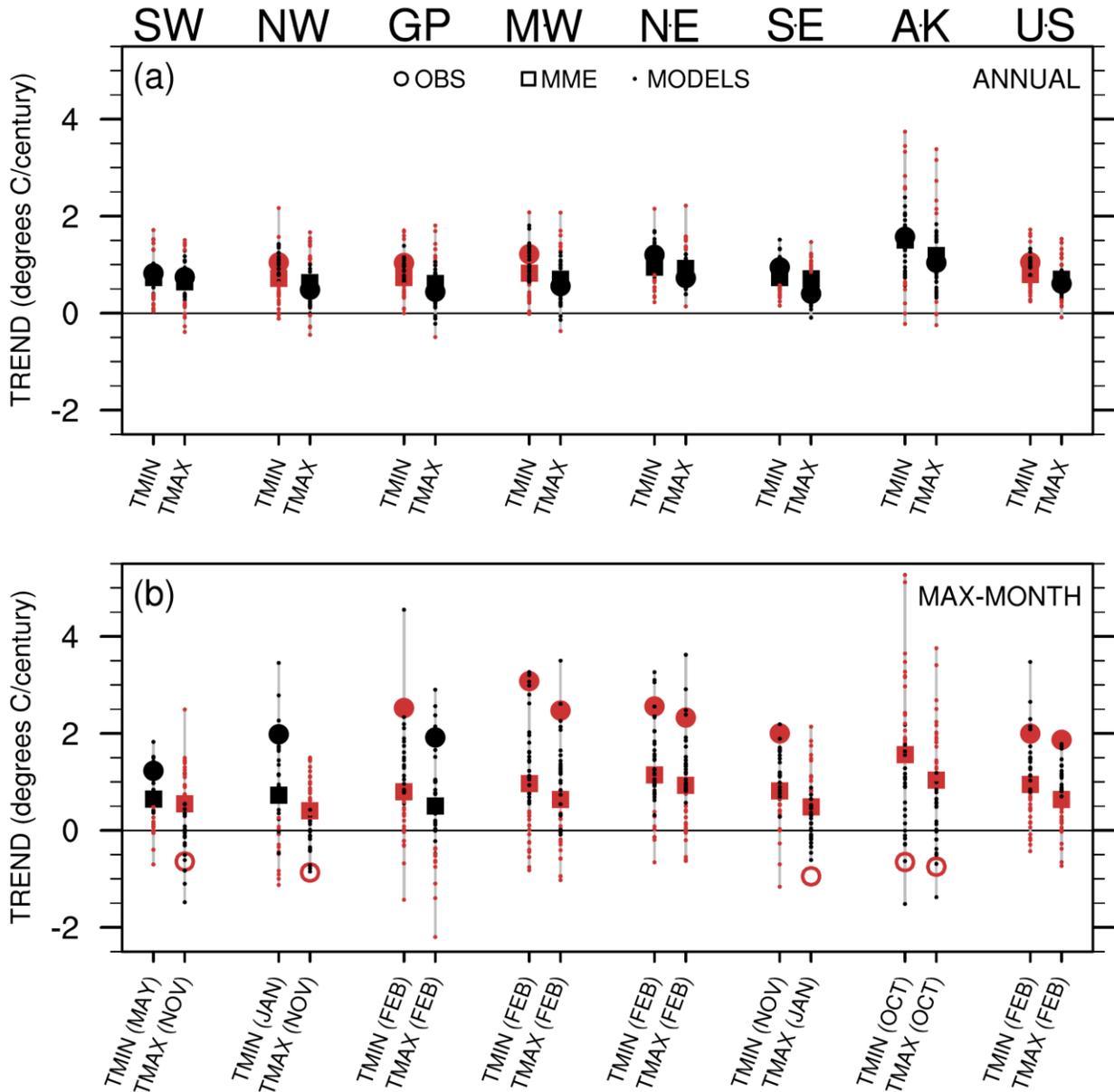
**Figure 1.** The eight study areas shown over ASTER global DEM topography on our  $1^{\circ} \times 1^{\circ}$  common grid. AK=Alaska, NW=northwest, SW=southwest, GP=great plains, MW=midwest, SE=southeast, NE=northeast, and US=United States.



