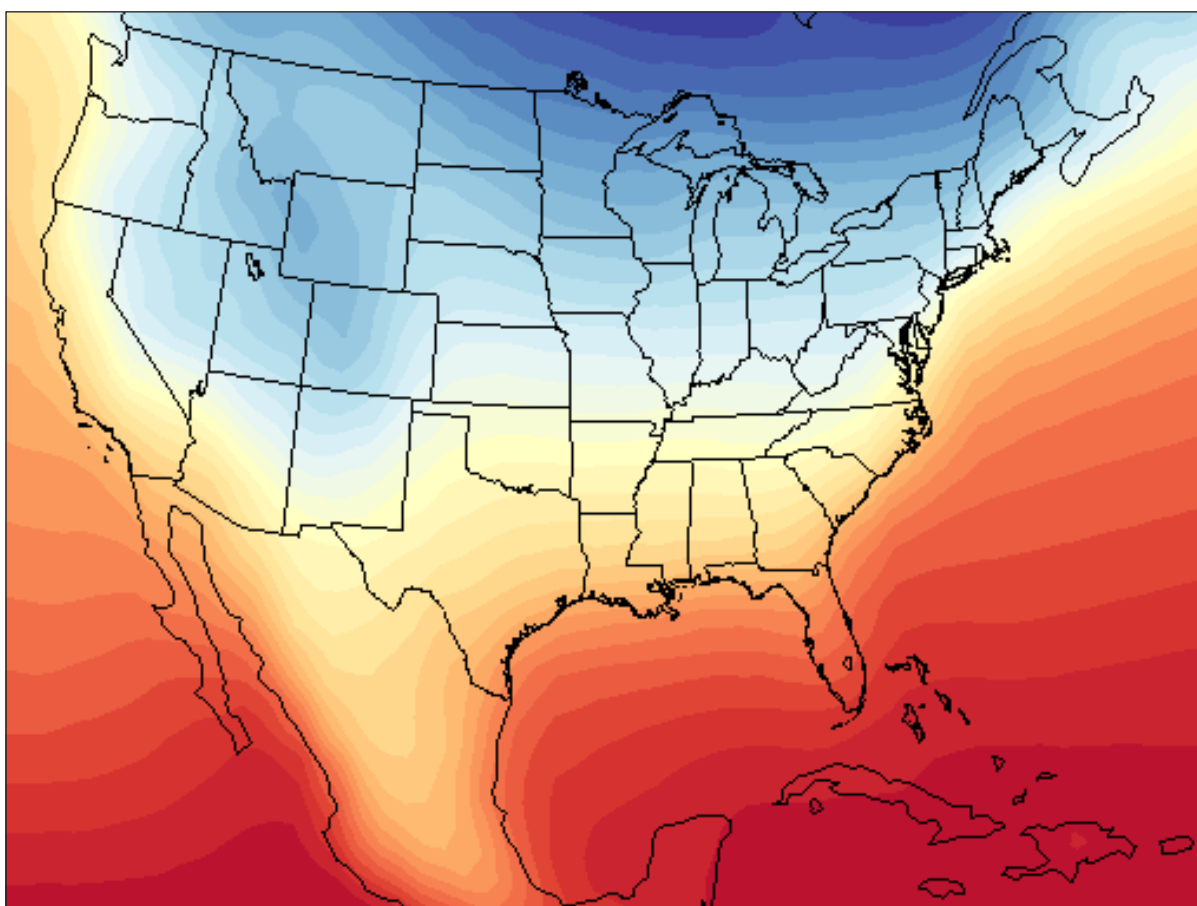


An Evaluation of 20th Century Climate for the Southeastern United States as Simulated by Coupled Model Intercomparison Project Phase 5 (CMIP5) Global Climate Models



Open-File Report 2016–1047

Cover. Map image showing mean winter temperature averaged over the global climate model simulations in CMIP5 (see figure 11C).

An Evaluation of 20th Century Climate for the Southeastern United States as Simulated by Coupled Model Intercomparison Project Phase 5 (CMIP5) Global Climate Models

By David E. Rupp, Oregon Climate Change Research Institute, College of Earth, Ocean, and
Atmospheric Sciences, Oregon State University, contractor to the U.S. Geological Survey (USGS),
work performed under USGS contract G13AC00407

Open-File Report 2016–1047

**U.S. Department of the Interior
U.S. Geological Survey**

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Director

U.S. Geological Survey, Reston, Virginia: 2016

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <http://www.usgs.gov> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <http://store.usgs.gov/>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Rupp, D.E., 2016, An evaluation of 20th century climate for the Southeastern United States as simulated by Coupled Model Intercomparison Project Phase 5 (CMIP5) global climate models: U.S. Geological Survey Open-File Report 2016–1047, 32 p., <http://dx.doi.org/10.3133/ofr20161047>.

ISSN 2331-1258 (online)

Contents

Abstract.....	1
1. Introduction.....	1
2. Data and Methods	2
2.1. Data	2
2.2. Performance Metrics	2
2.3. Model Ranking by Overall Performance	5
3. Results and Discussion of Model Ranking	7
Sensitivity to Observational Dataset	7
4. Model Performance by Metric	11
4.1. Climatologic Mean	11
4.2. Seasonal Cycle.....	11
4.3. Diurnal Temperature Range	14
4.4. Large-Scale Spatial Patterns.....	14
4.5. 20th Century Trend.....	14
4.6. Temporal Variability	14
4.7. Long-Term Persistence.....	25
4.8. ENSO Teleconnections	25
References Cited.....	31

Figures

1. Map showing the Southeastern United States as defined in this study	5
2. Graph showing relative error of the ensemble mean of each metric for each CMIP5 GCM	8
3. Graphs showing 41 CMIP5 GCMs ranked according to normalized error score from EOF analysis of performance metrics	9
4. Graphs showing loadings of the first four principal components from EOF analysis of 22 evaluation metrics and 41 CMIP5 GCMs	10
5. Graphs showing mean annual temperature and precipitation bias for 41 CMIP5 GCMs averaged over the Southeastern United States.....	12
6. Graphs showing mean seasonal cycle of temperature and relative precipitation averaged over the Southeastern United States	13
7. Graphs showing mean seasonal cycle amplitude in temperature and relative precipitation for the Southeastern United States	15
8. Maps showing mean seasonal cycle amplitude of temperature from CRU, ERA40, and the CMIP5 multi-model mean, and mean season cycle amplitude of relative precipitation from CRU, ERA40, and the CMIP5 multi-model mean	16
9. Graph showing mean seasonal cycle of diurnal temperature range averaged over the Southeastern United States	17
10. Graphs showing mean diurnal temperature range in winter and summer averaged over the Southeastern United States	18

11. Maps showing mean winter temperature from NCEP, ERA40, and the CMIP5 multi-model mean, and mean summer temperature from NCEP, ERA40, and the CMIP5 multi-model mean	19
12. Graphs showing normalized standard deviations and correlation coefficients by season for the climatological mean fields of temperature and precipitation from CMIP5	20
13. Maps showing mean winter precipitation from NCEP, ERA40, and the CMIP5 multi-model mean, and mean summer precipitation from NCEP, ERA40, and the CMIP5 multi-model mean	21
14. Graphs showing Southeastern United States-averaged trends in annual mean temperature and precipitation over the 20th century for all simulations and observations	22
15. Graphs showing variance of temperature anomalies and coefficient of variation of water year precipitation against temporal resolution of averaging for the Southeastern United States-averaged time series.....	23
16. Graphs showing standard deviation of temperature anomalies at temporal resolutions of 1 year and 8 years	24
17. Graphs showing coefficient of variation of precipitation at temporal resolutions of 1 water year and 8 water years	26
18. Graphs showing standard deviation of mean winter, spring, summer, and fall temperature anomalies.....	27
19. Graphs showing coefficient of variation of mean winter, spring, summer, and fall precipitation	28
20. Histogram of the Hurst exponent for Southeastern United States-averaged temperature in all CMIP5 simulations, and graph showing rescaled range against time scale calculated from the observations and simulations.....	29
21. Histogram of the Hurst exponent for Southeastern United States-averaged precipitation in all CMIP5 simulations, and graph showing rescaled range against time scale calculated from the observations and simulations.....	29
22. Graphs showing sensitivity to ENSO as the change in Southeastern United States winter temperature and precipitation for a one degree Celsius change in the Niño3.4 index averaged over November–March.....	30
23. Maps showing correlation of CRU winter temperature and precipitation with the Niño3.4 index averaged over November–March, and multi-model mean correlation of simulated winter temperature and precipitation to the same index.....	31

Tables

1. CMIP5 models used in this study and some of their attributes	3
2. Definitions of performance metrics, the confidence in the metrics for model ranking, and observational datasets used by the metrics.....	6
3. Loadings by metric of the leading five principal components	11

An Evaluation of 20th Century Climate for the Southeastern United States as Simulated by Coupled Model Intercomparison Project Phase 5 (CMIP5) Global Climate Models

By David E. Rupp, Oregon Climate Change Research Institute, College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, contractor to the U.S. Geological Survey (USGS), work performed under USGS contract G13AC00407

Abstract

The 20th century climate for the Southeastern United States and surrounding areas as simulated by global climate models used in the Coupled Model Intercomparison Project Phase 5 (CMIP5) was evaluated. A suite of statistics that characterize various aspects of the regional climate was calculated from both model simulations and observation-based datasets. CMIP5 global climate models were ranked by their ability to reproduce the observed climate. Differences in the performance of the models between regions of the United States (the Southeastern and Northwestern United States) warrant a regional-scale assessment of CMIP5 models.

1. Introduction

Simulations from the global climate models (GCMs) that are part of the Coupled Model Intercomparison Project Phase 5 (CMIP5) provide the basis for many of the conclusions in the recent Intergovernmental Panel on Climate Change (IPCC) Assessment Report #5 (AR5) (IPCC, 2013). The data from these simulations also are being used for local and regional climate change assessments across the United States either in direct form or following a statistical or dynamical downscaling transformation to bring the data to a spatial resolution compatible with local or regional impacts. In response, the question of GCM reliability (also termed fidelity, credibility, or veracity) at the regional scale is often raised by users of these data. To help address this question, CMIP5 GCMs were evaluated with respect to how well they reproduce the observed climate of the Southeastern United States.

Users should also consider how well GCMs reproduce properties of atmosphere, ocean, and land surface dynamics outside of the Southeastern United States and surrounding areas. Many assessments have already been done on various aspects of GCM performance. Chapter 9 of the IPCC AR5 (IPCC, 2013) provides a useful evaluation of many aspects of climate models, and Sheffield and others (2013) examined in particular the historical CMIP5 simulations over North America. Phillips and others (2014) calculated several atmospheric and ocean modes of variability (North Atlantic Oscillation, El Niño-Southern Oscillation, and Atlantic Multidecadal Oscillation, to name a few) as simulated by CMIP5 models and made their analysis models available to the public (http://webext.cgd.ucar.edu/Multi-Case/CVDP_ex/CMIP5-Historical/). Note that these are only a few examples from the large body of existing literature. The evaluation provided here is meant to augment the existing literature and provide information specifically about the Southeastern United States.

In this study, monthly temperature and precipitation data from 41 global climate models (GCMs) of the Coupled Model Intercomparison Project Phase 5 (CMIP5) were compared to observations for the 20th century, with a focus on the Southeastern United States and the surrounding areas. The methodology largely followed that used by Rupp and others (2013) for the Pacific Northwestern United States, for which a suite of statistics, or metrics, were calculated that characterize various aspects of the regional climate. Performance, or fidelity, was assessed on the basis of the GCMs' abilities to reproduce the observed metrics. GCMs were ranked in their fidelity using two methods. The first method simply treated all metrics equally. The second method considered two properties of the metrics: (1) redundancy of information (dependence) among metrics and (2) confidence in the reliability of an individual metric for accurately ranking models. Confidence was related to how robust the estimate of the metric was to ensemble size,

given that for most of the models only a small number of ensemble members (that is, realizations of the 20th century) were available.

The work in this study was done under U.S. Geological Survey (USGS) contract G13AC00407.

2. Data and Methods

The methodology used in the current study very closely follows Rupp and others (2013) but is repeated below for the reader's convenience. References to the methodology other than those changes stated explicitly in the text below should cite Rupp and others (2013) and not the current report.

2.1. Data

Simulated near surface temperature (T), daily minimum (Tmin) and daily maximum (Tmax) temperature, and precipitation rate (P) were acquired from 41 GCMs (see table 1) of the CMIP5 “historical” experiment (Taylor and others, 2012). The historical experiment included both natural and anthropogenic forcings for the years 1850–2005. For a given GCM, the number of members per ensemble varied from 1 to 10, differing only by initial conditions. Monthly data were obtained with the exception of Tmin and Tmax data for three GCMs (IPSL-CM5A-LR, IPSL-CM5A-MR, and IPSL-CM5B-MR), which had known problems with monthly mean Tmin and Tmax at the time the data were accessed from the CMIP5 archive. Monthly mean Tmin and Tmax data were calculated from daily data for these three GCMs.

For historical observations, we relied on five gridded datasets of monthly means of the following variables: near surface daily minimum, maximum, and average temperature, and surface precipitation rate. The observation-based datasets were

1. University of East Anglia Climatic Research Unit (CRU) TS3.10.01, $0.5^\circ \times 0.5^\circ$, 1901–2009 (Harris and others, 2014), accessed January 13, 2013, at <https://crudata.uea.ac.uk/cru/data/hrg/>.
2. Parameter-elevation Regressions on Independent Slopes Model (PRISM), $2.5' \times 2.5'$, 1895–2012 (Daly and others, 2008), accessed March 17, 2014, at <http://www.prism.oregonstate.edu>.
3. University of Delaware Air Temperature and Precipitation (UDelaware) v3.01, $0.5^\circ \times 0.5^\circ$, 1901–2010 (Matsuura and Willmott, 2012a, b), accessed July 14, 2013, at http://climate.geog.udel.edu/~climate/html_pages/download.html#ghcn_T_P_clim.
4. National Center for Environmental Prediction/National Center for Atmospheric Research Reanalysis (NCEP), $\sim 1.9^\circ \times 1.9^\circ$, 1948–2012 (Kalnay and others, 1996), accessed February 29, 2013, at <http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.surface.html>.
5. European Centre for Medium-Range Weather Forecasts 40 Year Re-analysis (ERA40), $\sim 2.5^\circ \times 2.5^\circ$ mid-1957 to mid-2002 (Uppala and others, 2005), accessed July 28, 2013, at <http://www.ecmwf.int/en/research/climate-reanalysis/browse-reanalysis-datasets/>.

CRU, PRISM, and UDelaware datasets are based on surface station observations, and NCEP and ERA40 are reanalysis datasets that are based on a numerical model of the atmosphere that assimilates observations to update model states.

CRU, UDelaware, NCEP, ERA40, and CMIP5 datasets were regridded to a common resolution of $1^\circ \times 1^\circ$ using an inverse-distance-weighting interpolation algorithm. PRISM datasets were regridded by averaging all native cells within the coarser $1^\circ \times 1^\circ$ cell. Grid cell centers were located on the whole degree.

2.2. Performance Metrics

The metrics used in the study consider properties of both the regionally averaged time series and larger-scale patterns having regional influence. The following metrics of temperature and precipitation were selected on the basis of having theoretical merits as well as being relevant for modeling the impacts of climate variability and change:

1. Climatological mean of annual value (Mean).
2. Mean seasonal amplitude (SeasonAmp).
3. Spatial standard deviation (SpaceSD) of the climatological mean field, by season.
4. Spatial correlation (SpaceCor) of the observed to modeled climatological mean fields, by season.
5. Linear trend of annual values (Trend).
6. Time series variance (TimeVar) of temperature and coefficient of variation (TimeCV) of precipitation: Calculated at frequencies ranging from 1 to 10 years.
7. Time series variance (TimeVar) of temperature and coefficient of variation (TimeCV) of precipitation of seasonal means.
8. Persistence (Hurst) measured using the Hurst exponent.
9. Strength of El Niño Southern Oscillation (ENSO) teleconnection in winter.

Also, for temperature only, one additional metric was calculated:

10. Mean diurnal temperature range (DTR), by season.

The above metrics are identical to those in Rupp and others (2013) except for the seventh metric listed, which was added after consultation with the USGS Southeast Climate Science

Table 1. CMIP5 models used in this study and some of their attributes.

