



## RESEARCH ARTICLE

10.1002/2017JD027194

## Key Points:

- Biases in the net shortwave, downward longwave fluxes, and evaporative fraction are contributors to warm surface air temperature biases
- Radiation and evaporative fraction biases are associated with clouds, precipitation, evaporation, and soil moisture biases
- Biases in evaporative fraction are generally more important than biases in radiation in explaining the temperature biases

## Correspondence to:

H.-Y. Ma,  
ma21@llnl.gov

## Citation:

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 United States. *Journal of Geophysical Research: Atmospheres*, 123. <https://doi.org/10.1002/2017JD027194>

Received 25 MAY 2017

Accepted 25 SEP 2017

Accepted article online 27 FEB 2018

## CAUSES: On the Role of Surface Energy Budget Errors to the Warm Surface Air Temperature Error Over the Central United States

H.-Y. Ma<sup>1</sup> , S. A. Klein<sup>1</sup> , S. Xie<sup>1</sup> , C. Zhang<sup>1</sup> , S. Tang<sup>1</sup> , Q. Tang<sup>1</sup> , C. J. Morcrette<sup>2</sup> , K. Van Weverberg<sup>2</sup> , J. Petch<sup>2</sup> , M. Ahlgrimm<sup>3</sup> , L. K. Berg<sup>4</sup> , F. Cheruy<sup>5</sup> , J. Cole<sup>6</sup> , R. Forbes<sup>3</sup> , W. I. Gustafson Jr<sup>4</sup> , M. Huang<sup>4</sup> , Y. Liu<sup>4</sup> , W. Merryfield<sup>6</sup> , Y. Qian<sup>4</sup> , R. Roehrig<sup>7</sup> , and Y.-C. Wang<sup>8</sup>

<sup>1</sup>Lawrence Livermore National Laboratory, Livermore, CA, USA, <sup>2</sup>Met Office, Exeter, UK, <sup>3</sup>European Centre for Medium-Range Weather Forecasts, Reading, UK, <sup>4</sup>Pacific Northwest National Laboratory, Richland, WA, USA, <sup>5</sup>Laboratoire de Météorologie Dynamique, Paris, France, <sup>6</sup>Canadian Centre for Climate Modelling and Analysis, Environment Canada, Victoria, British Columbia, Canada, <sup>7</sup>CNRM, Meteo-France/CNRS, Toulouse, France, <sup>8</sup>Research Center for Environmental Change, Academia Sinica, Taipei, Taiwan

**Abstract** Many weather forecast and climate models simulate warm surface air temperature ( $T_{2m}$ ) biases over midlatitude continents during the summertime, especially over the Great Plains. We present here one of a series of papers from a multimodel intercomparison project (CAUSES: Cloud Above the United States and Errors at the Surface), which aims to evaluate the role of cloud, radiation, and precipitation biases in contributing to the  $T_{2m}$  bias using a short-term hindcast approach during the spring and summer of 2011. Observations are mainly from the Atmospheric Radiation Measurement Southern Great Plains sites. The present study examines the contributions of surface energy budget errors. All participating models simulate too much net shortwave and longwave fluxes at the surface but with no consistent mean bias sign in turbulent fluxes over the Central United States and Southern Great Plains. Nevertheless, biases in the net shortwave and downward longwave fluxes as well as surface evaporative fraction (EF) are contributors to  $T_{2m}$  bias. Radiation biases are largely affected by cloud simulations, while EF bias is largely affected by soil moisture modulated by seasonal accumulated precipitation and evaporation. An approximate equation based upon the surface energy budget is derived to further quantify the magnitudes of radiation and EF contributions to  $T_{2m}$  bias. Our analysis ascribes that a large EF underestimate is the dominant source of error in all models with a large positive temperature bias, whereas an EF overestimate compensates for an excess of absorbed shortwave radiation in nearly all the models with the smallest temperature bias.

### 1. Introduction

Surface air temperature or 2 m temperature ( $T_{2m}$  hereafter) is one of the most important variables for weather forecasts and climate simulations as it is highly relevant to our daily activities. However, most contemporary regional or global weather forecast and climate models suffer from large systematic warm biases in  $T_{2m}$ , especially over the midlatitude continents (e.g., Cheruy et al., 2014; Ma, Xie, et al., 2014). They appear in just the first few days of hindcasts in climate models when these models are initialized with reanalysis data (Ma, Xie, et al. 2014; Xie et al., 2012). The warm bias is largest over the U.S. Great Plains and is most pronounced during the summertime. The warm bias is most significant in Atmospheric Model Intercomparison Project (AMIP)-type simulations and less severe in the coupled ocean-atmosphere simulations. Nevertheless, such warm biases can hinder model prediction skills and could potentially impact the fidelity of simulated future temperature changes (Cheruy et al., 2014).

This long-standing problem involves multiscale process interactions, including those related to clouds, radiation, precipitation, and the land surface. Past studies suggest several possible causes for the  $T_{2m}$  bias. Using a climate model hindcast approach (Ma et al., 2015; Phillips et al., 2004; Williams et al., 2013) with the Geophysical Fluid Dynamics Laboratory Atmospheric Model version 2, Klein et al. (2006) suggests that the lack of precipitation and soil moisture in the model, especially precipitation associated with eastward propagation of convection from the lee of the Rockies to the U.S. Great Plains (e.g., Carbone et al., 2002; Jiang et al., 2006), can enhance the  $T_{2m}$  bias through interactions between the atmosphere and the land surface's soil

©2018. The Authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

moisture. Through the analysis of Coupled Model Intercomparison Project Phase 5 (CMIP5) models (Taylor et al., 2012), Cheruy et al. (2014) further illustrate that the warm biases in most CMIP5 models are consistent with excessive shortwave radiative fluxes due to a lack of overall cloudiness, as well as with a lack of maintenance of evaporation at the surface. The warm bias is especially severe over the soil moisture-limited regime where the land-atmosphere interactions are most significant.

Indeed, other than precipitation and land-atmosphere interactions, clouds and associated radiation biases can also contribute to the  $T_{2m}$  bias. Using a hindcast approach, Van Weverberg et al. (2015) analyzed model time step output from the Met Office Unified Model (MetUM) and the Community Atmosphere Model version 5 (CAM5) to study the impact of clouds and radiation on  $T_{2m}$  bias during the U.S. Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) Program's Mid-latitude Continental Convective Clouds Experiment (MC3E, 22 April to 6 June 2011, Jensen et al., 2016). By compositing the  $T_{2m}$  bias and bias growth with coinciding biases in radiation (both longwave and shortwave fluxes) from an objective cloud regime classification (seven cloud regimes based on cloud occurrence in the low (0–3 km), middle (3–6 km), and high levels (6–20 km) of the troposphere), their analysis identifies certain cloud regimes that matter the most for the  $T_{2m}$  bias and bias growth. Additionally, they separate the identification by portions of the day to ascertain the bias behavior across the diurnal cycle. Interestingly, the cloud regimes that contribute to the largest  $T_{2m}$  bias growth are not identical in both models within the diurnal cycle, indicating different cloud processes at work for contributing to the  $T_{2m}$  bias.

To better address this issue of  $T_{2m}$  bias through process level understanding, a multimodel intercomparison project, with an observationally based focus, named Cloud Above the United States and Errors at the Surface (CAUSES, <http://portal.nersc.gov/project/capt/CAUSES/>) was organized to evaluate the role of cloud, radiation, and precipitation processes in contributing to the surface temperature biases over the Central United States. The project foci are on (1) the errors in clouds and radiation and (2) the errors in surface water and energy budgets, in contributing to the  $T_{2m}$  bias. The key feature of the project is to apply a hindcast approach (Ma et al., 2015; Phillips et al., 2004; Williams et al., 2013) to diagnose model biases using high temporal frequency model outputs with coincident observations collected from the U.S. DOE ARM Facility sites over the U. S. Southern Great Plain (SGP), as well as other available observations. The study period is from 1 April to 31 August 2011, which covers the entire MC3E period. The MC3E campaign provides very high-frequency sounding data, clouds retrievals, and surface meteorological measurements, which are especially suitable for model evaluation. The DOE ARM SGP sites are located within the region of large  $T_{2m}$  bias, and therefore, the analysis done at the location is especially representative for the region. Note that 2011 is a drier year over the SGP vicinity in observations compared to other years (e.g., Santanello et al., 2015).

This paper is one of a series of four initial papers from the CAUSES project. The first paper by Morcrette et al. (2018) summarizes the entire project design and documents the warm bias characteristics and their spatial extent and intensity. The second paper by Van Weverberg et al. (2018) focuses on the cloud and radiation attribution to  $T_{2m}$  errors. Another paper by Zhang et al. (2018) studies the  $T_{2m}$  errors and the connection to cloud radiation and surface energy budget analysis from the CMIP5/AMIP simulations. Finally, the goal for this paper is to document surface energy budget errors in relationship to  $T_{2m}$  errors in the models and to identify common robust features in model biases. This paper also aims to identify the relative error contributions (relative importance) in the surface radiation and evaporative fraction (EF: the ratio of latent heat flux to the sum of latent and sensible heat fluxes) to the  $T_{2m}$  bias.

The remainder of this manuscript is organized into four sections. Section 2 describes the validation data sets, intercomparison experiments, and participating models. Section 3 presents the surface energy budget analysis from participating models and their relationship to  $T_{2m}$  bias both over the contiguous United States (CONUS) domain and at DOE ARM SGP sites. Section 4 identifies the relative contribution of surface radiation and EF to  $T_{2m}$  bias through a theoretical analysis. Section 5 draws conclusions.

## 2. Validation Data Sets and Models

### 2.1. Validation Data Sets

Several observational data sets were obtained to evaluate  $T_{2m}$  and surface fluxes simulations. For CONUS domain analysis, the observed  $T_{2m}$  is from the National Oceanic and Atmospheric Administration (NOAA)

Quality Controlled Local Climatological Data (QCLCD, <https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/quality-controlled-local-climatological-data-qclcd>). It consists of hourly, daily, and monthly summaries for approximately 1,900 U.S. locations and is available beginning 1 January 2005 and continues to the present. The station data are further interpolated onto 1° latitude by 1° longitude in the horizontal for comparison with gridded model data (Morcrette et al., 2018). Surface shortwave and longwave radiative fluxes are obtained from Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced And Filled (EBAF) observations (Kato et al., 2013, Edition 3A). These fields are derived from the SYN1deg product with a horizontal resolution of 1° latitude by 1° longitude (Doelling et al., 2013). Surface latent and sensible heat fluxes were derived by upscaling observations from the current global network of eddy covariance towers (FLUXNET, May 2012 version, Jung et al., 2011).

