RESEARCH ARTICLE

CAUSES: Diagnosis of the Summertime Warm Bias in CMIP5 Climate Models at the ARM Southern Great Plains Site

Chengzhu Zhang[1] (ID), Shaocheng Xie[1] (ID), Stephen A. Klein[1] (ID), Hsi-yen Ma[1] (ID), Shuaiqi Tang[1] (ID), Kwinten Van Weverberg[2] (ID), Cyril J. Morcrette[2] (ID), and Jon Petch[2] (ID)

[1]Lawrence Livermore National Laboratory, Livermore, CA, USA, [2]Met Office, Exeter, UK

Abstract All the weather and climate models participating in the Clouds Above the United States and Errors at the Surface project show a summertime surface air temperature (T2 m) warm bias in the region of the central United States. To understand the warm bias in long-term climate simulations, we assess the Atmospheric Model Intercomparison Project simulations from the Coupled Model Intercomparison Project Phase 5, with long-term observations mainly from the Atmospheric Radiation Measurement program Southern Great Plains site. Quantities related to the surface energy and water budget, and large-scale circulation are analyzed to identify possible factors and plausible links involved in the warm bias. The systematic warm season bias is characterized by an overestimation of T2 m and underestimation of surface humidity, precipitation, and precipitable water. Accompanying the warm bias is an overestimation of absorbed solar radiation at the surface, which is due to a combination of insufficient cloud reflection and clear-sky shortwave absorption by water vapor and an underestimation in surface albedo. The bias in cloud is shown to contribute most to the radiation bias. The surface layer soil moisture impacts T2 m through its control on evaporative fraction. The error in evaporative fraction is another important contributor to T2 m. Similar sources of error are found in hindcast from other Clouds Above the United States and Errors at the Surface studies. In Atmospheric Model Intercomparison Project simulations, biases in meridional wind velocity associated with the low-level jet and the 500 hPa vertical velocity may also relate to T2 m bias through their control on the surface energy and water budget.

1. Introduction

A prominent summertime warm bias over the midlatitude continents summer has been identified in most weather and climate models (Cheruy et al., 2014; Klein et al., 2006; Ma et al., 2014; Merrifield & Xie, 2016; Mueller & Seneviratne, 2014), including all models participating the Clouds Above the United States and Errors at the Surface (CAUSES) project (Morcrette et al., 2018). Cheruy et al. (2014) showed that models with a larger present-day warm bias tended to predict stronger surface warming in simulations of 21st century climate change. Therefore, it is important to understand the causes of the warm bias present in models and identify the model deficiencies that possibly generate the bias.

Running climate models in short-range hindcasts initialized with analysis data have been proven to be an efficient way to isolate model errors to deficiencies in model physical parameterizations because the -produced large-scale circulation has not been drifted away from observations (Ma et al., 2014, 2015; Phillips et al., 2004). Many of the systematic model errors associated with precipitation, clouds, and other related fields were robust and generally independent of the choice of analysis data for initial conditions (Phillips et al., 2004; Williamson et al., 2005). By examining climate models in both short-term hindcasts and long-term climate simulations, Xie et al. (2012) and Ma et al. (2014) showed that the warm bias in climate simulations appears in the first 2 to 3 days of hindcasts, which indicates that the bias is likely related to problems in representing fast physics such as cloud, radiation, and precipitation in these models. Using the hindcast approach, Klein et al. (2006) hypothesized that the lack of precipitation results in a positive land-atmosphere feedback, which enhanced the warm bias that was initiated by the excess surface solar radiation. Building upon the hindcast approach, Van Weverberg et al. (2015) developed an objective cloud-regime classification method at the model time step level to identify the cloud regimes that contribute most to the bias growth with time. They showed that both cloud and land/surface-related errors contributed to the bias growth in the two models being evaluated. By examining short-range hindcast simulations from all models participating in CAUSES,

Van Weverberg et al. (2018) further identified that the overestimations of downwelling shortwave radiative flux are due to underestimations in the radiative effects of deep convective clouds arising from too infrequent deep convection and/or insufficiently reflective deep convective clouds.

With a broader selection of model simulations, the multimodel ensemble simulations from the Coupled Model Intercomparison Project Phase 5 (CMIP) project have also been assessed for understanding the warm bias in long-term climate. Mueller and Seneviratne (2014) showed that systematically underestimated evapotranspiration over regions that feature strong land-atmosphere coupling (so-called "hot spots" region (Koster et al., 2004)) is related to the summer surface temperature overestimation. By examining the spread of the warm bias among the CMIP5 models, Cheruy et al. (2014) concluded that a lack of cloud is also an important reason for the warm bias. However, the relative importance of cloud radiation errors and evaporation errors in the creation of the warm bias has not yet been identified.

With the goal to develop further understanding of the warm bias, four CAUSES studies including the present one have been conducted. The first three studies examine 5 day hindcast simulations from about 10 participating models and focus on (1) understanding the temperature bias characteristics and their spatial and temporal variations (Morcrette et al., 2018), (2) attributing radiation errors to errors in cloud, clear-sky absorption and scattering and surface albedo (Van Weverberg et al., 2018), and (3) determining the covariance of temperature biases with the terms of the surface energy budget and identifying relative importance of errors from atmosphere models and land models (Ma et al., 2018). In these three studies, the short-term hindcast experiments span a time period from 1 April 2011 to 31 August 2011 in order to utilize validation data sets collected during the U.S. Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) program's Midlatitude Continental Convective Clouds Experiment (Jensen et al., 2016) over the Southern Great Plains (SGP) site.

To understand the warm bias in long-term climate simulations in the wider context of models, in this paper the Atmospheric Model Intercomparison Project (AMIP) simulations produced by the CMIP5 climate models are assessed with long-term data collected at the ARM SGP site as well as with satellite observations and other ground-based observations. In our evaluation, we use the comprehensive collections of ground-based station data from ARM including measurements for atmospheric state, precipitation, turbulent fluxes, radiation, and the land surface. Most of these data sets are available for the recent few decades and are useful to evaluate long-term climate simulations. In this study, we examine multimodel (CMIP5) data sets, with the evaluated domain limited to a 3° × 3° size grid box centered at ARM SGP site in Oklahoma. This enables to make full use of ARM's surface network in climate model evaluations (Xie et al., 2010). Auxiliary products from satellite-based measurement and other ground-based measurements are also analyzed.

The goal of this study is to quantify and better understand the biases in various surface and atmospheric fields associated with the warm bias in the CMIP simulations. We achieve this goal by analyzing the surface energy budget, water budget, and large-scale circulation with ARM detailed field observations and other supplemental observations. This paper also provides perspectives as to relevance of the hindcast biases studied in the other CAUSES paper to the climate bias. Similar as other aforementioned evaluation studies using CMIP climate simulations, the direction of causality between those identified biases needs to be further investigated with more controlled experiments in follow-up studies.

The paper is organized as follows. The observational data sets and selected CMIP5 models are described in section 2. Section 3 shows the characteristics of the model biases and presents a bias diagnosis. Summary and discussion are provided in section 4.

2. Observations and Model Data Description

2.1. Observation Data Sets

The observational data used in this study are primarily from that collected at the DOE's ARM Climate Research Facility SGP site with its central facility located at Lamont, Oklahoma (36.6°N, 97.5°W). In order to compare with grid box mean variables output from climate models, the majority of the observational fields are from the ARM continuous forcing evaluation data sets (Xie et al., 2004) which attempts to determine the spatial average for a region of approximately 3° latitude-longitude centered on the central facility. The long-term continuous forcing data sets are available from 1999 to 2011 with limited missing time periods and its

temporal resolution is 1~h. The multiyear monthly mean climatology built from this data set is used in this study. In this data set, the vertical profiles of the atmospheric state variables (temperature and specific humidity) are from the National Oceanic and Atmospheric Administration rapid update cycle analysis but are adjusted to conserve the column-integrated mass, dry static energy, and moisture through a constrained variational analysis approach developed by Zhang and Lin (1997) and Zhang et al. (2001) using observed surface and Top-of-the-Atmosphere fluxes as the constraints. The surface quantities include both radiation and turbulence fluxes, which are first interpolated into 0.5° × 0.5° grids within the ARM SGP domain that covers a 3° × 3° area (see Figure 1 from Tang et al., 2016) before the domain mean is calculated.

Table 1 summarizes all data sets used and provides additional information on data sources and estimated uncertainties. From the continuous forcing product, the surface screen-level temperature and humidity are based on 9 Surface Meteorological Observation Stations, 127 Oklahoma, and 13 Kansas mesonet stations. Note that the number of stations varies with time. The precipitation rate is obtained from the Arkansas-Red Basin River Forecast Center precipitation product, which provides hourly gridded (4 km × 4 km) precipitation field by combining both WSR-88D Nexrad radar precipitation estimates and rain gauge reports, with the missing periods supplemented by the stational data (Breidenbach et al., 1998; Fulton et al., 1998). The precipitable water (a.k.a., total column water vapor, defined as the depth of liquid water that would result if all the water vapor in an atmosphere column were precipitated as rain) available in the continuous forcing data set is retrieved from microwave radiometers at up to five ARM stations.

The derived all-sky radiative fluxes, including downwelling/upwelling shortwave and longwave radiative fluxes in the continuous forcing data sets, are based on 14 radiometers in the Solar and Infrared Observation Stations. The Data Quality Assessment for ARM Radiation Data methodology is applied to use climatological analyses of the surface radiation measurements to control the quality of the data (Long & Shi, 2006). For the clear-sky surface radiation product (not included in continuous forcing), we use ARM's Radiative Flux Analysis data set that follows an empirical fitting method to determine the clear-sky fluxes at cloudy times (Long & Ackerman, 2000; Long & Turner, 2008). The measurements from about 27 ARM facilities at the SGP site are used to calculate the domain mean clear-sky fluxes climatology. The satellite-retrieved data product, the Clouds and the Earth's Radiant Energy System-Energy Balanced and Filled (CERES-EBAF) surface product ed2.8 is also used to evaluate both the all-sky and clear-sky surface radiative fluxes. A detailed comparison between these two products is presented in section 3.2.1.

The surface sensible heat and latent heat fluxes are measured at ARM's Energy Balance Bowen Ratio (EBBR) stations since 1993 and Quality Controlled Eddy Correlation fluxes (QCECOR) stations since 2003 (Berg & Lamb, 2016). The vertical fluxes of sensible and latent heat produced by the EBBR systems are estimated from the vertical temperature and humidity gradients. The Bulk Aerodynamic technique is applied to the EBBR data streams to address sunrise and sunset spikes in the fluxes data (Cook, 2011a: EBBR handbook). The ECOR technique estimates the vertical fluxes by correlating the vertical wind component with temperature (sensible heat flux) and humidity (latent heat flux) (Cook, 2011b: ECOR handbook). The EBBR stations are often deployed at stable land, such as pasture and grassland, while QCECOR stations are usually at disturbed land such as cropland and wooded land. The multiyear monthly climatology of surface latent and sensible heat fluxes is constructed by averaging over the measurements from available EBBR and QCECOR stations during the period from 1999 to 2011. Measurements from up to 19 EBBR and 13 QCECOR stations are used to calculate the domain mean.

Soil moisture data are from the Soil Water and Temperature Systems (SWATSs) (Bond, 2005: SWATS handbook). Two profiles of sensors are installed 1 m apart from each other that perform measurement at eight different depths. To calculate the soil moisture variable equivalent to the model output variable (mrsos: soil moisture-integrated over uppermost 10 cm layer), the volumetric soil moisture measured by two sensors for top 5 cm and 15 cm depths are averaged for each site. Data from a total of 22 sites are used for generating the domain mean climatology from 1999 to 2011.

The Aerosol Optical Depths (AODs) are available from MultiFilter Rotating Shadowband Radiometer (MFRSR) deployed at up to 31 ARM sites. A review of the utility of the narrowband radiometer can be found in Michalsky and Long (2016) and McComiskey and Ferrare (2016). To compare with model output of AOD output at 550 nm (od550aer), the monthly mean AOD_{500} is extrapolated to AOD_{550} following

$$AOD_{550} = AOD_{500}(550/500)^{-\alpha},$$

Table 1
Observed Quantities Considered in the Evaluation, Including the Quantity Names, the Data Sources, the Temporal and Spatial Information of the Data, the Original Data Sources or Instrument, and the Estimated Observational Uncertainties

Quantity	Data sources	Temporal/spatial resolution	Original data sources/instruments	Estimated uncertainty[a]
Surface screen-level temperature/humidity	Continuous forcing (1999–2011)	Hourly; averaged over 3° × 3° domain	Surface Meteorological Observation System, Oklahoma and Kansas mesonet stations	0.2–0.4 K for temperature 1–3% for relative humidity (Campos & Sisterson, 2015; Ritsche, 2008)
Temperature/humidity/ wind profile	Same as above	Same as above	NOAA/NCEP Rapid Update Cycle analysis data	N/A
Surface precipitation	Same as above	Same as above	Arkansas-Red Basin River Forecast Center Nexrad radar precipitation estimates w/ rain gauge (Breidenbach et al., 1998; Fulton et al., 1998)	40% (Xie et al., 2014)
Precipitable water	Same as above	Same as above	Microwave Water Radiometer: water liquid and vapor along line of sight path	0.5–0.7 mm (Campos & Sisterson, 2015)
Surface all-sky radiative fluxes	Same as above	Same as above	Data Quality Assessment for ARM Radiation Data (Long & Shi, 2006, 2008; Phillips & Klein, 2014)	10 Wm^{-2} (longwave down) 5 Wm^{-2} (longwave up) 20 Wm^{-2} (shortwave down) 5 Wm^{-2} (shortwave up)
	CERES-EBAF surface (2000–2015)	Monthly; 1° × 1°	Satellite-retrieved data (Kato et al., 2013)	17–19 Wm^{-2} (longwave) 12 Wm^{-2} (shortwave)
Surface clear-sky radiative fluxes	ARM: RADFLUXANAL (1999–2011)	60 s; site averaged	Radiative Flux Analysis for clear skies (Long & Ackerman, 2000; Long & Turner, 2008)	20 Wm^{-2} (shortwave) 4–5 Wm^{-2} (longwave)
	CERES-EBAF surface (2000–2015)	Monthly; 1° × 1°	Satellite-retrieved data (Kato et al., 2013)	17–19 Wm^{-2} (longwave) 12 Wm^{-2} (shortwave)
Surface latent/sensible heat	ARM: BAEBBR (1999–2011)	30 min; site averaged	Best-Estimate Fluxes from EBBR Measurements and Bulk Aerodynamics Calculations (BAEBBR) (Cook, 2011a)	± 10%
	ARM: QCECOR (1999–2011)	30 min; site averaged	Quality Controlled Eddy Correlation Flux Measurement (Cook, 2011b)	±5% (latent heat) ±6% (sensible heat)
Surface layer soil moisture content	ARM: SWATS (1999–2011)	Hourly; site averaged	Soil Water and Temperature System (SWATS) (Bond, 2005)	±5 mm for 10 cm depth
Aerosol optical depth 550 nm	ARM: MFRSRAOD1MICH (1999–2011) AERONET Cloud and Radiation Testbed (cart) sgp (1999–2011)	15 s; two sites 15 min; single site	Multifilter Rotating Shadowband Radiometer (MFRSR) The CIMEL Sun radiometer	~0.01 (Knootz et al., 2013) ~0.01 (Eck et al., 1999)

[a]The estimated observational uncertainty listed here is referred from the data product handbooks and/or pertinent references. The uncertainties reported usually consider systematic errors including instrument errors/ retrieval errors at observational frequencies. Because climatology monthly mean of each variable is evaluated in this study, we assume that the random errors from sampling both in space and time are negligible in comparison to these estimated observational uncertainties.

where the Ångström exponent, α, is estimated using the AODs measured at the wavelengths 415 nm and 615 nm. The AOD measurements taken near by the SGP central facility from the National Aeronautics and Space Administration (NASA) Aerosol Robotic Network (AERONET) are also included as an evaluation data set. This single-site measurement is compared with data sets obtained from two ARM's MFRSR stations that are almost colocated with the AERONET station. By evaluating the annual cycles of AOD, we also note that the values obtained from the central facility can well represent the domain mean climatology calculated by averaging all available MFRSR stations.