[T, temperature; P, precipitation; Tmin, daily minimum; Tmax, daily maximum]

Model	Center	Number of ensemble members: T/ P/ Tmin/ Tmax/	Atmospheric resolution (longitude x latitude)	Vertical levels in atmosphere
BCC-CSM1-1	Beijing Climate Center, China Meteorological Administration	3/ 3/ 3/ 3	2.8x2.8	26
BCC-CSM1-1-M	Beijing Climate Center, China Meteorological Administration	3/ 3/ 3/ 3	1.12x1.12	26
BNU-ESM	College of Global Change and Earth System Science, Beijing Normal University, China	1/ 1/ 1/ 1	2.8x2.8	26
CanESM2	Canadian Centre for Climate Modeling and Analysis	5/ 5/ 5/ 5	2.8x2.8	35
CCSM4	National Center of Atmospheric Research, USA	6/ 6/ 6/ 6	1.25x0.94	26
CESM1-BGC	Community Earth System Model Contributors	1/ 1/ 1/ 1	1.25x0.94	26
CESM1-CAM5	Community Earth System Model Contributors	3/ 3/ 3/ 3	1.25x0.94	26
CESM1-FASTCHEM	Community Earth System Model Contributors	3/ 3/ 3/ 3	1.25x0.94	26
CESM1-WACCM	Community Earth System Model Contributors	1/ 1/ 1/ 1	2.5x1.89	66
CMCC-CESM	Centro Euro-Mediterraneo per I Cambiamenti Climatici	1/ 1/ 1/ 1	3.75x3.71	39
CMCC-CM	Centro Euro-Mediterraneo per I Cambiamenti Climatici	1/ 1/ 1/ 1	0.75x0.75	31
CMCC-CMS	Centro Euro-Mediterraneo per I Cambiamenti Climatici	1/ 1/ 1/ 1	1.88x1.87	95
CNRM-CM5	National Centre of Meteorological Research, France	10/ 10/ 10/ 10	1.4x1.4	31
CNRM-CM5-2	National Centre of Meteorological Research, France	1/ 1/ 1/ 1	1.4x1.4	31
CSIRO-Mk3-6-0	Commonwealth Scientific and Industrial Research Organization/Queensland Climate Change Centre of Excellence, Australia	10/ 10/ 10/ 10	1.8x1.8	18
EC-EARTH	EC-EARTH consortium	5/ 7/ 4/ 4	1.13x1.12	62
FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences	5/ 5/ 5/ 5	2.8x2.8	26
FGOALS-s2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences	3/ 3/ 3/ 3	2.8x1.7	26
FIO-ESM	The First Institute of Oceanography, SOA, China	3/ 3/ 3/ 3	2.81x2.79	26
GFDL-CM3	NOAA Geophysical Fluid Dynamics Laboratory, USA	5/ 5/ 5/ 5	2.5x2.0	48
GFDL-ESM2G	NOAA Geophysical Fluid Dynamics Laboratory, USA	3/ 3/ 1/ 1	2.5x2.0	48

4 An Evaluation of 20th Century Climate for the Southeastern United States

Table 1. CMIP5 models used in this study and some of their attributes.—Continued

[T, temperature; P, precipitation; Tmin, daily minimum; Tmax, daily maximum]

Model	Center	Number of ensemble members: T/ P/ Tmin/ Tmax/	Atmospheric resolution (longitude x latitude)	Vertical levels in atmosphere
GFDL-ESM2M	NOAA Geophysical Fluid Dynamics Laboratory, USA	1/ 1/ 1/ 1	2.5x2.0	48
GISS-E2-H	NASA Goddard Institute for Space Studies, USA	5/ 5/ 5/ 5	2.5x2.0	40
GISS-E2-H-CC	NASA Goddard Institute for Space Studies, USA	1/ 1/ 1/ 1	2.5x2.0	40
GISS-E2-R	NASA Goddard Institute for Space Studies, USA	3/ 3/ 3/ 3	2.5x2.0	40
GISS-E2-H-CC	NASA Goddard Institute for Space Studies, USA	1/ 1/ 1/ 1	2.5x2.0	40
HadCM3	Met Office Hadley Center, UK	10/ 10/ 10/ 10	3.75x2.5	19
HadGEM2-AO	Met Office Hadley Center, UK	1/ 1/ 1/ 1	1.88x1.25	38
HadGEM2-CC	Met Office Hadley Center, UK	1/ 1/ 1/ 1	1.88x1.25	60
HadGEM2-ES	Met Office Hadley Center, UK	5/ 5/ 5/ 5	1.88x1.25	38
INMCM4	Institute for Numerical Mathematics, Russia	1/ 1/ 1/ 1	2.0x1.5	21
IPSL-CM5A-LR	Institut Pierre Simon Laplace, France	6/ 6/ 1/ 1	3.75x1.8	39
IPSL-CM5A-MR	Institut Pierre Simon Laplace, France	3/ 3/ 1/ 1	2.5x1.25	39
IPSL-CM5B-LR	Institut Pierre Simon Laplace, France	1/ 1/ 1/ 1	3.75x1.8	39
MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	5/ 5/ 5/ 5	1.4x1.4	40
MIROC-ESM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	3/ 3/ 3/ 3	2.8x2.8	80
MIROC-ESM-CHEM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	1/ 1/ 1/ 1	2.8x2.8	80
MPI-ESM-LR	Max Planck Institute for Meteorology, Germany	3/ 3/ 3/ 3	1.88x1.87	47
MPI-ESM-MR	Max Planck Institute for Meteorology, Germany	3/ 3/ 3/ 3	1.88x1.87	95
MRI-CGCM3	Meteorological Research Institute, Japan	5/ 5/ 5/ 5	1.1x1.1	48
NorESM1-M	Norwegian Climate Center, Norway	3/ 3/ 3/ 3	2.5x1.9	26

Center. A full list of the metrics, along with the observational datasets used to evaluate each metric, is given in table 2.

Most metrics were evaluated as spatial averages over the entire Southeastern United States, defined here as the land area shown in figure 1. However, because the climate of the Southeastern United States is affected by large-scale oceanic and atmospheric patterns, the spatial variance and correlation metrics were examined over a larger domain (115°W – 50°W, 15°N – 55°N). This expanded domain covers a large portion of North America and the northwestern Atlantic Ocean.

Several metrics (Mean, SeasonAmp, SpaceSD, SpaceCor) were calculated over the latter four decades of the 20th century (1960–99), and DTR was calculated over 1950–99 in order to include the shorter NCEP and ERA40 datasets in the analysis. However, those metrics that are more sensitive to record length (that is, those that do not simply describe the mean state of the time series) were calculated over the 20th century (1901–99) and consequently only for CRU, PRISM, and UDelaware.

In addition to calculating each metric for each ensemble member of each model, we also calculated a “multi-model mean” value. Given that models have different numbers of ensemble members, those models with larger ensembles will generally give a more reliable estimate (that is, a smaller standard error) of a particular statistic than those models with smaller ensembles. For simplicity, however, we gave each model equal weight when calculating a multi-model mean.

2.3. Model Ranking by Overall Performance

A large number of metrics may help to elucidate the different strengths and weaknesses of models. At the same time, a suite of metrics also presents a challenge for selecting a subset of more credible models, for at least two reasons. For one, some metrics may be more relevant than others for a particular application, and the rankings may depend on which

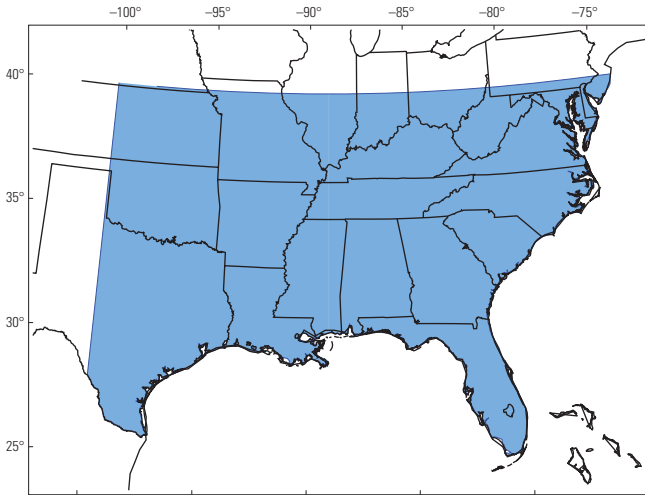


Figure 1. The Southeastern United States as defined in this study.

set of metrics are applied (for example, Pierce and others, 2009). For another, there may be redundancy among metrics given that not all are independent. In either case, treating all metrics equally might be inadvisable. We, therefore, applied two methods for ranking the models, as described below. The first simply treated all metrics equally; the second did not.

The first method included all performance metrics and assigned equal weight to each metric. For a given model i and metric j , we defined an error $E_{i,j}$ as

$$E_{i,j} = |x_{obs,j} - x_{i,j}| \quad (1)$$

where x_{obs} and x_i are the observed and simulated ensemble mean metric, respectively. For x_{obs} , we used the mean of the ensemble of observations, where more than one observed dataset was examined. Application of equation 1 included correlations (where x_{obs} necessarily equaled 1). Furthermore, we defined a relative error $E_{i,j}^*$ as

$$E_{i,j}^* = \frac{E_{i,j} - \min(E_{i,j})}{\max(E_{i,j}) - \min(E_{i,j})} \quad (2)$$

and then summed the relative error across all m metrics

$$E_{i,tot}^* = \sum_{j=1}^m E_{i,j}^* \quad (3)$$

to get the total relative error $E_{i,tot}^*$ per model. Ordering the models by their respective total relative error determined the ranking.

The second method for ranking the models took into account both the redundancy in information among metrics and the confidence in the rankings of the individual metrics. To address the latter, we first excluded those metrics that were identified as not being robust. This exclusion of metrics is described in detail in section 3.2 of Rupp and others (2013). Briefly, those metrics that show high *intra*-model (that is, *intra*-ensemble) spread relative to *inter*-model spread were identified as not being robust metrics. We defined four categories of robustness, or confidence, in rankings: “highest,” “higher,” “lower,” “lowest.” Table 2 lists the category within which each metric falls. Those metrics categorized as “lowest” were excluded from the following analysis. Also, so as not to so heavily weight those metrics calculated for each of four seasons, we used only the winter (DJF for December, January, February) and summer (JJA for June, July, August) values.

To address the matter of information redundancy, we conducted an empirical orthogonal function (EOF) analysis on the remaining metrics. This allowed us to reduce the large number of metrics, some of which co-vary and others of which add little information, down to a reduced number of orthogonal

Table 2. Definitions of performance metrics, the confidence in the metrics for model ranking, and observational datasets used by the metrics.

Metric ^a	Confidence category	Description	Observational datasets
Mean-T Mean-P	Highest Highest	Mean annual temperature (T) and precipitation (P), 1960–99.	CRU, PRISM, UDelaware, ERA40 ^b , NCEP ^b
DTR-MMM ^c	Highest	Mean diurnal temperature range, 1950–99.	CRU ^d , PRISM ^d , NCEP
SeasonAmp-T SeasonAmp-P	Highest Higher	Mean amplitude of seasonal cycle as the difference between warmest and coldest month (T) or wettest and driest month (P). Monthly precipitation calculated as percentage of mean annual total, 1960–99.	CRU, PRISM, UDelaware, ERA40 ^b , NCEP ^b
SpaceCor-MMM-T ^{c,e} SpaceCor-MMM-P ^{c,e}	Highest Higher	Correlation of simulated with observed the mean spatial pattern, 1960–99.	ERA40, NCEP ^d
SpaceSD-MMM-T ^{c,e} SpaceSD-MMM-P ^{c,e}	Highest Higher	Standard deviation of the mean spatial pattern, 1960–99. All standard deviations are normalized by the standard deviation of the observed pattern.	ERA40, NCEP ^d
TimeVar.1-T ^f TimeVar.8-T ^f	Lower Lowest	Variance of temperature calculated at frequencies (time periods of aggregation) ranging for $N = 1$ and 8 years, 1901–99.	CRU, PRISM, UDelaware
TimeCV.1-P ^g TimeCV.8-P ^g	Lower Lowest	Coefficient of variation (CV) of precipitation calculated at frequencies (time periods of aggregation) ranging for $N = 1$ and 8 water years, 1902–99.	CRU, PRISM, UDelaware
TimeVar-MMM-T ^c	Lower	Variance of seasonal mean temperature, 1901–99.	CRU, PRISM, UDelaware
TimeCV-MMM-P ^c	Lower	Coefficient of variation of seasonal mean precipitation, 1901–99.	CRU, PRISM, UDelaware
Trend-T Trend-P	Lower Lowest	Linear trend of annual temperature and precipitation, 1901–99.	CRU, PRISM, UDelaware
ENSO-T ENSO-P	Lower Lowest	Correlation of winter temperature and precipitation with Niño3.4 index, 1901–99.	CRU, PRISM, UDelaware
Hurst-T Hurst-P	Lowest Lowest	Hurst exponent using monthly difference anomalies (T) or fractional anomalies (P), 1901–99.	CRU, PRISM, UDelaware

^aUnless otherwise noted, metrics are averaged over the Southeastern United States.^bTemperature only used in ranking, not precipitation.^cMMM is the season designation: DJF (December, January, February), MAM (March, April, May), JJA (June, July, August), and SON (September, October, November).^dNot used in ranking.^eExpanded domain: 115°W – 50°W, 15°N – 55°N.^fTemperature calculations are for calendar years.^gPrecipitation calculations are for water years, which is the period from October 1 to September 30 and is designated by the year in which the period ends.

and more consequential metrics. We treated the observations as if they were from another model, such that the EOF analysis was done on all values x , which include both observed (x_{obs}) and simulated (x_i) values for each metric. Note that for the EOF analysis we normalized the metric values by subtracting their mean and dividing by their standard deviation.

The leading EOFs provide a greatly reduced number of new orthogonal metrics, which can be examined separately. However, to arrive at a single metric from which to rank overall model performance, we simply calculated the Euclidean distance from the observations to each modeled value in EOF space across all dimensions of the leading EOFs. We used this distance as the overall error score per GCM and normalized it to range from 0 (least error) to 1 (most error).

3. Results and Discussion of Model Ranking

In this section we focus on the results of the model ranking. A discussion of the models' performance for each metric is provided in section 4.

The ranking of models using the simple method on all 42 metrics is given in figure 2. Also shown are the relative errors for the individual metrics. Each model scored well in at least one metric, and several models scored poorly in only a few metrics. Overall, the highest ranked models include CNRM-CM5/CNRM-CM5-2 pair of models, the CESM1/CCSM4 family of models (with the exception of CESM1-WACCM), and the CMCC-CM/CMCC-CMS pair of models. Other high scoring models are MPI-ESM-LR, the "CC" versions of the GISS family of models, and HadGEM2-ES. There were some differences between the models that scored near the top for the Southeastern United States and those that scored best for the Pacific Northwest United States as reported in Rupp and others (2013). For example, while the CESM1/CCSM4 and CNRM-CM5 families fared well in both regions, the GISS family of models scored well in the Southeast but poorly overall in the Pacific Northwest.

From the EOF analysis on 22 of the full 42 metrics, the leading 5 principal components (PC) cumulatively explained 22, 40, 52, 62, and 69 percent of the variance, respectively. The models, ranked in order using the first five PCs, are shown in figure 3A. Figure 3A also shows the effect of using only the first two or four PCs. In a few cases, using only the first two PCs makes a large change to a model's ranking (for example, see CMCC-CESM in figure 3), which illustrates the sensitivity in ranking to how model metrics are weighted.

To examine the influence of the additional metrics added for this evaluation of the Southeastern United States (TimeVar-DJF-T, TimeVar-JJA-T, TimeCV-DJF-P, TimeCV-JJA-P), we reran the EOF analysis with the same 18 metrics used in Rupp and others (2013). The inclusion of the additional metrics had only minor effect on the ranking of the models (compare figure 3A and 3B). For example, CMCC-CM and GISS-E2-H-CC

occupied the top 2 positions using 22 metrics, but were in fifth and fourth positions, respectively, when the set of 18 metrics were used.

Comparing the results from our initial simpler ranking to the more complex EOF-based analysis reveals minor differences. Nearly no models occupied precisely the same position in each method, but the general order was similar. For example, of the top 12 models resulting from the EOF method, 9 placed in the top 12 using the simple ranking method.

How the models scored in EOF "space" is shown in figure 4 for the four leading PCs (PC2 versus PC1 and PC4 versus PC3). The loadings, or weights, given to each metric within each of the leading five PCs is given in table 3, where the metrics given the most weighting are highlighted. PC1 placed more weight on metrics related to temporal variability, though annual mean temperature and precipitation were also weighted heavily. In contrast, PC2 was more controlled by metrics quantifying the larger scale spatial patterns similar to that seen by Rupp and others (2013) for the Pacific Northwest.

Sensitivity to Observational Dataset

It is important to note that we have not quantified the sensitivity of the overall rankings to the choice of observational dataset, but simply averaged over observational datasets when more than one dataset was considered. However, we expect that the effect of observational dataset selection is not insignificant given that the spread among observational datasets is large when compared to intermodel differences for a few of the metrics. As discussed further in section 4, the metrics with noteworthy spread among the observational datasets are the mean annual precipitation (Mean-P), the amplitude of the seasonal cycle of precipitation (SeasonAmp-P), the diurnal temperature range, most notably in summer (DTR-JJA), and the spatial pattern of precipitation, again most notably in summer (SpaceCorr-JJA-P, SpaceVar-JJA-P). By far the largest differences occur between the station-based datasets (CRU, PRISM, and UDelaware) and the reanalysis datasets (ERA40 and NCEP), though sizable discrepancies also exist between ERA40 and NCEP for summer precipitation. If we assume that the station-based datasets provide more reliable estimates of precipitation, this raises the question of whether reanalysis datasets should be used to evaluate precipitation from climate models in regions where gridded station-based data of high quality (that is, high station density and lengthy records) are available. In fact, in this study we departed from Rupp and others (2013) in that we excluded ERA40 and NCEP when taking the average of observation datasets for the regionally averaged metrics Mean-P and SeasonAmp-P, relying solely on CRU, PRISM, and UDelaware. However, we still used ERA40 and NCEP for evaluating large-scale spatial patterns that cover both land and ocean.