Over the ARM SGP, in order to make faithful comparison between model grid box mean variables and ARM observations, observed T<sub>2m</sub> and surface fluxes are from the U.S. DOE ARM continuous forcing evaluation data sets (Xie et al., 2004). Surface measurements from stations centered around the ARM Central Facility at Lamont, Oklahoma (36.6°N, 97.5°W), are first interpolated onto a 0.5° latitude × 0.5° longitude grid within the ARM SGP domain, and then an approximate 3° latitude × 3° longitude domain mean is calculated (see Figure 1 in Tang et al., 2016). The long-term continuous forcing data sets are available from 1999 to 2011 and are being updated to the present (version 2). For the CAUSES project, the surface turbulent fluxes are reprocessed to include measurements from both ARM Energy Balance Bowen Ratio (EBBR) and Eddy Correlation fluxes (ECOR) sites (Berg & Lamb, 2016). The EBBR turbulent flux estimates are calculated from observations of net radiation, soil surface heat flux, and the near-surface vertical gradients of temperature and relative humidity. The Bulk Aerodynamic technique is applied to the EBBR data streams to address sunrise and sunset spikes in the fluxes data (EBBR handbook, [https://www.arm.gov/publications/tech\\_reports/handbooks/ebbr\\_handbook.pdf](https://www.arm.gov/publications/tech_reports/handbooks/ebbr_handbook.pdf)). The ECOR technique estimates the vertical fluxes by correlating the vertical wind component with temperature (sensible heat flux) and humidity (latent heat flux) (ECOR handbook, [https://www.arm.gov/publications/tech\\_reports/handbooks/ecor\\_handbook.pdf](https://www.arm.gov/publications/tech_reports/handbooks/ecor_handbook.pdf)). One feature of the ECOR stations is that they do not close the surface energy budget. The ARM EBBR stations are located in areas with pasture and grassland, while ECOR stations are deployed in cropland and a single wooded location. In our analysis, up to 19 EBBR sites and 13 quality controlled ECOR sites are analyzed to calculate the areal mean. For soil moisture, measurements are taken from the Soil Water and Temperature System (SWATS, Schneider et al., 2003, Bond, 2005) with additional stringent data quality controls described in Xie et al. (2010). Supplementing the two sensors at the ARM SGP Central Facility, up to 10 measurements from other SWATS stations surrounding the Central Facility are further synthesized into a domain average value (~3° latitude × 3° longitude) for better comparison with model grid point values (version 1.3, doi:10.5439/1178331). More technical details are provided in Tang and Xie (2015). Note that none of the measurements from NOAA QCLCD, CERES, and FLUXNET are used in creating the ARM continuous forcing data sets. Therefore, the gridded data sets from NOAA QCLCD, CERES, and FLUXNET are independent of ARM observations. The observational uncertainty and limitations of above data sets are discussed in Kato et al. (2013), Jung et al. (2011), Zhang et al. (2018), and references therein. In general, model biases in the T<sub>2m</sub> and surface fluxes are much larger than the observational uncertainty, especially over the SGP.

## 2.2. Model and Experiments

For the surface energy budget analysis, model output from the short-term hindcast experiments 1 (Experiments 1c and 1d in particular) were analyzed. The detailed experiment design is described in Morcrette et al. (2018) and also documented on the project website (<http://portal.nersc.gov/project/capt/CAUSES/>). In short, each participating modeling center performed a series of 5 day long hindcasts starting every day in the period from 00Z 1 April 2011 to 00Z 31 August 2011. The initial conditions for atmospheric state variables are from the ERA-Interim (Dee et al., 2011). Experiments 1c and 1d include output variables interpolated to a 1° latitude × 1° longitude horizontal grid with hourly temporal resolutions for the entire CONUS domain. Simulations of T<sub>2m</sub>, surface upward and downward shortwave and longwave radiative fluxes, latent, and sensible heat fluxes are the main focus although precipitation and soil moisture simulations are also examined since they are relevant to the surface turbulent flux simulations. The nine (9) participating modeling centers that contributed output in the required format to perform the spatial

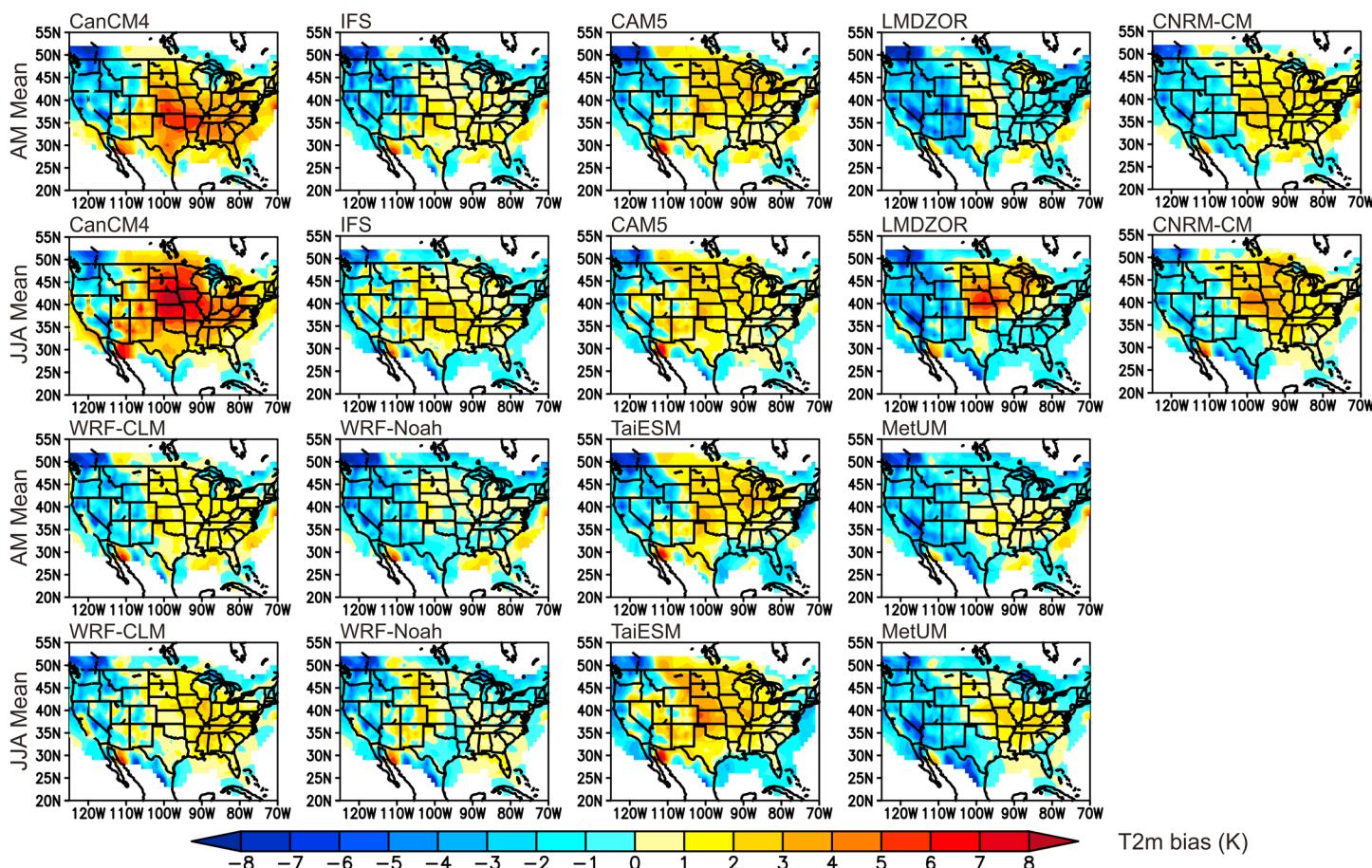
**Table 1***Summary of Participating Models Analyzed in This Study*

Model name	Submitting institute	Resolution	References
CanCM4	Canadian Centre for Climate Modelling and Analysis	T63, L36	Merryfield et al. (2013)
IFS	European Centre for Medium-Range Weather Forecasts	T511, L91	ECMWF, CY41R1
CAM5	Lawrence Livermore National Laboratory	1.25° × 0.9°, L30	Neale et al. (2012)
LMDZOR	Laboratoire de Météorologie Dynamique	LONxLAT: 144 × 142, L39	Hourdin et al. (2012)
WRF-CLM	Pacific Northwest National Laboratory	dx = 36 km, L35	Ma, Rasch, et al. (2014)
WRF-Noah	Pacific Northwest National Laboratory	dx = 36 km, L35	Ma, Rasch, et al. (2014)
TaiESM	Academia Sinica, Taiwan	1.25° × 0.9°, L30	Neale et al. (2012), Lee et al. (2013), and Wang et al. (2015)
MetUM	Met Office	N512(~dx = 30 km), L52	Brown et al. (2012) and Walters et al. (2014)
CNRM-CM	CNRM, Meteo-France/CNRS	T127, L31	Voldoire et al. (2013) and Masson et al. (2013)

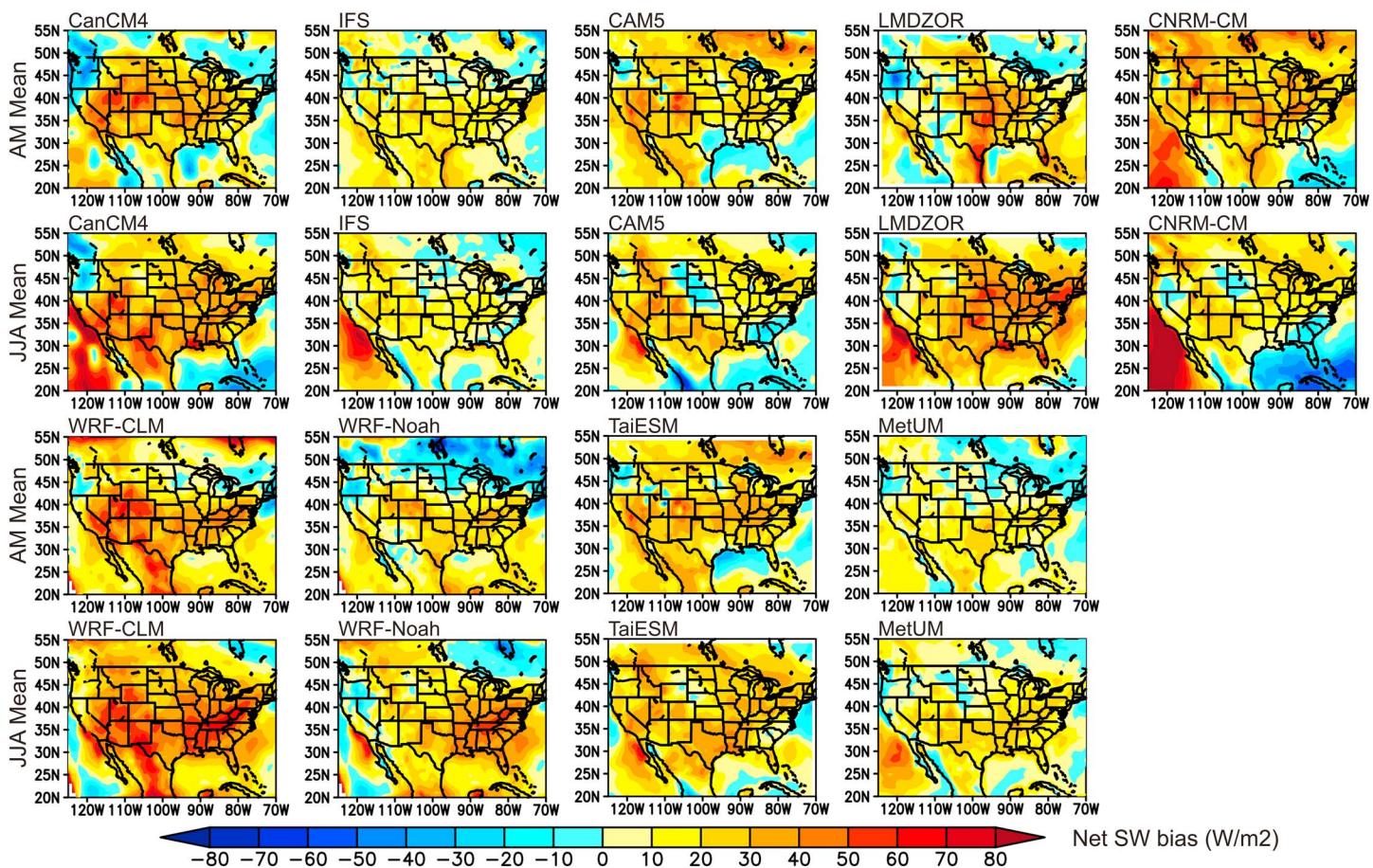
analysis over the CONUS are listed in Table 1. Due to model spin-up issue (Ma et al., 2013; Ma, Xie, et al., 2014), our analysis only focuses on hindcasts from 24 to 120 hindcast lead time (Day 2 to 5).

### 3. Errors in Surface Energy Budget and Their Connection to Surface Air Temperature ( $T_{2m}$ ) Bias

We first demonstrate the characteristics of errors in the  $T_{2m}$  and surface energy budget terms over the CONUS domain as well as over the ARM SGP as we plan to identify the common error features across all the models. We then seek the connection of surface energy budget errors to the  $T_{2m}$  error by examining daily,



**Figure 1.** April–May and June–August 2011 mean  $T_{2m}$  bias (K). Model values are averaged over Day 2 to 5 (24 to 120 h) hindcast lead time. The reference field is from NOAA QLCD. AM = April and May; JJA = June, July, and August.