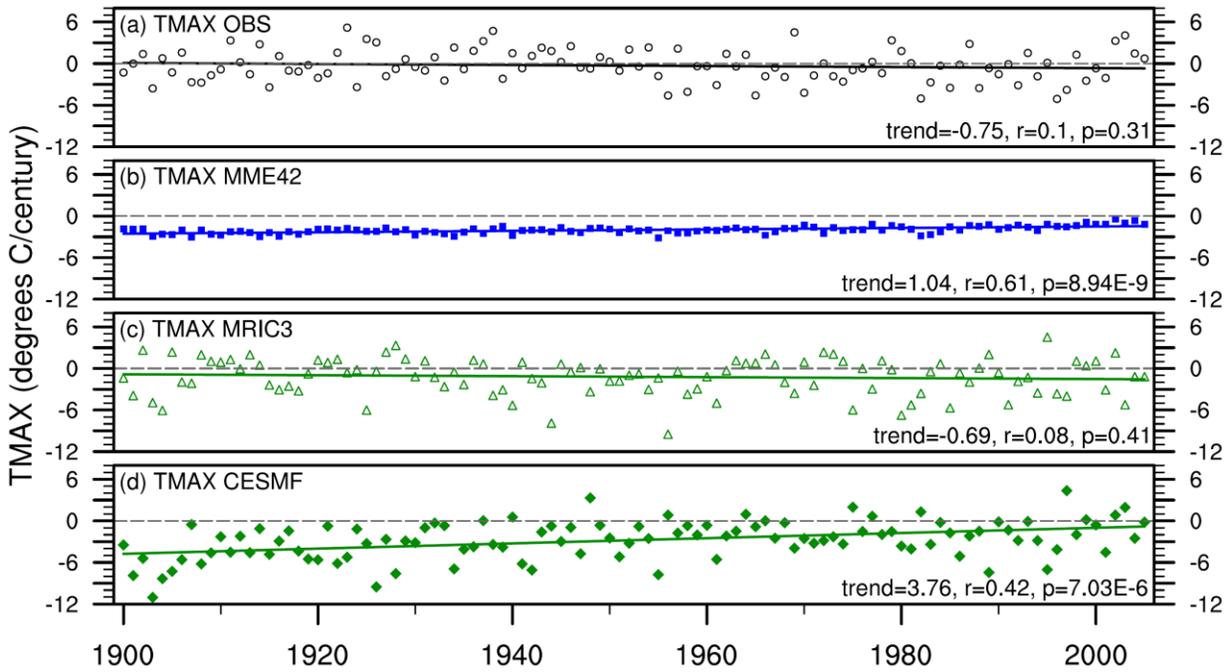
**Figure 2.** Bias in seasonal and annual mean Tmin and Tmax for the period 1900-2005 (where W, S, S, F, and A on the x-axis stand for winter, spring, summer, fall, and annual, respectively). The MME-average is shown with a square marker and individual models are shown with dots along grey vertical model spread bars. Colored markers indicate bias significance at the 90% confidence level after adjusting for serial lag-1 autocorrelation.