In regard to DTR, it is worth noting that near-coast grid cells of the Southeastern United States domain contained some influence of ocean cells in both the models and in NCEP,

unlike CRU and PRISM. We might expect this ocean influence to have suppressed DTR somewhat in both the models and NCEP as compared to CRU and PRISM; in fact, DTR from the NCEP and the multimodel ensemble mean was 2–3 degrees Celsius ($^{\circ}\text{C}$) lower than CRU and PRISM. Rupp and others (2013) noted this and, as a test, removed the coastal grid cells from their domain and recalculated the regionally averaged DTR. While they found that this slightly increased DTR across all datasets, it only negligibly affected the relative values of DTR, thus the ranking of models based on DTR alone did not change. Also, the discrepancy between NCEP and CRU/PRISM remained unchanged, implying that the differences among observational datasets were not related to the influence of ocean cells. We conducted a similar test for the Southeastern region, excluding the $1^{\circ}\times 1^{\circ}$ grid cells near the Atlantic coastline and all grid cells south of 32° , therefore excluding Florida and southern Texas entirely. Similar to Rupp

and others (2013), we found that whereas reducing the domain slightly increased DTR, it did not markedly affect the seasonal variations shown by the models and observational datasets.

The results imply other causes for the large differences between the station- and reanalysis-based values of DTR. One factor may be that observations of T_{min} and T_{max} are relatively instantaneous, whereas simulated T_{min} and T_{max} have been averaged over some time step that varies by GCM and reanalysis dataset and, therefore, are effectively biased toward a lower DTR. If this time-average smoothing is a major factor, then it is flawed to expect the modeled values to match the station-based instantaneous values; in effect, the reanalysis becomes a better reference against which to judge a model's ability to simulate DTR. In fact, when determining the above rankings, we used NCEP as the observational dataset to calculate the error in DTR, unlike Rupp and others (2013), who used the average of NCEP, CRU, and PRISM.

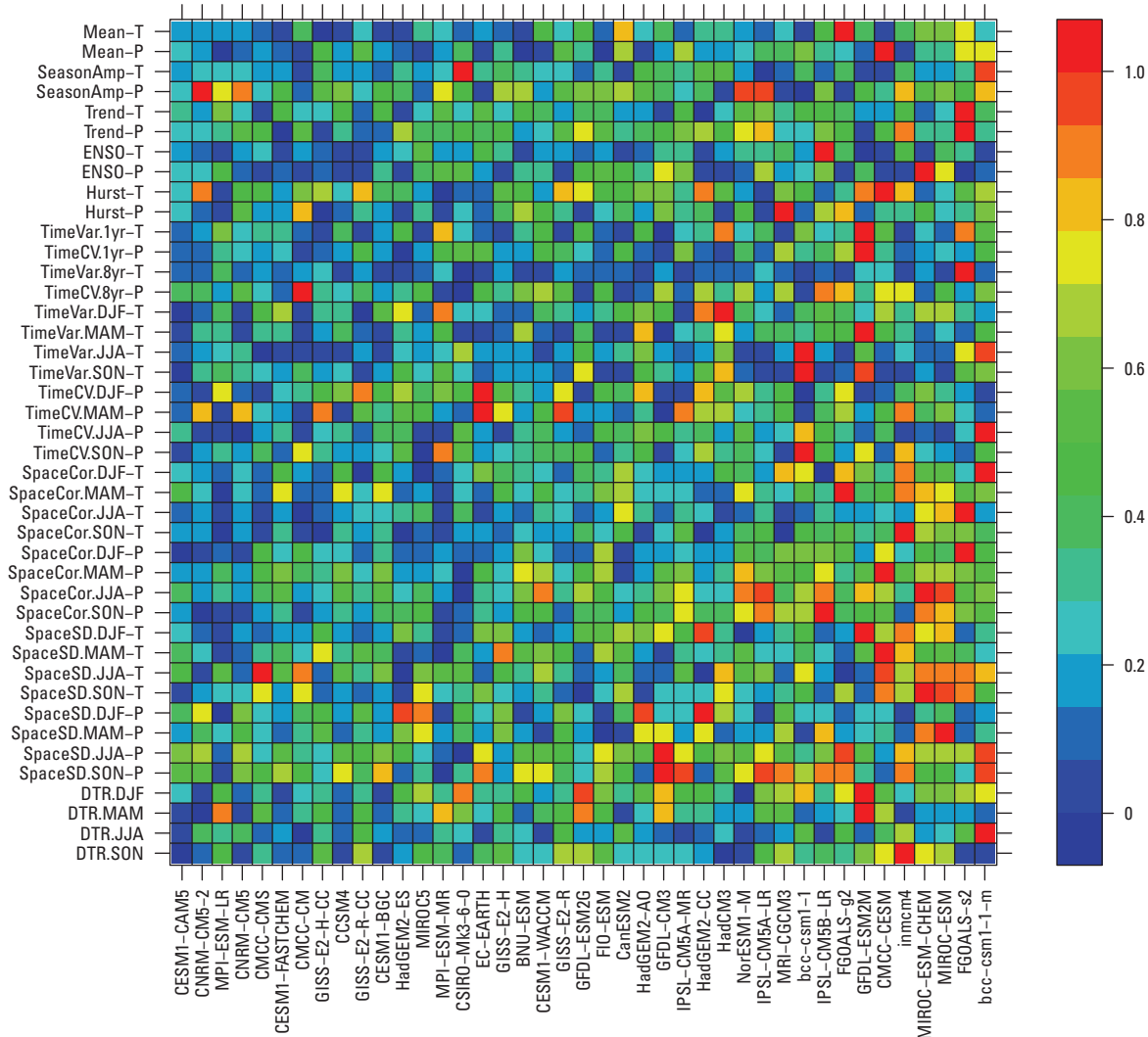


Figure 2. Relative error of the ensemble mean of each metric for each CMIP5 GCM. Models are ordered from least (left) to most (right) total relative error, where total relative error is the sum of relative errors from all metrics.

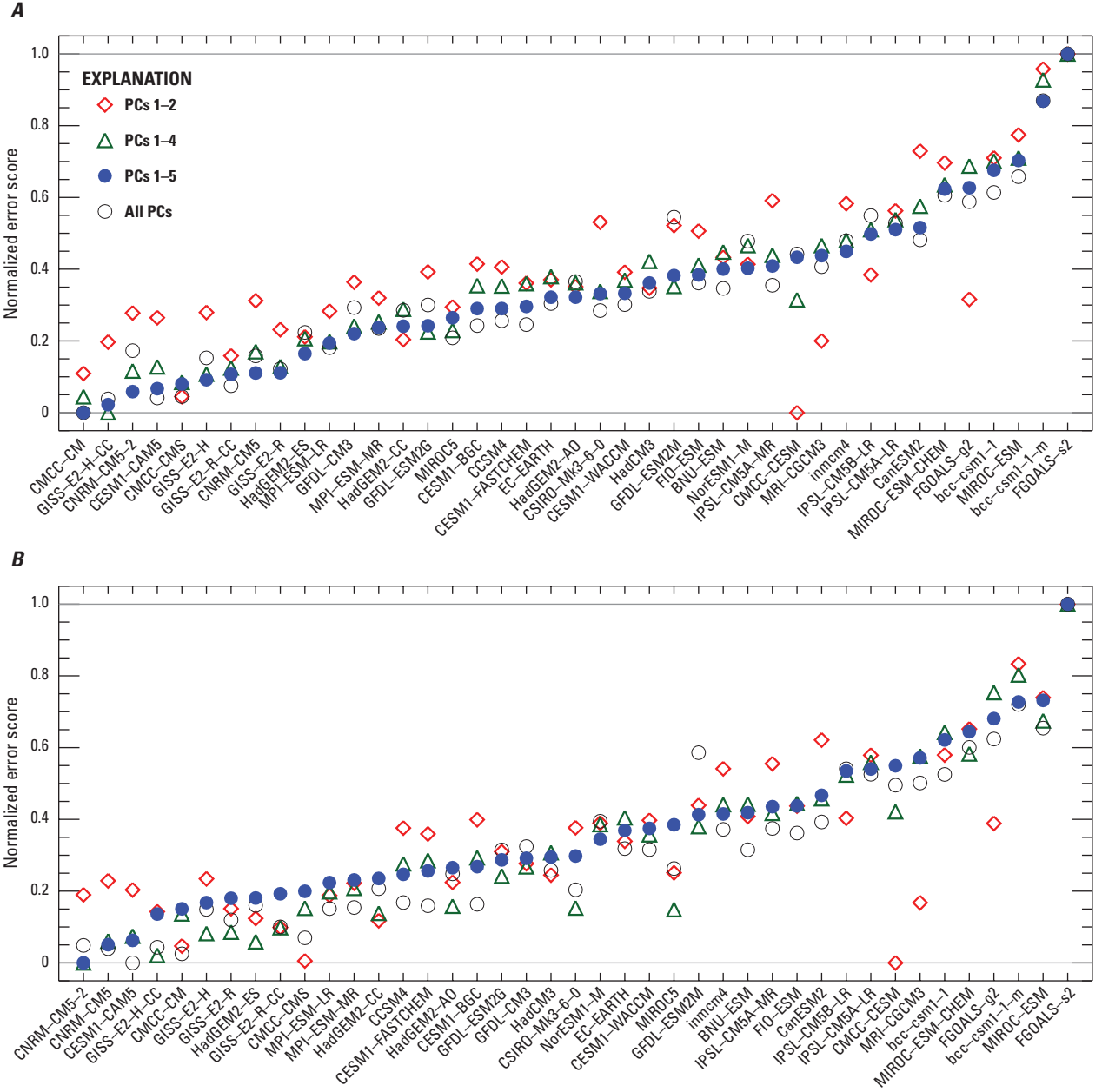


Figure 3. Forty-one CMIP5 GCMs ranked according to normalized error score from EOF analysis of performance metrics. Ranking is based on the first five principal components (filled blue circles). The open symbols show the models' error scores using the first 2, 4, and all 22 principal components (PCs). The best scoring model has a normalized error score of 0. Ranks are plotted using *A*, the 22 metrics and observational datasets described in this study, and *B*, the 18 metrics and observational datasets used in Rupp and others (2013).

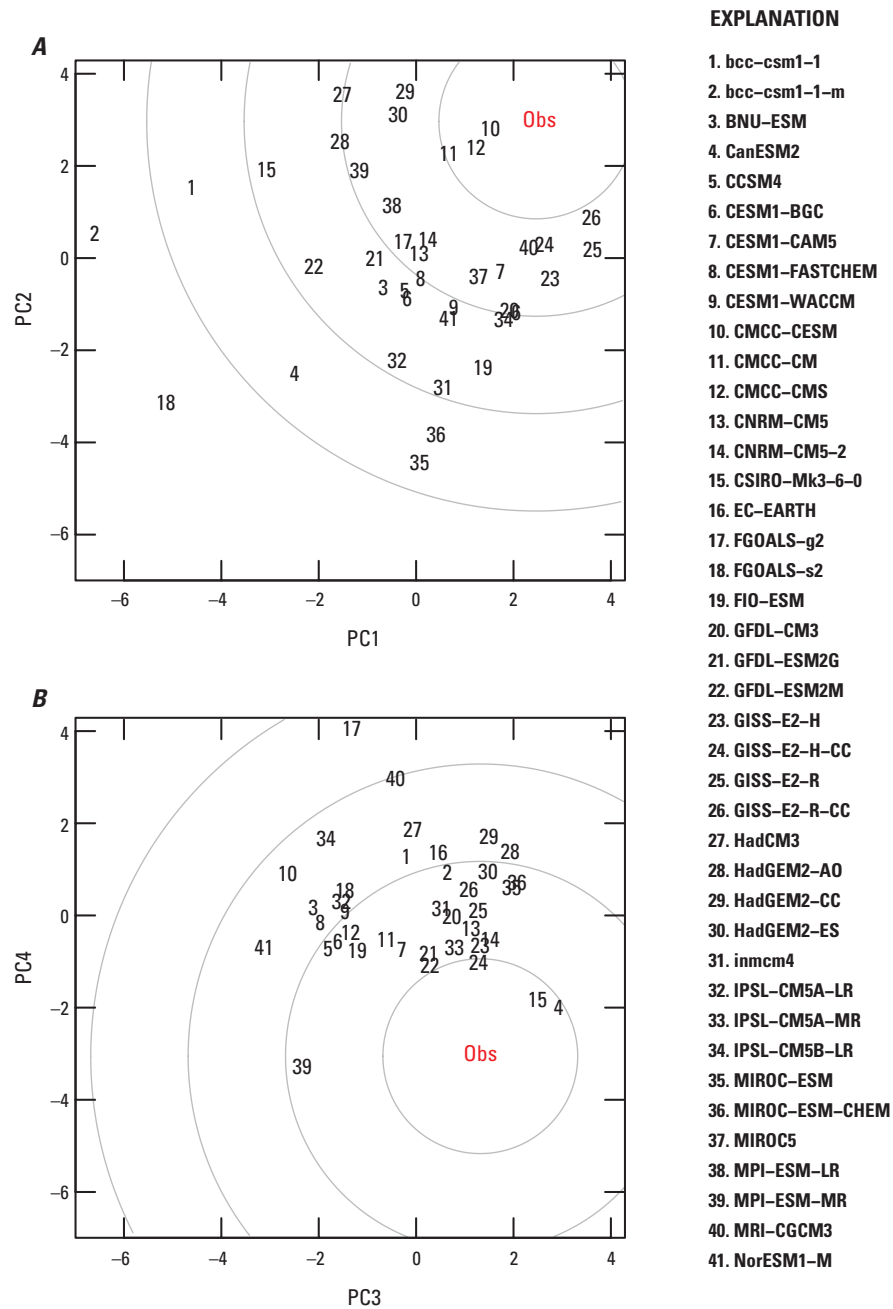


Figure 4. Loadings of the first four principal components (PCs) from EOF analysis of 22 evaluation metrics and 41 CMIP5 GCMs: A, PC2 versus PC1 and B, PC4 versus PC3. “Obs” indicates the observation dataset.

Table 3. Loadings by metric of the leading five principal components (PCs). Absolute values of loadings greater than 0.20 and 0.30 have been shaded in yellow and red, respectively.

Metric	PC 1	PC 2	PC 3	PC 4	PC 5
Mean-T	0.34	0.20	-0.05	0.12	0.17
Mean-P	-0.33	0.19	0.12	0.16	-0.09
SeasonAmp-T	-0.08	-0.17	-0.04	-0.26	-0.12
SeasonAmp-P	-0.18	-0.24	-0.35	0.00	-0.01
Trend-T	0.06	-0.20	0.11	0.10	0.14
ENSO-T	-0.28	0.19	-0.27	0.00	0.14
TimeVar.1-T	-0.31	0.05	0.12	-0.36	0.13
TimeCV.1-P	-0.10	0.30	-0.34	0.14	-0.06
TimeVar-DJF-T	-0.16	-0.04	-0.29	-0.45	-0.07
TimeCV-DJF-P	-0.40	0.12	0.07	0.00	0.01
TimeVar-JJA-T	-0.31	0.15	0.33	-0.01	0.04
TimeCV-JJA-P	0.19	0.15	0.01	-0.29	0.22
DTR-DJF	0.15	0.28	-0.19	-0.10	-0.27
DTR-JJA	0.12	0.14	0.34	-0.31	-0.28
SpaceCor-DJF-T	0.08	0.28	0.22	-0.21	-0.26
SpaceCor-JJA-T	-0.07	0.35	-0.20	0.22	-0.07
SpaceCor-DJF-P	-0.10	0.39	-0.07	-0.02	0.10
SpaceCor-JJA-P	0.07	0.20	0.38	0.21	0.05
SpaceSD-DJF-T	0.08	0.28	-0.03	-0.30	0.34
SpaceSD-JJA-T	0.16	-0.01	-0.13	-0.17	-0.55
SpaceSD-DJF-P	-0.32	-0.09	0.13	0.18	-0.37
SpaceSD-JJA-P	0.34	0.20	-0.05	0.12	0.17

4. Model Performance by Metric

4.1. Climatologic Mean

The simulated mean annual temperature of the South-eastern United States ranged by 7 °C from the coolest to the warmest model (fig. 5A), though the coolest model was a full 2 °C cooler than the next coolest model. The observational datasets differed only slightly among themselves, with a range of about 0.5 °C between the warmest and coolest datasets. Taken as an average, the five observational datasets were less than 0.5 °C warmer than the median of the simulated mean annual temperatures. (Note: Henceforth the observed values will be reported as the average of the observational datasets used, unless specifically stated otherwise.)

For mean annual precipitation, the range across models was large: 65 centimeters per year (cm year⁻¹) difference between the wettest and driest model (fig. 5B). However, the observational datasets differed greatly with a range of about 45 cm year⁻¹. The three station-based datasets (CRU, PRISM, and UDelaware) were comparably similar with the reanalysis

datasets giving the most (NCEP) and least (ERA40) precipitation of the five observational datasets. The average of the three station-based datasets were only about 5 cm year⁻¹ greater than the median of the simulated mean annual precipitation. Implications of the discrepancies among the observational datasets are discussed in section 3.

4.2. Seasonal Cycle

All the models reproduced the phase and general shape of the seasonal cycle of temperature (fig. 6A), though the amplitude of the seasonal cycle varied widely among models (fig. 7A), ranging from 19.4 °C to 28.9 °C. The median of the modeled amplitude was 22.5 °C, which was within 1 °C of the observed amplitude.