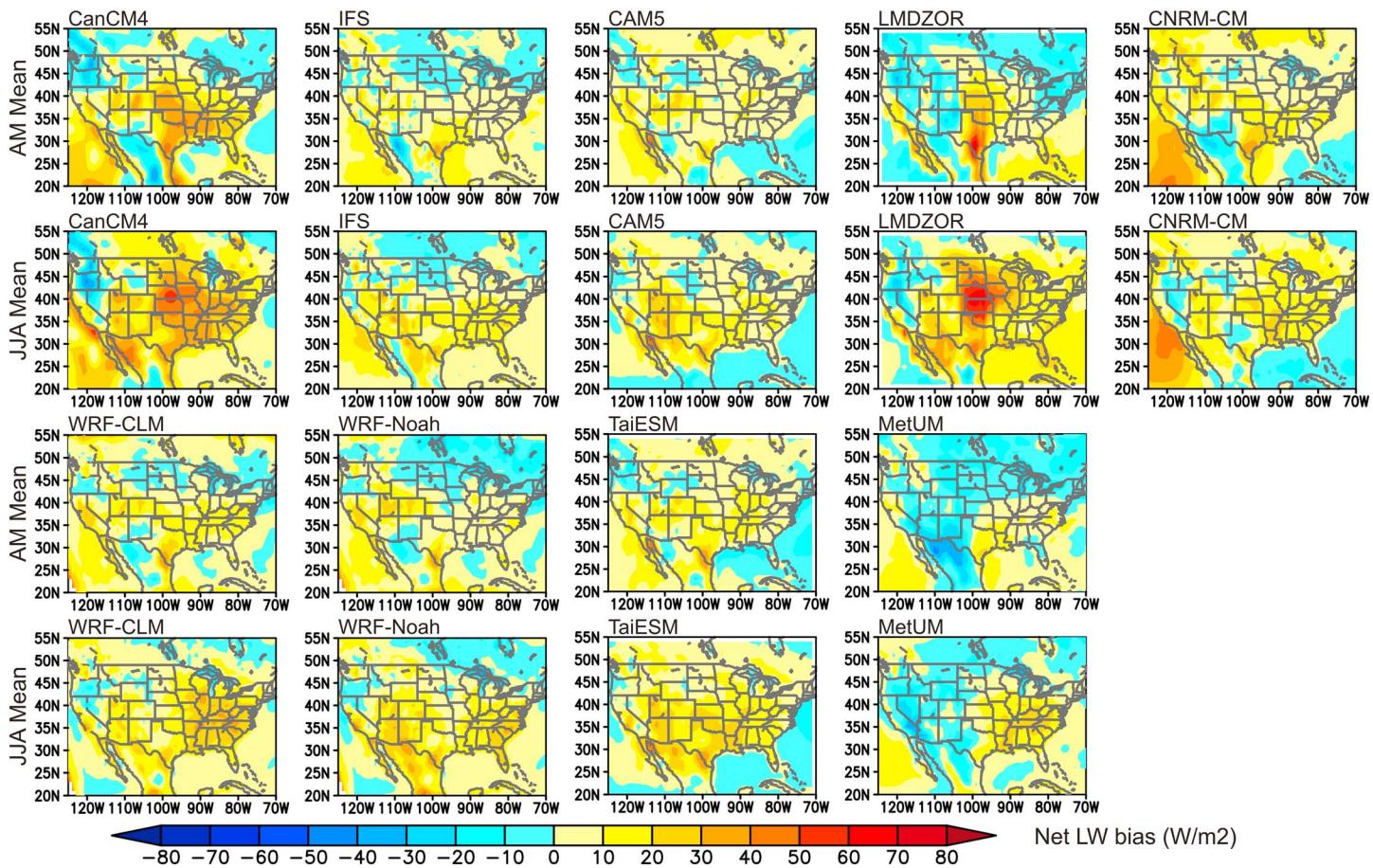
**Figure 2.** Same as Figure 1 except for net surface shortwave flux ( $\text{W m}^{-2}$ ). The reference field is from CERES EBAF. The mean biases are averaged for the full 24 h in a diurnal cycle.

monthly, and seasonal mean error quantity from model hindcasts both in space and time correlation analyses. Furthermore, we will compare error characteristics for certain model pairs that have a lot in common, such as CAM5 and TaiESM or WRF-Noah and WRF-CLM. The major differences that could affect  $T_{2m}$  bias simulations between CAM5 and TaiESM are the deep convection trigger function and cloud microphysics (Wang et al., 2015), while the only difference between WRF-CLM and WRF-Noah is the land model. All the initial conditions and boundary forcings between these model pairs are otherwise the same. Therefore, the comparison for the former pair can demonstrate the role of cloud parameterization changes on  $T_{2m}$ , and the comparison for the latter can demonstrate the role of land surface processes on  $T_{2m}$ .

We further define the sign convention for surface fluxes used in this study. Surface upward and downward shortwave radiative fluxes (SWUP and SWDN, respectively), as well as upward and downward longwave radiative fluxes (LWUP and LWDN, respectively), are all positive regardless their directions. Net surface shortwave radiative flux or absorbed shortwave radiative flux (SWNET) is defined as SWDN minus SWUP. Net surface longwave radiative flux (LWNET) is defined as LWUP minus LWDN. Surface latent heat (LH) and sensible heat (SH) fluxes are positive from the land surface into the atmosphere and negative from the atmosphere into the land. Therefore, positive (negative) SWNET indicates heat gain (loss) for the land surface, while positive (negative) LWNET, LH, and SH indicate heat loss (gain) for the land surface.

### 3.1. Error Characteristics Over the CONUS

Figure 1 shows the April–May (spring) and June–August (summer) 2011 mean  $T_{2m}$  bias (averages from Day 2 to 5 hindcast lead time) in reference to NOAA QCLCD. In general, all models simulate a warm  $T_{2m}$  bias over Central United States with the largest warm bias and spatial extent from CanCM4.  $T_{2m}$  simulations from

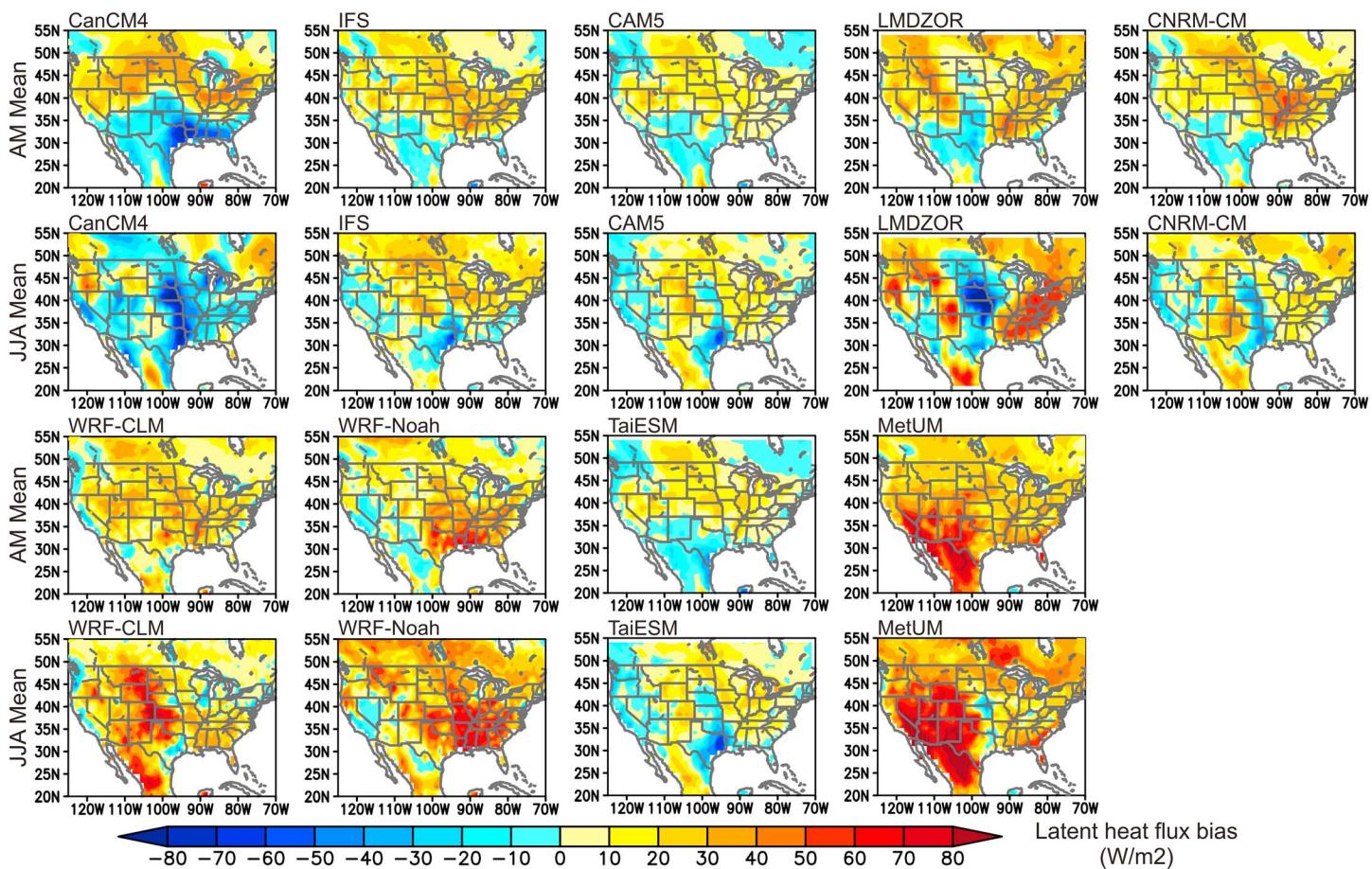


**Figure 3.** Same as Figure 1 except for net surface longwave flux ( $\text{W m}^{-2}$ ). The reference field is from CERES EBAF.

WRF-Noah and MetUM show the least bias magnitude and extent. Comparing spring and summer seasons, all models simulate larger bias magnitude and broader extent in summer and bias magnitude can be as large as 8 K. The bias pattern between CAM5 and TaiESM are very similar. The bias magnitude is slightly large in TaiESM in June, July, and August. The WRF-CLM simulate larger and broader warm bias compared to WRF-Noah. Readers are referred to Morcrette et al. (2018) for more discussion on warm bias extent and their characteristics.

Figures 2–5 present the spring and summer mean biases of surface energy budget terms: SWNET, LWNET, LH, and SH, respectively. For SWNET (Figure 2), all models simulate positive biases (excessive warming to the surface), mainly associated with biases in the cloud simulations (see Van Weverberg et al., 2018). The positive bias magnitude and extent over CONUS is slightly larger and broader in summer for most models except for CAM5 and CNRM-CM, which show slightly negative bias over the Midwest and Southeast of the United States. It is also worth knowing that the bias magnitude and extent in WRF-CLM is largest among all the models although the warm  $T_{2m}$  bias is much smaller (cf. Figure 1) indicating that there are other factors contributing more to its  $T_{2m}$  bias. LWNET biases (Figure 3) are generally positive (excessive cooling of the surface) in the Central United States with similar bias patterns to those of  $T_{2m}$ . This is mostly because the LWNET bias is mainly contributed by LWUP bias associated with warmer skin temperature.