The multiyear monthly climatology is constructed for all observed variables analyzed in this paper. The climatology of the observational data sets is formed for the period from 1999 to 2011, except for that of the variables from CERES-EBAF product which uses data available from 2000 to 2015.

2.2. CMIP5 AMIP Simulations

Simulations of 23 models contributing to the CMIP5 (Taylor et al., 2012) multimodel experiments have been used (see Table 2 for details). We evaluate the multiyear monthly mean climatology data from these models from the CMIP5 atmospheric only (AMIP) experiments from years 1979 to 2008. In AMIP experiments, atmosphere and land models are integrated freely subject to the boundary conditions of observed sea surface temperatures and sea ice content. All data have been linearly interpolated to 3° × 3° resolution with center located at the SGP central facility to create a unified grid that can be compared with the continuous forcing product. We note that only two CMIP5 models that we analyzed here (CESM1-CAM5 and CanAM4) participated in the CAUSES hindcast intercomparison project.

3. Results

3.1. Climate Model Bias in Summer at the ARM SGP Site

Consistent with earlier work that examined the simulation across whole continents (Cheruy et al., 2014), we find a summertime surface air temperature bias of a similar magnitude over the much smaller domain at the ARM SGP site (Figure 1a). All 23 evaluated models exhibit a positive June–July–August (JJA) mean 2 m air temperature (T2 m) bias ranging from 0.7 to 10.4 K, with a multimodel mean of 5.1 K and a standard deviation of 1.9 K. Because the years used to form climatology are different between model output and observational data, in all the annual cycle figures presented in this paper, we provide error bars as 1 standard deviation from each monthly mean as a measure of the dispersion of observed monthly mean data from the mean climatology.

Concurrently with the T2 m bias, model simulations also show systematic bias in the fields of surface precipitation, relative humidity (with respect to liquid), and column-integrated water vapor in JJA (Figures 1b–1d). All models underestimate the JJA surface precipitation with a multimodel mean bias of −1.7 mm/d. Most models predict a May precipitation peak instead of the observed June peak. The multimodel mean JJA relative humidity is underestimated by 21% (Note here that surface relative humidity is only available for 18 models). Underestimation is also found in simulated specific humidity at the surface (not shown). Except for the two MPI models: MPI-ESM-MR and MPI-ESM-LR, all models tend to predict drier atmospheres with a multimodel mean water vapor deficit of 5.4 mm in JJA.

Figure 2 displays the vertical profiles of temperature and specific humidity from the models and the continuous forcing. The model biases for both variables tend to decrease with increasing altitude. Below 850 mb, all models have a warm bias except for GISS-E2-R, which only shows slight warm bias near the surface; while specific humidity is underestimated by all the models. Shown in Figures 1 and 2, the surface warm and dry bias during SGP summer is robust for almost all the models although intermodel spread is considerable.

3.2. Surface Energy Budget

The near-surface temperature is driven primarily by the exchange of energy between land surface and atmosphere. To better understand the near-surface temperature bias, we examine the multimodel mean and cross model bias spread in each surface energy budget component for JJA mean. The surface energy conservation requires

$$(1 - \alpha)\,\text{SWDN} + \text{LWDN} = \text{LWUP} + \text{LH} + \text{SH} + \text{G}, \tag{1}$$

Table 2
Information of Models Used in the Evaluation, Including the Modeling Institutions, Model Standard Names and Pertinent References

| Modeling institution	Model name	Original ATM grids	References
Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorolog, Australia	ACCESS1.0	144 × 192	Bi et al. (2013)
	ACCESS1.3	144 × 192	
Beijing Climate Center, China Meteorological Administration	BCC-CSM1.1	64 × 128	Wu et al. (2010)
	BCC-CSM1.1(m)	160 × 320	
College of Global Change and Earth System Science, Beijing Normal University	BNU-ESM	64 × 128	Ji et al. (2014)
Canadian Centre for Climate Modeling and Analysis	CanAM4	64 × 128	Chylek et al. (2011)
National Center for Atmospheric Research Community Earth System Model Contributors	CCSM4	192 × 288	Gent et al. (2011)
	CESM1-CAM5	192 × 288	Neale et al. (2010)
Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence	CSIRO-Mk3-6-0	96 × 192	Rotstayn et al. (2012)
LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and CESS, Tsinghua University	FGOALS-g2	64 × 128	Xu et al. (2013)
	FGOALS-s2	108 × 128	Bao et al. (2013)
NOAA Geophysical Fluid Dynamics Laboratory	GFDL-HIRAM-C360	360 × 360	Zhao (2009)
	GFDL-HIRAM-C180	180 × 180	
NASA Goddard Institute for Space Studies	GISS-E2-R	90 × 144	Schmidt et al. (2006, 2014)
Met Office Hadley Centre	HadGEM2-A	144 × 192	Martin et al. (2006) and Collins et al. (2011)
Institute Pierre-Simon Laplace	IPSL-CM5A-LR	96 × 96	Dufresne et al. (2013) and Hourdin et al. (2013)
	IPSL-CM5B-LR	96 × 96	
	IPSL-CM5A-MR	142 × 144	
Institute for Numerical Mathematics	Inmcm4	120 × 180	Volodin et al. (2010)
Atmosphere and Ocean Research Institute, National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	MIROC5	128 × 256	Watanabe et al. (2011)
Max Planck Institute for Meteorology	MPI-ESM-MR	96 × 192	Jungclaus et al. (2013)
	MPI-ESM-LR	96 × 192	Stevens et al. (2013)
Norwegian Climate Centre	NorESM1-M	96 × 144	Bentsen et al. (2013)
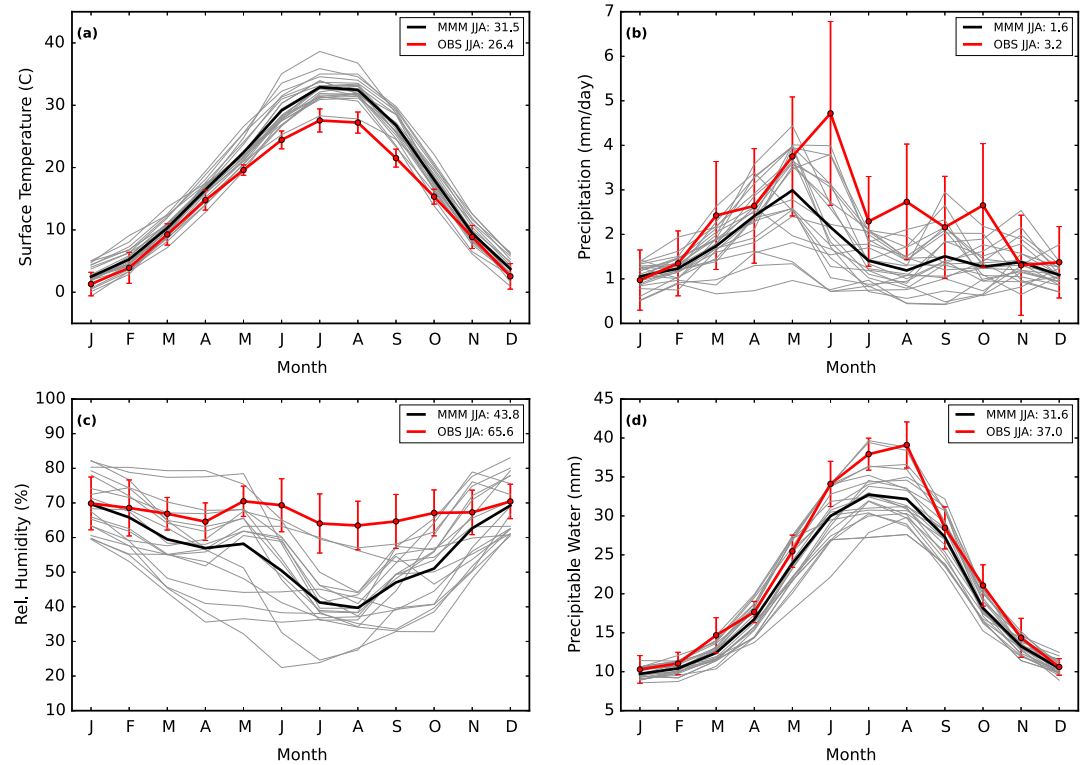
Figure 1. Annual cycle of monthly mean of (a) surface air temperature, (b) precipitation, (c) surface air relative humidity, and (d) precipitable water over the Atmospheric Radiation Measurement (ARM) Southern Great Plains domain (averaged over 35–38°N, 99–96°W) from ARM observation averaged over 1999 to 2011(red line with error bars representing 1 standard deviation of interannual variability) and Coupled Model Intercomparison Project Phase 5 simulations averaged over 1979 to 2008 (gray lines for individual Coupled Model Intercomparison Project Phase 5 models and black line for multimodel mean). JJA mean values are shown in the legend.
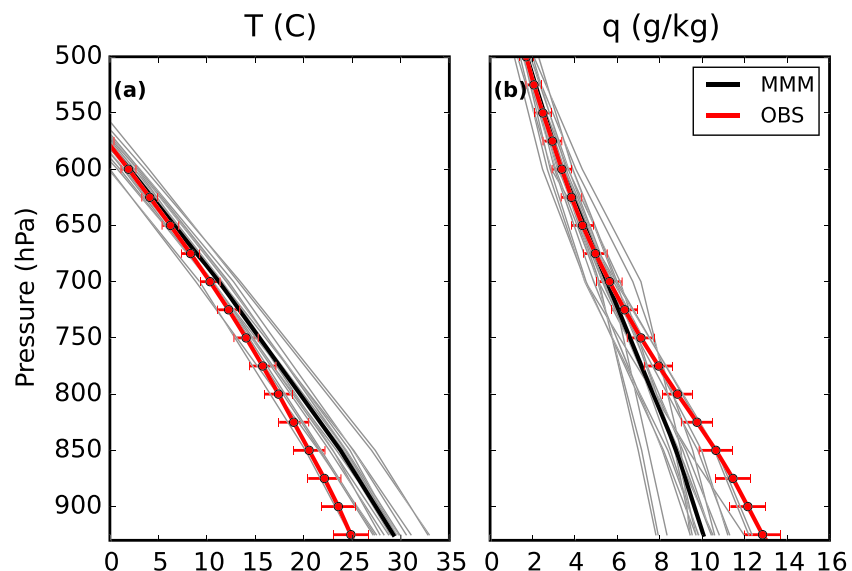
Figure 2. Vertical profiles of JJA mean over the Atmospheric Radiation Measurement Southern Great Plains domain (a) temperature and (b) specific humidity with same line types and temporal averaging as Figure 1.

Table 3

JJA Mean of the Surface Energy Budget Components From the Observations (With Standard Deviation of Interannual Variability) and the Models (Positive Downward)

Surface energy budget components (W/m²)	ARM observation (W/m²)	Multimodel mean (W/m²)	Mean bias (W/m²)
Shortwave down	273.5 ± 18.2	300.6	27.1
Shortwave up	−52.8 ± 4.5	−50.4	2.4
Shortwave net	220.7 ± 13.9	250.2	29.5
Longwave down	398.8 ± 8.0	401.1	2.3
Longwave up	−460.6 ± 11.9	−500.5	−39.9
Longwave net	−61.8 ± 7.4	−99.4	−37.6
Sensible heat flux	−52.3 ± 11.6	−80.5	−27.8
Latent heat flux	−101.4 ± 13.3	−61.2	40.4

Note. The mean bias is calculated from the multimodel means and the observed means. ARM = Atmospheric Radiation Measurement.

where α is surface albedo, SWDN is downward shortwave flux, LWDN (UP) is longwave downward (upward) fluxes, LH and SH are latent and sensible heat fluxes, G is ground heat flux, respectively. The ground heat flux is small relative to the other terms when averaged over long time scales (such as monthly means).

Table 3 shows the JJA multimodel mean and observed values for the surface energy fluxes terms. The CMIP5 simulations on average overestimate the shortwave radiation absorbed by the surface by about 30 W/m² . The net longwave bias of about −38 W/m² mostly stems from the upward flux due to a stronger longwave radiative cooling of an overly warm surface. The sensible heat flux is overestimated by 28 W/m² , and the latent heat flux is underestimated by 40 W/m² , partially the result of the warmer and drier surface. The uncertainties from turbulent fluxes measurements are known to be greater than those of the radiation measurements (Brotzge & Crawford, 2003; Teixeira & Bastiaanssen, 2012). In our analysis, the magnitude of the multimodel mean bias is greater than the observational uncertainties of both sensible and latent heat fluxes. More rigorous uncertainty estimates that separate systematic bias and random errors will be useful to quantify the observational uncertainty in the surface turbulence fluxes.

The bias distribution of surface energy components varies greatly among the CMIP5 models (Figure 3). The downward solar radiation at the surface is overestimated by most models. Only two models simulate a negative shortwave bias, of which the GISS-2E-R model underestimates flux by as large as −28 W/m² due to a strongly overpredicted cloud radiative effect, as will be shown later. For the downward longwave radiation, although the multimodel mean bias is only 2.3 W/m² , the majority of the models (14 out of 23) show a positive bias. The relatively smaller bias in LWDN shown in these models is partially due to the compensating effect between the underestimation of precipitable water and clouds and the overestimation of lower troposphere temperature as shown earlier. The former acts to reduce the amount of downward longwave fluxes reaching the surface, while the latter acts oppositely.

The multimodel mean upward shortwave is underestimated by 2.4 W/m² , with most models (13 out of 23) predicting too little reflected solar radiation (Figure 3). Corresponding to the warm bias presented, all CMIP5 models show a positive upward longwave radiation. For the turbulent fluxes, most models overestimate the sensible heat flux and underestimate the latent heat flux, with the bias spread of 61 W/m² and 68 W/m² , and the standard deviation of 18.0 and 23.3 W/m² for these two fields, respectively.