Most models generated the general pattern of increased precipitation in the months of March through August that is evident in CRU, PRISM, and UDelaware (fig. 6B). Interestingly, large differences exist between the three station-based datasets and two reanalysis datasets; both ERA40 and NCEP show a strong seasonal cycle of dry winter and wet summer.

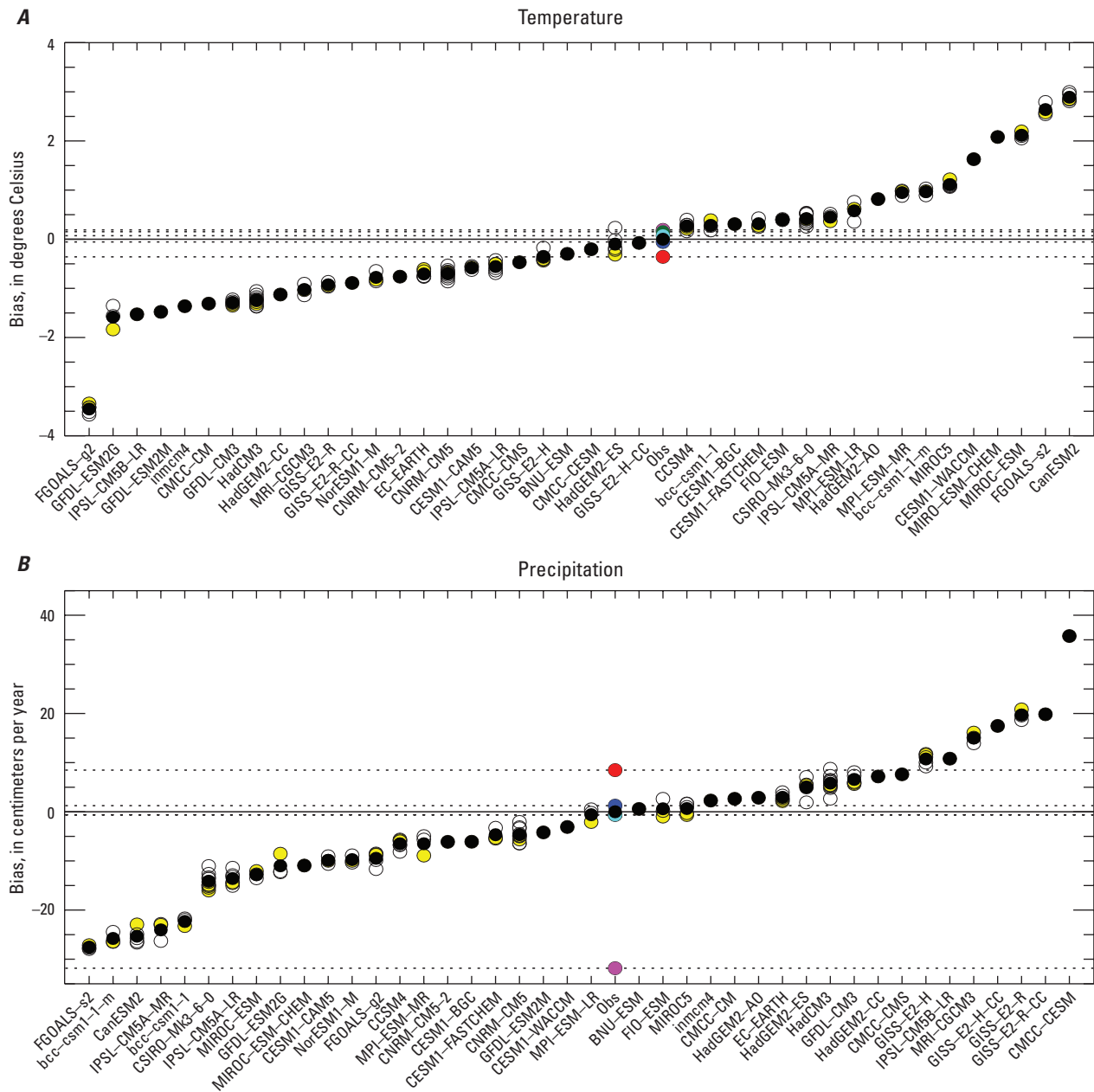


Figure 5. Mean annual *A*, temperature and *B*, precipitation bias for 41 CMIP5 GCMs averaged over the Southeastern United States. For each GCM, black-filled circles show the ensemble average, yellow-filled circles show the first ensemble member, and the open circles show the remaining ensemble members. Observed (Obs) values are from NCEP (red), ERA40 (magenta), CRU (dark green), PRISM (blue), UDelaware (cyan), and average of observations (black). For precipitation, only CRU, PRISM and UDelaware are calculated in the average.

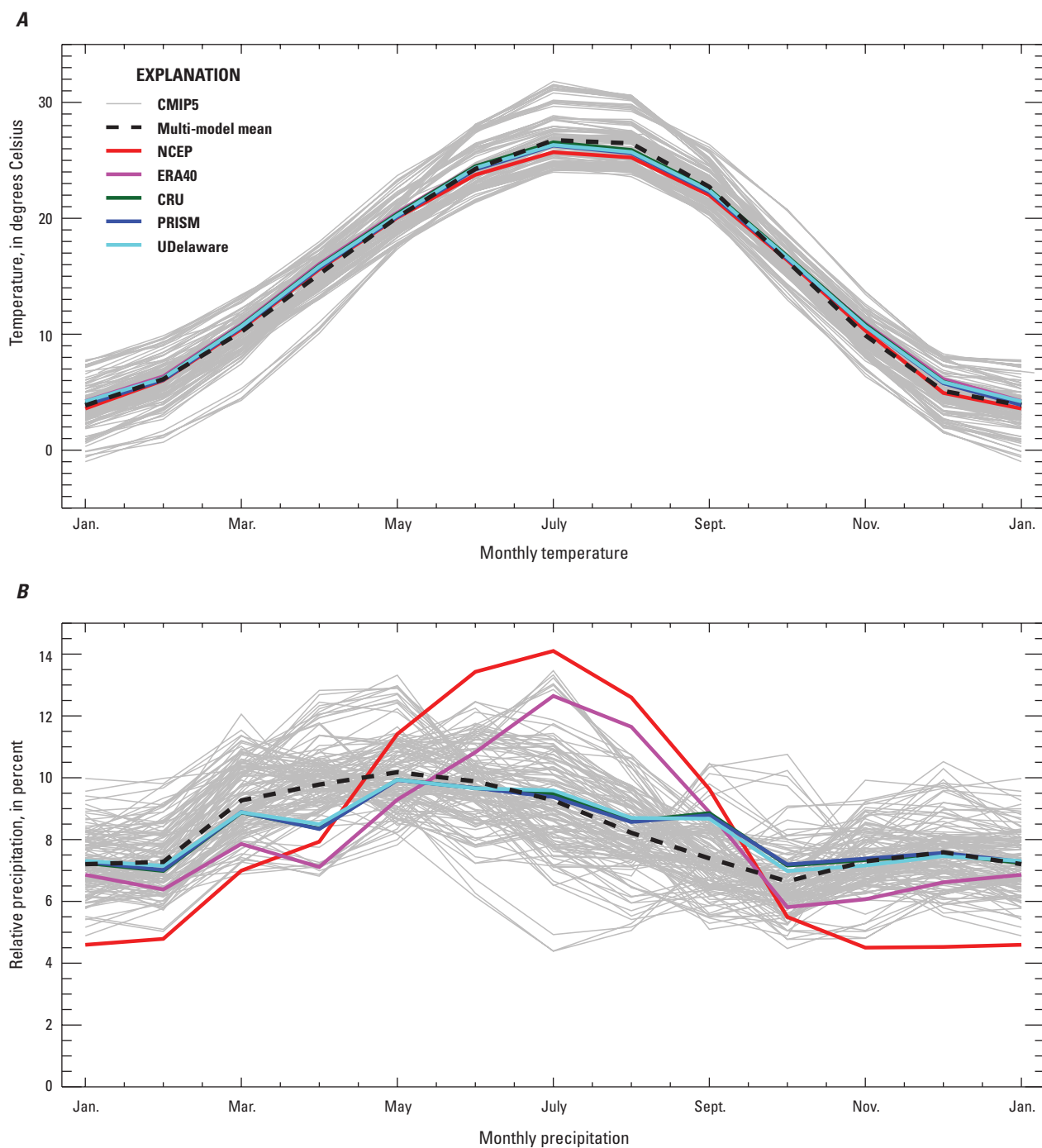


Figure 6. Mean seasonal cycle of *A*, temperature and *B*, relative precipitation (as a percentage of total annual precipitation) averaged over the Southeastern United States. Monthly means are calculated from gridded observation datasets (NCEP, ERA40, CRU, PRISM, UDelaware) and from all ensemble members from 41 CMIP5 GCMs.

This seasonal cycle in the reanalysis is both exaggerated in amplitude and shifted in phase by a few months with respect to the station-based datasets (and, interestingly, also to the overall pattern of the CMIP5 models).

Calculating mean monthly precipitation as a percentage of the mean annual total precipitation, simulated seasonal precipitation amplitude ranged from as small as 2.5 percent to as large as 7.2 percent (fig. 7B). In comparison, the three station-based datasets gave precipitation amplitude of about 3 percent, implying the models on average are exaggerating the strength of the seasonal cycle. (Note that with mean monthly precipitation calculated as a percentage of annual, the percentages reported above are the differences of percentage precipitation between the wettest and driest months.)

Though the statistics above were averaged over the entire Southeastern United States, there are regional spatial gradients in the amplitudes of the seasonal cycle of both temperature and precipitation. Visual inspection showed good agreement between the observed (CRU and ERA40) and the multi-model mean spatial pattern of the temperature cycle amplitude (fig. 8, left panels). The models as a whole accurately reproduced the strong coast-to-interior gradient.

Most of the Southeastern United States has a relatively weak seasonal-cycle amplitude in precipitation, with increased magnitude in Florida and the westernmost part of the Southeastern region. The multi-model mean reproduces this spatial pattern, but to a lesser extent than the observational record (fig. 8, right panels), and, not surprisingly, lacks some finer-scale features. In general, the multi-model mean amplitude overestimates the region compared to CRU and underestimates the amplitude over Florida and western part of the region.

4.3. Diurnal Temperature Range

The Southeastern region as a whole exhibited small variations in the mean diurnal temperature range (DTR) throughout the year, with two maximums in April and October, and two minimums in January and July. This feature was present in both the station-based and reanalysis datasets (fig. 9). Some models reproduced this pattern, though others showed a unimodal cycle with a maximum in summer and minimum in winter.

Simulated mean DTR ranged from 6.2 to 11.5 °C in winter and from 7.3 to 15.2 °C in summer (fig. 10). For winter, DTR from all models was lower than observed DTR from CRU and PRISM; for summer, DTR from all but three models was lower than observed DTR from CRU and PRISM. In contrast, the multi-model mean simulated DTR for both seasons was more consistent with NCEP. Implications of the discrepancy between the station-based and reanalysis datasets are discussed in section 3.

4.4. Large-Scale Spatial Patterns

The multi-model mean temperature field over North America and the northwest Atlantic Ocean accurately reproduced the ERA40 climatological fields (fig. 11), with correlation coefficients (r) of 0.997 and 0.9741 in winter and summer,

respectively (for NCEP, $r = 0.998$ and 0.983 , respectively). Much of this high correlation resulted from simply matching the general latitudinal temperature gradient, but other continental features of the climatological fields were also reproduced. Individually, all models were very highly correlated to observations in winter ($0.982 \leq r \leq 0.996$) and highly correlated in summer ($0.84 \leq r \leq 0.98$). The variances of the modeled fields were also similar to the observed variance, though more so in winter when all standard deviations were within ± 15 percent of the observed standard deviations. In summer, all models were within ± 25 percent of observations (fig. 12A).

The multi-model mean precipitation field over North America and the northwest Atlantic Ocean (fig. 13) generally reproduced the main large-scale climatological features of the ERA40 field in winter ($r = 0.70$), though correlation was lower in summer ($r = 0.60$) (for NCEP, $r = 0.75$ and 0.48 , respectively). Individually, the spatial patterns of most, but not all, models correlated reasonably well with the spatial pattern of ERA40 precipitation in winter ($0.49 \leq r \leq 0.89$) although the correlations weakened in spring and fall and were weakest in summer ($0.00 \leq r \leq 0.71$). Normalized standard deviations ranged from about 0.5 to 1.5 across all simulations and all seasons, with a large majority of models simulating too much spatial variability in winter and spring and too little variability in summer and fall (fig. 12B).

4.5. 20th Century Trend

The average annual temperature in the Southeastern region decreased during the 20th century by an estimated 0.27 °C, calculated as the average from CRU, PRISM and UDelaware. Of 41 CMIP5 models, only 2 produced a downward trend; the multi-model mean trend was +0.60 °C over the 20th century, with models ranging from 0.06 °C to 1.96 °C (fig. 14A). This downward trend in observed temperatures over the region, known as the “warming hole,” is discussed in the context of CMIP5 results by Kumar and others (2013).

The linear trend in observed regional mean annual precipitation was +7 percent over the 20th century. Models produced ensemble-average trends ranging from –5 to +8 percent per century, while only two individual ensemble members from two GCMs exceeded the observed +7 percent per century (fig. 14B). The multi-model mean trend in annual precipitation was +1.5 percent per century.

4.6. Temporal Variability

Overall, the CMIP5 models tended to overproduce interannual variability in regionally averaged times series of temperature relative to the observations, though this bias is not large and the overall bias decreased as the temporal aggregation increased from the annual scale to the decadal scale (fig. 15A). At the annual scale, simulated standard deviations ranged by a factor of about 2, from 0.43 to 0.83 °C (fig. 16A). At the octadal (that is, 8-year) scale, simulated values ranged from 0.17 to 0.68 °C, or a factor of 4 (fig. 16B).

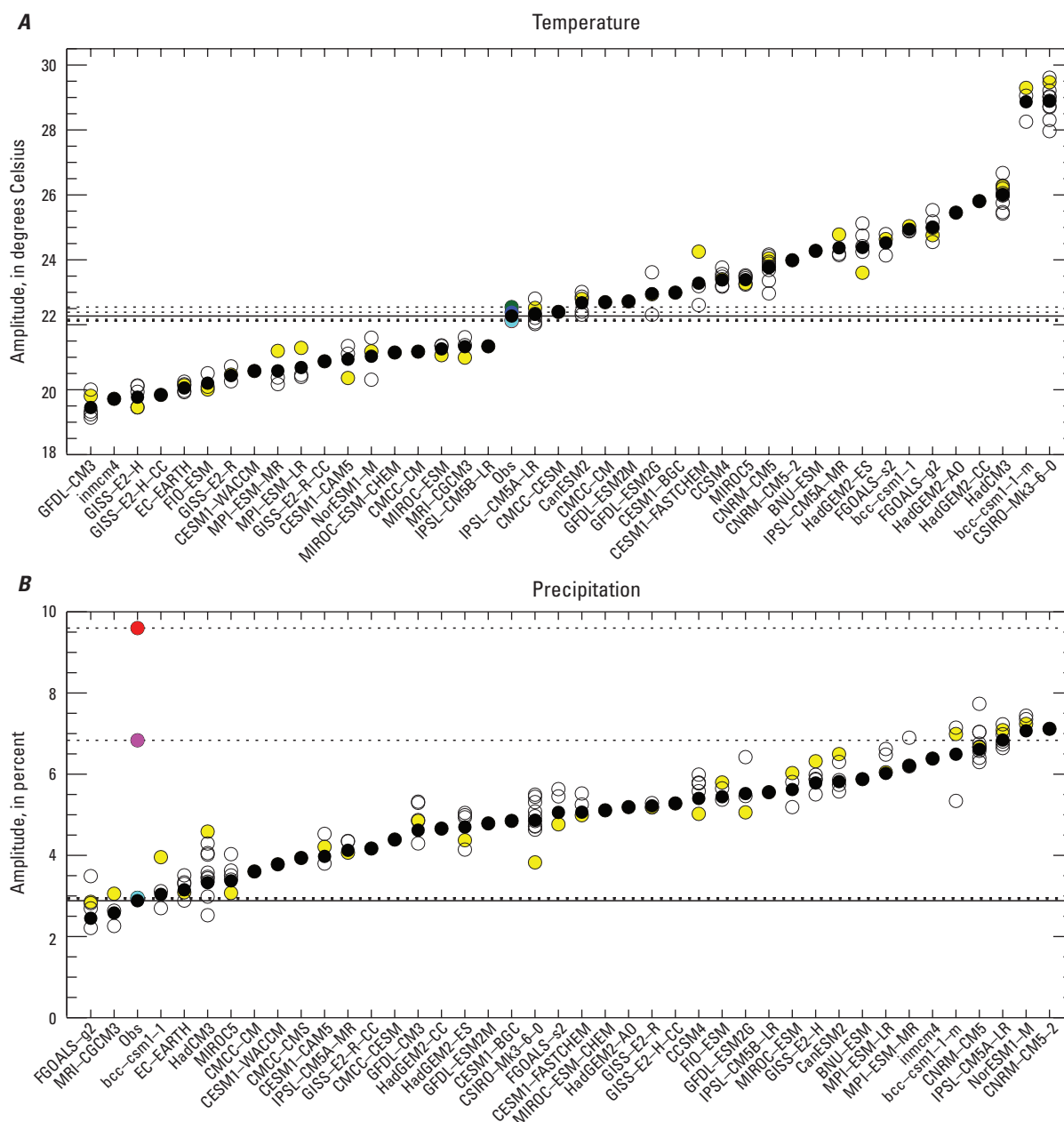


Figure 7. Mean seasonal cycle amplitude in *A*, temperature and *B*, relative precipitation for the Southeastern United States. For each CMIP5 GCM, black-filled circles show the ensemble average, yellow-filled circles show the first ensemble member, and the open circles show the remaining ensemble members. Observed (Obs) values are from NCEP (red), ERA40 (magenta), CRU (dark green), PRISM (blue), UDelaware (cyan), and average of observations (black). Monthly precipitation is calculated as a percentage of the mean annual total, so the amplitude is the difference of percentages. For precipitation, only CRU, PRISM, and UDelaware are calculated in the average.