Unlike surface radiative flux biases, there are no consistent mean bias signs and magnitude among the participating models for LH (Figure 4). In spring, most models simulate positive LH biases (too much evaporation) in the Central United States except for CanCM4 and LMDZOR which simulate negative LH biases over Kansas, Oklahoma, and Texas. In summer, most model simulate negative LH biases in the Central or Southern United States ( $\sim 105^{\circ}\text{W}$ – $95^{\circ}\text{W}$  and  $30^{\circ}\text{N}$ – $45^{\circ}\text{N}$ ) except for the two WRF model versions and MetUM, which even show

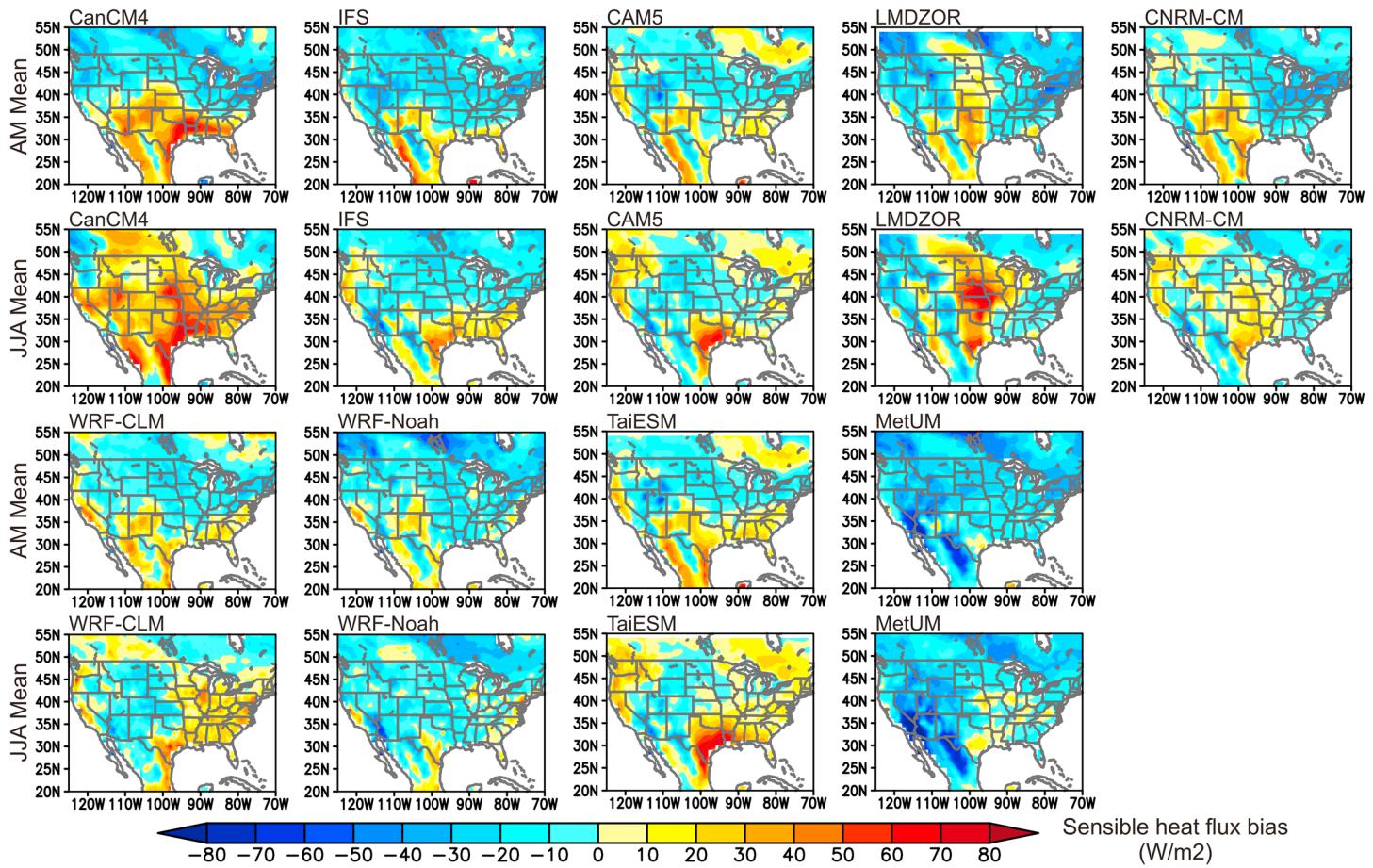


**Figure 4.** Same as Figure 1 except for surface latent heat flux ( $\text{W m}^{-2}$ ). The reference field is from FLUXNET.

slightly larger positive biases. For SH biases (Figure 5), there are also no consistent mean bias signs and magnitude among the participating models in spring and summer. In spring, most models simulate negative biases in the Central United States except for CanCM4 and LMDZOR. In summer, the magnitude for negative biases in most models either is smaller or becomes positive biases. One exception is WRF-Noah, which shows similar negative bias magnitude.

Comparing CAM5 to TaiESM, we find that changes in the convective triggering and cloud microphysics lead to slightly larger SWNET bias in the TaiESM. Through the radiation attribution study by Van Weverberg et al. (2018), TaiESM is shown to underestimate the frequency of occurrence of deep clouds as deep convection is triggered less frequently by the model due to the imposed new trigger function. TaiESM also misses low- and middle-tropospheric cloud amount especially in the afternoon. Both of which lead to larger positive SWNET bias during the daytime. The LH biases from both models are very similar for both spring and summer, while SH bias in both CAM5 and TaiESM is slightly positive in the Central United States and has a broader extent in summer.

Comparing WRF-CLM to WRF-Noah, it is interesting to point out that WRF-CLM simulates much larger positive SWNET bias even with the same atmospheric dynamics and physics, as well as initial and boundary conditions. Both WRF versions simulate positive LH biases over the Central United States although larger positive LH bias is more toward the west in WRF-CLM and more toward the east in WRF-Noah. Both WRF version simulate slightly negative SH biases over the Central United States. WRF-Noah shows broader negative bias extent over the CONUS domain. It is clear that land-atmosphere interactions have significant impact on the cloud and radiation simulations during the study period. However, the pathways of how the land processes affect the planetary boundary layer, cloud, and radiation processes are not clear and beyond the scope of this study.



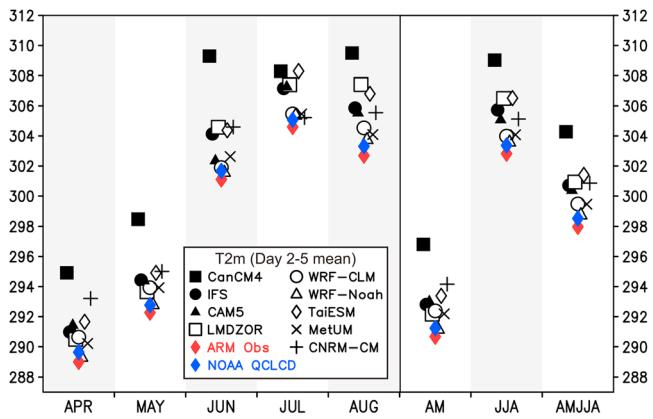
**Figure 5.** Same as Figure 1 except for surface sensible heat flux ( $\text{W m}^{-2}$ ). The reference field is from FLUXNET.

### 3.2. Error Characteristics Over the ARM SGP

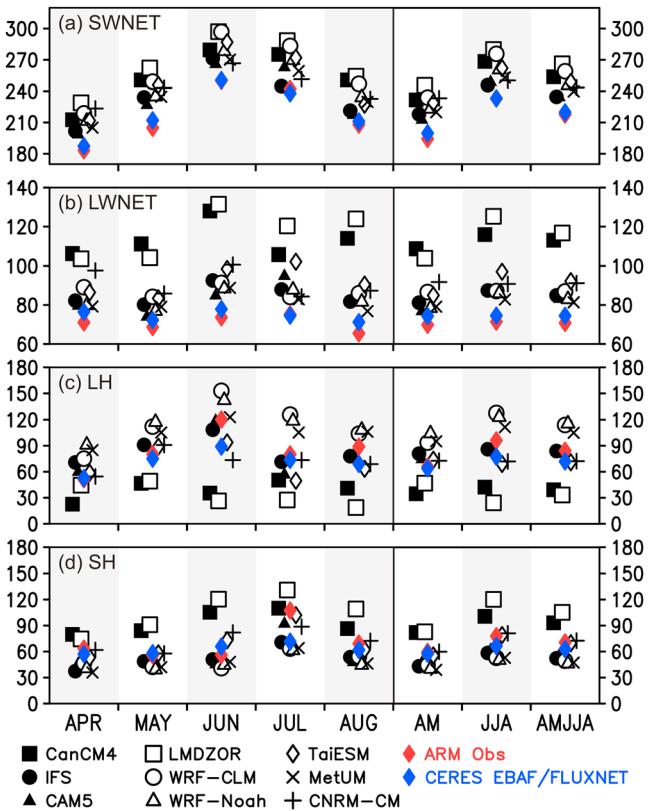
We further focus on the error analysis at the ARM SGP and see whether similar bias features found over the Central United States are also present here. Figure 6 shows the 2011 monthly and seasonal mean  $T_{2m}$  from observations and simulations at ARM SGP. To consider potential uncertainties in the observations, we also plotted the  $T_{2m}$  values from NOAA QCLCD, and they are  $\sim 0.5$  K warmer than the ARM observations.

Compared to ARM observations (red diamond), all models simulate warmer  $T_{2m}$  with largest bias magnitude from CanCM4 for  $\sim 6$  K and smallest from WRF-Noah for  $\sim 0.5$  K (overlapped with QCLCD). Several models, such as CAM5, IFS, LMDZOR, and TaiESM, show slightly larger  $T_{2m}$  values in summer compared to those in spring.

Figure 7 shows 2011 monthly and seasonal mean surface energy budget terms at SGP. Similar to what were presented from the CONUS domain, all models simulate larger SWNET and LWNET compared to ARM observations or CERES EBAF (blue diamond). SWNET biases do not seem to change with month or season, while LWNET biases are slightly larger in summer mostly due to larger LWDN. For LH and SH, there are no consistent mean bias sign across all models, even when considering different observed data sets (ARM versus FLUXNET). The characteristics of  $T_{2m}$  and surface energy budget simulations shown at the SGP are consistent with the discussion earlier for the CONUS domain.



**Figure 6.** Year 2011 monthly and seasonal mean  $T_{2m}$  (K) over the ARM SGP domain (averaged over  $35^{\circ}\text{--}38^{\circ}\text{N}$ ,  $99^{\circ}\text{--}96^{\circ}\text{W}$ ). Model values are averaged over Day 2 to 5 (24 to 120 h) hindcast lead time. The symbols are staggered in time just for clarity.



**Figure 7.** Same as Figure 6 except for surface net shortwave (SWNET), net longwave (LWNET), latent (LH) and sensible (SH) heat fluxes ( $\text{W m}^{-2}$ ). The symbols are staggered in time just for clarity.

after sunrise and gradually become positive biases in the afternoon. In summer, IFS, CAM5, TaiESM, MetUM, and CNRM-CM simulate larger negative LH biases during the daytime, while the two WRF versions still simulate positive LH biases. The diurnal evolutions are similar between CAM5 and TaiESM except TaiESM shows slightly larger diurnal variations. For WRF-CLM and WRF-Noah, WRF-CLM simulates larger  $T_{2m}$ , SWNET, and SH biases, while WRF-Noah simulates larger LH bias. The turbulent fluxes indicate larger EF in WRF-Noah consistent with larger soil moisture as will be demonstrated later in Figure 14.  $T_{2m}$  biases in both WRF versions are larger during nighttime in spring, while  $T_{2m}$  bias is larger during the daytime in WRF-CLM in summer.