The detailed analysis of the individual components of the radiative and nonradiative (turbulent) fluxes is analyzed in sections 3.2.1 and 3.2.2.

3.2.1. Surface Radiative Flux Biases

To consider uncertainty in the surface radiative fluxes measurements, data from both ground-based (ARM) and satellite-based (CERES-EBAF, ed2.8, (Kato et al., 2013)) products are used. The JJA mean net shortwave biases in the CMIP5 multimodel mean are 30 W/m² and 26 W/m² against the two observations, respectively (second column of Table 4). Figure 4 shows that there is a positive correlation among models between the near-surface air temperature bias and the net shortwave bias. Except for the GISS model, all models exhibit a positive net shortwave bias for JJA based on both ARM and CERES-EBAF. This figure also shows that there would be a warm bias present even if the shortwave radiation error were zero, indicating other processes also contributing to the surface temperature bias.

To quantify the size of radiation errors in the shortwave from different sources, the net shortwave error can be partitioned into three sources—clear-sky absorption/reflection, cloud effects, and surface albedo—from the following short derivation. The net shortwave error $NetSW_{err}$ can be written as follows:

$$NetSW_{err} = SWDN_{mod} - SWDN_{obs} - SWUP_{mod} + SWUP_{obs}, \qquad (2)$$
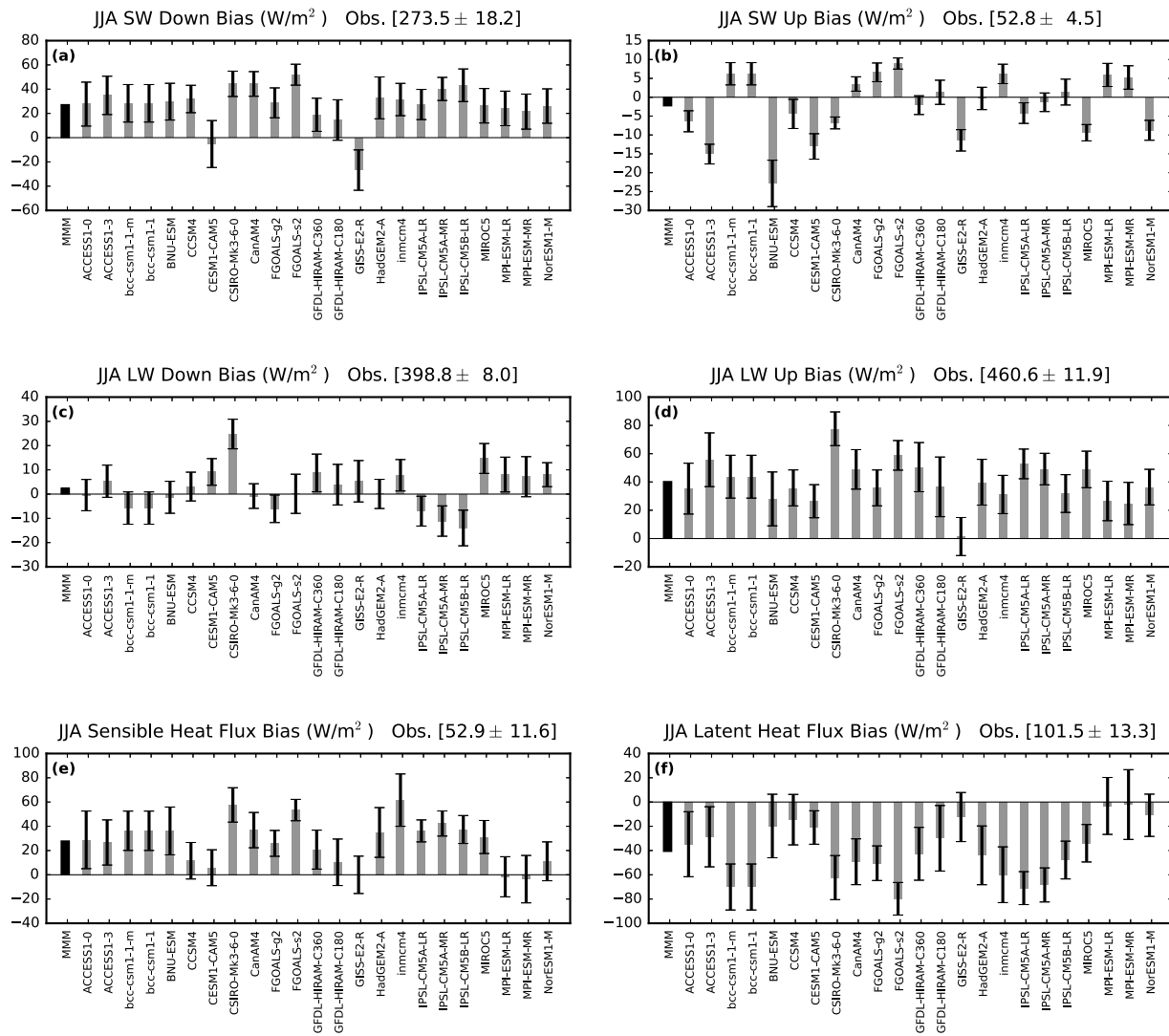
Figure 3. The Coupled Model Intercomparison Project Phase 5 model bias of the 1979–2008 JJA mean surface energy budget components over the Atmospheric Radiation Measurement Southern Great Plains domain, including (a) shortwave downwelling flux, (b) shortwave upwelling flux, (c) longwave downwelling flux, (d) longwave downwelling flux, (e) sensible heat flux, and (f) latent heat flux. Biases are calculated as absolute value of models minus that of observed, without considering the direction of the fluxes. Gray bars are for individual models with error bars indicating interannual standard deviation, black bar for Multimodel Mean. Mean observed values and interannual standard deviations are shown in the brackets above the plots.

where $SWDN_{mod/obs}$ and $SWUP_{mod/obs}$ are downwelling and upwelling shortwave fluxes from models and observations, respectively. Given that $SWUP = \alpha SWDN$, where α is surface albedo, equation (2) can be written as follows:

$$NetSW_{err} = (1 - \alpha_{mod})SWDN_{mod} - (1 - \alpha_{obs})SWDN_{obs}. \qquad (3)$$

Table 4

Net SW Error Attribution of JJA Mean Net Shortwave Fluxes: From Left to Right, Each Column Gives Observation Data Sources, Net Shortwave Error, Albedo Weighted Error From Clear Sky, Albedo Weighted Error From Cloud Radiative Effect, and Albedo Error Including Both Weighted and Absolute Errors and Error Covariance Term

Observational data set	Net SW error (W/m^2)	SW$_{clr}$ error (W/m^2)	SW$_{cre}$ error (W/m^2)	Albedo error (W/m^2) [absolute albedo error]	Error covariance
ARM	29.5 ± 15.4	4.9 ± 4.2	17.0 ± 11.6	6.9 ± 7.0 [−0.025 ± 0.025]	0.7 ± 1.2
CERES-EBAF	26.2 ± 15.4	12.5 ± 4.3	12.2 ± 11.9	1.3 ± 7.0 [−0.005 ± 0.025]	0.1 ± 1.0

Note. JJA mean bias and ±1 standard deviation of multimodel values. The downward direction is considered positive for fluxes. ARM = Atmospheric Radiation Measurement; CERES-EBAF = Clouds and the Earth's Radiant Energy System-Energy Balanced and Filled.
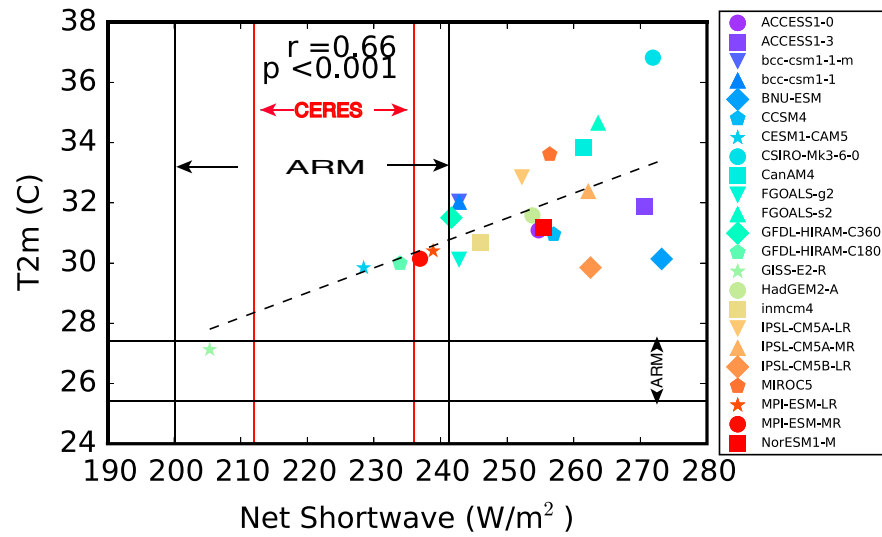
Figure 4. The relationship between JJA mean T2 m and net shortwave radiation at the surface. Individual models are shown as scatter symbols. The range of observational uncertainty for each observation is estimated by the larger of (1) the calculated 95% Confidence Interval based on interannual variability of observed values and (2) the estimated observational uncertainty from Table 1 (if available). This method is applied to all other scatter plots in this paper. The observational data periods are from 1999 to 2011 for Atmospheric Radiation Measurement (ARM) and from 2000 to 2015 for Clouds and the Earth's Radiant Energy System (CERES).

To isolate the surface albedo error ($\alpha_{err} = \alpha_{mod} - \alpha_{obs}$), equation (3) is rearranged as follows:

$$\text{NetSW}_{err} = (1 - \alpha_{obs})\text{SWDN}_{err} - \alpha_{err}\text{SWDN}_{obs} - \alpha_{err}\text{SWDN}_{err}, \tag{4}$$

In order to further partition the SWDN_{err} into errors from clear sky and cloud radiative effect, by using the definition, $\text{SWDN} = \text{SWDN}_{cre} + \text{SWDN}_{clr}$, we arrive with the final equation for net shortwave error partitioning:

$$\text{NetSW}_{err} = (1 - \alpha_{obs})\text{SWDN}_{clr,err} + (1 - \alpha_{obs})\text{SWDN}_{cre,err} - \alpha_{err}\text{SWDN}_{obs} - \alpha_{err}\text{SWDN}_{err} \tag{5}$$

where the first two terms on the right side of the equation are the surface absorption-weighted downward shortwave clear-sky error and cloud radiative effect error at the surface, respectively, the third term is the albedo error weighted by observed SWDN, and the fourth is the covariance term from both albedo and SWDN errors, which is negligible (shown in the last column of Table 4). Note that the calculation of cloud radiative effect term involves clear-sky fluxes, so the cloud radiative effect error is not purely a cloud error. The JJA mean of these error terms is displayed in Table 4, while the bias components from individual models are illustrated in Figure 5.

For the multimodel mean, the cloud radiative error accounts for 58% (with respect to ARM) and 47% (to CERES-EBAF) of the net solar radiation error at the surface. A cloud deficit results in excess solar radiation absorbed at the surface. Based on the satellite observations, the clear-sky error has about the same magnitude as the cloud error. However, the clear-sky error is much smaller if the calculation is based on the ground-based observations. Models also tend to underestimate the surface albedo, which contributes to a positive bias in the net solar radiation at the surface. This error is smaller based on the satellite observations as compared to the ARM observations. Understanding the causes of surface albedo biases in climate models is challenging since different models may use different surface albedo parameterizations. According to studies from Li et al. (2016) and Levine and Boos (2017), large source of land surface albedo error comes from the representation of land use type (i.e., vegetation and soil properties). The intermodel correlation coefficients between bias in summertime T2 m and three NetSW_{err} components are the following: 0.65, 0.39, and 0.09 for cloud, clear sky, and surface albedo errors, respectively.

We note here that caution is required in quantifying relative errors from clear sky and cloudy sky, because the observations and climate models usually have different ways to calculate clear-sky values. The ARM clear-sky
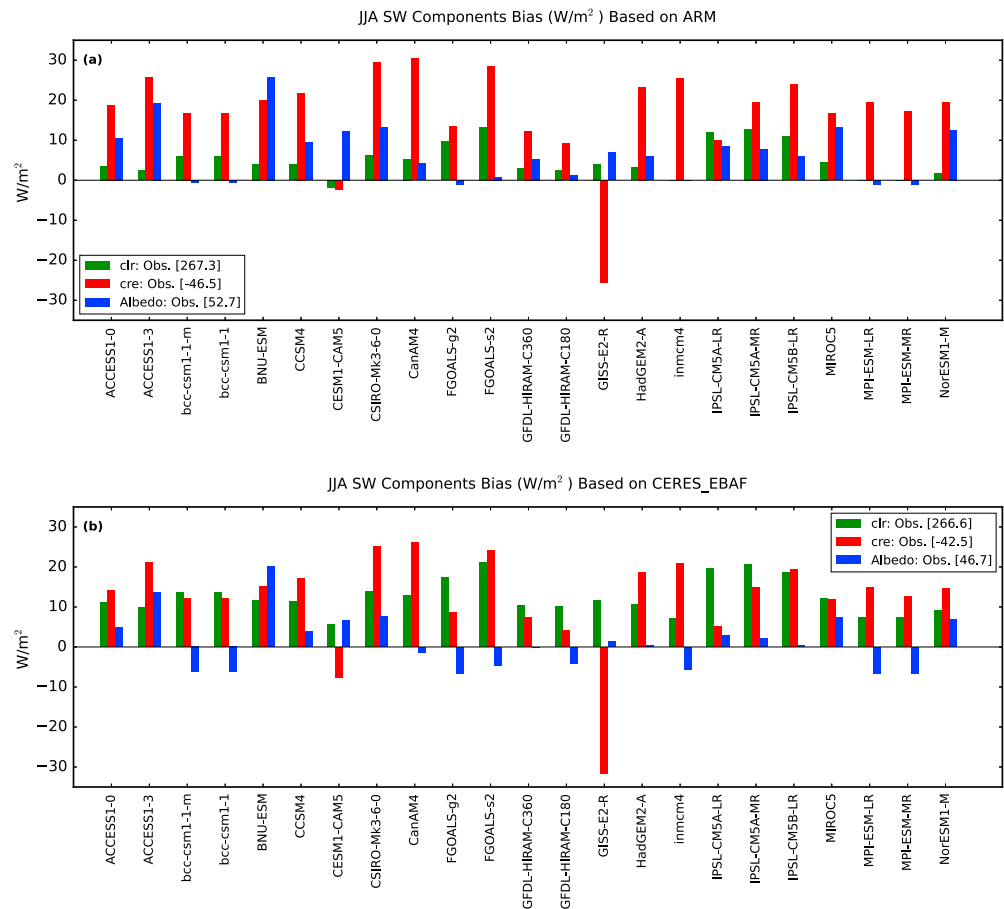
Figure 5. JJA mean net shortwave bias components including the surface absorption-weighted shortwave clear-sky error, the surface absorption-weighted cloud radiative effect error at the surface, and the albedo error weighted by observed downwelling shortwave flux, corresponding to the first three terms in equation (5), calculated based on (a) Atmospheric Radiation Measurement (ARM) and (b) Clouds and the Earth's Radiant Energy System-Energy Balanced and Filled (CERES-EBAF) (positive bias means too much energy gain at surface) over the ARM Southern Great Plains domain. The corresponding JJA mean observed values are shown in the legend box.

flux value is calculated following the Radiative Flux Analysis (Long & Ackerman, 2000), for which clear-sky fluxes are identified from clear-sky-only periods. An empirical cosine fit based on solar zenith angle is applied to fill in cloudy time periods in order to generate continuous data stream. For the CERES-EBAF data, the clear-sky fluxes are calculated by removing the cloud at all times (including cloudy times), followed by a weighting based on the clear-sky fraction. However, in climate models, clear-sky fluxes are calculated by removing cloud with no clear-sky fraction weighting applied. It is important to note that after applying a similar clear-sky sampling to the satellite data and ARM clear-sky fluxes, the two data sets showed excellent agreement (D. Rutan, personal communication). It is recommended that same sampling criteria are necessary when comparing model data and observations.