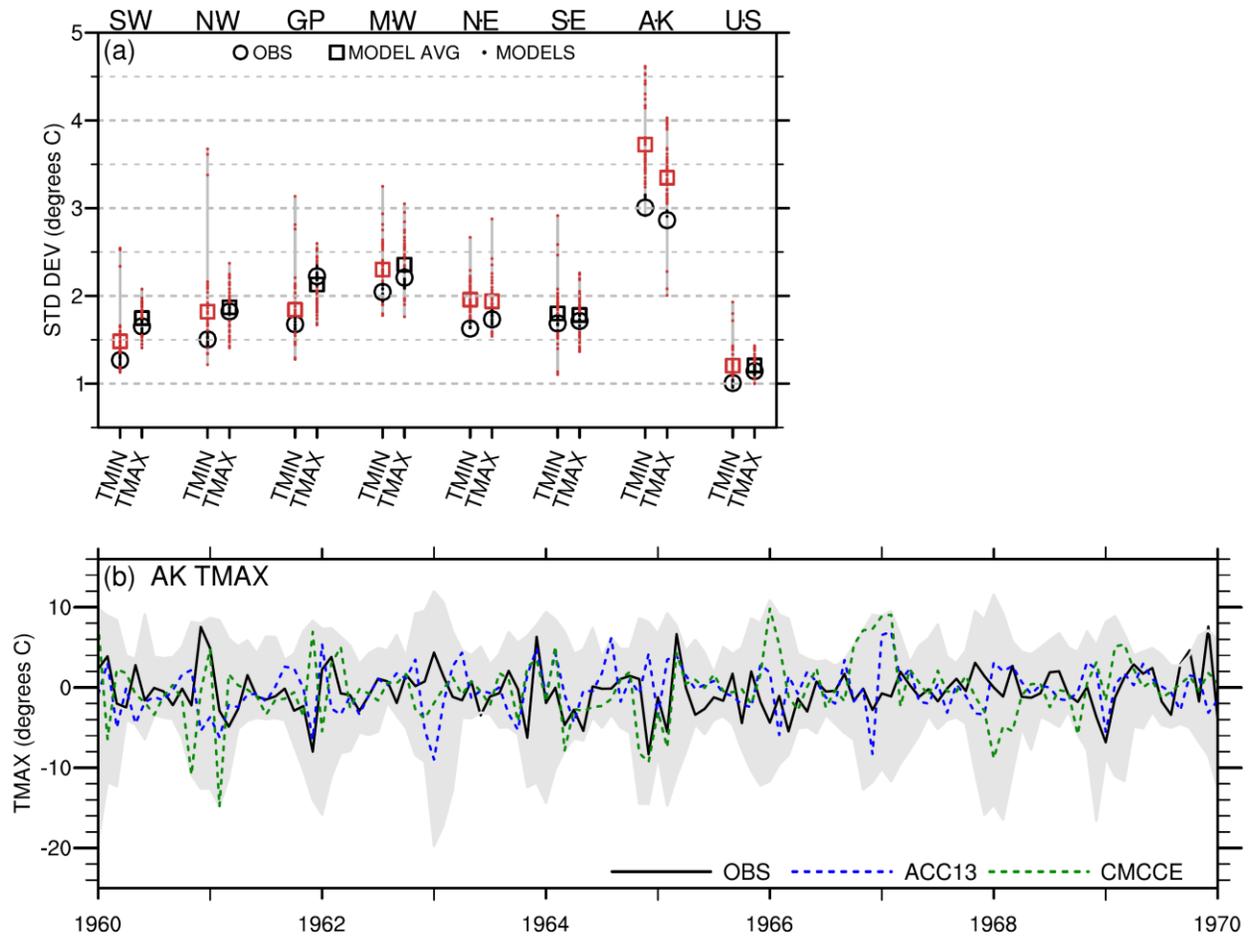
**Figure 3.** May time series of Tmin for the southwest region, where the solid line shows the observations, the blue dash shows the MME-average, and the gray shade shows the model spread. The green dash shows an example of one individual model.