### 3.3. Connection of the Surface Energy Budget Errors to $T_{2m}$ Error

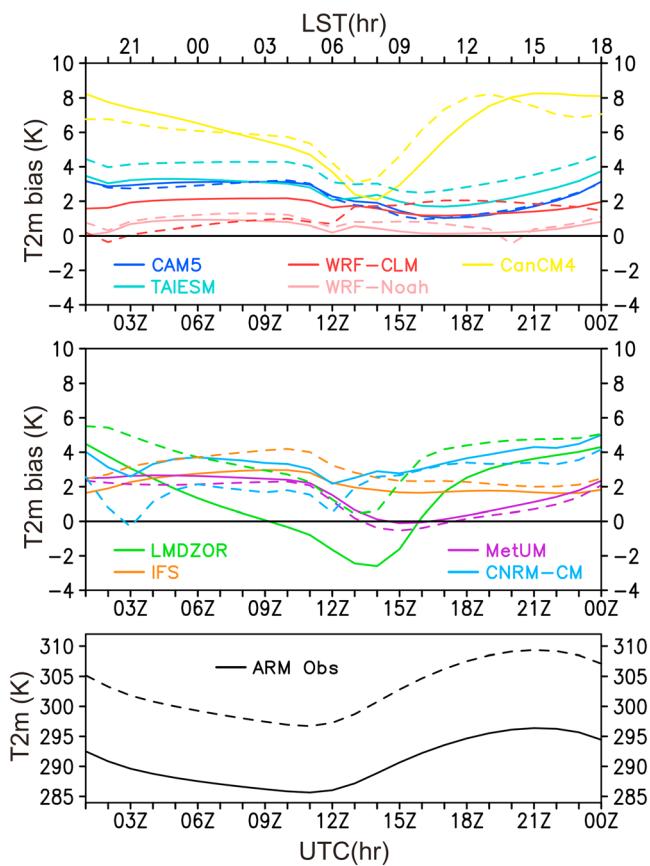
In this section, we identify the connection of the surface energy budget errors to  $T_{2m}$  error through performing correlation analysis both in space over the CONUS domain and in time at the ARM SGP. The goal is to distinguish the main contributors to the  $T_{2m}$  error among each of the energy budget terms.

As most models show very broad extent of warm bias over the Central United States, we first identify for each model whether there is a common and robust connection between warm  $T_{2m}$  bias and surface energy budget term biases spatially over the portion of the Central United States ( $30\text{--}50^\circ\text{N}$ ,  $85\text{--}105^\circ\text{W}$ ) where most models simulate the largest  $T_{2m}$  bias. Large spatial error correlations indicate large spatial coherent patterns (similar error contributors) between  $T_{2m}$  and the selected fields.

Figure 10 shows the spatial bias correlation coefficients between monthly or seasonal mean  $T_{2m}$  bias and biases from surface energy budget terms or other related fields, such as precipitation. Positive correlations between  $T_{2m}$  bias and SWNET or LWNET biases are found for most models in spring. However, in summer, a few models show negative correlations. For LWNET, large positive correlations are found for most

We further present in Figures 8 and 9 the spring and summer mean diurnal cycles of  $T_{2m}$  and energy budget term biases at the ARM SGP, respectively. It is clear that biases in both  $T_{2m}$  and energy budget terms show strong diurnal cycles in their magnitude. For  $T_{2m}$  (Figure 8), most models show largest bias in the early morning between 09 and 12Z (local time is UTC-6), while CanCM4, LMDZOR, TaiESM, and CNRM-CM show largest bias around 00-01Z. The diurnal cycle amplitude for  $T_{2m}$  is larger in summer (dashed lines) than in spring (solid lines) for IFS, LMDZOR, TaiESM, WRF-Noah, and CNRM-CM, while CanCM4, CAM5, WRF-CLM and MetUM show similar bias amplitude.

SWNET biases (Figure 9) are largest between 18 and 21Z for all the models with slightly larger magnitude in spring for most models. Right around sunrise and sunset, all models simulate slightly negative SWNET biases because radiative schemes in most models are usually called with a larger time interval (more than one model time step) compared to other parameterizations. This leads to slightly underestimation of shortwave flux when compared to high-temporal frequency of ARM measurements (subhourly). LWNET biases are relatively constant throughout the diurnal cycle with slightly larger magnitude in the late afternoon for most models. Diurnal cycles of LH and SH biases show very different behaviors across all models as seen from the monthly and seasonal mean bias plots earlier. For individual models, the bias magnitude for LH and SH is generally largest during the daytime and close to zero during the nighttime. This is, of course, unsurprising as LH and SH are  $\sim 0$  at night in both models and observations. Models that simulate larger SWNET biases also show larger LH and SH biases in the diurnal cycle amplitude, such as CanCM4 and LMDZOR. In spring, CanCM4 and LMDZOR simulate negative LH biases but positive SH biases between 16 and 20Z. The bias magnitude is larger in summer. For the other seven models in spring, negative LH biases are present



**Figure 8.** April–May 2011 (solid lines) and June–August 2011 (dashed lines) mean diurnal cycles of T<sub>2m</sub> (K) over the ARM SGP domain (averaged over 35–38°N, 99–96°W). Model values are averaged over Day 2 to 5 (24 to 120 h) hindcast lead time. The time is UTC on the bottom and local time is UTC–6 h.

Finally, we performed correlation analysis on the seasonal mean biases of individual energy budget terms to T<sub>2m</sub> bias across all the models at the ARM SGP. The goal is to identify whether there is any linear relationship between the magnitude of seasonal mean T<sub>2m</sub> bias and the magnitude of individual energy budget term biases across all the models. Figures 13 and 14 show scatterplots of seasonal mean bias between T<sub>2m</sub> and selected surface energy and water budget terms for spring and summer across all the models. We also separate the analysis into daytime and nighttime analysis. Across all the models at the ARM SGP, large seasonal mean bias correlations are found between T<sub>2m</sub> and EF, sensible, or latent heat fluxes for both daytime and nighttime in spring and summer. Large correlation is found between T<sub>2m</sub> and downward longwave bias for nighttime. It is interesting to note that the bias correlations for SWNET or downward shortwave flux are not as large as those for LH, SH, or EF. This suggests that models with larger positive SH or negative LH and EF biases are likely to simulate larger T<sub>2m</sub> mean bias on seasonal time scales.

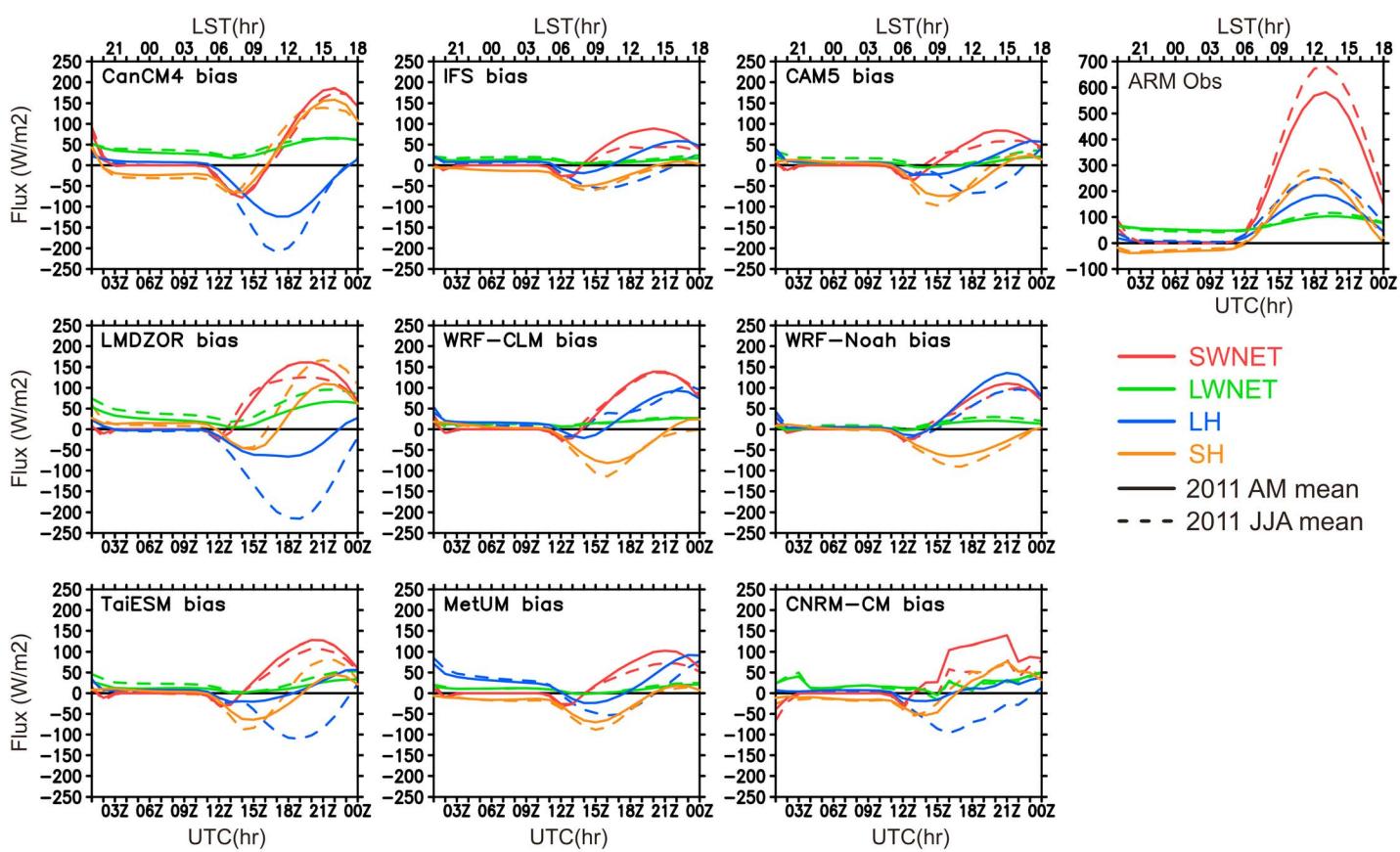
### 3.4. Water Budget Analysis

We identified EF bias as a major contributor to the T<sub>2m</sub> bias from the previous section. Here we further performed analysis of accumulated precipitation (P), accumulated evaporation (E), and 10 cm soil moisture in the models as an attempt to understand why models have errors in the EF. We will also identify whether some

models and this is mostly due to the surface upward longwave bias associated with warmer skin temperature and T<sub>2m</sub>. Positive correlations are also found between T<sub>2m</sub> and LWBN biases for most models. For surface turbulent fluxes in most models, there are negative correlations between T<sub>2m</sub> and EF biases, and in contrast, there are large positive correlations between T<sub>2m</sub> and SH. For precipitation and LH biases, no consistent sign in the spatial correlation coefficients is found in spring or summer.

We next demonstrate qualitatively for each model the connection between errors in T<sub>2m</sub> and surface energy and water budgets on daily time scales over the ARM SGP. The idea here is to identify the major contributors of the surface energy budget errors to T<sub>2m</sub> day-to-day error variations and also to identify if there are common features across the models. Figures 11 and 12 show daily bias correlation coefficients between T<sub>2m</sub> and selected surface energy and water budgets for 2011 spring and summer. We further separate the analysis into daytime and nighttime since SWNET, LH, and SH biases are most dominant during the day, while LWBN biases are most dominant at night. For daytime in spring (Figure 11a), T<sub>2m</sub> bias shows largest positive correlations with SWNET, SWUP, and SWDN with values between ~0.45 and ~0.65 for all the models. T<sub>2m</sub> bias also show positive correlations with LH and SH biases for all the models except for LH bias from CanCM4. Yet T<sub>2m</sub> bias shows negative correlations with EF, precipitation, and 10 cm soil moisture biases for most models. For daytime in summer (Figure 11b), the positive correlations for SWNET, SWUP, and SWDN are smaller for some models (LMDZ, WRF-Noah, WRF-CLM, and CAM5). The negative correlations for EF, precipitation, and 10 cm soil moisture biases are larger for most models. The correlations for LWBN become more positive compared to spring values. The correlations for LH become negative from positive values for most models.

During the nighttime (Figure 12), T<sub>2m</sub> bias shows positive correlations with downward longwave flux bias and negative correlations for sensible heat flux bias for all the models in both spring and summer. The signs of the correlation coefficients are consistent with overly warm near-surface air adding heat to the surface through radiation and turbulent processes. The correlations of T<sub>2m</sub> bias with those of precipitation and 10 cm soil moisture become negative from spring to summer.