Figure 5 shows the cross-model distribution of error components in net shortwave, based on the ARM and CERES data, respectively. Even though the multimodel mean clear-sky error calculated based on CERES is approximately equal in size to the cloud radiative effect error, the majority of the investigated models show larger cloud error than clear-sky error no matter which observation is used. The standard deviations of the error components across models are about 12 Wm^{-2} for the cloud radiative effect error, 7 Wm^{-2} for the albedo error, and 4 Wm^{-2} for the clear-sky error.

Because atmospheric absorption of solar radiation by water vapor affects the amount of solar radiation reaching the surface under clear skies, we performed intermodel correlation analysis to explore the contribution of errors in water vapor absorption of solar radiation to errors in downwelling shortwave radiation at the surface.
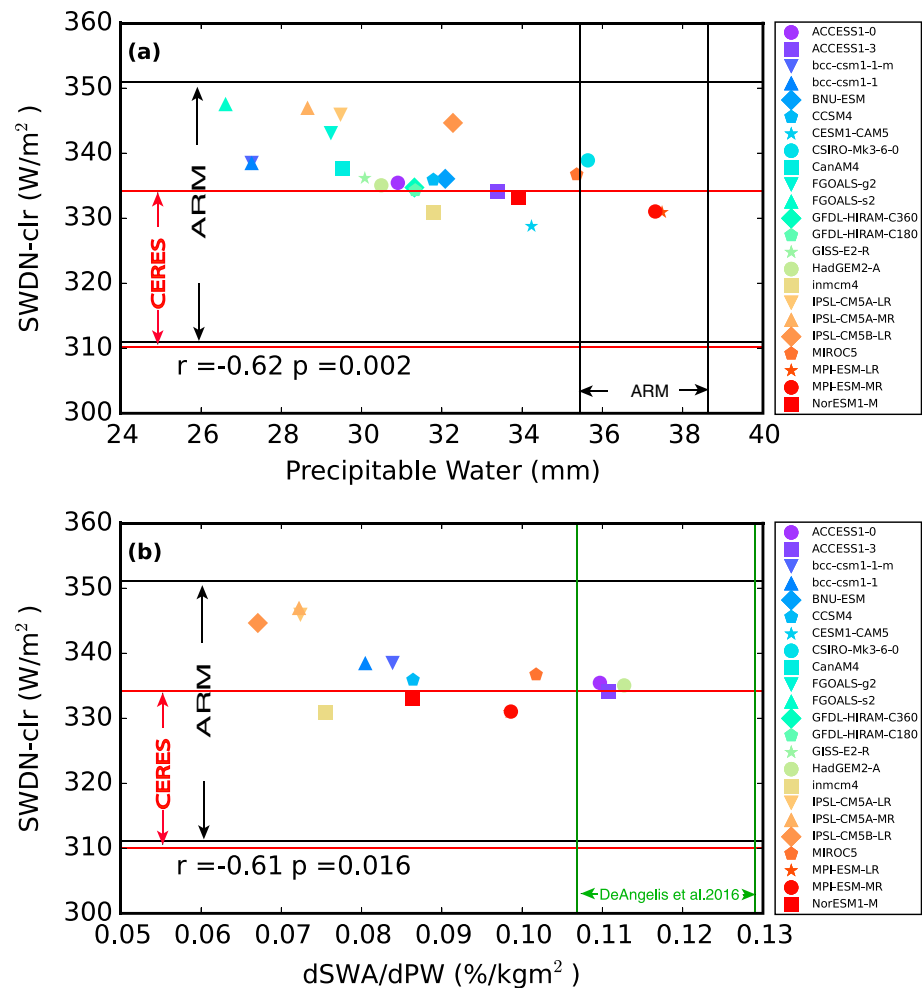
Figure 6. Same as Figure 4 but for the following: (a) JJA mean SWDN$_{clr}$ versus precipitable water and (b) JJA mean SWDN$_{clr}$ versus dSWA/dPW. Observational uncertainties are estimated following method described in Figure 4 caption. The observation range of dSWA/dPW is obtained from DeAngelis et al. (2015) (width of green vertical lines).

First, there is the expected anticorrelation among the models between biases in the amount of precipitable water (PW) and clear-sky shortwave downward radiation fluxes. Less water vapor would cause less absorption of solar radiation within the atmosphere allowing more solar radiation to reach the surface under clear skies (Figure 6a). However, even if models have the correct amount of water vapor, they could underestimate the amount of absorption of solar radiation by water vapor if their radiation schemes were in error. We quantify this error by examining the sensitivity of shortwave absorption within the atmosphere (SWA) to increases in PW (DeAngelis et al., 2015). We find the expected anticorrelation between dSWA/dPW and the error in the downward shortwave radiation at the surface under clear skies (Figure 6b) with models tending to underestimate the sensitivity of solar absorption to variation of atmospheric water vapor (dSWA/dPW). This leads to an underestimation in the atmospheric shortwave absorption within the atmosphere, and hence too much solar radiation reaching the surface. We note that the value of dSWA/dPW is taken from DeAngelis et al. (2015) for the models in common with our study. The true value of dSWA/dPW is estimated to be 0.11–0.13% kg^{-1} m^2 (DeAngelis et al., 2015).

While off-line radiation calculations would more conclusively be able to demonstrate these points (Fildier & Collins, 2015; Pincus et al., 2015), these correlations are suggestive of a role for errors in shortwave absorption in the atmosphere by water vapor (due to both errors in the amount of water vapor and the radiation scheme itself) contributing to too much solar radiation reaching the surface. For instance, the ACCESS1.0, ACCESS1.3 and HadGEM2 models are all using the same Edwards and Slingo radiation code (Edwards & Slingo, 1996) for
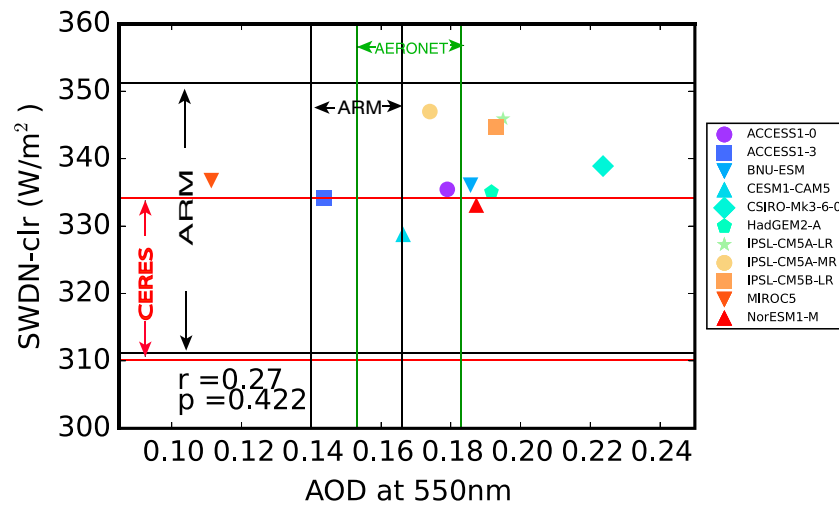
Figure 7. Same as Figure 4 but for JJA mean SWDN$_{clr}$ versus aerosol optical depth at 550 nm wavelength. The Aerosol Robotic Network (AERONET) observation is a single-site measurement from near the Southern Great Plains central facility, while the Atmospheric Radiation Measurement (ARM) observation is taken from two almost colocated sites. Measurements from 1999 to 2011 are used for both sites.

radiative transfer, with the calculated dSWA/dPW clustered and fallen within the observed range. However, the clear-sky shortwave fluxes from these three models can be still underestimated due to a lack of atmospheric water vapor.

The overestimation of solar radiation incident at the Earth's surface for clear sky is a long-standing issue in climate modeling (Liang et al., 2010; Wild, 2008; Wild et al., 2015). While radiation codes are improved in CMIP5 models relative to CMIP3 models (Wild et al., 2006), significant radiation errors remain in many CMIP5 models (Freidenreich & Ramaswamy, 2011). The errors in radiation could also be due to the fact that 3-D radiation effects are not represented in global models (Hogan & Shonk, 2013).

Errors in aerosol concentration or optical properties would also contribute to errors in the clear-sky radiation. Figure 7 displays the model simulations of aerosol optical depth in comparison to ARM and AERONET data. The models simulate a wide range of AOD at 550 nm (0.11–0.22), with most models overestimating the AOD. However, no clear correlation is found between AOD and clear-sky shortwave at surface. A positive bias in AOD would cause too much absorption and scattering of shortwave radiation and partially offset the positive downward bias due to weak water vapor absorption.

Similar to the net shortwave error, the net longwave error is partitioned into errors in upward radiation, downward clear-sky radiation, and cloud radiative effects (LWDN$_{cre}$ = LWDN − LWDN$_{clr}$) as following:

$$NetLW_{err} = LWUP_{err} + LWDN_{clr,err} + LWDN_{cre,err}, \tag{6}$$

corresponding to columns in Table 5. Note that we do not evaluate the much smaller errors arising from errors in the surface emissivity (Feldman et al., 2014; Huang et al., 2016). The upwelling longwave error due to warmer surface accounts for the majority of the net LW error in the multimodel mean. The upward longwave fluxes from

Table 5
Error Attribution of JJA Mean Net Longwave Fluxes: From Left to Right, Each Column Gives Observation Data Sources, Net Longwave Error, Upwelling Longwave Error, Downwelling Longwave Error From Clear Sky, and Downwelling Longwave Error From Cloud Radiative Effect

Observational data set	Net LW error (W/m^2)	LWUP error (W/m^2)	LWDN$_{clr}$ error (W/m^2)	LWDN$_{cre}$ error (W/m^2)
ARM	−37.6 ± 15.7	−40.0 ± 14.8	12.0 ± 9.3	−10.0 ± 3.7
CERES-EBAF	−40.3 ± 15.7	−33.0 ± 14.8	−4.3 ± 9.3	−3.4 ± 3.7

Note. The downward direction is considered positive. Positive values indicate too much energy gain. ARM = Atmospheric Radiation Measurement; CERES-EBAF = Clouds and the Earth's Radiant Energy System-Energy Balanced and Filled.
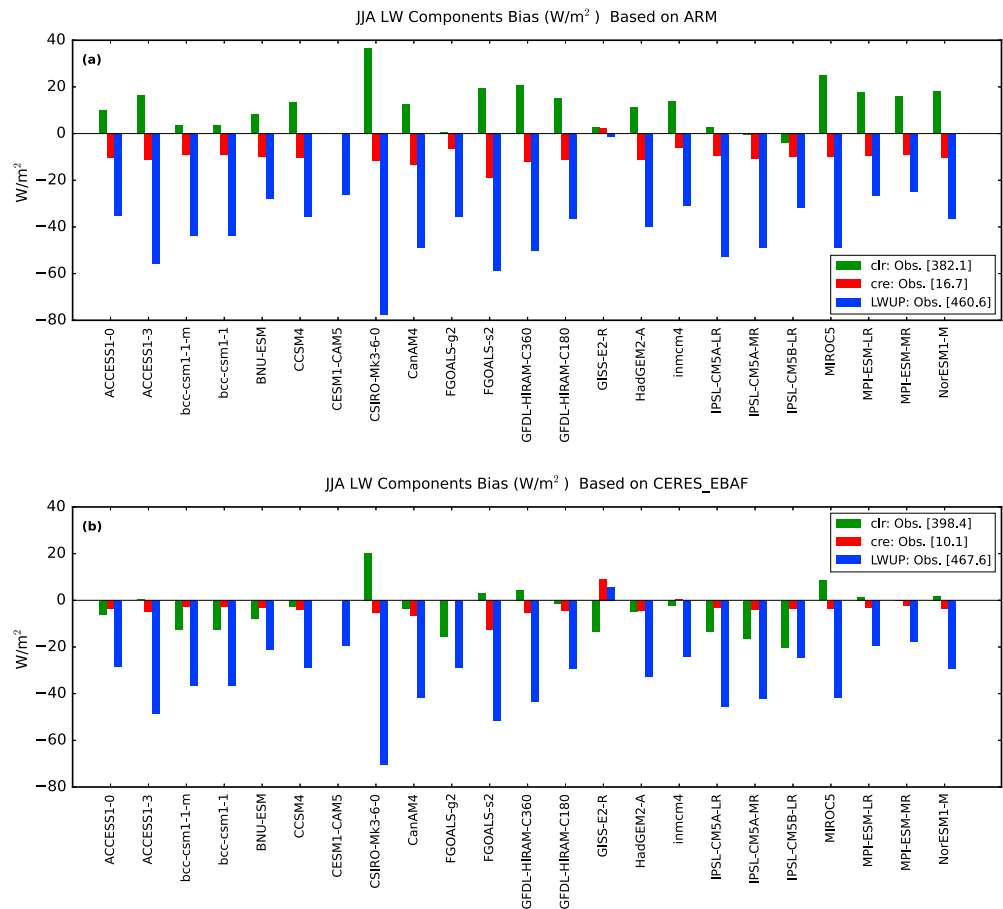
Figure 8. Same as Figure 5 but for longwave bias components: error from LWUP, $LWDN_{clr}$, and $LWDN_{cre}$. Note that $LWDN_{clr}$ is not available from CESM-CAM5, therefore $LWDN_{cre}$ missing.

both ARM and CERES-EBAF show a difference of 7 W/m^2 which suggests that there would be small surface temperature difference (~1 K, approximated from the Stefan-Boltzmann law) if one estimates temperature from the upward longwave fluxes using these two observational data sets.