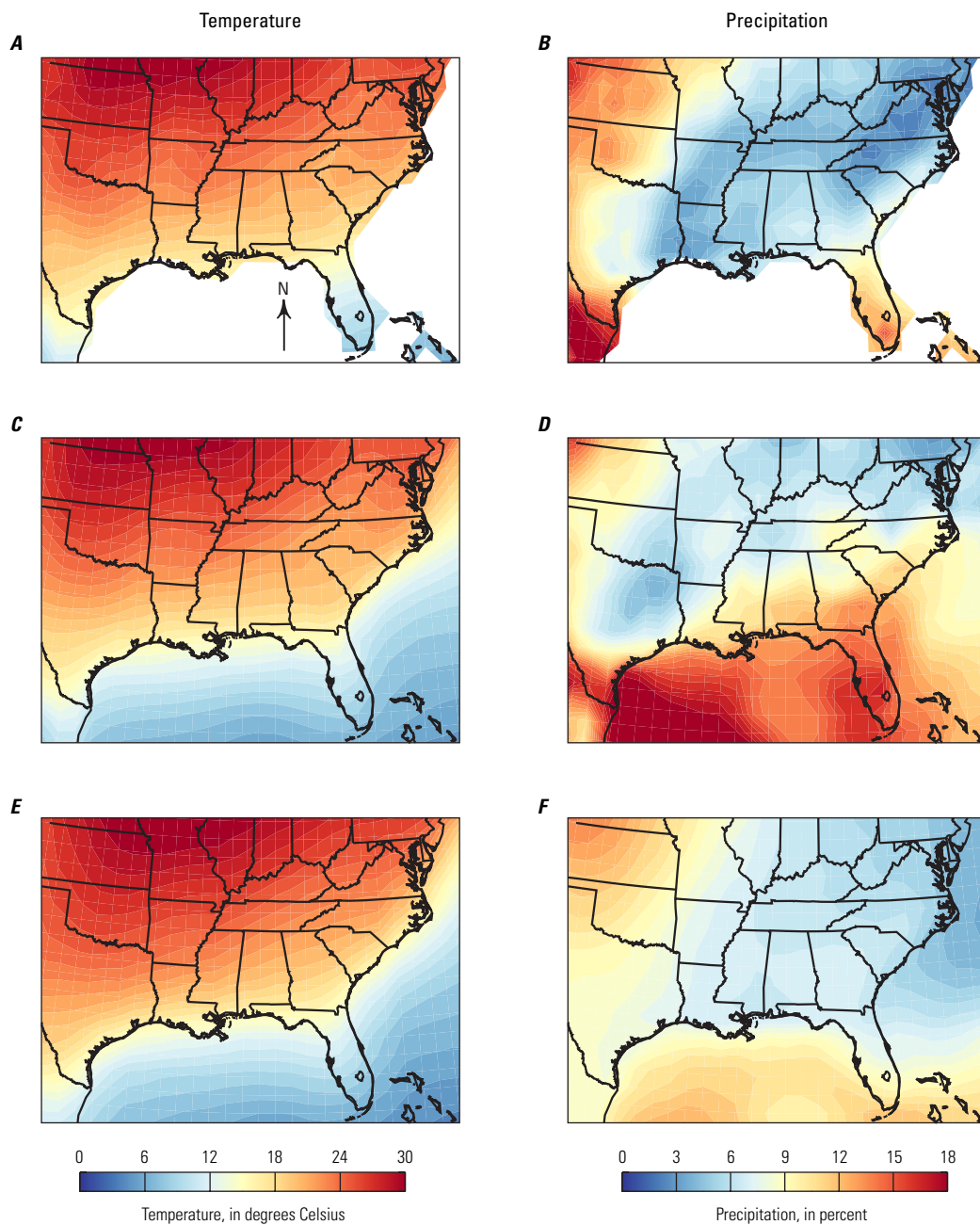


Figure 8. Mean seasonal cycle amplitude of temperature from *A*, CRU, *C*, ERA40, and *E*, the CMIP5 multi-model mean, and mean season cycle amplitude of relative precipitation from *B*, CRU, *D*, ERA40, and *F*, the CMIP5 multi-model mean.

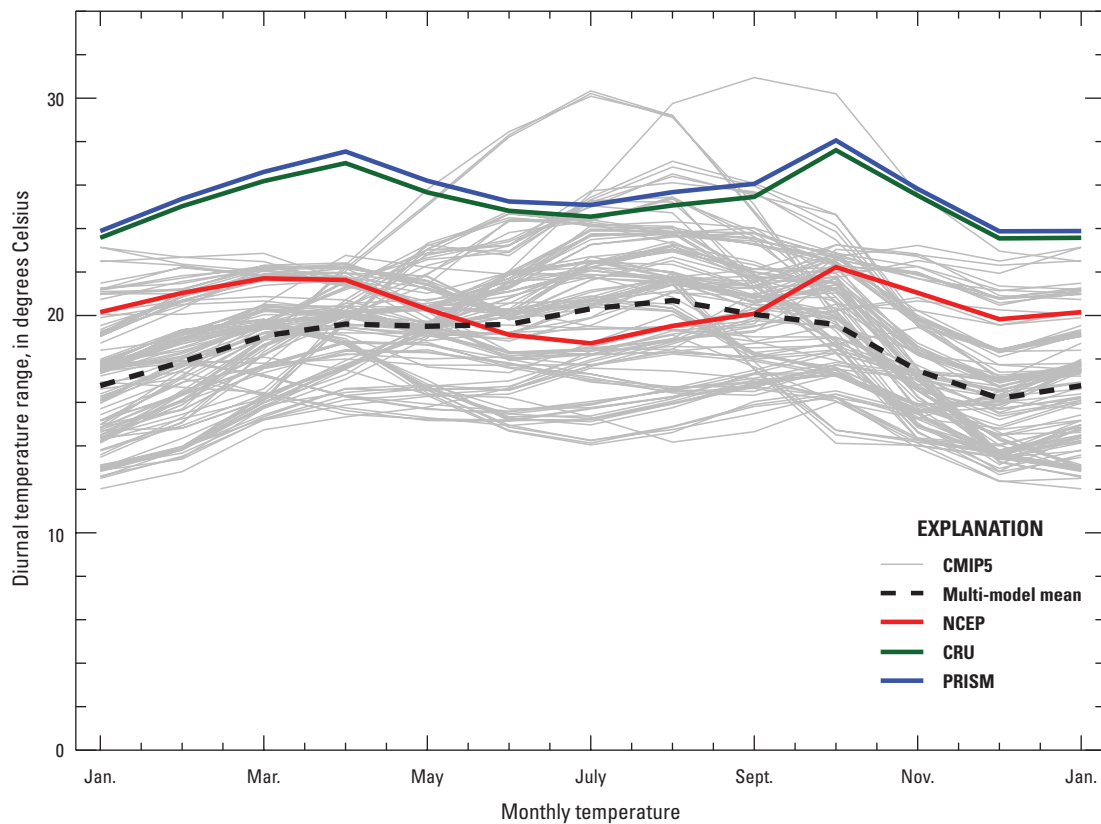


Figure 9. Mean seasonal cycle of diurnal temperature range averaged over the Southeastern United States. Monthly means are calculated from gridded observation datasets (NCEP, CRU, and PRISM) and from all ensemble members from 41 CMIP5 GCMs.

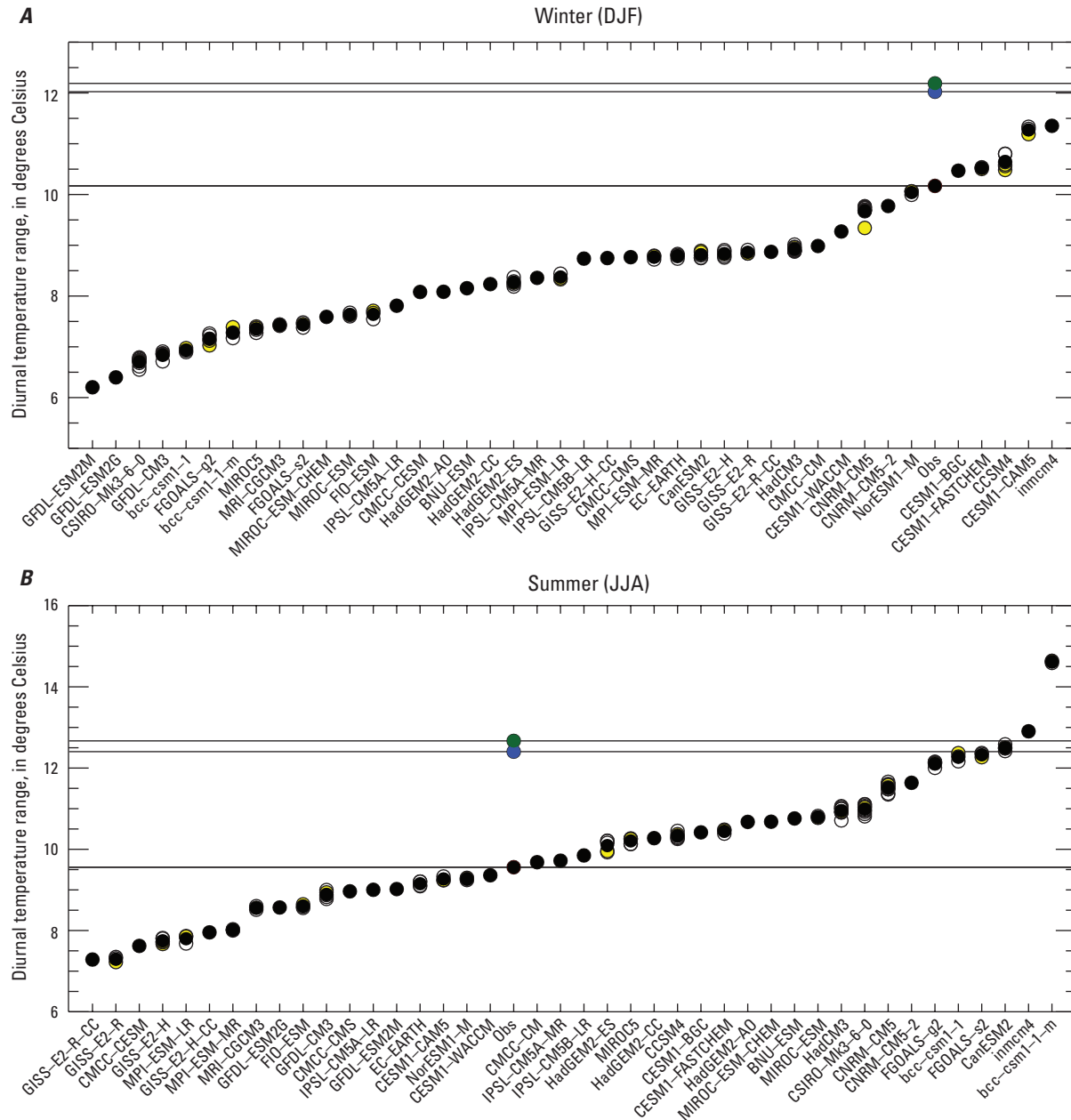


Figure 10. Mean diurnal temperature range (DTR) in winter (DJF) and summer (JJA) averaged over the Southeastern United States. For each of 41 CMIP5 models, black-filled circles show the ensemble average, yellow-filled circles show the first ensemble member, and the open circles show the remaining ensemble members. Observed (Obs) values are from CRU (dark green), PRISM (blue), and NCEP (black).

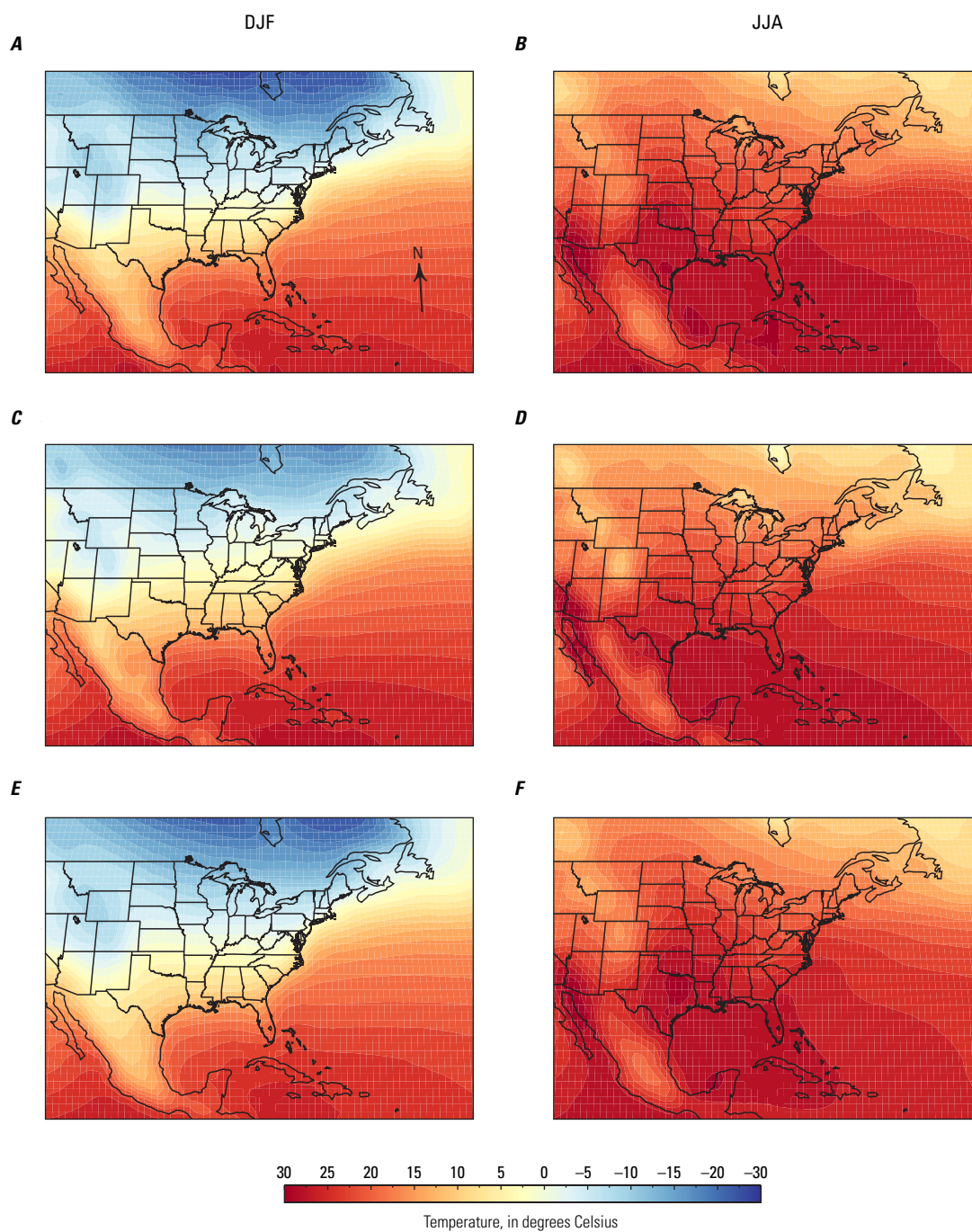


Figure 11. Mean winter (DJF) temperature from *A*, NCEP, *C*, ERA40, and *E*, the CMIP5 multi-model mean, and mean summer (JJA) temperature from *B*, NCEP, *D*, ERA40, and *F*, the CMIP5 multi-model mean.