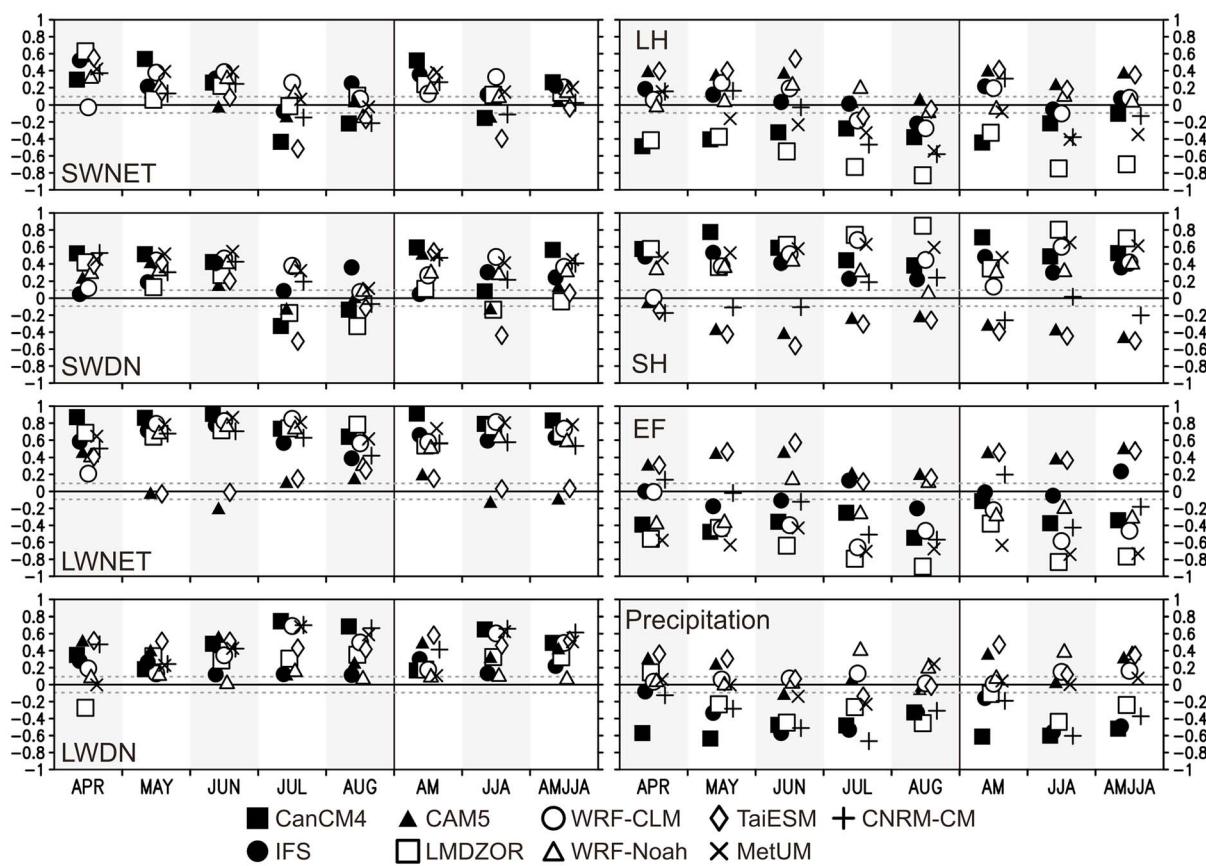


**Figure 9.** April–May 2011 (solid lines) and June–August 2011 (dashed lines) mean diurnal cycles of surface net shortwave (SWNET), net longwave (LWNET), latent (LH) and sensible (SH) heat fluxes ( $\text{W m}^{-2}$ ) over the ARM SGP domain (averaged over  $35\text{--}38^\circ\text{N}$ ,  $99\text{--}96^\circ\text{W}$ ). Model values are averaged over Day 2 to 5 (24 to 120 h) hindcast lead time. The time is UTC on the bottom and local time is UTC–6 h.

models have EF errors even with reasonable soil moisture. The former condition might reflect incorrect atmospheric precipitation forcing (although land model errors may still contribute), and the latter might indicate land model errors.

Figure 15 shows accumulated P, E, and P-E from 1 April to 31 August for both ARM observations and model hindcasts. MetUM and the two WRF model versions overestimate P and E by ~25–40%, while CanCM4 and LMDZOR underestimate P and E by ~25–50%. CNRM-CM underestimate P after late spring. CAM5, IFS, and TaiESM slightly underestimate P after early August. CAM5 and IFS simulate very good E, while TaiESM and CNRM-CM simulate slightly lower E. From observations, P-E become overall negative after June in 2011 suggesting that the land is drying out from early summer. This could be one of the reasons for different model behaviors in the simulated surface energy budget between the spring and summer seasons. In terms of absolute values of P-E, CAM5, TaiESM, and MetUM simulate reasonable P-E and IFS and CNRM-CM slightly overestimate P-E. The two WRF versions significantly overestimate the absolute magnitude of P-E mostly due to excessive evaporation, while CanCM4 and LMDZOR simulate much smaller P-E mostly due to low evaporation, suggesting a drier land condition (as will be shown later in the 10 cm soil moisture plot).

Figure 15d shows the 10 cm soil moisture from SWATS measurements and model simulations. Although there can be large uncertainties in SWATS measurements, particularly for the extreme higher and lower values of soil moisture, as discussed in Phillips et al. (2017), SWATS provides a reference state when comparing simulated 10 cm soil moisture and its relationship to precipitation and evaporation. We find that all models simulate lower 10 cm soil moisture for most days compared to SWATS. Much lower 10 cm soil moisture is simulated by CanCM4, IFS, LMDZOR, and CNRM-CM. The lower soil moisture in these four model hindcasts are mainly due to their land initial conditions. As indicated in Ma et al. (2015), initial soil moisture in the land model depends highly on the method that is used to generate the land initial conditions. For example, an



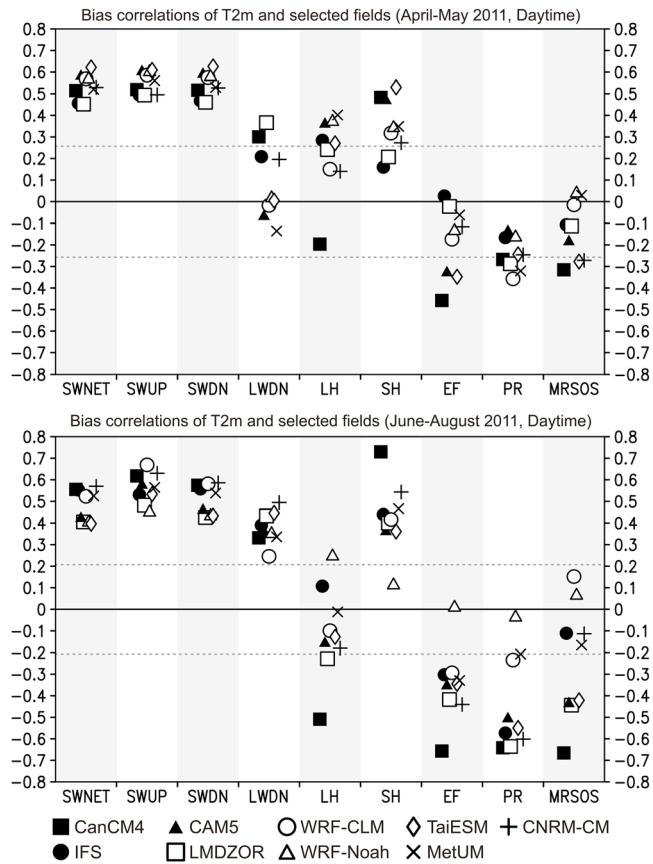
**Figure 10.** Spatial bias correlation coefficients between monthly or seasonal mean  $T_{2m}$  (K) bias and biases from selected fields over the Central United States (domain of correlation: 30–50°N, 85–105°W). Model values are averaged over Day 2 to 5 (24 to 120 h) hindcast lead time. The symbols are staggered in time just for clarity. The gray dotted lines indicate the level of statistical significance for the correlations with probability  $p = 0.05$ . SWNET = net surface shortwave flux; SWDN = surface downward shortwave flux; LWNET = net surface longwave flux; LWDN = surface downward longwave flux; LH = latent heat; SH = sensible heat; EF = evaporative fraction.

atmospheric state nudging method could provide erroneous initial soil moisture because of the erroneous precipitation in the host atmospheric model. This is the case for the CanCM4 and LMDZOR initialization methods. It is not clear why there is low soil moisture in the IFS and CNRM-CM initial conditions. Another factor for the soil moisture variability across all the models is the uncertainty due to heterogeneity in soil properties even if the models have the right physics.

Nevertheless, we find that the daily 10 cm soil moisture varies similarly to the accumulated P-E to the first order (with temporal correlations between 0.6 and 0.7 for individual models) even though we do not consider the surface runoff and subsurface infiltration information from the models. This suggests that 10 cm soil moisture in the model is largely controlled by the day-to-day changes in accumulated P-E, indicating both atmospheric and land components can affect the simulations of the soil moisture. For the EF bias (Figure 15e), models with negative EF bias generally show lower 10 cm soil moisture and models with positive EF bias generally show higher 10 cm soil moisture. One exception is IFS, which shows reasonable EF but with relatively lower 10 cm soil moisture. Figure 15f demonstrates the relationship between seasonal mean EF bias and 10 cm soil moisture bias (in reference to SWATs). We can find that the bias correlations for both seasons are fairly high with  $R = 0.86$  for spring and  $R = 0.78$  for summer. This suggests a strong influence of 10 cm soil moisture on surface EF in the models.

#### 4. Theoretical Interpretation of the $T_{2m}$ Bias

The model-observational comparisons in both hindcast mode (the present paper) and in climate integrations (Cheruy et al., (2014); Zhang et al. (2018)) establish that the warm  $T_{2m}$  bias is associated with two factors: (1)

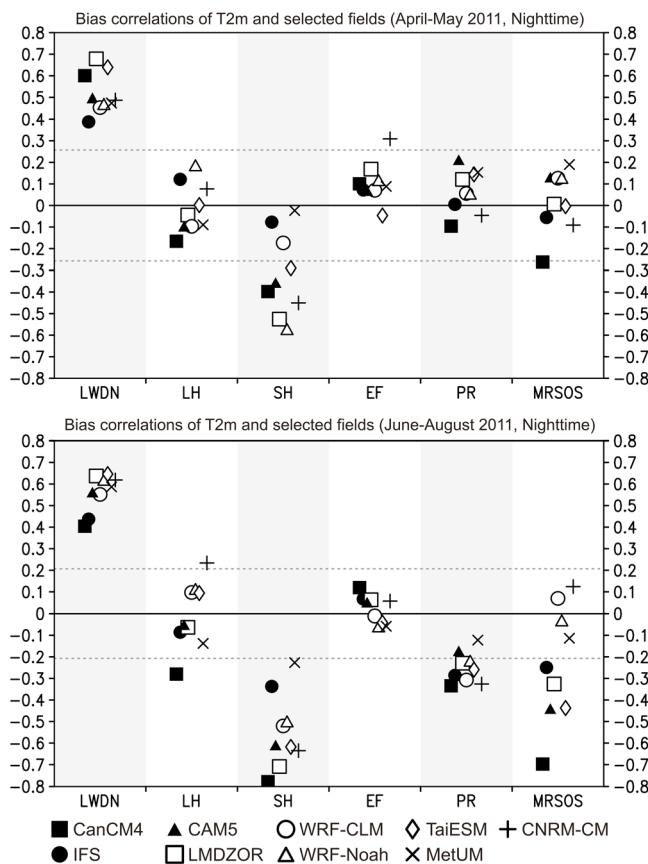


**Figure 11.** Daily bias correlation coefficients between 2011 daytime mean  $T_{2m}$  (K) bias and biases from selected fields for individual models over the ARM SGP domain (averaged over  $35^{\circ}$ – $38^{\circ}$ N,  $99^{\circ}$ – $96^{\circ}$ W). Model values are averaged over Day 2 to 5 (24 to 120 h) hindcast lead time. SWNET = net surface shortwave flux; SWUP = surface upwelling shortwave flux; SWDN = surface downward shortwave flux; LWDN = surface downward longwave flux; LH = surface latent heat flux; SH = surface sensible heat flux; EF = evaporative fraction; PR = precipitation; MRSOS = 10 cm soil moisture). The gray dotted lines indicate the level of statistical significance for the correlations with probability  $p = 0.05$ .

an excess of SWNET and (2) an underestimate of the fraction of surface heat loss that is in the form of evaporation. Physically, we can think of each factor as contributing to the warm bias. For the first factor, excess radiation absorbed directly heats the surface causing a warm bias. For the second factor, if evaporation is suppressed, radiative heating (even if unbiased) will be used to heat the surface instead of evaporating water causing a warm bias. These can be viewed as two independent sources of model error: even if the EF is correct, an excess of radiation heating will cause a warm bias; and even if the radiative heating is correct, an underestimate of evaporation will cause excess radiation to be used to heat the surface so that it gives off an excess of sensible heat and longwave radiation.