For downward longwave error, the models tend to underestimate the LWDN in cloudy sky. The cloud deficit contributes to a negative cloud radiative effect error and therefore causes less warming at surface. Again, the clear-sky flux errors show large differences depending on the choice of observational data set. This outcome is consistent with the clear-sky shortwave flux analysis. Too little water vapor in the atmosphere would lead to an underestimate in the downward longwave that is partially counteracted by a too warm lower atmosphere that would lead to an overestimate in the downward longwave. The standard deviation of the error components across models is 14.8 Wm^{-2} for upward error, 9.3 Wm^{-2} for clear-sky error, and 3.7 Wm^{-2} for the cloud error. The intermodel standard deviation for cloud error (3.7 Wm^{-2}) is small relative to that in the shortwave. However, it is considered reasonable since the JJA mean cloud radiative effect on downward longwave has values of 15.7 and 10.1 Wm^{-2} from ARM and CERES, respectively (Figure 8), while $SWDN_{cre}$ has values around -50 Wm^{-2} (Figure 5).

3.2.2. Turbulent Flux and Soil Moisture Biases

The nonradiative terms in the surface energy budget are characterized by an underestimated latent heat flux and overestimated sensible heat flux (Table 3). The partitioning of surface turbulent energy fluxes into sensible versus latent heat has strong influence on surface air temperature. When Evaporative Fraction (EF = LH/(LH + SH)) decreases, the surface cannot cool by evaporation as much, so the surface temperature rises until the sensible heat and upward longwave radiation increase from increased surface temperature (plus increased ground heat flux) balances the absorbed shortwave and downwelling longwave radiation
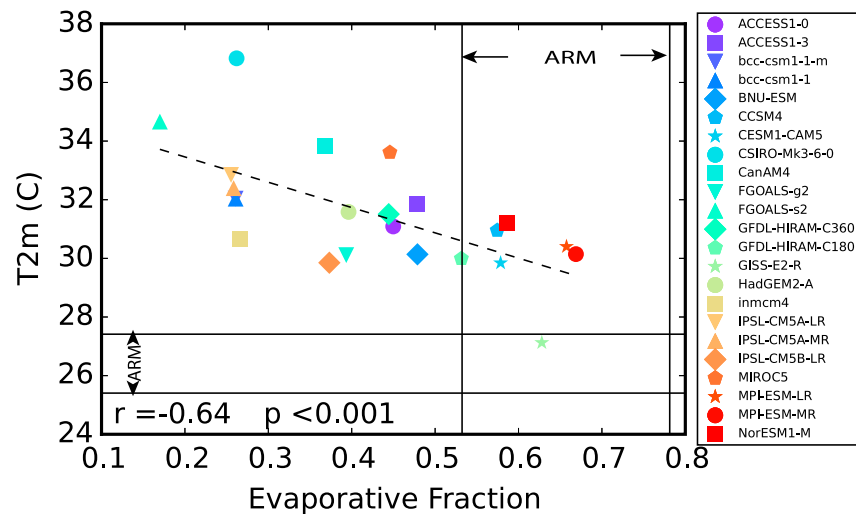
Figure 9. Same as Figure 4 except for different set of variables: JJA mean T2 m (*y* axis) versus evaporative fraction (*x* axis).

(Seneviratne et al., 2010). The multimodel JJA mean EF (0.43) is lower relative to ARM observations (0.66). Figure 8 shows that the cross-model spread of EF ranges from 0.16 to 0.67 with the majority of the models showing a negative bias. Given that the 95% confidence interval range is narrower, the observed uncertainty displayed in Figure 9 is estimated to be 0.53–0.78, considering the maximum of instrument uncertainty (±10% for latent and sensible heat fluxes). The conclusion that EF is underestimated in most models is robust.

Consistent with our physical explanation, models with smaller evaporative fraction tend to predict higher T2 m with intermodel correlation coefficient of −0.64 between evaporative fraction and T2 m. This finding is consistent with the Figure 2k from Cheruy et al. (2014), in which their correlation coefficient is −0.58. The difference in the correlation arises partly from a larger domain size and a slightly different selection of models used in Cheruy et al. (2014). In addition, the observed EF used in their study was estimated from a group of satellite-based products, reanalysis data, and off-line land surface models (Jimenez et al., 2011). Those data sources may be less reliable than the in situ ARM observations from a network of stations that we use.

The intermodel correlation coefficient between evaporative fraction and T2 m has approximately the same magnitude as that of the net surface solar radiation, indicating that both radiation and land-surface processes are closely related to the warm bias. For individual models, the GISS-E2-R model and the MPI-ESM models simulate EFs that are most consistent with observed value. Because the GISS-E2-R model underestimates the net surface solar radiation by simulating too strong cloud radiative effects, it has a much smaller T2 m bias than the MPI-ESM models, which overestimate net surface solar radiation.

Previous studies showed that for regions where the atmosphere and land interact effectively, soil moisture plays an important role in the partition of available surface energy and therefore has strong influence on surface temperature (Cheruy et al., 2013; Mueller & Seneviratne, 2014; Seneviratne et al., 2010). The soil moisture used for analyses in these previous studies was from model data, remote sensing satellite retrievals, or derived from the precipitation index (Mueller & Seneviratne, 2012). In our study, we use directly measured soil moisture provided by the ARM SWATS data to evaluate the models. We note that only a subset of models (see legend of Figure 10) is included due to the availability of soil moisture output.

Figure 10 shows the comparison of surface layer soil moisture content (vertically integrated moisture in the top 10 cm of the soil) between the observations and the models. All models significantly underestimate the soil moisture throughout the year (Figure 10a). For JJA mean values, the model spread ranges from 2.9 to 25.6 mm, with a multimodel mean of 13.1 mm compared to the observed value of 28.9 mm. Across the models being analyzed, there is an insignificant negative correlation coefficient between T2 m and soil moisture content (Figure 10b). This is not unexpected, even that T2 m correlates strongly to EF (Figure 9)
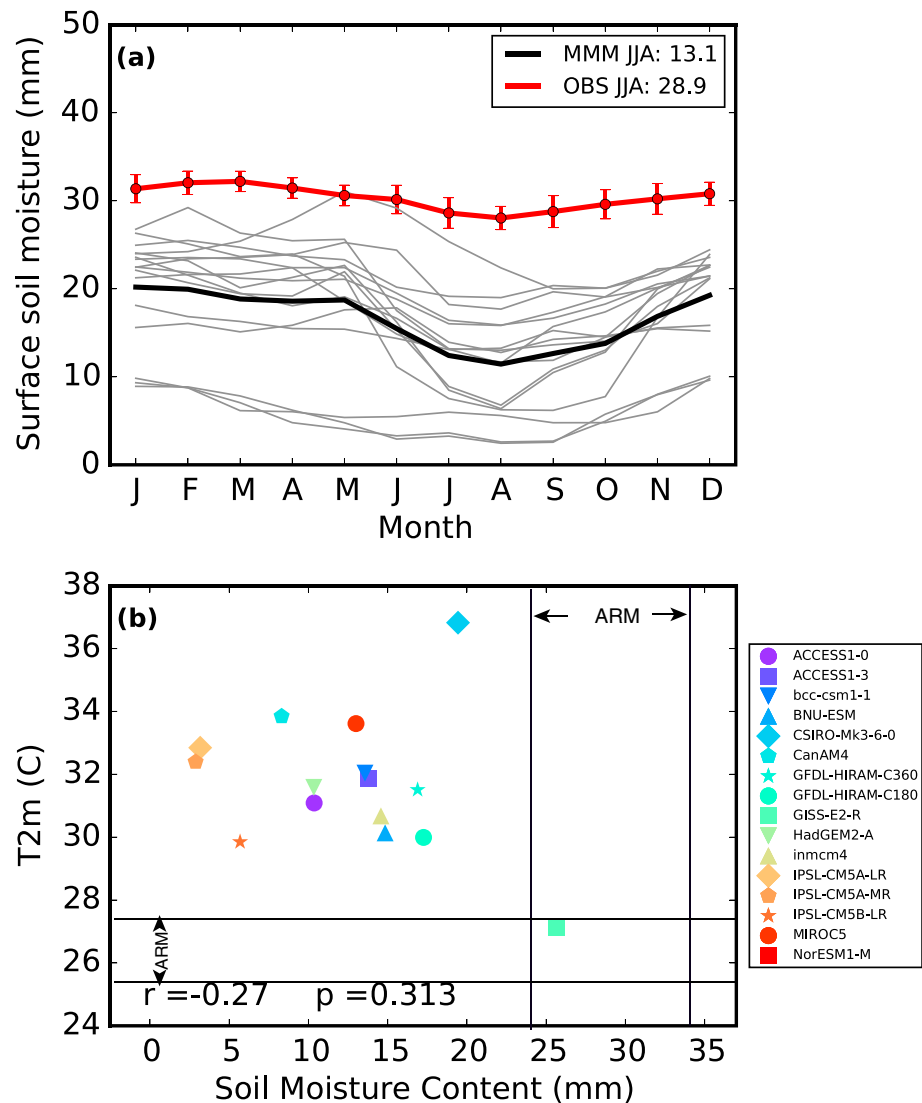
Figure 10. (a) is the same as Figure 1 except for surface 10 cm soil moisture content. (b) is the same as Figure 4 but showing the relationship between JJA mean T2 m and the soil moisture content. (A total of 16 models analyzed has surface soil moisture output available.)

and EF correlates strongly to soil moisture (not shown). Unlike net shortwave and EF, soil moisture does not affect the surface air temperature directly. The low soil moisture content limits the availability of water for evaporation and thus reduces the evaporative fraction and contributes to a warmer surface temperature indirectly. A recent study from Williams et al. (2016) shows that weakening the relationships between EF and soil moisture could damp the warm bias. The observed range displayed in Figure 10b is based on estimated observational uncertainty (±5 mm) from the SWATS handbook instead of the 95% CI range, which is narrower. We note that ARM has soil moisture measurements from other sensors; this study only uses SWATs due to better data availability, that is, less missing values and more years and sites available. Based on discussion provided by Phillips et al. (2017), the observation uncertainty for soil moisture can be larger if one considers the data from other sensors.

3.3. Atmospheric Water Cycle

To address the soil moisture deficit is challenging because the local interactions between soil moisture, evaporation, and precipitation are complex. Nonetheless, the atmospheric moisture transport also plays an
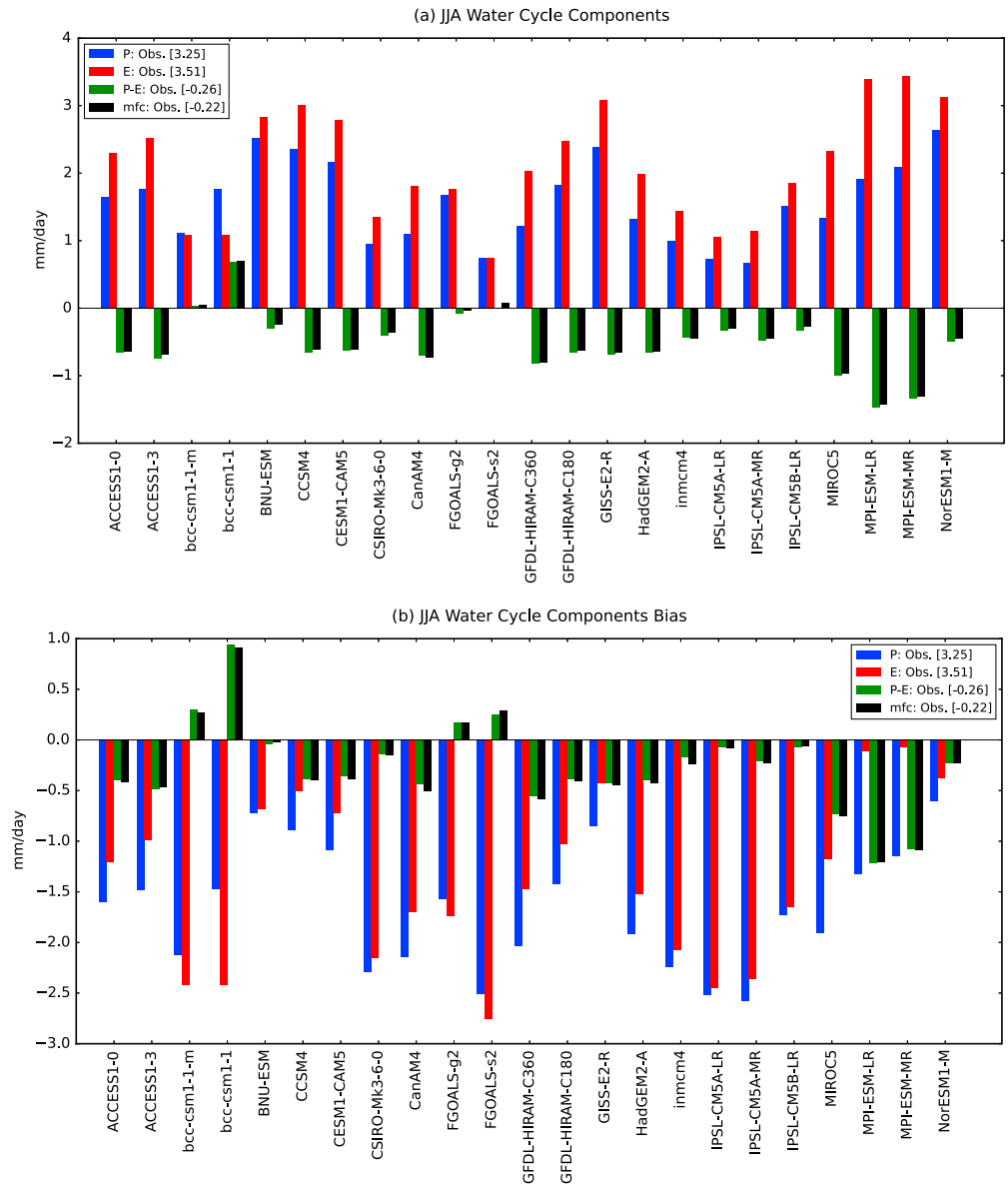
Figure 11. JJA mean water cycle components from CMIP5 models: including precipitation (*P*), evaporation (*E*), *P* − *E*, and Moisture Flux Convergence. The mean observed values (mm/d) are shown in legend. (a) for the JJA mean and (b) for JJA mean bias.

important role in the hydrological cycle and impact local precipitation and evapotranspiration (Jin et al., 2011; Seager et al., 2010; Wei et al., 2016). In this section, we examine the atmospheric water budget:

$$\frac{\partial \, \text{PW}}{\partial \, t} = E - P + \text{MFC} \tag{7}$$

In equation (7), MFC is the Moisture Flux Convergence that is computed as the residual of $P - E + \frac{\partial \, \text{PW}}{\partial \, t}$, and P is precipitation, E is evaporation, and $\frac{\partial \, \text{PW}}{\partial \, t}$ is the local change of precipitable water. The mean change rate of each month is approximated by the PW change between the following and the preceding month. Since $\frac{\partial \, \text{PW}}{\partial \, t}$ is much smaller than P and E at long time scales, the total MFC approximately equals to $P - E$ (Figure 11).