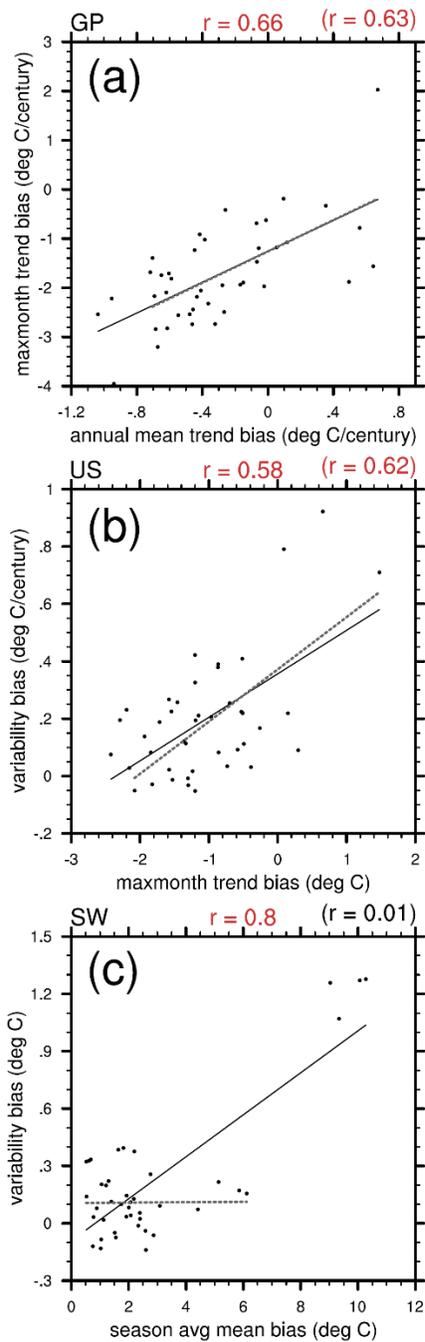
**Figure 4.** Annual average (a) and monthly average trends (b) in Tmin and Tmax. Only the month with the greatest MME-average trend bias is shown for each region in (b). Large circles are observations, large squares are the MME-average values, and individual models are shown with dots along grey vertical model spread bars. Filled large markers indicate MME-average and observed trends that are significant at the 90% confidence level, after adjusting for serial lag-1 autocorrelation. Color indicates that the modeled and observed trends are significantly different from each other at the 90% confidence level, after adjusting for serial lag-1 autocorrelation.



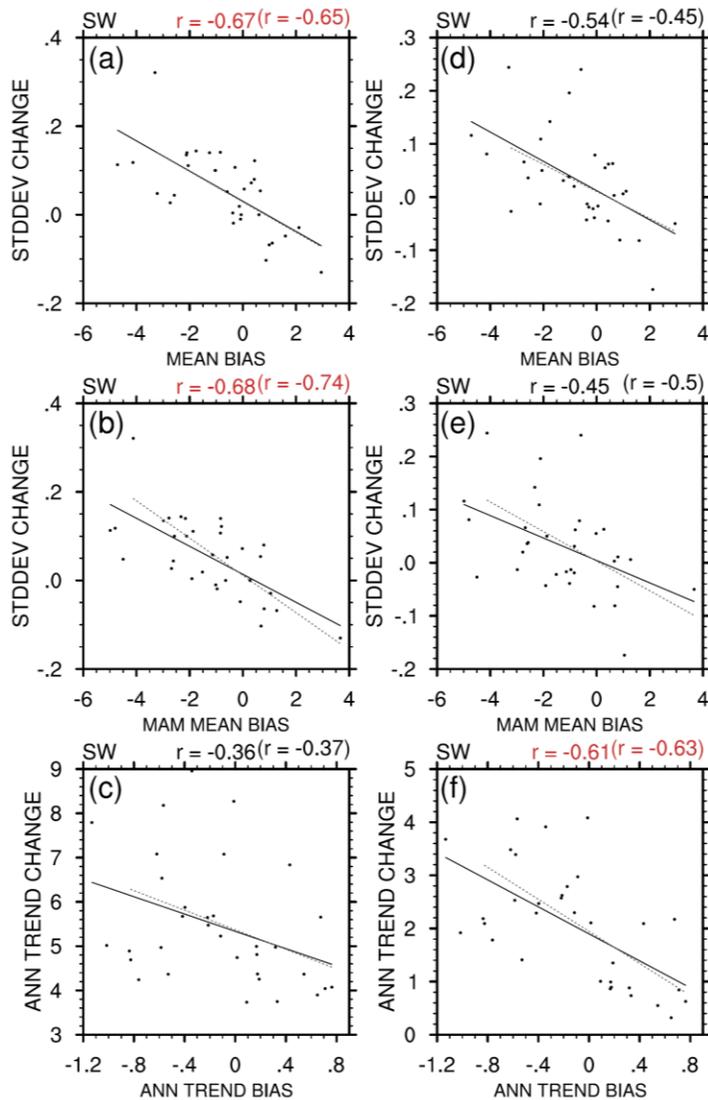
**Figure 5.** Alaska max-month (October) trend in Tmax for (a) observations, (b) the 42-model MME-average, (c) the model with the smallest trend bias (MRIC3), and (d) the model with the largest trend bias (CESMF). Filled markers indicate that the trend is significant at the 90% confidence level, after adjusting for serial lag-1 autocorrelation ( $p < 0.1$ ), while open markers are used for insignificant trends ( $p \geq 0.1$ ). The Pearson linear correlation coefficient ( $r$ ) of each trend is shown for reference.



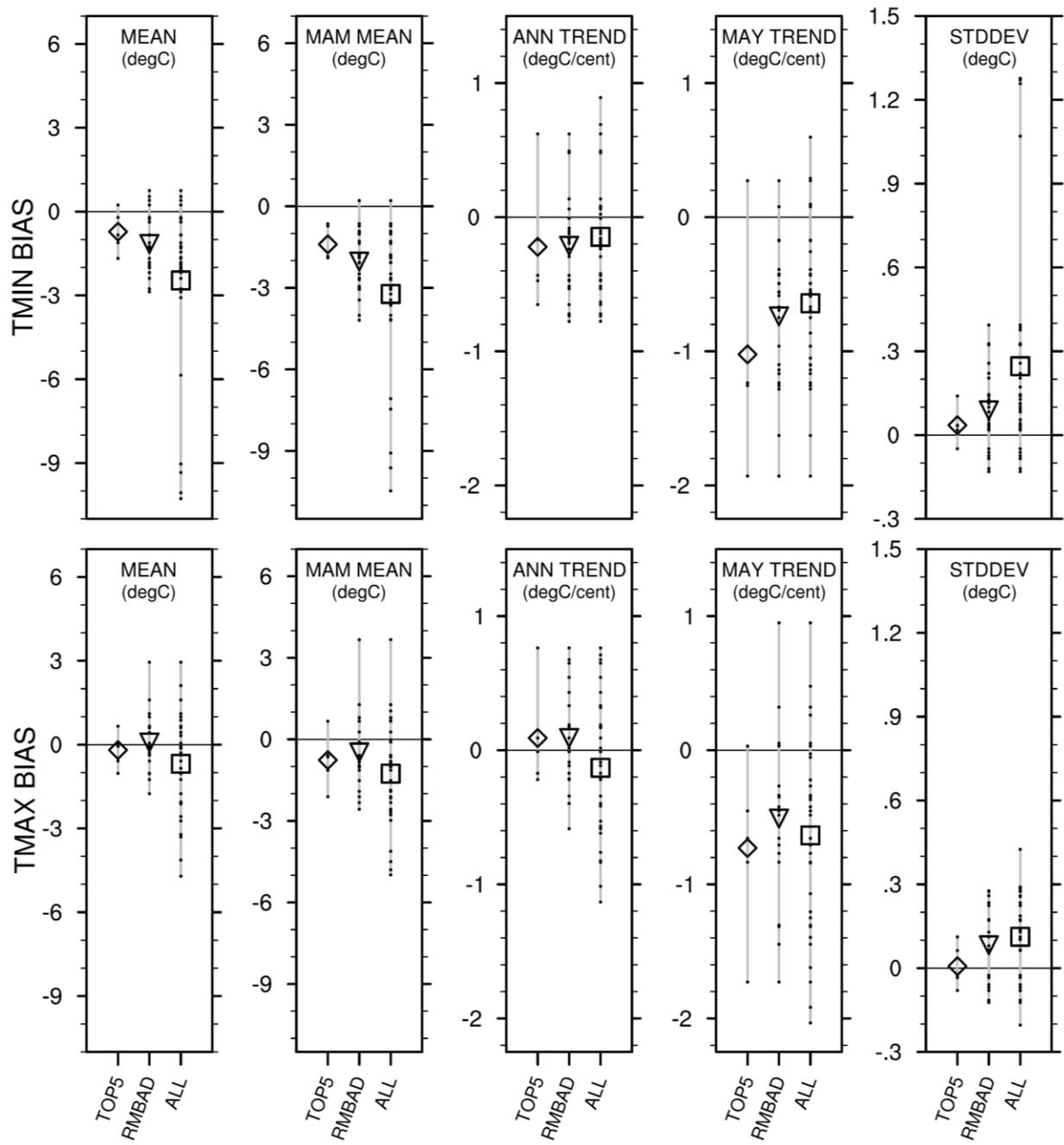