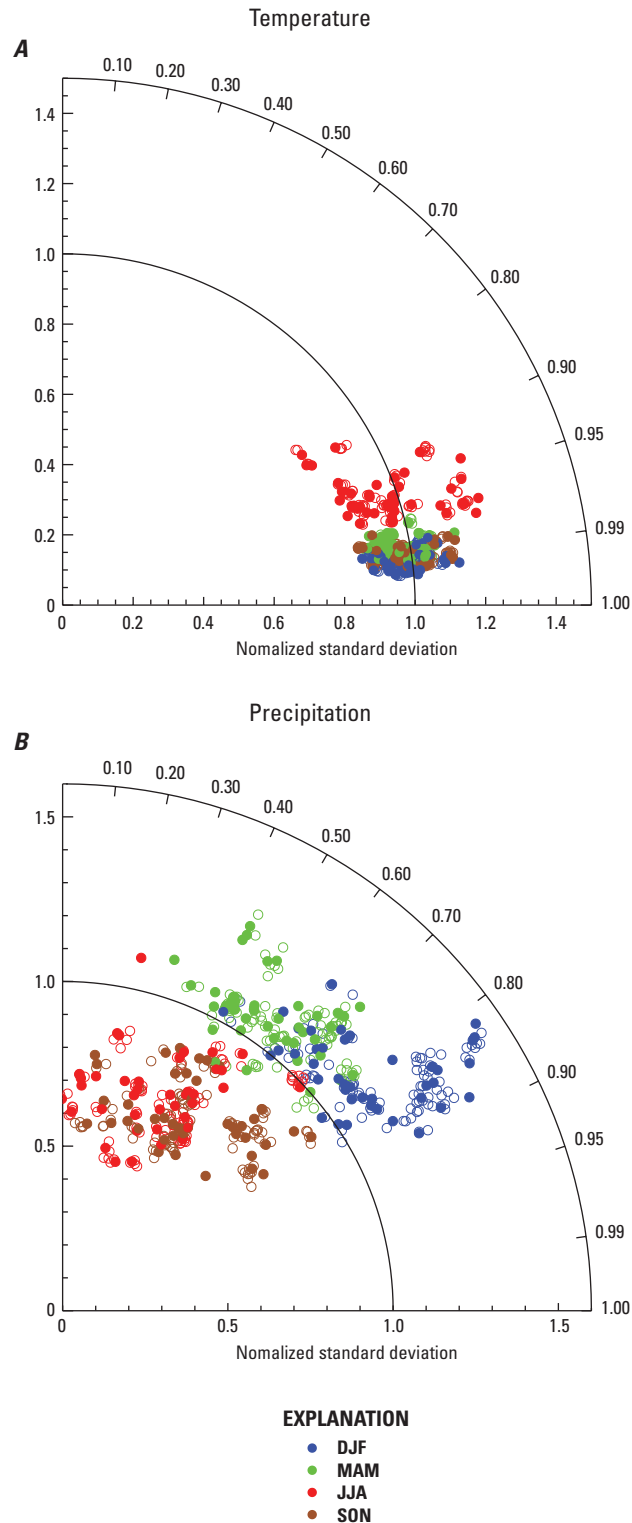


Figure 12. Normalized standard deviations (radius) and correlation coefficients (angle) by season for the climatological mean fields of *A*, temperature and *B*, precipitation from CMIP5. The spatial domain is approximately that shown in figure 11. For each variable, the reference field for the normalization and the correlation is ERA40. Filled circles show the first ensemble members from each model, and open circles show remaining ensemble members. Note that a perfect simulation would have both a normalized standard deviation and a correlation coefficient equal to unity.

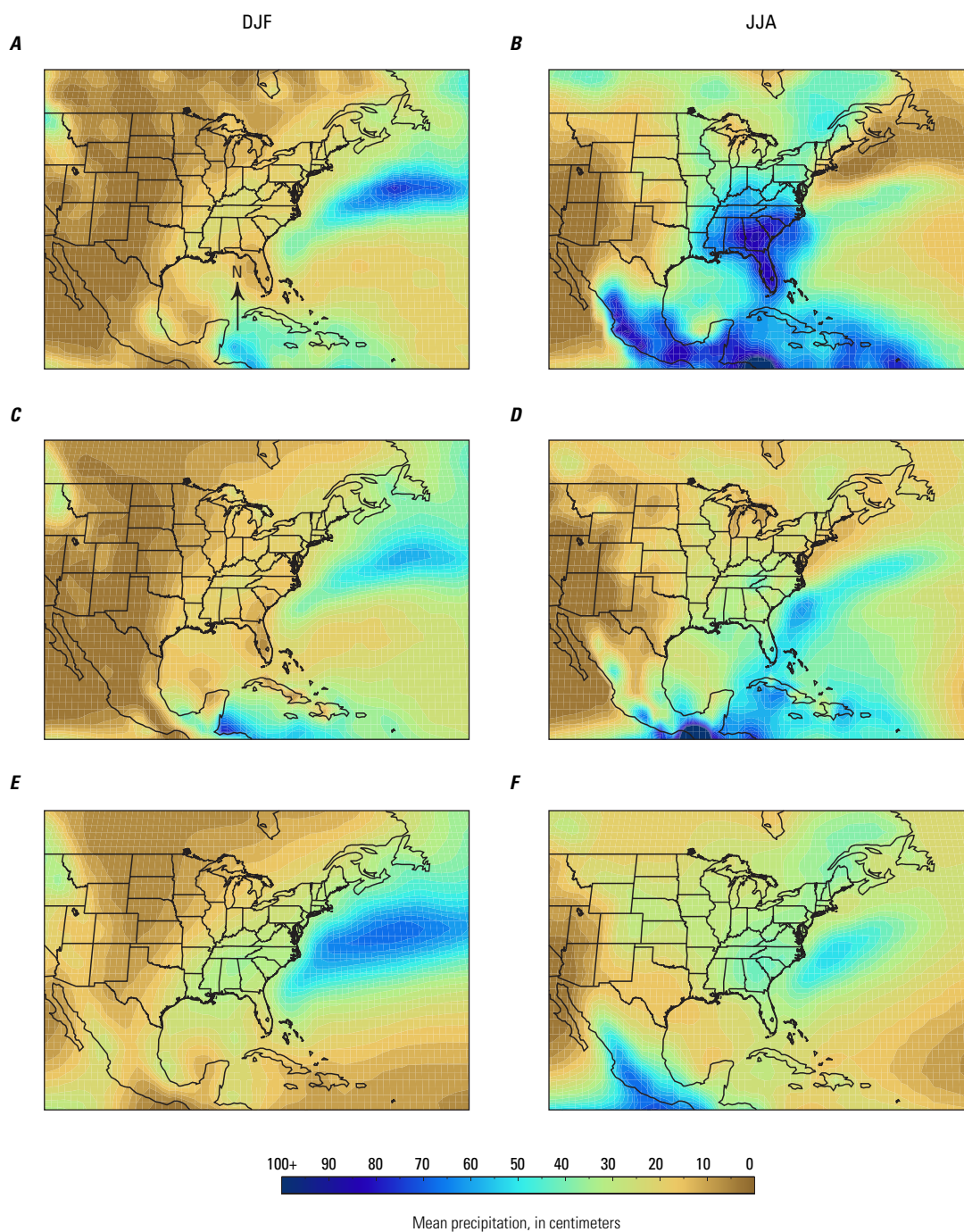


Figure 13. Mean winter (DJF) precipitation from *A*, NCEP, *C*, ERA40, and *E*, the CMIP5 multi-model mean, and mean summer (JJA) precipitation from *B*, NCEP, *D*, ERA40, and *F*, the CMIP5 multi-model mean.

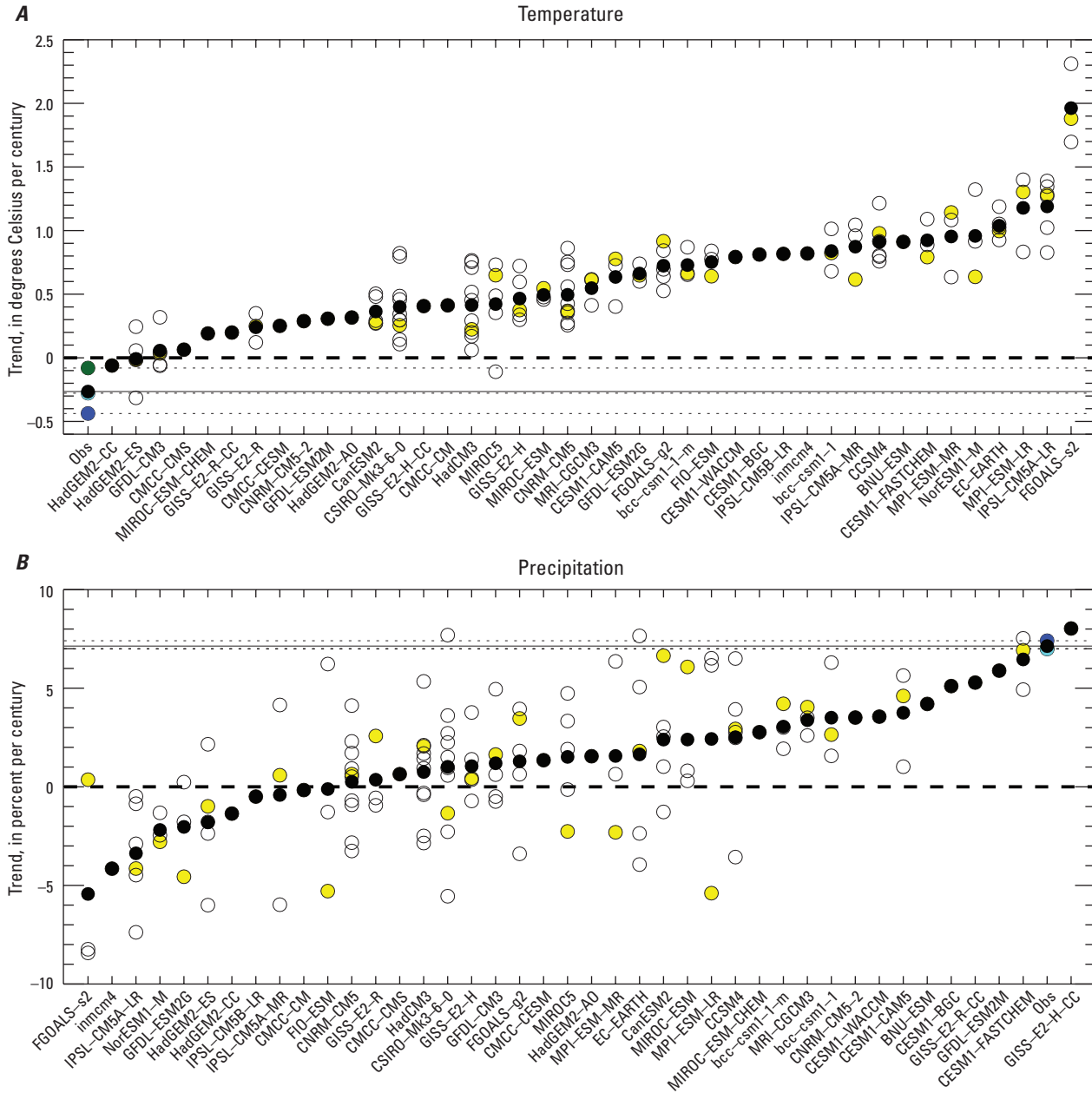


Figure 14. Southeastern United States-averaged trends in annual mean *A*, temperature and *B*, precipitation over the 20th century for all simulations and observations. For each of 41 CMIP5 GCMs, black-filled circles show the ensemble average, yellow-filled circles show the first ensemble member, and the open circles show the remaining ensemble members. Observed (Obs) values are from CRU (dark green), PRISM (blue), UDelaware (cyan), and average of observations (black).

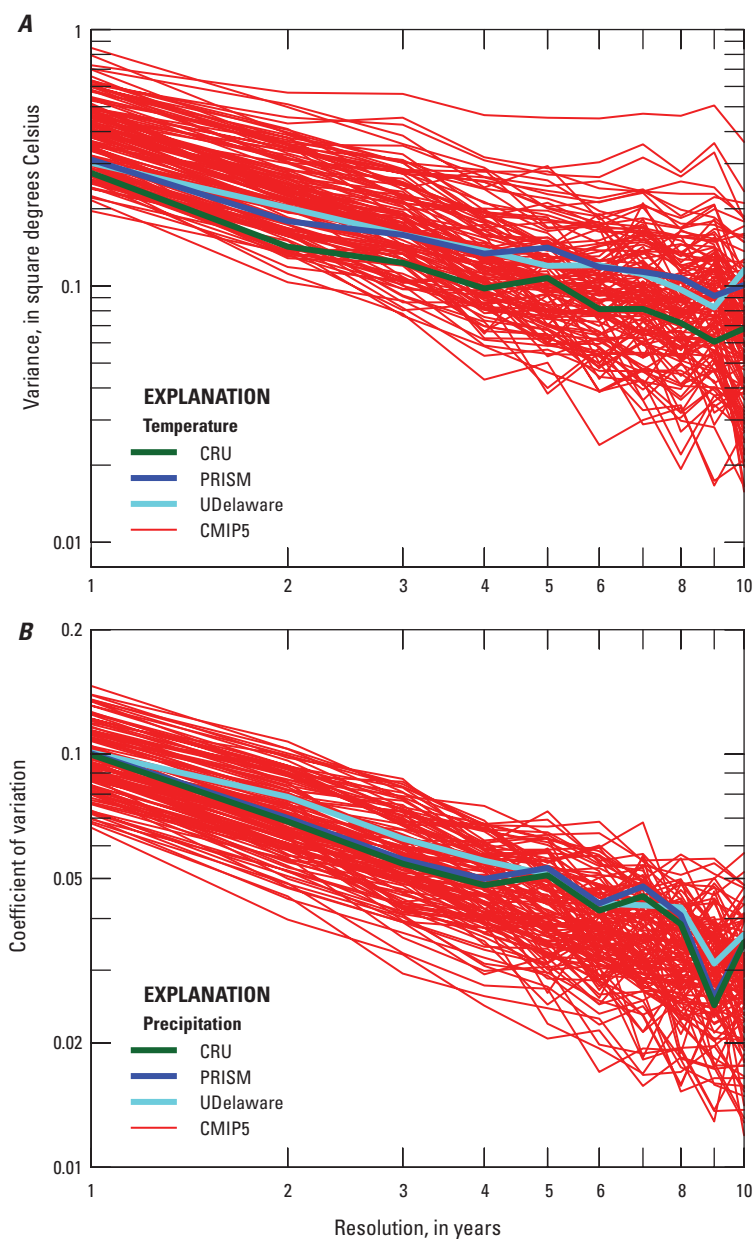


Figure 15. *A*, Variance of temperature anomalies and *B*, coefficient of variation of water year precipitation against temporal resolution of averaging for the Southeastern United States-averaged time series. Red lines show results from all ensemble members from 41 CMIP5 GCMs.

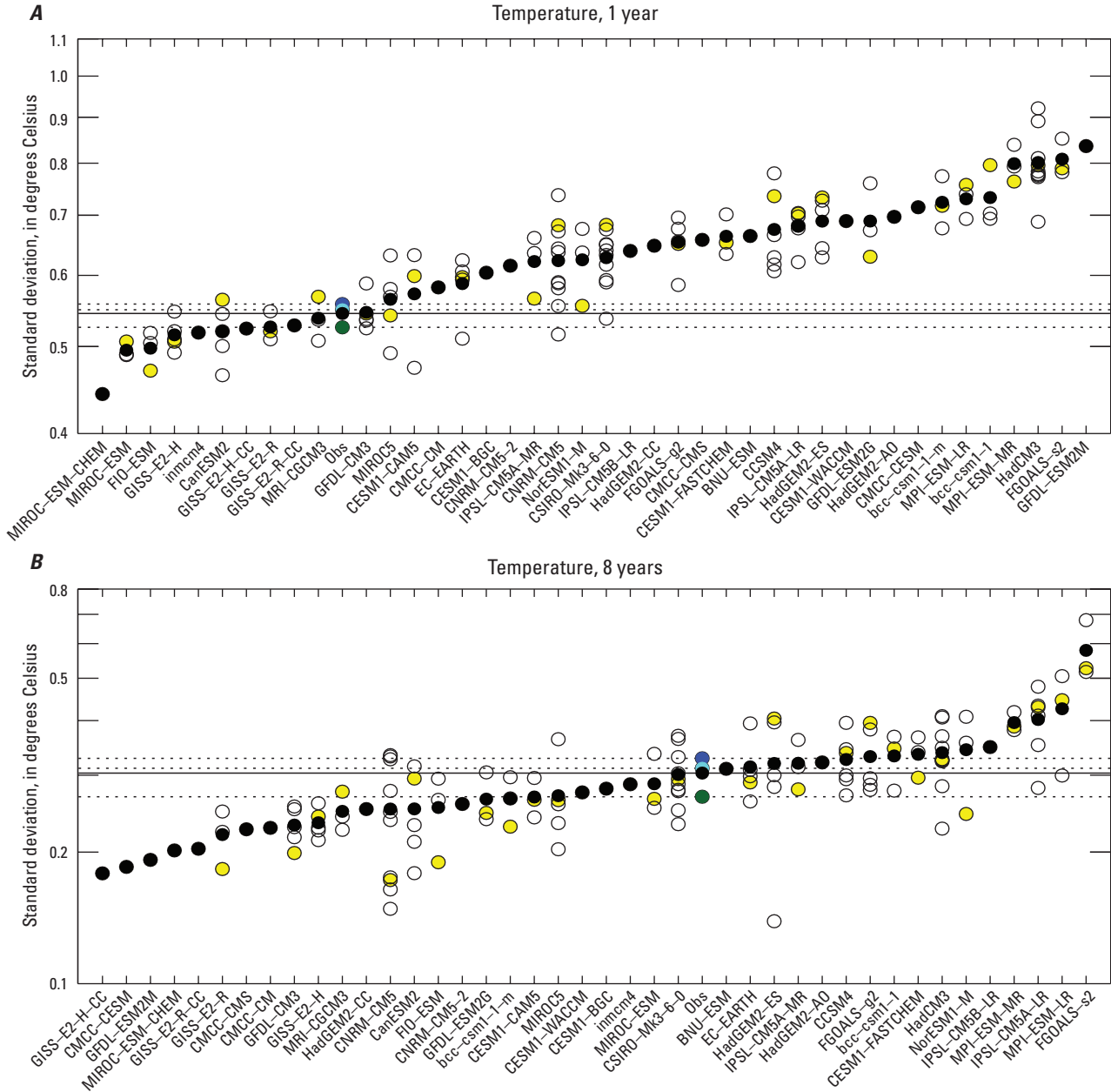


Figure 16. Standard deviation of temperature anomalies at temporal resolutions of *A*, 1 calendar year and *B*, 8 calendar years. Values were averaged over the Southeastern United States. For each of 41 CMIP5 models, black-filled circles show the ensemble average, yellow-filled circles show the first ensemble member, and the open circles show the remaining ensemble members. Observed (Obs) values are from CRU (dark green), PRISM (blue), UDelaware (cyan), and average of observations (black).