Over the summer months, the mean moisture flux convergence estimated from the ARM data is negative (−0.22 mm/d). The magnitudes of JJA mean evaporation and precipitation are 3.51 and 3.25 mm/d,

Table 6
Intermodel Correlations Between JJA Mean T2 m and Water Cycle Budget Components Averaged Over 1979–2008, With p Values Provided

Variable	P	E	P − E	MFC
T2 m	−0.66 (<0.001)	−0.59 (0.002)	0.21 (0.35)	0.21 (0.35)

respectively, which suggests that precipitation is dominated by moisture provided locally by evaporation instead of through remote moisture sources. Hence, precipitation is recycled vigorously during the summertime at SGP. This result supports the "hot spots" argument from Koster et al. (2004), which found that strong atmosphere-land coupling occurs over SGP summer. Most of the models capture the dominance of evaporation's contribution to precipitation (i.e., moisture flux divergence). It is interesting to note that although the MPI-ESM models simulate most accurate evaporation and EF (see also Figures 9 and 11), they simulate the strongest moisture flux divergence (most negative $P − E$) among all models.

Consistent with Figure 1b, all models underestimate precipitation amount with bias ranging from −1.61 to −0.28 mm/d and a multimodel mean bias of −0.74 mm/d. Across models, precipitation correlates to T2 m slightly stronger than evaporation with a negative correlation coefficient of −0.66 versus −0.59 (Table 6). This indicates that the lack of precipitation may be a more important contributor to the warm bias. Because land-surface interactions between soil moisture, evaporation, and precipitation are complex, this study only provides some insight on possible contributors. For instance, winter and spring precipitation deficit may also be a cause of JJA precipitation bias through the impact on soil moisture going into the summer. Ma et al. (2018) suggest that modeled soil moisture is largely controlled by the day-to-day changes in accumulated $P − E$ in the hindcast runs. Therefore, a deficit in the precipitation in spring could lead to a deficit in accumulated P going into the summer season. A lack of nocturnal precipitation could be an important contributor to the soil moisture deficit and can amplify the surface temperature bias through the positive land-atmosphere feedback (Klein et al., 2006). Nevertheless, it is difficult to isolate its relative contribution due to various feedbacks and interactions in the climate integrations in the present paper. To reveal the relative importance of each physical process contributing to the warm bias, more complex attribution techniques are required.

3.4. Impact of the Low-Level Jet

We now focus on other aspects of the large-scale circulation that could impact the local physical processes at the SGP site. One of the most important circulation features that affects SGP in summer is the low-level jet (LLJ), which brings moisture northward from the Gulf of Mexico over the SGP (Berg et al., 2015; Sheffield et al., 2013; Stensrud, 1996). Figure 12 compares the vertical profile of JJA mean meridio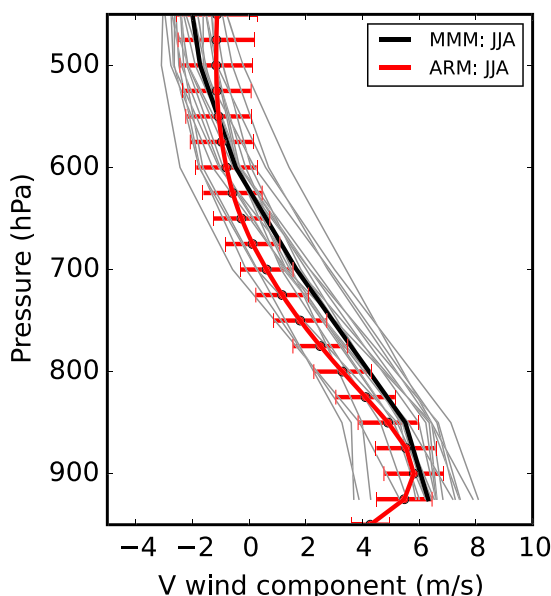nal wind components of models to the ARM data. The CMIP5 models stimulate the maximum JJA mean meridional wind at the 925 hPa, while ARM shows the maximum at 900 hPa. Both are consistent with the LLJs height range reported from Berg et al. (2015). A peak around 900 hPa in meridional wind is only shown in ARM observation. This could be a result of higher vertical resolution (25 hPa) of ARM data. In contrast, the CMIP data have a much lower resolution in the vertical, which varies from 75 hPa near the surface to over 150 hPa in the troposphere. Although models reasonably simulate the height of the maximum JJA mean meridional wind, the intermodel spread is large and the wind speeds tend to be overestimated by most models at 925 mb (Figures 11 and 13a). The causes for model spread are difficult to attribute because both the large-scale circulation patterns and boundary layer properties (i.e., surface roughness and vertical mixing strength) do affect the low-level wind speed (Song et al., 2005; Svensson & Lindvall, 2015).

Figure 13 shows the annual cycle of monthly mean meridional wind V, $V*q$, horizontal wind convergence at 925 hPa, and vertical pressure velocity (ω) at 500 hPa obtained from ARM data and models. The model spread is large for all analyzed monthly mean fields. The multimodel mean simulates the 925 hPa V and $V*q$ peaks in July while the ARM observation peaks in June. For both observation and models, there are
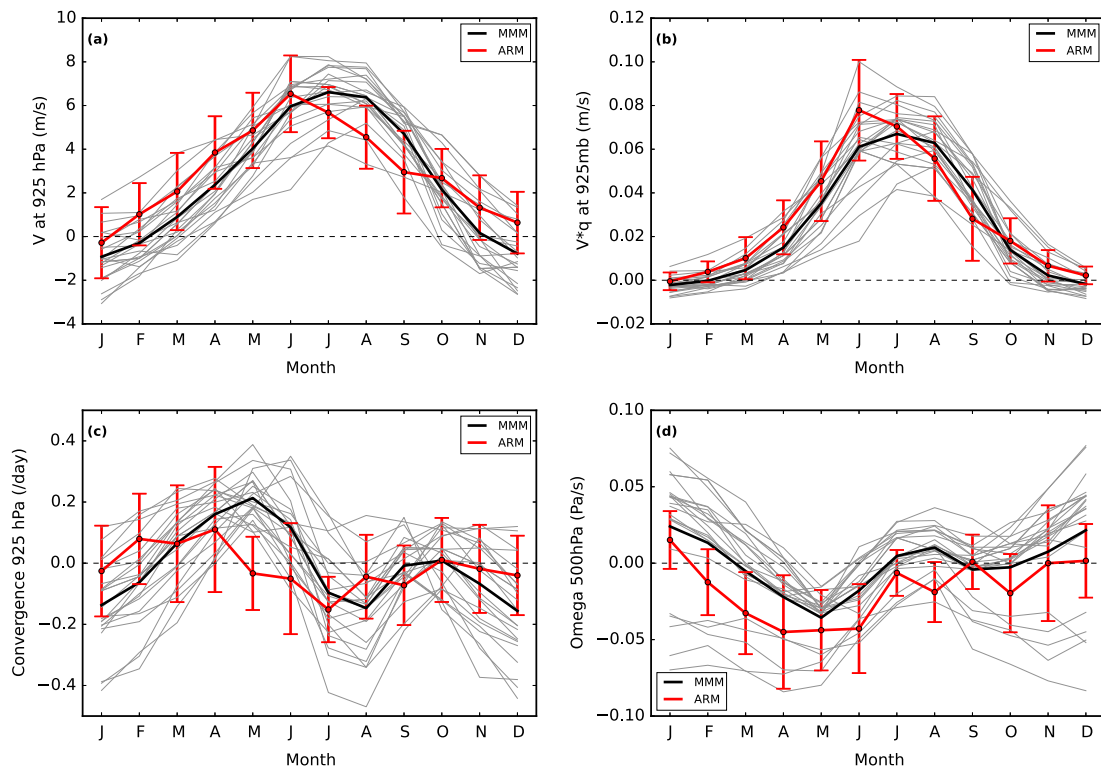

Figure 12. Vertical profile of climatological JJA mean meridional wind component from CMIP5 models, with same line types as Figure 1.

Figure 13. Same as Figure 1 but for annual cycle of monthly mean (a) meridional wind V at 925 hPa, (b) $V*q$ at 925 hPa, (c) horizontal wind convergence at 925 hPa, and (d) vertical pressure velocity (ω) at 500 hPa.

stronger V (Figure 13a) and $V*q$ (Figure 13b) in JJA associated with the summer time LLJ, and in general stronger divergence at 925 hPa (Figure 13c) than other seasons of the year. The JJA mean divergence corresponds to the negative MFC shown in Figure 11. It suggests that the appearance of LLJ in summer over the SGP site does not help create local moisture convergence. One possible explanation for this is that the SGP site is located where the jets accelerate. According to Berg et al. (2015), when summer time LLJs are observed over the SGP site, enhanced precipitation and moisture convergence tend to appear farther north over Iowa and surrounding regions. Focusing on conditions at Kansas/Oklahoma border, Song et al. (2005) drew a similar conclusion that strong southerly jets do not necessarily lead to more moisture convergence and precipitation locally; rather precipitation is better related to the passage of frontal systems. For JJA mean vertical motion at 500 hPa, ARM data estimate a mean ascending motion with the magnitude of −0.023 Pa/s, while models simulate much weaker ascending motion (−0.002 Pa/s). The underestimate of ascending motion during the summer is consistent with the deficit in the model-produced precipitation.

Across models, there is positive intermodel T2 m correlation to meridional wind speed (Figure 14a) and ω at 500 hPa (Figure 14b). Models that strongly overestimate low-level V and underestimate 500 hPa ascending motion tend to have large T2 m bias. To provide more context to these correlations, Table 7 lists the correlation coefficients between the aspects of the large-scale circulation and components of surface energy and water budgets. V at 925 hPa and ω at 500 hPa have significant positive correlation with net shortwave radiation and negative correlation with low-level convergence. Higher meridional wind and stronger subsidence is likely to link to stronger divergence and suppressed cloud and convection initiation, which in turn can result in underestimation of net surface shortwave which leads to surface temperature warm bias. The underestimation of convective cloud is found to contribute most to the radiation bias from the hindcast study performed by Van Weverberg et al. (2018). The large-scale biases in climate simulation can possibly amplify surface air temperature though their impact on surface energy budget and water cycle budget. It is also noted that precipitation has insignificant correlation to the large-scale properties being evaluated so its bias does not seem to be simply related to biases in the large-scale circulation.
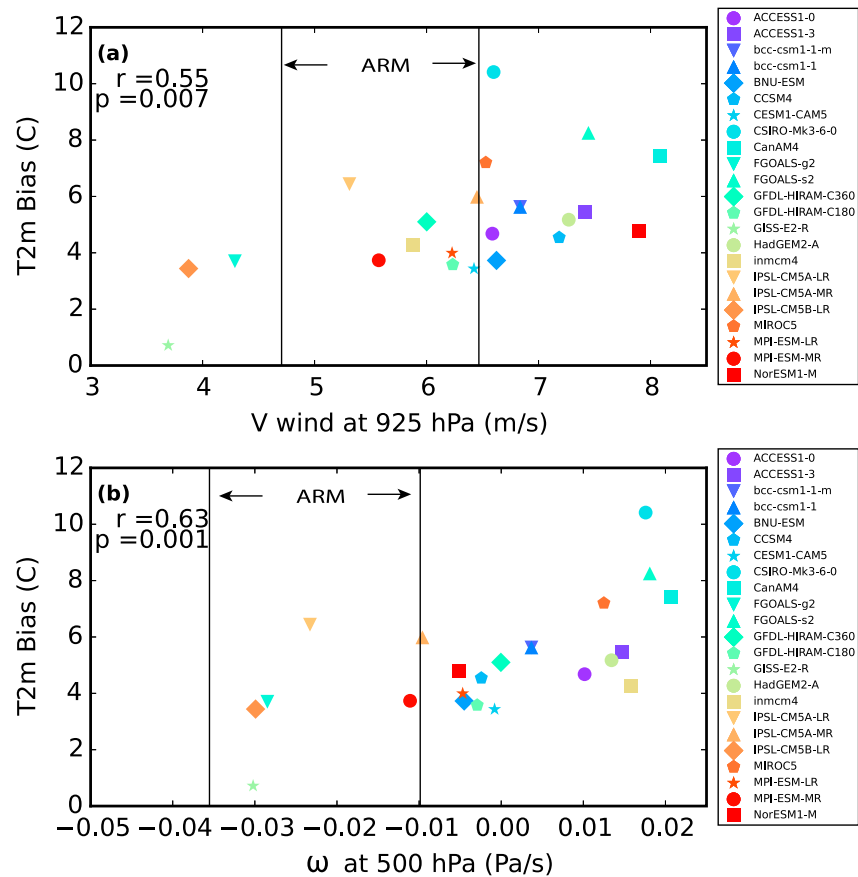
Figure 14. Same as Figure 4 but for JJA mean T2 m bias versus (a) V wind speed at 925 hPa and (b) ω at 500 hPa for CMIP5 models. The width of vertical lines represents the 95% Confidence Interval (CI) for an estimate of V wind at 925 hPa and ω at 500 hPa based on interannual variability from ARM data.

This analysis shows that both large-scale and local physical processes act together in the generation of the warm bias. Bias from large-scale circulation can complicate the attribution process for pinpointing deficiency in model physical processes. More advanced simulation techniques, such as hindcast approach (which prescribes the large-scale circulation) used in CAUSES project, may be more useful in addressing this issue.

Table 7
Summary of Intermodel Correlation Between JJA Mean Variables Selected From Aspects of the Circulations, Water Cycle Budget, and Surface Energy Budget

Variable	V at 925 mb	$V*q$ at 925 mb	Convergence at 925 mb	ω at 500 mb
T2 m	**0.55**	0.19	−0.31	**0.63**
Net SW	**0.50**	0.28	−0.32	**0.43**
EF	−0.14	0.40	−0.05	−0.31
Precipitation	−0.06	0.31	−0.02	−0.32
Evaporation	−0.06	**0.48**	−0.10	−0.23
$P - E$	0.03	**−0.46**	0.16	−0.01
PW	−0.04	**0.45**	−0.05	0.02
V at 925 mb	**1.0**	**0.80**	**−0.81**	**0.81**
$V*q$ at 925 mb	**0.80**	**1.0**	**−0.74**	**0.57**
Convergence at 925 mb	**−0.81**	**−0.74**	**1.0**	**−0.77**
ω at 500 mb	**0.81**	**0.57**	**−0.77**	**1.0**

Note. Numbers in bold indicate that the correlations are statistically significant with *p* value less than 0.05.

4. Summary and Discussion

This paper is part of the CAUSES project with the emphasis on analyzing the warm bias in climate simulations and providing perspective as to relevance of the hindcast biases studied in the other CAUSES papers (i.e., Ma et al., 2018; Morcrette et al., 2018; Van Weverberg et al., 2018) to the climate bias. In this study, we evaluate the CMIP5 AMIP-type simulations and identify the climate biases at the ARM SGP site using comprehensive data sets from in situ ARM observations, satellite observations, and other ground-based measurements. The simulated summer season mean state climate at the ARM SGP site by the CMIP5 models features a significant warm surface air temperature bias along with an underestimation of surface relative humidity, precipitation, and precipitable water. Through the examination of surface energy budget components, we find that most models tend to overestimate the surface absorbed solar radiation. The excessive downward shortwave flux is due to model biases in cloud, clear sky, and surface albedo. Based on both ARM and CERES measurements, downward shortwave bias associated with cloud deficit is generally the largest contributor to the error in the net-absorbed surface shortwave radiation. While the relative importance of the clear-sky downward shortwave bias and surface albedo error varies between observational data sets, they are both important factors to be considered. The underestimated clear-sky shortwave absorption is shown to be related to both biased low precipitable water and the low sensitivity of a model's radiation code to the variation of water vapor. Surface albedo appears to be too low in certain models.