**Figure 6.** The time series standard deviation (a) where circles are the observations, dots along grey vertical model spread bars are individual models, and squares represent the average of all individual models. Color indicates models with variability that is significantly different than observations at the 95% confidence level, adjusted for serial lag-1 autocorrelation. (b) The detrended Tmax AK time series with mean removed for the 1960's, where the black line is the observed time series, the blue dashed line is the model with the smallest variability bias (ACC13) for the 106 year study period, and the green dashed line is the model with the largest variability bias (CMCCE). Gray shading shows the model spread.



**Figure 7.** Linear regression between 20<sup>th</sup> century biases for 42 models. (a) Regression between annual trend bias and max-month trend bias for the GP region. (b) Regression between max-month trend bias and variability bias for the US region. (c) Regression between seasonal average mean bias and variability bias for the SW region. The regression line is shown in solid black with the corresponding correlation coefficient,  $r$ , located at the top center of each plot. A second regression that excludes the 10% of models (4) with the largest magnitude x-axis variable bias is shown with a dashed line and the corresponding  $r$  value is in parentheses at the top right of each plot. Colored  $r$  values represent significance at the 95% confidence level, adjusted for reduced model independence.



**Figure 8.** Linear regression of 20<sup>th</sup> century bias to 21<sup>st</sup> century RCP 8.5 (a,b,c) and RCP 4.5 (d,e,f) projected change for the southwestern US using 33 models. Regression mean bias versus variability change (a,d), spring mean bias versus variability change (b,e), and annual trend bias vs annual trend change (c,f). The regression line is shown in solid black with the corresponding correlation coefficient,  $r$ , located at the top center of each plot. A second regression that excludes the 10% of models (3) with the largest magnitude x-axis variable bias is shown with a dashed line and the corresponding  $r$  value is in parentheses at the top right of each plot. Colored  $r$  values represent significance at the 95% confidence level, adjusted for reduced model independence.



**Figure 9.** Comparison of historical biases between the TOP5, RMBAD and ALL-model MMEs, for Tmin (top) and Tmax (bottom) using annual mean, March-April-May seasonal mean, annual trend, May trend, and standard deviation bias metrics. Large markers indicate the MME-average and individual models are shown with dots along grey model spread bars.