In the case of precipitation, the observed annual variability was similar to the mean of the variability from all models (fig. 15B), with the coefficient of variation (CV) ranging from 0.07 to 0.15 for simulated annual precipitation (fig. 17A). Though both the simulations and observations showed apparent power-law scaling of the CV, the simulated CVs in general decreased too rapidly with increasing scale. A consequence is that by the octadal scale, most of the models were generating too little variability (fig. 17B).

Separated by season, the above characteristics in year-to-year variability remain similar: small overall bias in standard deviation of temperature and coefficient of variability of precipitation (see figs. 18 and 19). What is particularly notable, however, is that some models rank very differently across seasons. For example, of all models, NorESM1-M has the lowest CV of precipitation in summer but the second highest CV in winter.

4.7. Long-Term Persistence

The Hurst exponent (Hurst, 1951) of the observed temperature anomalies ranged from 0.70 to 0.73, depending on the dataset (CRU, PRISM, or UDelaware). Though the causes of observed Hurst exponent are not explored here, these values could, for example, indicate long-term memory or non-stationarity in the mean (Klemes, 1974). In either case, the Hurst exponent >0.5 implies that the processes that determine temperature over the region occur over a wide range of scales (Tessier and others, 1996). The mean Hurst exponent averaged over all models was 0.68. Individual simulations showed Hurst exponents all greater than 0.5 ($0.60 \leq H \leq 0.79$) with 90 percent of values falling between 0.63 and 0.75 (fig. 20A).

The estimated Hurst exponent of the observed precipitation anomalies was 0.64 for all three datasets and slightly less than that for temperature. The mean simulated Hurst exponent was, remarkably, also 0.64, with 90 percent of values falling between 0.55 and 0.68 (fig. 21A) and all were greater than 0.5.

4.8. ENSO Teleconnections

Consistent with observations, a negative regional temperature response to El Niño Southern Oscillation (ENSO) was apparent in the models: all but two models had a negative response of winter (JFM) temperature to ENSO (fig. 22A). The multi-model mean response was a 0.43°C decrease in winter temperature for every 1°C increase in the Niño3.4 index, which is slightly weaker than the observed decrease of $0.58^\circ\text{C } ^\circ\text{C}^{-1}$. The agreement in the spatial pattern of the ENSO response is remarkable, though the location of the observed transition from negative to positive temperature response over the United States occurs $2\text{--}3^\circ$ latitude southward of the multi-model mean transition (fig. 23A, B).

A precipitation response to ENSO was also apparent in the simulations, with all models showing increased JFM precipitation with increasing Niño3.4 index (fig. 22B). The multi-model mean response was $7.4\text{ percent } ^\circ\text{C}^{-1}$, compared with the observed response of $3.7\text{ percent } ^\circ\text{C}^{-1}$. The spatial patterns of the observed and mean simulated ENSO precipitation response were generally similar, though the CRU observations showed a negative precipitation response over the Appalachians, which was not apparent in the multi-model mean (fig. 23C, D). Because the precipitation response to ENSO varied in sign across the Southeastern region (as given by CRU), the regionally averaged response has limitations as a performance metric.

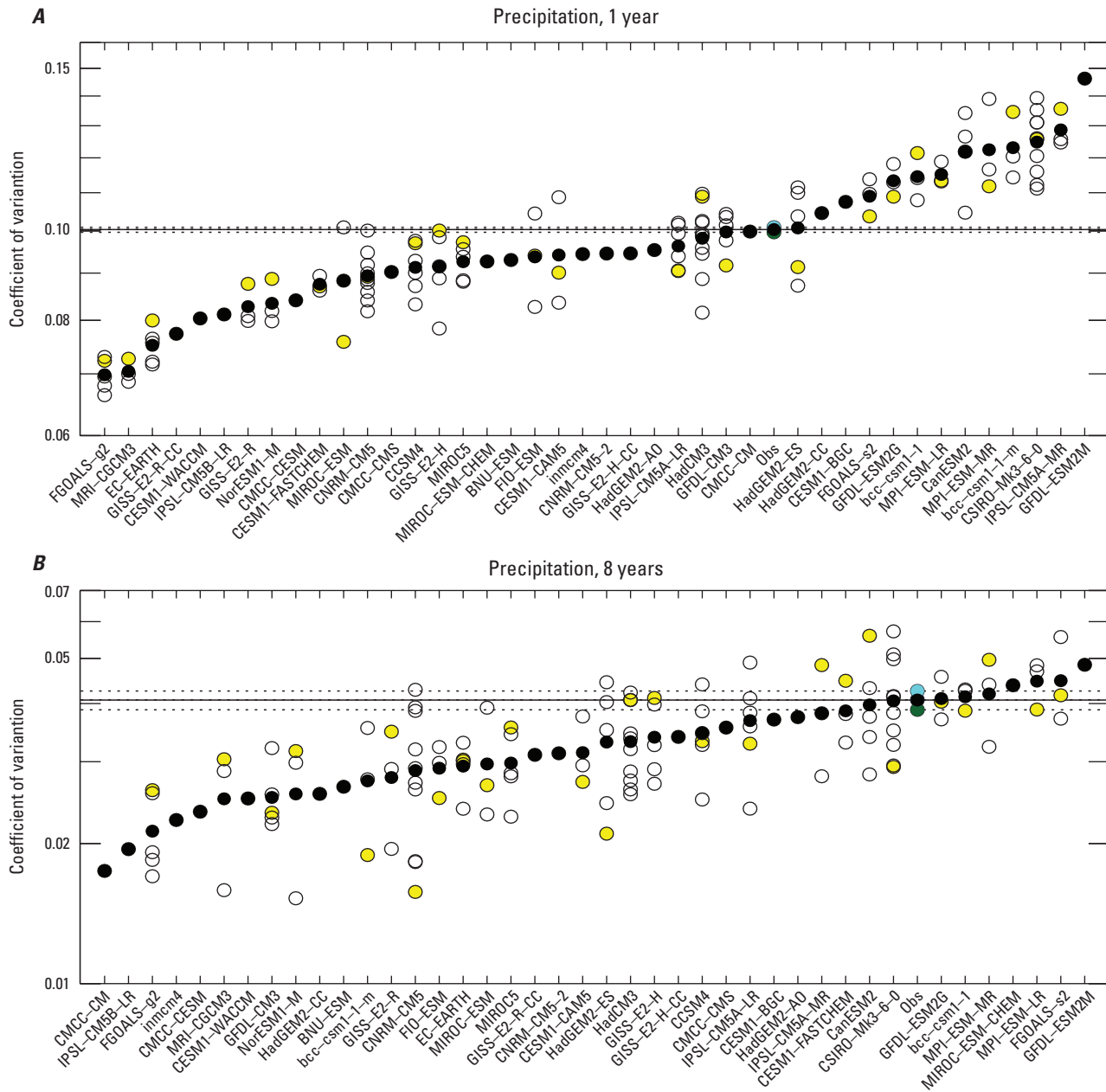


Figure 17. Coefficient of variation of precipitation at temporal resolutions of *A*, 1 water year and *B*, 8 water years. (Water year is the period from October 1 to September 30 and is designated by the year in which the period ends.) Values were averaged over the Southeastern United States. For each of 41 CMIP5 models, black-filled circles show the ensemble average, yellow-filled circles show the first ensemble member, and the open circles show the remaining ensemble members. Observed (Obs) values are from CRU (dark green), PRISM (blue, masked by other symbols), UDelaware (cyan), and average of observations (black).

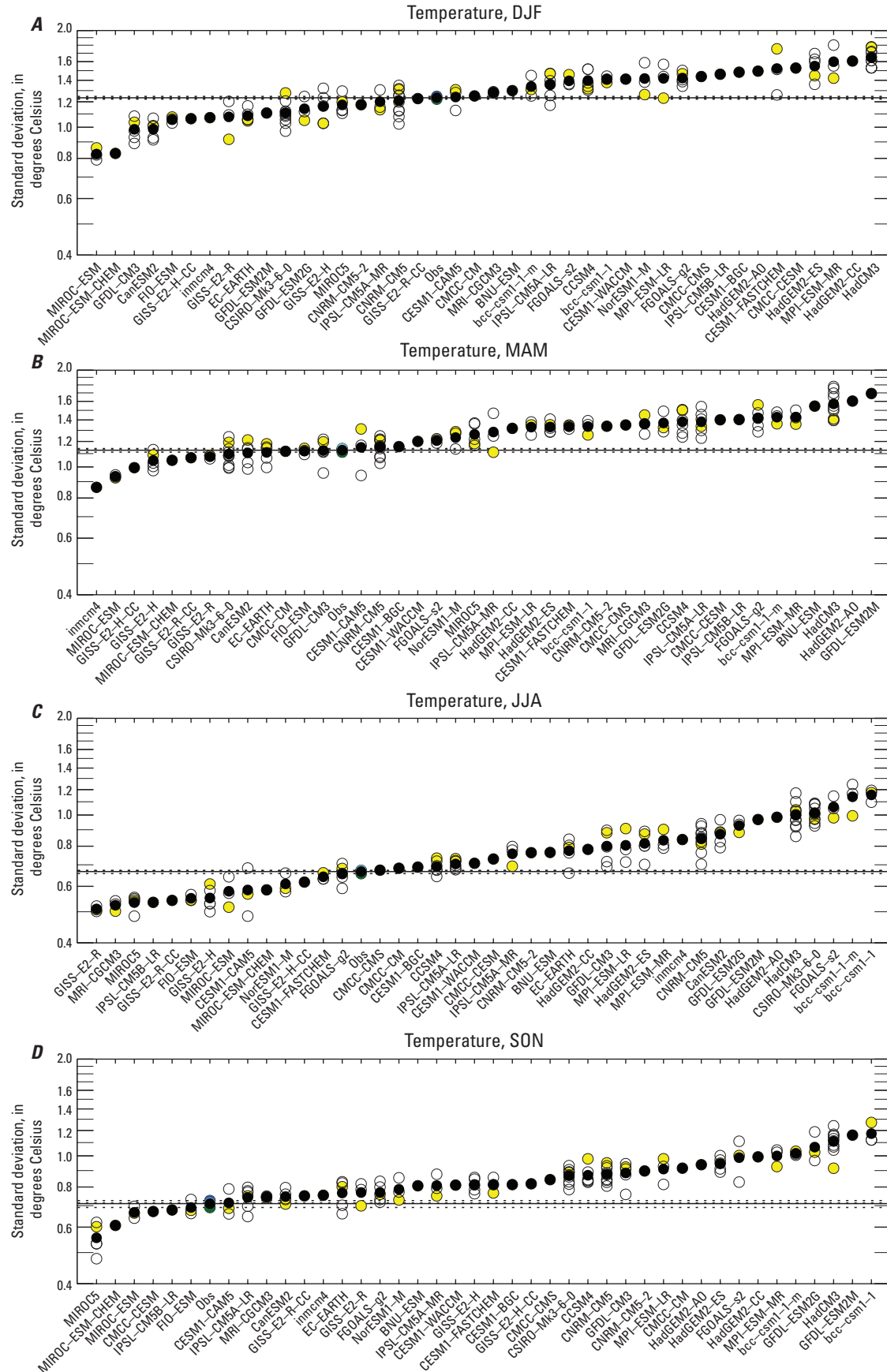


Figure 18. Standard deviation (degrees Celsius) of mean *A*, winter (DJF, December, January, February), *B*, spring (MAM, March, April, May), *C*, summer (JJA, June, July, August), and *D*, fall (SON, September, October, November) temperature anomalies. Values were averaged over the Southeastern United States domain. For each of 41 CMIP5 models, black-filled circles show the ensemble average, yellow-filled circles show the first ensemble member, and the open circles show the remaining ensemble members. Observed (Obs) values are from CRU (dark green), PRISM (blue), UDelaware (cyan), and average of observations (black).

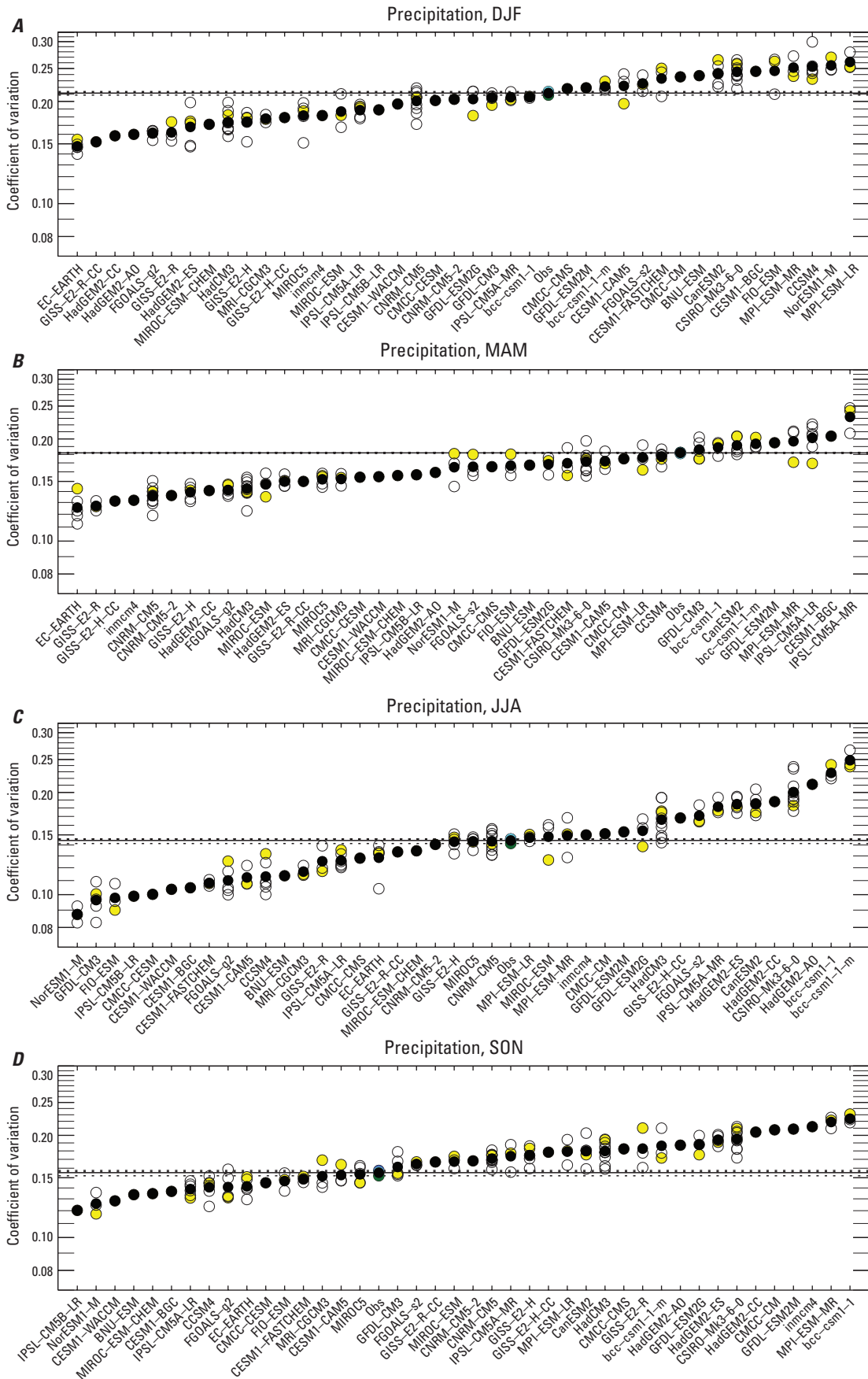


Figure 19. Coefficient of variation of mean *A*, winter (JJA), *B*, spring (MAM), *C*, summer (JJA), and *D*, fall (SON) precipitation. Values were averaged over the Southeastern United States. For each of 41 CMIP5 models, black-filled circles show the ensemble average, yellow-filled circles show the first ensemble member, and the open circles show the remaining ensemble members. Observed (Obs) values are from CRU (dark green), PRISM (blue), UDelaware (cyan), and average of observations (black).