The multimodel-mean turbulent fluxes also feature a bias that is consistent with the warm bias; namely, an underestimation of latent heat flux and an overestimation of sensible heat. The biased low evaporative fraction causes too much of the available radiation to be used to heat the surface instead of evaporating water, leading to a higher surface temperature. The evaporative fraction bias partially arises through the control of soil moisture. A negative intermodel correlation (although it is weak) is found between biases in the summertime T2 m and soil moisture. The soil moisture deficit is associated with a lack of precipitation, which correlates strongly with surface temperature across models. Furthermore, lack of groundwater or irrigation parameterizations (e.g., Lo et al., 2010; Qian et al., 2013) in the land models could also contribute to the soil moisture deficit. Therefore, both atmospheric and land surface components of the models have deficiencies that can lead to an underestimate of evaporative fraction.

The surface energy budget analysis shows that both surface solar radiation and evaporative fraction errors are contributors to the warm bias. These findings based on long-term climate simulations are similar to those discovered from short-term hindcast experiments (Ma et al., 2018; Van Weverberg et al., 2018), although short-term hindcasts in general have smaller warm bias which is associated with smaller errors in radiation and evaporative fraction compared to climate simulations. The relative magnitude of biases between hindcast and climate simulations is consistent with an interaction between large-scale circulation and physical processes that amplifies the warm bias at the surface in climate integrations. One may wonder about the relative contribution of these errors to the temperature bias. In the CAUSES paper of Ma et al. (2018), which emphasized on understanding the warm bias from the hindcast approach, a theoretical analysis is used to identify the relative contribution of surface radiation and evaporative fraction errors to the temperature bias. From this theoretical analysis, Ma et al. (2018) find that EF errors play a larger role than radiation errors in determining a model's temperature bias for both hindcast and AMIP simulations.

As aforementioned, the warm bias found in the CMIP5 AMIP simulations could be associated with large-scale circulation biases, including the climatically important LLJ which transports moisture to the SGP site. Although there is large intermodel spread, the multimodel mean moisture transport agrees reasonably well with observations. However, the transported moisture does not tend to converge and promote precipitation over the SGP site efficiently. Too strong meridional wind and weaker upward motion are associated with the warm bias. High meridional wind and subsidence are also correlated with large surface wind divergence, moisture divergence, and net shortwave flux at the surface. This implies a plausible pathway that large-scale circulation biases could contribute to the warm bias in AMIP-type climate simulations, making it difficult to isolate physical parameterization problems in these models. To address this issue, the CAUSES intermodel comparison project applies the hindcast approach to isolate the contribution of the warm bias from model physics (Ma et al., 2018; Morcrette et al., 2018; Van Weverberg et al., 2018).

In AMIP simulations, the positive land-atmosphere feedback loop can amplify the temperature bias and may be responsible for part of the temperature bias. As pointed by the hindcast study from Klein et al. (2006),

feedbacks between precipitation and the land surface can amplify the temperature bias. A recent study by Williams et al. (2016) shows that weakening the coupling between evaporative fraction and soil moisture has the effect to damp the warm bias. Performing more controlled hindcast experiments in follow-up studies could be helpful for pinpointing the causes of warm biases by isolating various components of the land-atmosphere coupling feedback loop. These controlled experiments could involve atmosphere-only model tests in hindcast mode (which exclude long-term land-model biases), land-only model tests with prescribed atmospheric forcing, or coupled land-atmosphere model tests in which specific flux component feedbacks are disabled and overridden with observed values (e.g., radiation or precipitation fluxes to the land, or evaporative fraction for the atmosphere).

The analysis method and data sets developed in this work and other CAUSES studies could be useful for the proposed controlled hindcasts in follow-up studies. By analyzing the characteristics of the warm bias at the SGP site and other locations over the American Midwest where tend to be too warm in many atmospheric models, Morcrette et al. (2018) pointed out that detailed model evaluation using data collected at SGP site is representative of model biases over a wider region. The analyses carried out in CAUSES studies therefore may be indicative for model errors over a broader area over American Midwest and other midlatitude continental areas.

References

Bao, Q., Lin, P., Zhou, T., Liu, Y., Yu, Y., Wu, G., et al. (2013). The flexible global ocean-atmosphere-land system model, spectral version 2: FGOALS-s2. *Advances in Atmospheric Sciences*, *30*(3), 561–576. https://doi.org/10.1007/s00376-012-2113-9

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., et al. (2013). The Norwegian Earth System Model, NorESM1-M—Part 1: Description and basic evaluation of the physical climate. *Geoscientific Model Development*, *6*(3), 687–720. https://doi.org/10.5194/gmd-6-687-2013

Berg, L. K., & Lamb, P. J. (2016). Surface properties and interactions: Coupling the land and atmosphere within the ARM program. The Atmospheric Radiation Measurement (ARM) Program: The First 20 Years, Meteor. Monogr., No. 57, Amer. Meteor. Soc. https://doi.org/10.1175/AMSMONOGRAPHS-D-15-0044.1

Berg, L. K., Riihimaki, L. D., Qian, Y., Yan, H., & Huang, M. (2015). The low-level jet over the Southern Great Plains determined from observations and reanalyses and its impact on moisture transport. *Journal of Climate*, *28*(17), 6682–6706. https://doi.org/10.1175/JCLI-D-14-00719.1

Bi, D., Dix, M., Marsland, S., O'Farrell, S., Rashid, H., Uotila, P., et al. (2013). The ACCESS coupled model: Description, control climate and evaluation. *Australian Meteorological and Oceanographic Journal*, *63*(1), 41–64. https://doi.org/10.22499/2.6301.004

Bond, D. (2005). *Soil water and temperature system (SWATS) handbook*. Washington, DC: ARM Technical Report TR-063, U.S. Department of Energy.

Breidenbach, J., Seo, D. J., & Fulton, R. (1998). *Stage II and III post processing of NEXRAD precipitation estimates in the modernized weather service*. Arizona: AMS 78th Annual Meeting Phoenix. January 1998

Brotzge, J. A., & Crawford, K. C. (2003). Examination of the surface energy budget: A comparison of eddy correlation and Bowen ratio measurement systems. *Journal of Hydrometeorology*, *4*(2), 160–178. https://doi.org/10.1175/1525-7541(2003)4%3C160:EOTSEB%3E2.0.CO;2

Campos, E., & Sisterson, D. L. (2015). A unified approach for reporting ARM measurement uncertainties technical report. No. DOE/SC-ARM-TR-170. DOE ARM Climate Research Facility, Pacific Northwest National Laboratory; Richland, Washington.

Cheruy, F., Campoy, A., Dupont, J. C., Ducharne, A., Hourdin, F., Haeffelin, M., et al. (2013). Combined influence of atmospheric physics and soil hydrology on the simulated meteorology at the SIRTA atmospheric observatory. *Climate Dynamics*, *40*(9–10), 2251–2269. https://doi.org/10.1007/s00382-012-1469-y

Cheruy, F., Dufresne, J. L., Hourdin, F., & Ducharne, A. (2014). Role of clouds and land-atmosphere coupling in midlatitude continental summer warm biases and climate change amplification in CMIP5 simulations. *Geophysical Research Letters*, *41*, 6493–6500. https://doi.org/10.1002/2014GL061145

Chylek, P., Li, J., Dubey, M. K., Wang, M., & Lesins, G. (2011). Observed and model simulated 20th century Arctic temperature variability: Canadian Earth System Model CanESM2. *Atmospheric Chemistry and Physics Discussions*, *11*(8), 22,893–22,907. https://doi.org/10.5194/acpd-11-22893-2011

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., et al. (2011). Development and evaluation of an Earth-System model— HadGEM2. *Geoscientific Model Development*, *4*(4), 1051–1075. https://doi.org/10.5194/gmd-4-1051-2011

Cook, D. R. (2011a). Energy balance Bowen ratio station (EBBR) handbook (No. DOE/SC-ARM/TR-037). DOE Office of Science Atmospheric Radiation Measurement (ARM) Program (United States).

Cook, D. R. (2011b). Eddy correlation flux measurement system (ECOR) handbook (No. DOE/SC-A/TR-052). DOE Office of Science Atmospheric Radiation Measurement (ARM) Program (United States).

DeAngelis, A. M., Qu, X., Zelinka, M. D., & Hall, A. (2015). An observational radiative constraint on hydrologic cycle intensification. *Nature*, *528*(7581), 249–253. https://doi.org/10.1038/nature15770

Dufresne, J.-L., Foujols, M. A., Denvil, S., Caubel, A., Marti, O., Aumont, O., et al. (2013). Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5. *Climate Dynamics*, *40*(9-10), 2123–2165. https://doi.org/10.1007/s00382-012-1636-1

Eck, T. F., Holben, B. N., Reid, J. S., Dubovik, O., Smirnov, A., O'Neill, N. T., et al. (1999). Wavelength dependence of the optical depth of biomass burning, urban and desert dust aerosols. *Journal of Geophysical Research*, *104*(31), 333–350.

Edwards, J. M., & Slingo, A. (1996). Studies with a flexible new radiation code. I: Choosing a configuration for a large-scale model. *Quarterly Journal of the Royal Meteorological Society*, *122*(531), 689–719. https://doi.org/10.1002/qj.49712253107

Feldman, D. R., Collins, W. D., Pincus, R., Huang, X., & Chen, X. (2014). Far-infrared surface emissivity and climate. *Proceedings of the National Academy of Sciences*, *111*(46), 16,297–16,302. https://doi.org/10.1073/pnas.1413640111

Fildier, B., & Collins, W. D. (2015). Origins of climate model discrepancies in atmospheric shortwave absorption and global precipitation changes. *Geophysical Research Letters*, *42*, 8749–8757. https://doi.org/10.1002/2015GL065931

Freidenreich, S. M., & Ramaswamy, V. (2011). Analysis of the biases in the downward shortwave surface flux in the GFDL CM2.1 general circulation model. *Journal of Geophysical Research*, *116*, D08208. https://doi.org/10.1029/2010JD014930

Fulton, R. A., Breidenbach, J. P., Seo, D. J., Miller, D. A., & O'Bannon, T. (1998). The WSR-88D rainfall algorithm. *Weather and Forecasting*, *13*(2), 377–395. https://doi.org/10.1175/1520-0434(1998)013%3C0377:TWRA%3E2.0.CO;2

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., et al. (2011). The community climate system model version 4. *Journal of Climate*, *24*(19), 4973–4991. https://doi.org/10.1175/2011JCLI4083.1

Hogan, R. J., & Shonk, J. K. P. (2013). Incorporating the effects of 3D radiative transfer in the presence of clouds into two-stream radiation schemes. *Journal of the Atmospheric Sciences*, *70*(2), 708–724. https://doi.org/10.1175/JAS-D-12-041.1

Hourdin, F., Grandpeix, J. Y., Rio, C., Bony, S., Jam, A., Cheruy, F., et al. (2013). LMDZ5B: The atmospheric component of the IPSL climate model with revisited parameterizations for clouds and convection. *Climate Dynamics*, *40*(9-10), 2193–2222. https://doi.org/10.1007/s00382-012-1343-y

Huang, X., Chen, X., Zhou, D. K., & Liu, X. (2016). An observationally based global band-by-band surface emissivity dataset for climate and weather simulations. *Journal of the Atmospheric Sciences*, *73*(9), 3541–3555. https://doi.org/10.1175/JAS-D-15-0355.1

Jensen, M. P., Petersen, W. A., Bansemer, A., Bharadwaj, N., Carey, L. D., Cecil, D. J., et al. (2016). The midlatitude continental convective clouds experiment (MC3E). *Bulletin of the American Meteorological Society*, *97*(9), 1667–1686. https://doi.org/10.1175/BAMS-D-14-00228.1

Ji, D., Wang, L., Feng, J., Wu, Q., Cheng, H., Zhang, Q., et al. (2014). Description and basic evaluation of Beijing Normal University Earth system model (BNU-ESM) version 1. *Geoscientific Model Development*, *7*(5), 2039–2064. https://doi.org/10.5194/gmd-7-2039-2014

Jimenez, C., Prigent, C., Mueller, B., Seneviratne, S. I., McCabe, M. F., Wood, E. F., et al. (2011). Global intercomparison of 12 land surface heat flux estimates. *Journal of Geophysical Research*, *116*, D02102. https://doi.org/10.1029/2010JD014545

Jin, F. J., Kitoh, A., & Alpert, P. (2011). Climatological relationships among the moisture budget components and rainfall amounts over the Mediterranean based on a super-high-resolution climate model. *Journal of Geophysical Research*, *116*, D09102. https://doi.org/10.1029/2010JD014021

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., et al. (2013). Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. *Journal of Advances in Modeling Earth Systems*, *5*, 422–446. https://doi.org/10.1002/jame.20023

Kato, S., Loeb, N. G., Rose, F. G., Doelling, D. R., Rutan, D. A., Caldwell, T. E., et al. (2013). Surface irradiances consistent with CERES-derived top-of-atmosphere shortwave and longwave irradiances. *Journal of Climate*, *26*(9), 2719–2740. https://doi.org/10.1175/jcli-d-12-00436.1

Klein, S. A., Jiang, X., Boyle, J., Malyshev, S., & Xie, S. (2006). Diagnosis of the summertime warm and dry bias over the U.S. Southern Great Plains in the GFDL climate model using a weather forecasting approach. *Geophysical Research Letters*, *33*, L18805. https://doi.org/10.1029/2006GL027567

Koontz, A., Flynn, C., Hodges, G., Michalsky, J., & Barnard, J. (2013). *Aerosol optical depth value-added product*. Washington, DC: DOE/SC-ARM/TR-129, USDOE Office of Science.

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., et al. (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, *305*(5687), 1138–1140. https://doi.org/10.1126/science.1100217

Levine, X. J., & Boos, W. R. (2017). Land surface albedo bias in climate models and its association with tropical rainfall. *Geophysical Research Letters*, *44*, 6363–6372. https://doi.org/10.1002/2017GL072510

Li, Y., Wang, T., Zeng, Z., Peng, S., Lian, X., & Piao, S. (2016). Evaluating biases in simulated land surface albedo from CMIP5 global climate models. *Journal of Geophysical Research: Atmospheres*, *121*, 6178–6190. https://doi.org/10.1002/2016JD024774

Liang, S., Wang, K., Zhang, X., & Wild, M. (2010). Review on estimation of land surface radiation and energy budgets from ground measurement, remote sensing and model simulations. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *3*(3), 225–240. https://doi.org/10.1109/jstars.2010.2048556

Lo, M.-H., Famiglietti, J. S., Yeh, P. J.-F., & Syed, T. H. (2010). Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated base flow data. *Water Resources Research*, *46*(5), W05517. https://doi.org/10.1029/2009WR007855

Long, C. N., & Ackerman, T. P. (2000). Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects. *Journal of Geophysical Research*, *105*(D12), 15,609–15,626. https://doi.org/10.1029/2000JD900077

Long, C. N., & Shi, Y. (2006). The QCRad value added product: Surface radiation measurement quality control testing, including climatologically configurable limits, Atmospheric Radiation Measurement Program Technical Report, ARM TR-074, 69 pp.