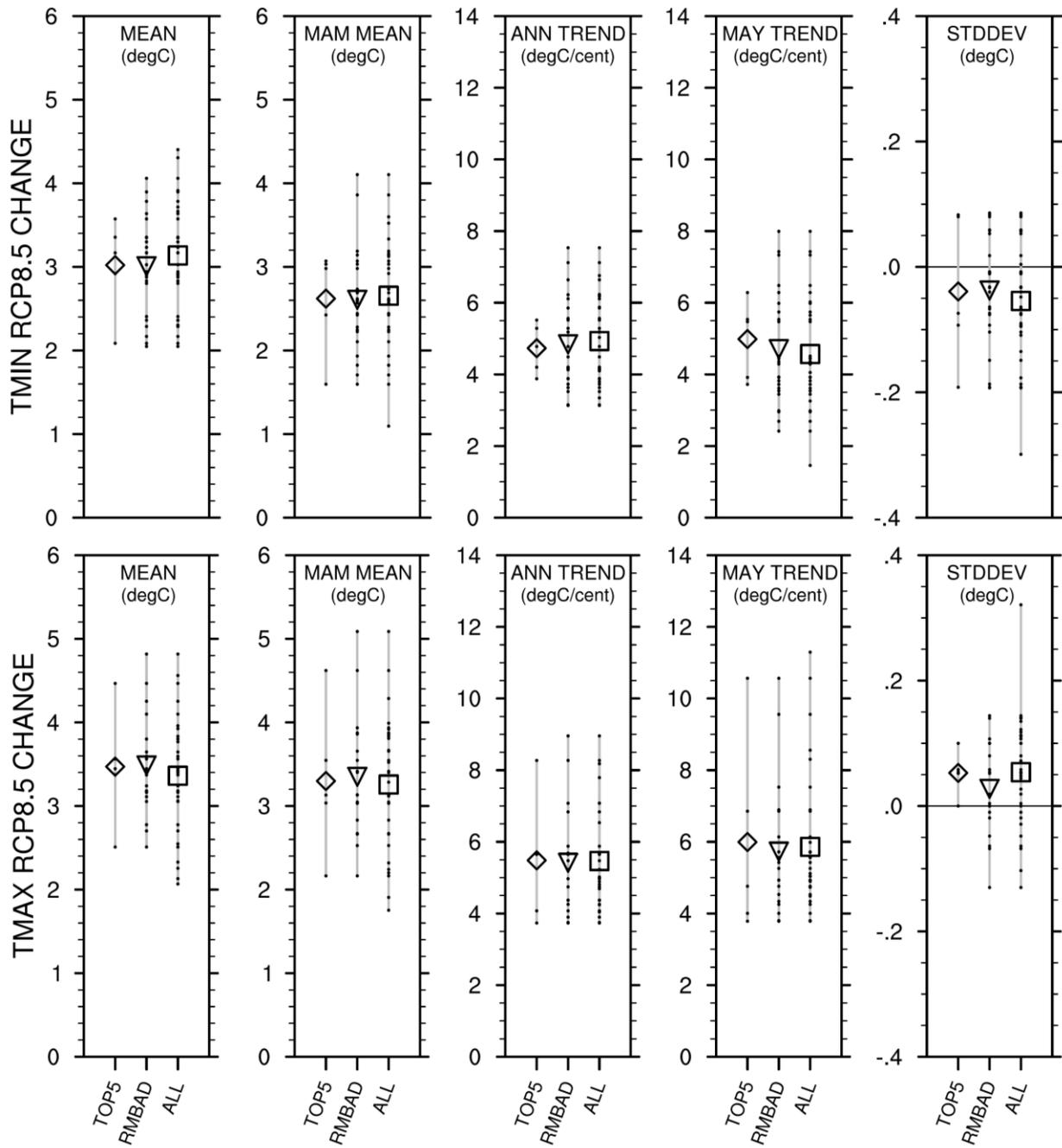


Figure 10. Same as Figure 9, except for RCP 8.5 projected change.