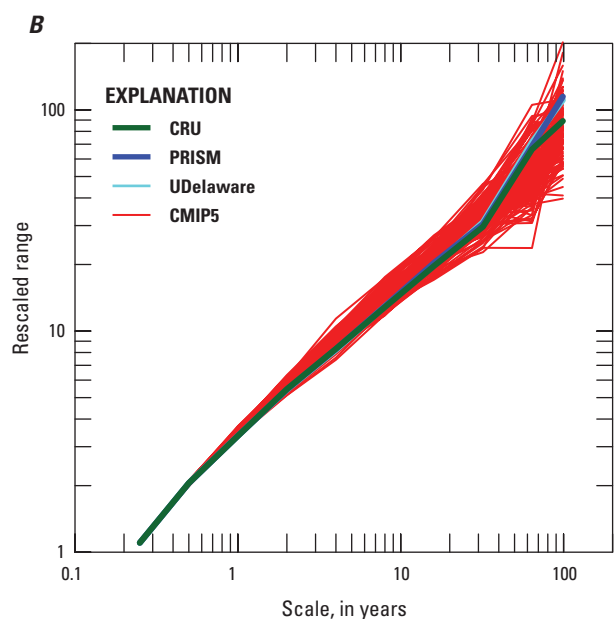
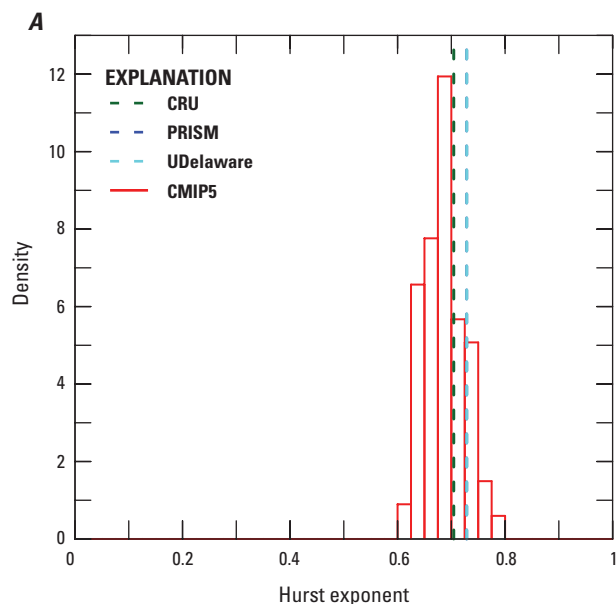


Figure 20. A, Histogram of the Hurst exponent for Southeastern United States-averaged temperature in all CMIP5 simulations. The vertical dashed lines indicate the Hurst exponent estimated from observations. B, Rescaled range against time scale calculated from the observations (heavy lines) and simulations (thin red lines).

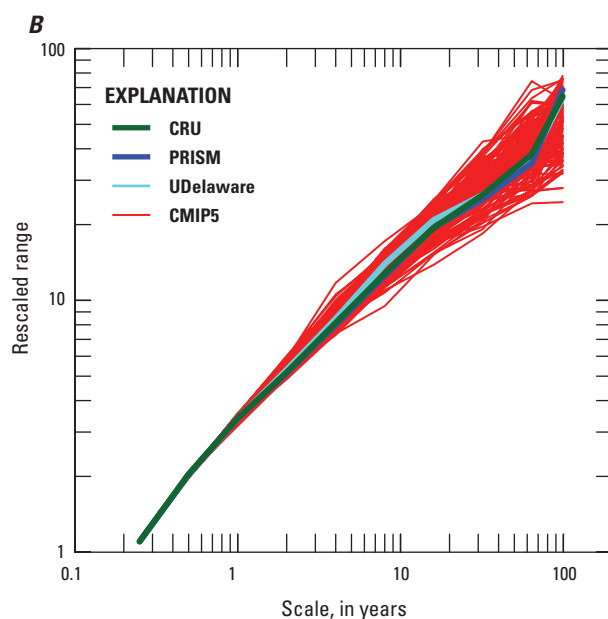
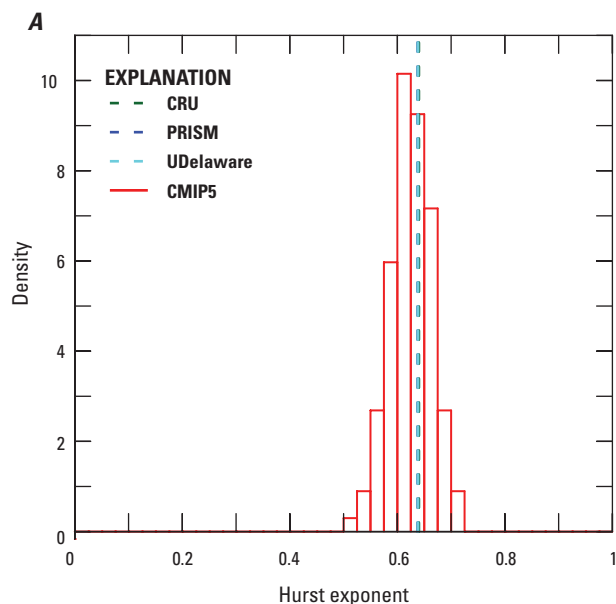


Figure 21. A, Histogram of the Hurst exponent for Southeastern United States-averaged precipitation in all CMIP5 simulations. The vertical dashed lines indicate the Hurst exponent estimated from observations. B, Rescaled range against time scale calculated from the observations (heavy lines) and simulations (thin red lines).

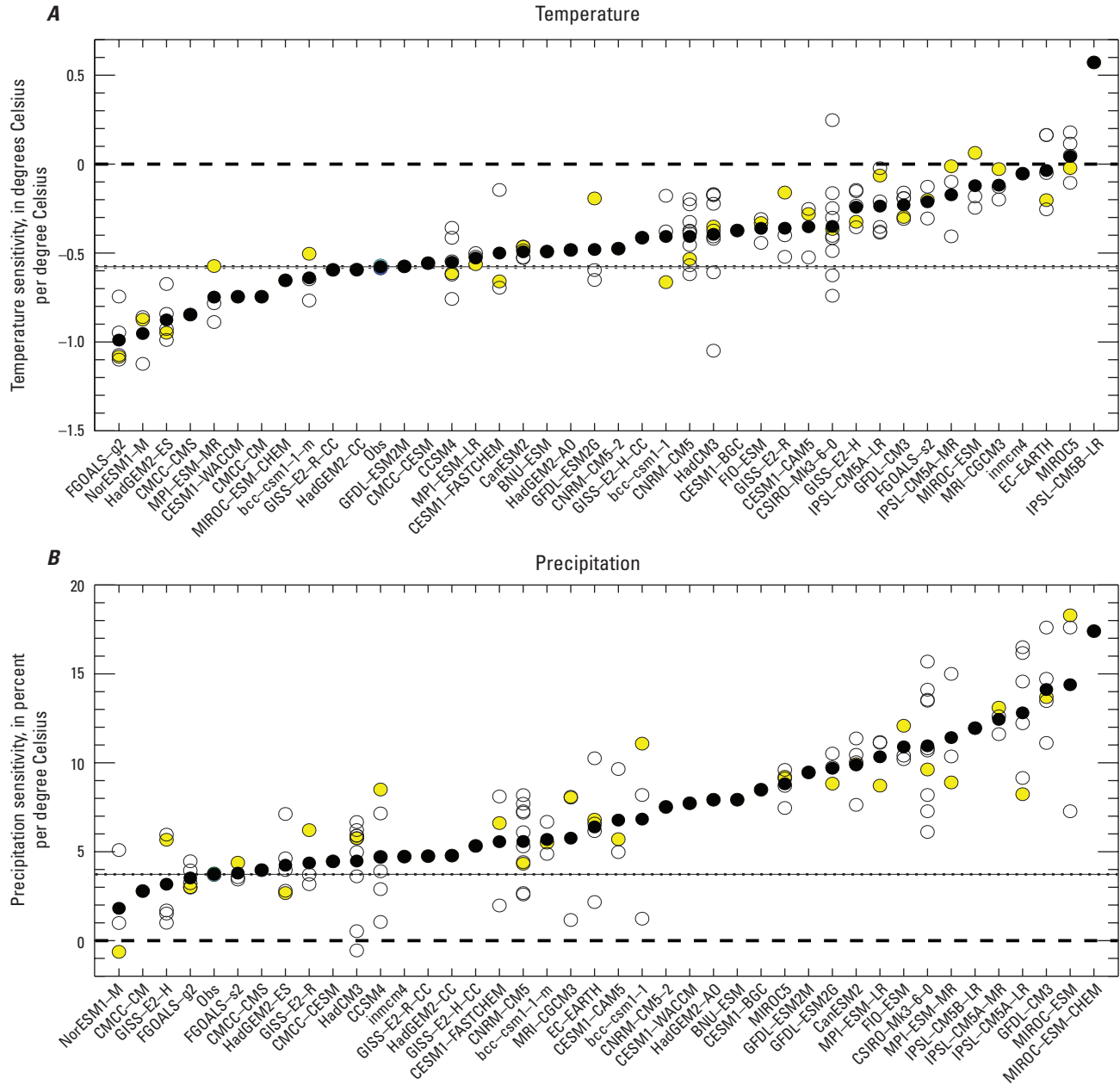


Figure 22. Sensitivity to ENSO as the change in Southeastern United States winter (JFM) *A*, temperature and *B*, precipitation for a one degree Celsius change in the Niño3.4 index averaged over November–March. For each of 41 CMIP5 models, black-filled circles show the ensemble average, yellow-filled circles show the first ensemble member, and the open circles show the remaining ensemble members. Observed (Obs) values are from CRU (dark green), PRISM (blue), UDelaware (cyan), and average of observations (black).

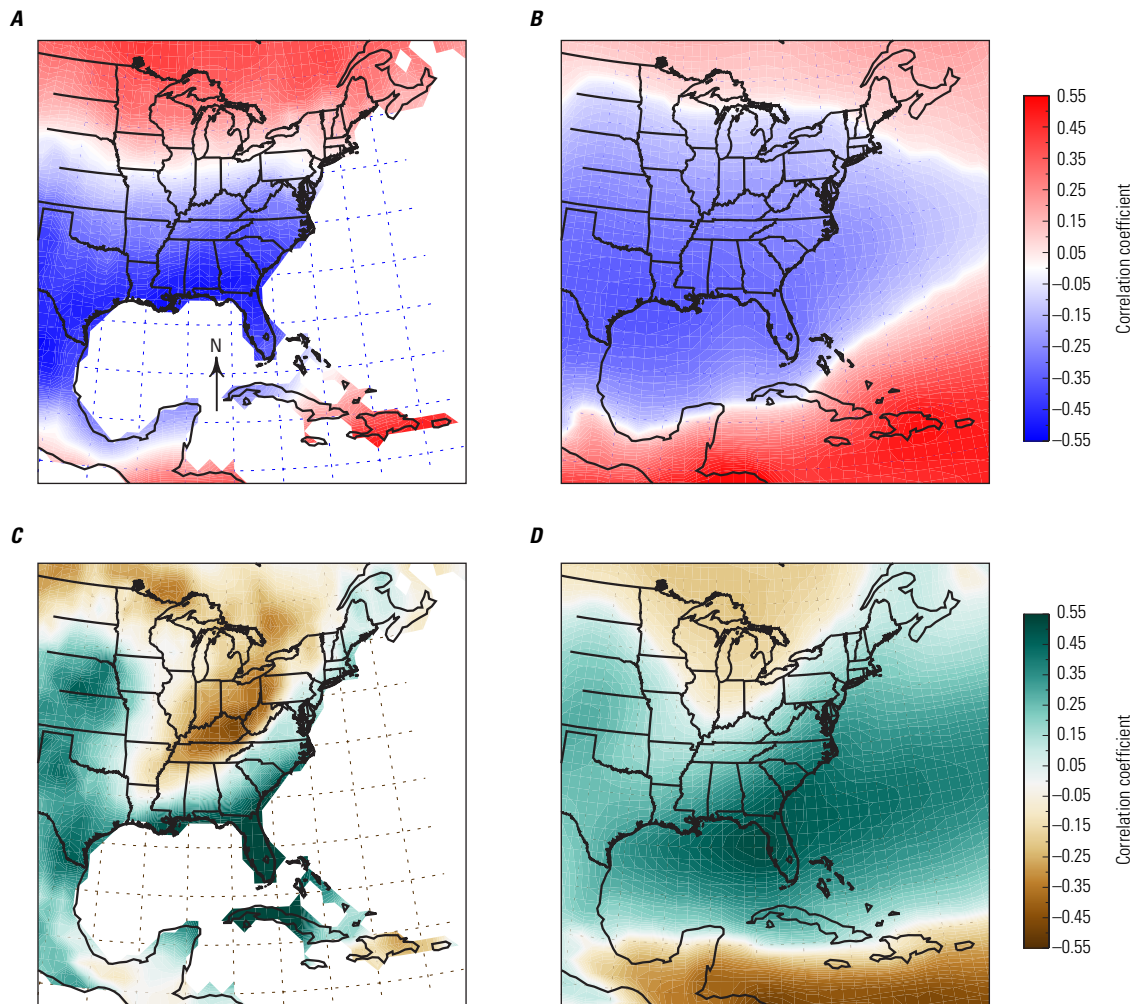


Figure 23. Correlation of CRU winter (JFM) *A*, temperature and *C*, precipitation with the Niño3.4 index averaged over November–March, and multi-model mean correlation of simulated winter *B*, temperature and *D*, precipitation to the same index. Simulations were from 41 CMIP5 models.

References Cited

- Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., and Pasteris, P.A., 2008, Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States: *International Journal of Climatology*, v. 28, no. 15, p. 2031–2064.
- Harris, I., Jones, P.D., Osborn, T.J., and Lister, D.H., 2014, Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 dataset: *International Journal of Climatology*, v. 34, no. 3, p. 623–642.
- Hurst, H.E., 1951, Long-term storage capacity of reservoirs: *Transactions of the American Society of Civil Engineers*, v. 116, p. 770–799.
- IPCC [Intergovernmental Panel on Climate Change], 2013, Climate change 2013—The physical science basis, *in* Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P.M., eds., Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change: Cambridge, United Kingdom and New York, Cambridge University Press, 1,535 p.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Jenne, Roy, and Joseph Dennis, 1996, The NCEP/NCAR 40-year reanalysis project: *Bulletin of the American Meteorological Society*, v. 77, no. 3, p. 437–470.

- Klemes, V., 1974, The Hurst phenomenon—A puzzle?: *Water Resources Research*, v. 10, p. 675–688.
- Kumar, S., Kinter, J., Dirmeyer, P.A., Pan, Z., and Adams, J., 2013, Multidecadal climate variability and the “Warming Hole” in North America—Results from CMIP5 twentieth- and twenty-first-century climate simulations: *Journal of Climatology*, v. 26, p. 3511–3527.
- Matsuura, K., and Willmott, C.J., 2012a, Terrestrial air temperature—1900–2010 gridded monthly time series, version 3.01, accessed September 28, 2013, at http://climate.geog.udel.edu/~climate/html_pages/Global2011/README.GlobalTsT2011.html.
- Matsuura, K., and Willmott, C.J., 2012b, Terrestrial precipitation—1900–2010 gridded monthly time series, version, 3.02, accessed September 28, 2013, at http://climate.geog.udel.edu/~climate/html_pages/Global2011/Precip_revised_3.02/README.GlobalTsP2011.html.
- Phillips, A.S., Deser, C., and Fasullo, J., 2014, Evaluating modes of variability in climate models: *Eos Transactions of the American Geophysical Union*, v. 95, no. 49, p. 453, 455.
- Pierce, D.W., Barnett, T.P., Santer, B.D., and Gleckler, P.J., 2009, Selecting global climate models for regional climate change studies: *Proceedings of the National Academy of Sciences U.S.A.*, v. 106, p. 8444–8446.
- Rupp, D.E., Abatzoglou, J.T., Hegewisch, K.C., and Mote, P.W., 2013, Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest U.S.A.: *Journal of Geophysical Research—Atmospheres*, v. 118, p. 10884–10906.
- Sheffield, J., Barrett, A.P., Colle, B., Fernando, D.N., Fu, R., Geil, K.L., Hu, Q., Kinter, J., Kumar, S., Langenbrunner, B., Lombardo, K., Long, L.N., Maloney, E., Mariotti, A., Meyerson, J.E., Mo, K.C., Neelin, J.D., Nigam, S., Pan, Z., Ren, T., Ruiz-Barradas, A., Serra, Y.L., Seth, A., Thibeault, J.M., Stroeve, J.C., Yang, Z., and Yin, L., 2013, North American climate in CMIP5 experiments, Part 1—Evaluation of historical simulations of continental and regional climatology: *Journal of Climate*, v. 26, no. 23, p. 9209–9245.
- Taylor, K.E., Stouffer, R.J., and Meehl, G.A., 2012, An overview of CMIP5 and the experiment design: *Bulletin of the American Meteorological Society*, v. 93, no. 4, p. 485–498.
- Tessier, Y., Lovejoy, S., Hubert, P., Schertzer, D., and Pecknold, S., 1996, Multifractal analysis and modeling of rainfall and river flows and scaling, causal transfer functions: *Journal of Geophysical Research—Atmospheres*, v. 101, p. 26427–26440.
- Uppala, S.M., K  llberg, P.W., Simmons, A.J., Andrae, U., Da Costa Bechtold, V., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E., Arpe, K., Balmaseda, M.A., Beljaars, A.C.M., Van De Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., H  lm, E., Hoskins, B.J., Isaksen, I., Janssen, P.A.E.M., Jenne, R., McNally, A.P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J., 2005, The ERA-40 re-analysis: *Quarterly Journal of the Royal Meteorological Society*, v. 131, no. 612, p. 2961–3012.

For additional information regarding this publication, contact:

Director
DOI Southeast Climate Science Center
North Carolina State University, Campus Box 7617
Raleigh, NC 27695
(919) 515-2229

Or visit the DOI Southeast Climate Science Center Web site at <https://www.doi.gov/csc/southeast/>

Prepared by:
USGS Science Publishing Network
Raleigh Publishing Service Center
3916 Sunset Ridge Road
Raleigh, NC 27607