Long, C. N., & Shi, Y. (2008). An automated quality assessment and control algorithm for surface radiation measurements. *TOASJ*, *2*(1), 23–37. https://doi.org/10.2174/1874282300802010023

Long, C. N., & Turner, D. D. (2008). A method for continuous estimation of clear-sky downwelling longwave radiative flux developed using ARM surface measurements. *Journal of Geophysical Research*, *113*, D18206. https://doi.org/10.1029/2008JD009936

Ma, H.-Y., Chuang, C. C., Klein, S. A., Lo, M.-H., Zhang, Y., Xie, S., et al. (2015). An improved hindcast approach for evaluation and diagnosis of physical processes in global climate models. *Journal of Advances in Modeling Earth Systems*, *7*, 1810–1827. https://doi.org/10.1002/2015MS000490

Ma, H.-Y., Klein, S. A., Xie, S., Zhang, C., Tang, S., Tang, Q., et al (2018). CAUSES: On the role of surface energy budget errors to the warm surface air temperature error over the Central U.S. *Journal of Geophysical Research: Atmospheres*, *123*. https://doi.org/10.1002/2017JD027194

Ma, H. Y., Xie, S., Klein, S. A., Williams, K. D., Boyle, J. S., Bony, S., et al. (2014). On the correspondence between mean forecast errors and climate errors in CMIP5 models. *Journal of Climate*, *27*(4), 1781–1798. https://doi.org/10.1175/jcli-d-13-00474.1

Martin, G. M., Ringer, M. A., Pope, V. D., Jones, A., Dearden, C., & Hinton, T. J. (2006). The physical properties of the atmosphere in the new Hadley Centre Global Environmental Model (HadGEM1). Part I: Model description and global climatology. *Journal of Climate*, *19*(7), 1274–1301. https://doi.org/10.1175/JCLI3636.1

McComiskey, A., & Ferrare, R. A. (2016). Aerosol physical and optical properties and processes in the ARM Program. The Atmospheric Radiation Measurement (ARM) program: The first 20 years, Meteor. Monogr., No. 57, Amer. Meteor. Soc. https://doi.org/10.1175/AMSMONOGRAPHS-D-15-0028.1

Merrifield, A. L., & Xie, S. P. (2016). Summer US surface air temperature variability: Controlling factors and AMIP simulation biases. *Journal of Climate*, *29*(14), 5123–5139. https://doi.org/10.1175/JCLI-D-15-0705.1

Michalsky, J. J., & Long, C. N. (2016). ARM solar and infrared broadband and filter radiometry. The Atmospheric Radiation Measurement (ARM) program: The first 20 years, Meteor. Monogr., No. 57, Amer. Meteor. Soc. https://doi.org/10.1175/AMSMONOGRAPHS-D-15-0031.1

Morcrette, C. J., Van Weverberg, K., Ma, H.-Y., Ahlgrimm, M., Bazile, E., Berg, L. K., et al. (2018). Introduction to CAUSES: Description of weather and climate models and their near-surface temperature errors in 5 day hindcasts near the Southern Great Plains. *Journal of Geophysical Research: Atmospheres*, *123*. https://doi.org/10.1002/2017JD027199

Mueller, B., & Seneviratne, S. I. (2012). Hot days induced by precipitation deficits at the global scale. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(31), 12,398–12,403. https://doi.org/10.1073/pnas.1204330109

Mueller, B., & Seneviratne, S. I. (2014). Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophysical Research Letters*, *41*, 128–134. https://doi.org/10.1002/2013GL058055

Neale, R. B., Richter, J. H., Conley, A. J., Park, S., Lauritzen, P. H., Gettelman, A., et al. (2010). Description of the NCAR Community Atmosphere Model (CAM 4.0). Tech. Rep. NCAR/TN-485+ STR. Boulder, CO: National Center for Atmospheric Research.

Phillips, T. J., & Klein, S. A. (2014). Land-atmosphere coupling manifested in warm-season observations on the US southern great plains. *Journal of Geophysical Research: Atmospheres*, *119*, 509–528. https://doi.org/10.1002/2013JD020492

Phillips, T. J., Potter, G. L., Williamson, D. L., Cederwall, R. T., Boyle, J. S., Fiorino, M., et al. (2004). Evaluating parameterizations in general circulation models: Climate simulation meets weather prediction. *Bulletin of the American Meteorological Society*, *85*(12), 1903–1916. https://doi.org/10.1175/BAMS-85-12-1903

Phillips, T. J., Klein, S. A., Ma, H.-Y., Tang, Q., Xie, S., Williams, I. N., et al. (2017). Using ARM observations to evaluate climate model simulations of land-atmosphere coupling on the U.S. Southern Great Plains. *Journal of Geophysical Research: Atmospheres*, *122*, 11,524–11,548. https://doi.org/10.1002/2017JD027141

Pincus, R., Mlawer, E. J., Oreopoulos, L., Ackerman, A. S., Baek, S., Brath, M., et al. (2015). Radiative flux and forcing parameterization error in aerosol-free clear skies. *Geophysical Research Letters*, *42*, 5485–5492. https://doi.org/10.1002/2015GL064291

Qian, Y., Huang, M., Yang, B., & Berg, L. K. (2013). A modeling study of irrigation effects on surface fluxes and land–air–cloud interactions in the Southern Great Plains. *Journal of Hydrometeorology*, *14*(3), 700–721. https://doi.org/10.1175/JHM-D-12-0134.1

Ritsche, M. T. (2008). *Surface meteorological observation system handbook*. Washington, DC: ARM Technical Report TR-031, USDOE Office of Science.

Rotstayn, L. D., Jeffrey, S. J., Collier, M. A., Dravitzki, S. M., Hirst, A. C., Syktus, J. I., & Wong, K. K. (2012). Aerosol- and greenhouse gas-induced changes in summer rainfall and circulation in the Australasian region: A study using single-forcing climate simulations. *Atmospheric Chemistry and Physics*, *12*(14), 6377–6404. https://doi.org/10.5194/acp-12-6377-2012

Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., et al. (2014). Configuration and assessment of the GISS Model E2 contributions to the CMIP5 archive. *Journal of Advances in Modeling Earth Systems*, *6*(1), 141–184. https://doi.org/10.1002/2013MS000265

Schmidt, G. A., Ruedy, R., Hansen, J. E., Aleinov, I., Bell, N., Bauer, M., et al. (2006). Present-day atmospheric simulations using GISS Model E: Comparison to in situ, satellite, and reanalysis data. *Journal of Climate*, *19*(2), 153–192. https://doi.org/10.1175/JCLI3612.1

Seager, R., Naik, N., & Vecchi, G. A. (2010). Thermodynamic and dynamic mechanisms for large-scale changes in the hydrological cycle in response to global warming. *Journal of Climate*, *23*(17), 4651–4668. https://doi.org/10.1175/2010jcli3655.1

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., et al. (2010). Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3–4), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Sheffield, J., Barrett, A. P., Colle, B., Nelun Fernando, D., Fu, R., Geil, K. L., et al. (2013). North American climate in CMIP5 experiments. Part I: Evaluation of historical simulations of continental and regional climatology. *Journal of Climate*, *26*(23), 9209–9245. https://doi.org/10.1175/JCLI-D-12-00592.1

Song, J., Liao, K., Coulter, R. L., & Lesht, B. M. (2005). Climatology of the low-level jet at the southern Great Plains atmospheric boundary layer experiments site. *Journal of Applied Meteorology*, *44*(10), 1593–1606. https://doi.org/10.1175/JAM2294.1

Stensrud, D. J. (1996). Importance of low-level jets to climate: A review. *Journal of Climate*, *9*(8), 1698–1711. https://doi.org/10.1175/1520-0442(1996)009%3C1698:IOLLJT%3E2.0.CO;2

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., et al. (2013). ECHAM6. *Journal of Advances in Modeling Earth Systems*, *5*(2), 146–172. https://doi.org/10.1002/jame.20015

Svensson, G., & Lindvall, J. (2015). Evaluation of near-surface variables and the vertical structure of the boundary layer in CMIP5 models. *Journal of Climate*, *28*(13), 5233–5253. https://doi.org/10.1175/jcli-d-14-00596.1

Tang, S., Zhang, M., & Xie, S. (2016). An ensemble constrained variational analysis of atmospheric forcing data and its application to evaluate clouds in CAM5. *Journal of Geophysical Research: Atmospheres*, *121*, 33–48. https://doi.org/10.1002/2015JD024167

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498.

Teixeira, A. D. C., & Bastiaanssen, W. G. M. (2012). Five methods to interpret field measurements of energy fluxes over a micro-sprinkler-irrigated mango orchard. *Irrigation Science*, *30*(1), 13–28. https://doi.org/10.1007/s00271-010-0256-y

Van Weverberg, K., Morcrette, C. J., Petch, J., Klein, S. A., Ma, H.-Y., Zhang, C., et al. (2018). CAUSES: Attribution of Surface Radiation Biases in NWP and Climate Models near the U.S. Southern Great Plains. *Journal of Geophysical Research*, *123*. https://doi.org/10.1002/2017JD027188

Van Weverberg, K., Morcrette, C. J., Ma, H.-Y., Klein, S. A., & Petch, J. C. (2015). Using regime analysis to identify the contribution of clouds to surface temperature errors in weather and climate models. *Quarterly Journal of the Royal Meteorological Society*, *141*(693), 3190–3206. https://doi.org/10.1002/qj.2603

Volodin, E. M., Dianskii, N. A., & Gusev, A. V. (2010). Simulating present day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations. *Izvestiya Atmospheric and Oceanic Physics*, *46*(4), 414–431. https://doi.org/10.1134/S000143381004002X

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., et al. (2011). MIROC-ESM 2010: Model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development*, *4*(4), 845–872. https://doi.org/10.5194/gmd-4-845-2011

Wei, J. F., Su, H., & Yang, Z. L. (2016). Impact of moisture flux convergence and soil moisture on precipitation: A case study for the southern United States with implications for the globe. *Climate Dynamics*, *46*(1–2), 467–481. https://doi.org/10.1007/s00382-015-2593-2

Wild, M. (2008). Short-wave and long-wave surface radiation budgets in GCMs: A review based on the IPCC-AR4/CMIP3 models. *Tellus Series a-Dynamic Meteorology and Oceanography*, *60*(5), 932–945. https://doi.org/10.1111/j.1600-0870.2008.00342.x

Wild, M., Folini, D., Hakuba, M. Z., Schaer, C., Seneviratne, S. I., Kato, S., et al. (2015). The energy balance over land and oceans: An assessment based on direct observations and CMIP5 climate models. *Climate Dynamics*, *44*(11–12), 3393–3429. https://doi.org/10.1007/s00382-014-2430-z

Wild, M., Long, C. N., & Ohmura, A. (2006). Evaluation of clear-sky solar fluxes in GCMs participating in AMIP and IPCC-AR4 from a surface perspective. *Journal of Geophysical Research*, *111*, D01104. https://doi.org/10.1029/2005JD006118

Williams, I. N., Lu, Y., Kueppers, L. M., Riley, W. J., Biraud, S. C., Bagley, J. E., & Torn, M. S. (2016). Land-atmosphere coupling and climate prediction over the US Southern Great Plains. *Journal of Geophysical Research: Atmospheres*, *121*, 12,125–12,144. https://doi.org/10.1002/2016JD025223

Williamson, D. L., Boyle, J., Cederwall, R., Fiorino, M., Hnilo, J., Olson, J., et al. (2005). Moisture and temperature balances at the atmospheric radiation measurement Southern Great Plains site in forecasts with the Community Atmosphere Model (CAM2). *Journal of Geophysical Research*, *110*, D15S16. https://doi.org/10.1029/2004JD005109

Wu, T., Yu, R., Zhang, F., Wang, Z., Dong, M., Wang, L., et al. (2010). The Beijing climate center atmospheric general circulation model: Description and its performance for the present-day climate. *Climate Dynamics*, *34*(1), 123–147. https://doi.org/10.1007/s00382-008-0487-2

Xie, S., Ma, H. Y., Boyle, J. S., Klein, S. A., & Zhang, Y. (2012). On the correspondence between short-and long-time-scale systematic errors in CAM4/CAM5 for the year of tropical convection. *Journal of Climate*, *25*(22), 7937–7955. https://doi.org/10.1175/JCLI-D-12-00134.1

Xie, S., McCoy, R. B., Klein, S. A., Cederwall, R. T., Wiscombe, W. J., Jensen, M. P., et al. (2010). ARM climate modeling best estimate data a new data product for climate studies. *Bulletin of the American Meteorological Society*, *91*(1), 13–20. https://doi.org/10.1175/2009bams2891.1

Xie, S., Zhang, Y., Giangrande, S. E., Jensen, M. P., McCoy, R., & Zhang, M. (2014). Interactions between cumulus convection and its environment as revealed by the MC3E sounding array. *Journal of Geophysical Research: Atmospheres*, *119*, 11,784–11,808. https://doi.org/10.1002/2014JD022011

Xie, S. C., Cederwall, R. T., & Zhang, M. H. (2004). Developing long-term single-column model/cloud system-resolving model forcing data using numerical weather prediction products constrained by surface and top of the atmosphere observations. *Journal of Geophysical Research*, *109*, D01104. https://doi.org/10.1029/2003JD004045

Xu, S., Song, M., Liu, J., Wang, B., Li, L., Huang, W., et al. (2013). Simulation of sea ice in FGOALS-g2: Climatology and late 20th century changes. *Advances in Atmospheric Sciences*, *30*(3), 658–673. https://doi.org/10.1007/s00376-013-2158-4

Zhang, M., & Lin, J. (1997). Constrained variational analysis of sounding data based on column-integrated budgets of mass, heat, moisture, and momentum: Approach and application to ARM measurements. *Journal of the Atmospheric Sciences*, *54*(11), 1503–1524. https://doi.org/10.1175/1520-0469(1997)054%3C1503:CVAOSD%3E2.0.CO;2

Zhang, M., Lin, J., Cederwall, R. T., Yio, J. J., & Xie, S. C. (2001). Objective analysis of ARM IOP data: Method and sensitivity. *Monthly Weather Review*, *129*(2), 295–311. https://doi.org/10.1175/1520-0493(2001)129%3C0295:OAOAID%3E2.0.CO;2

Zhao, M., Held, I. M., Lin, S.-J., & Vecchi, G. A. (2009). Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *Journal of Climate*, *22*(24), 6653–6678. https://doi.org/10.1175/2009JCLI3049.1