Distinguishing the contributions of each factor to a model's warm bias would be helpful in identifying which model components need fixing. The radiation error can roughly be viewed as an atmospheric problem as its biases are primarily associated with clouds (Qian et al., 2012; Van Weverberg et al., 2018; Zhang et al., 2018). The EF error depends on the model representation of the land surface including how water moves vertically through the land surface and how vegetation makes transpiration possible. This division is not perfect of course; errors in surface albedo makes a small contribution to the radiation error and errors in EF might be due to incorrect atmospheric forcing which does not give enough precipitation to the land surface. This is particularly the case for climate integrations (Klein et al., 2006, Zhang et al. (2018)). Nevertheless, for purposes of model development it is helpful to distinguish these two sources of error.

To this end, we derived an approximate equation (equation (1), taken from equation (A8) in Appendix A) to quantify the magnitudes of radiation and EF contributions to the warm bias (see Appendix A for the detailed



**Figure 12.** Same as Figure 11 except the correlations are calculated for nighttime period only. LWDN = surface downward longwave flux; LH = latent heat; SH = sensible heat; EF = evaporative fraction; PR = precipitation; MRSOS = 10 cm soil moisture.

state, whereas for climate models, the EF error is likely associated with a precipitation deficit, although land model physics are also likely important.

## 5. Summary and Discussion

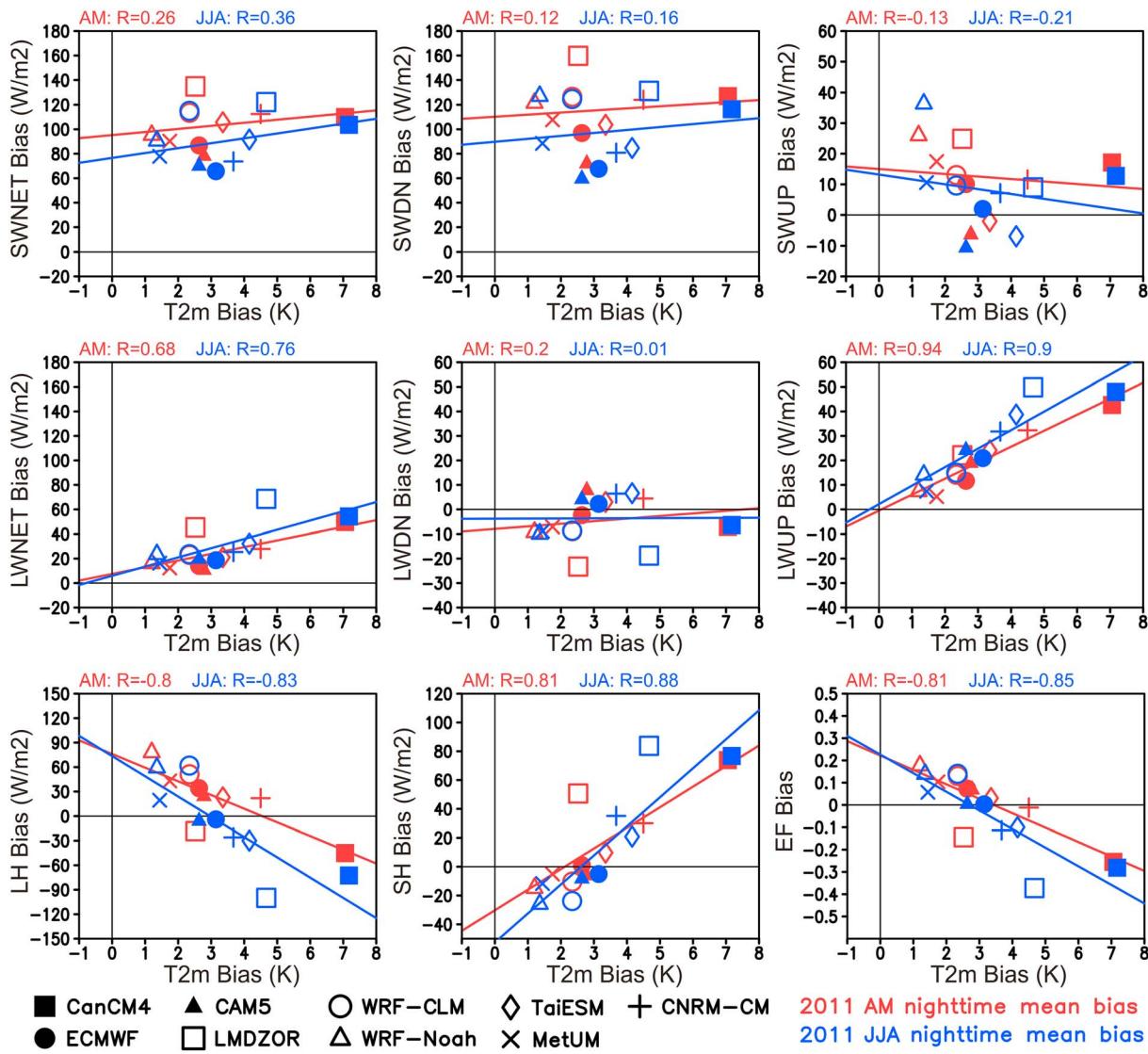
In this study, we examine the error contribution of surface energy budgets to spring and summertime warm  $T_{2m}$  bias over the Central United States using short-term hindcasts from nine weather forecast and climate models with observations from the ARM SGP sites as well as other available sources. This paper is one of the CAUSES intercomparison project components (Morcrette et al., 2018; Van Weverberg et al., 2018; Zhang et al., 2018), and the overall research foci for the project are (1) on the errors in clouds and radiation and (2) on the errors in surface water and energy budgets, in contributing to the spring and summertime  $T_{2m}$  bias over the Central United States.

Our analysis demonstrates that the warm  $T_{2m}$  bias is present within the first few days of model integration, and the warm bias is very robust across all the participating models over the Central United States and at the ARM SGP during the whole study period (April 2011 to August 2011). This is consistent with the findings in Morcrette et al. (2018). For the surface energy budget errors over both the Central United States and the ARM SGP, all models simulate too much net shortwave and net longwave fluxes at the surface but there is no consistency on the bias signs of LH and SH for both spring and summer seasons. In general, SWNET bias is slightly larger in spring, while the  $T_{2m}$ , LH, SH, and EF biases are larger in the summer. Errors in the surface energy budget terms also show large diurnal variations with largest bias magnitude during the daytime. We further performed analysis on the EF bias. Models with negative EF bias generally show lower 10 cm soil moisture and models with positive EF bias generally show higher 10 cm soil moisture. Further, day-to-day

derivation). Equation (1) is based upon the surface energy balance and expresses the physical mechanisms discussed above.

$$T'_{2m} = (SWNET' + LWDN')/30 - 2.5 \times EF' \times (SH_{mod} + LH_{mod})/30. \quad (1)$$

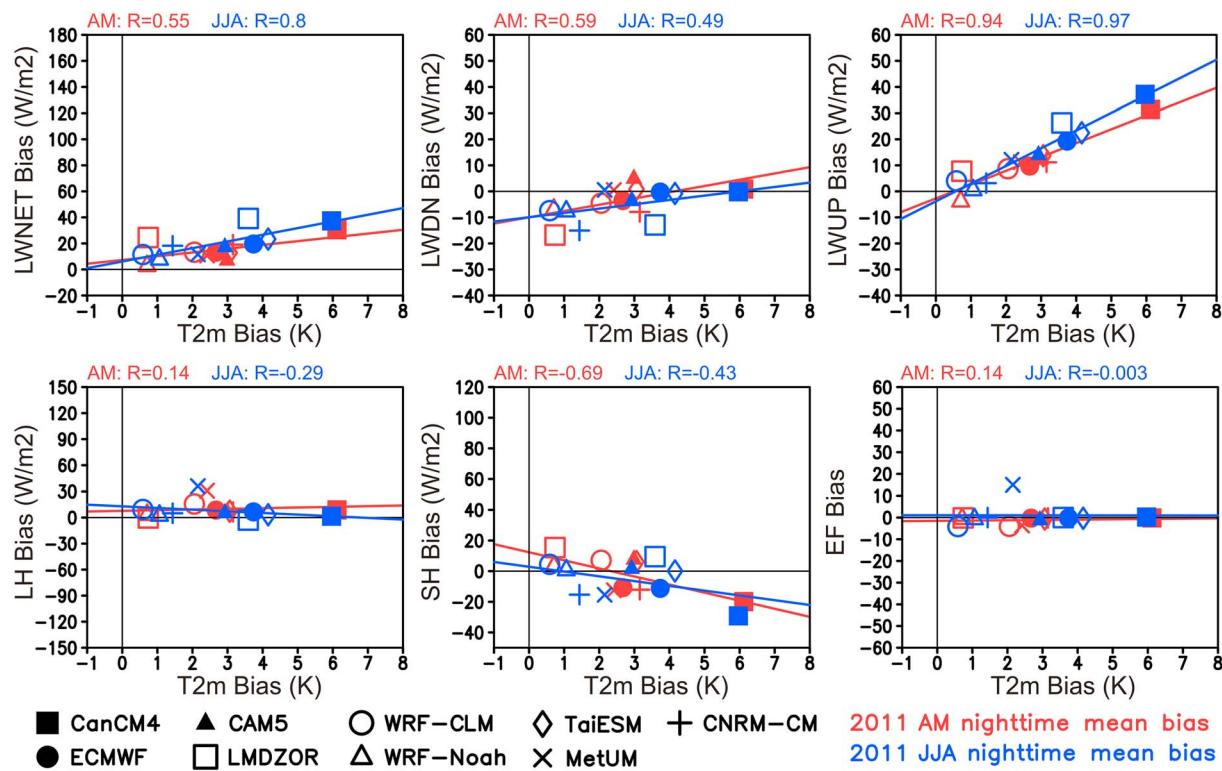
In (1),  $'$  represents the model variable bias in reference to observation, and  $_{mod}$  represents model mean value. We now use (1) to approximately identify separate contributions to  $T_{2m}$  error from errors in radiative input (first term on the right-hand side of (1)) and from errors in EF (second term on the right-hand side of (1)). These are shown in the scatterplots (Figure 16) of these components for the hindcast models and the CMIP5/AMIP counterpart (the CMIP5/AMIP data are from; Zhang et al., 2018). Their values are also listed in Tables 2 and 3. In doing so we find interesting similarities and differences between the hindcast and AMIP models. As for similarities, radiation errors cluster around 0.5–1.5 K with only two outlier models, whereas EF errors have a much wider intermodel spread ranging from −2.5 K in the WRF hindcast models to +5 K in some AMIP models. In both ensembles, EF and radiation errors are only weakly correlated; the correlation coefficients between the two terms are +0.45 and +0.35 for the hindcast and AMIP model ensembles, respectively. As for differences,  $T_{2m}$  errors due to radiation are more positive in the AMIP model ensemble than in the hindcast ensemble, but the increase in error due to radiation going from the hindcast ensemble to the AMIP ensemble is small, only about 0.5 K in the multimodel mean. Overall, this highlights the important role of the EF error in the presence of a relatively uniform but modest positive radiation error; namely, an underestimate of EF is the larger source of error for all models with a large temperature bias ( $> 4$  K), whereas in a few mostly hindcast models an overestimate of EF compensates for the radiation error and results in a small temperature bias. For hindcast models, the EF error is likely dominated by the initial condition of the land surface



**Figure 13.** Scatterplots of 2011 daytime seasonal mean (April–May mean for red and June–August mean for blue)  $T_{2m}$  (K) bias (abscissa) versus bias of selected fields (ordinate) across the participating models over the ARM SGP domain (averaged over 35–38°N, 99–96°W). Model values are averaged over Day 2 to 5 (24 to 120 h) hindcast lead time. The critical value for correlation coefficient with nine models is  $\pm 0.67$  with probability  $p = 0.05$ . LH = latent heat; SH = sensible heat; SWNET = net surface shortwave flux; LWNET = net surface longwave flux; SWUP = surface upward shortwave flux; SWDN = surface downward shortwave flux; LWUP = surface upward longwave flux; LWDN = surface downward longwave flux; EF = evaporative fraction; AM = April and May; JJA = June, July, and August.

changes in the 10 cm soil moisture are largely affected by the seasonal accumulated precipitation minus accumulated evaporation, indicating both atmospheric and land components can affect the simulations of the soil moisture.

We also focus on the comparison between two model pairs, CAM5 and TaiESM, as well as WRF-CLM and WRF-Noah since the major difference in the former pair is the deep convection trigger function, which can demonstrate the role of cloud parameterization changes on  $T_{2m}$ , while the only difference between the latter pair is the land model, which can demonstrate the role of land processes and land-atmosphere interactions on  $T_{2m}$ . Changes in the deep convection triggering function and cloud microphysics affect cloud simulations which leads to slightly larger SWNET bias in the TaiESM. On the other hand, land processes and land-atmosphere interactions have significant impact on the surface energy budget and  $T_{2m}$  simulations between the two WRF versions. WRF-CLM simulates larger positive SWNET bias, while WRF-Noah simulates larger positive LH and negative SH biases consistent with larger 10 cm soil moisture and EF. Both of which lead to smaller



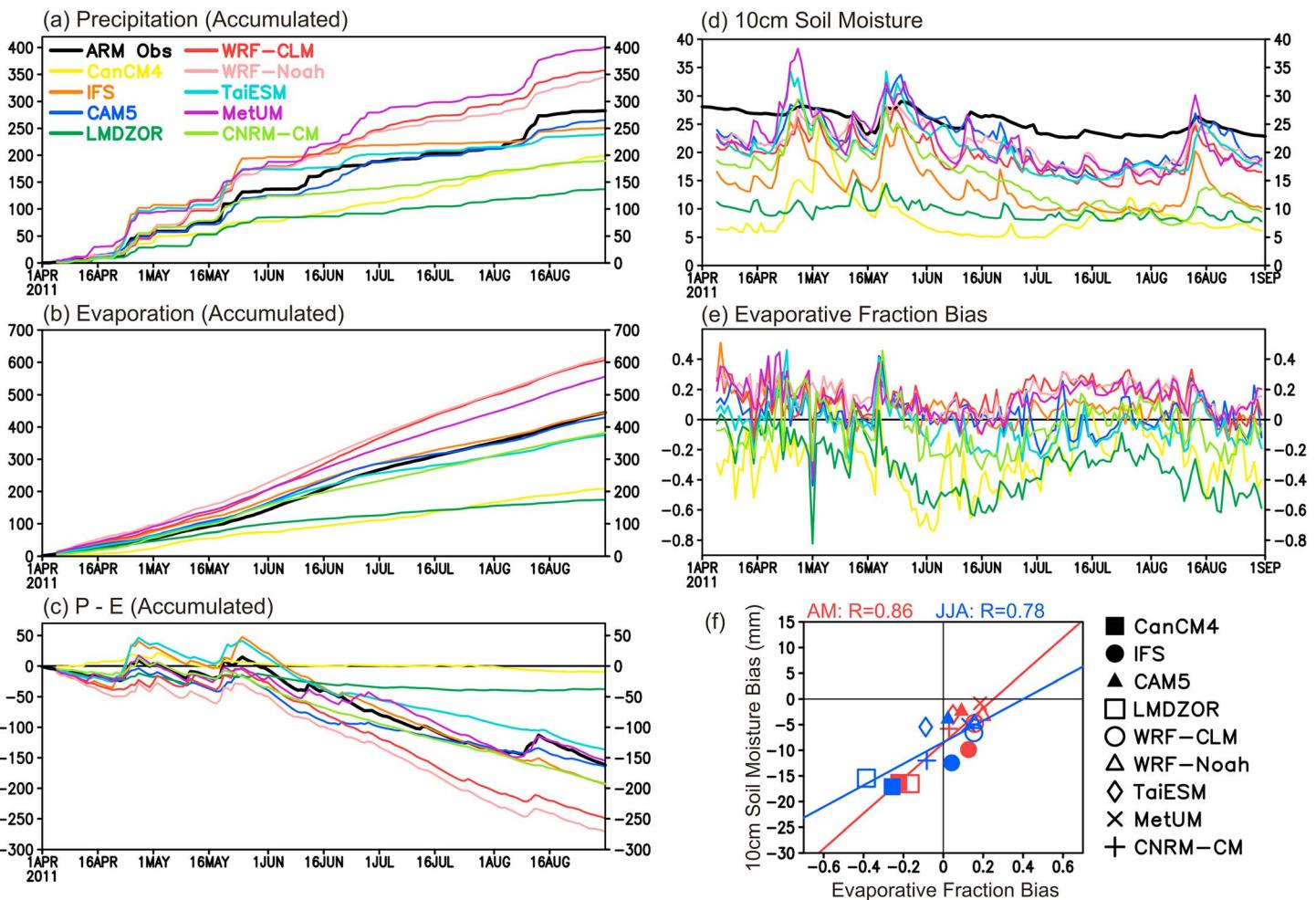
**Figure 14.** Same as Figure 13 except for the nighttime mean biases. LH = latent heat; SH = sensible heat; SWNET = net surface shortwave flux; LWNET = net surface longwave flux; SWUP = surface upward shortwave flux; SWDN = downward shortwave flux; LWUP = surface upward longwave flux; LWDN = surface downward longwave flux; EF = evaporative fraction; AM = April and May; JJA = June, July, and August.

$T_{2m}$  bias in WRF-Noah. Nevertheless, the pathway of how the land processes affect the planetary boundary layer, cloud, and radiation processes is not clear and will require further investigation.

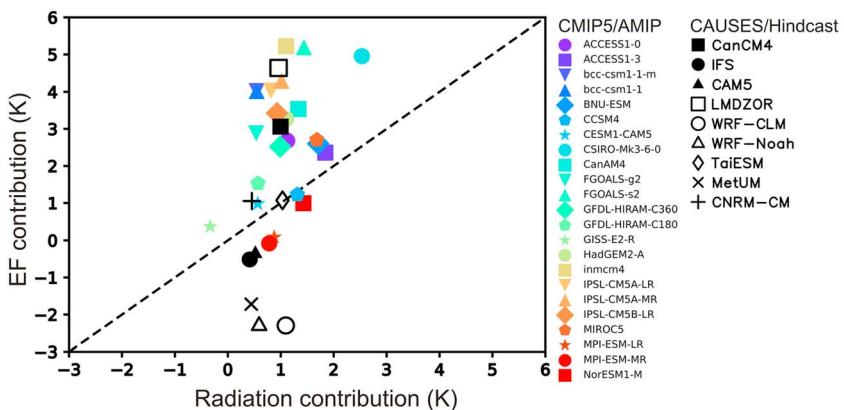
To further identify the connection of the surface energy budget errors to  $T_{2m}$  error, we performed correlation analyses both in space over the CONUS domain and in time over the ARM SGP with the goal to distinguish among each of the energy budget terms the main contributors to the  $T_{2m}$  error. Over the Central United States, we find that regions with positive seasonal  $T_{2m}$  mean bias are generally positively correlated with biases of SWNET, surface downward shortwave flux, and SH, but negatively correlated with biases of LH and EF for most models indicating that warm  $T_{2m}$  biases over this region shares similar biases in the surface energy budget. Over the SGP, the daily variation of  $T_{2m}$  bias from individual model is positively correlated with the daily variations of SWNET and downward shortwave flux biases during the daytime and is also positively correlated with downward longwave flux bias during the nighttime in both AM and JJA, while the daily variation of  $T_{2m}$  bias is negatively correlated with EF bias during the daytime. By examining the seasonal mean biases of  $T_{2m}$  and other energy budget terms across all the models over the SGP, we further find that models with larger positive SH or negative LH and EF bias magnitudes are likely to simulate larger  $T_{2m}$  mean bias on seasonal time scales.

Based upon the surface energy balance and the physical mechanisms discussed in this paper, we further derived an approximate equation to quantify the magnitudes of radiation and evaporative fraction contributions to the  $T_{2m}$  bias. Our analysis confirms the dominant role of EF errors; namely, a large underestimate of EF is the dominant reason for a large positive temperature bias, whereas an overestimate of EF may compensate for the excess absorbed solar radiation in the models with the smallest temperature bias. This theoretical analysis is consistent with the fact that  $T_{2m}$  biases are better correlated to EF errors than to radiation errors in the hindcast models examined in this paper.

While our primary focus here is on the surface energy budget analysis, we do recognize the importance of the surface water budget (e.g., precipitation or soil moisture) that could impact the simulations of surface fluxes as



**Figure 15.** Accumulated daily (a) precipitation (mm), (b) evaporation (mm), (c) precipitation – evaporation (mm), and daily (d) 10 cm soil moisture ( $\text{kg m}^{-2}$ ), (e) evaporative fraction bias (unitless) over the ARM SGP domain (averaged over  $35\text{--}38^\circ\text{N}$ ,  $99\text{--}96^\circ\text{W}$ ). Also in (f) is the scatterplot of surface evaporative fraction bias (abscissa) versus 10 cm soil moisture bias (ordinate). The critical value for correlation coefficient with nine models is  $\pm 0.67$  with probability  $p = 0.05$ . Model values are averaged over Day 2 to 5 (24 to 120 h) hindcast lead time.



**Figure 16.** Scatterplot of the predicted June–August mean error in surface air temperature from radiation input errors (equation (A8)) (abscissa) versus the predicted June–August mean error in surface air temperature error from evaporative fraction (EF) errors (ordinate) from 9 CAUSES hindcast models and 23 CMIP5/AMIP models (data source is from Zhang et al., 2018). The radiation and evaporative fraction errors are calculated from equation (1). See Tables 2 and 3 for a summary of the values.

**Table 2**

June–August 2011  $T_{2m}$  Mean Bias (K), As Well As Predicted  $T_{2m}$  Bias (K) and Bias Contributions From Radiation (K) and Evaporative Fraction (K) Derived From Equation (1) at the ARM SGP

Model	$T_{2m}$ bias	Predicted $T_{2m}$ bias	Predicted $T_{2m}$ bias contributed from radiation	Predicted $T_{2m}$ bias contributed from EF
CanCM4	6.21	4.06	1.00	3.06
IFS	2.90	-0.10	0.41	-0.51
CAM5	2.25	0.22	0.52	-0.30
LMDZOR	3.64	5.59	0.96	4.63
WRF-CLM	1.14	-1.22	1.10	-2.32
WRF-Noah	0.76	-1.67	0.59	-2.26
TaiESM	3.67	2.10	1.03	1.07
MetUM	1.24	-1.28	0.44	-1.72
CNRM-CM	2.30	1.51	0.46	1.05

we demonstrated in Section 3.4. Also, the biases in the  $T_{2m}$  can feed back to the simulations of surface energy and water budgets. The precipitation process is undoubtedly one of the major integrating contributors to  $T_{2m}$  bias through the interactions with the cloud and radiation processes, as well as interactions with soil moisture and surface evaporation (e.g., Cheruy et al., 2014; Klein et al., 2006). Yet one of the outstanding issues that we did not address in the present paper is the relative contribution of precipitation errors to the temperature errors. We demonstrated the relative importance of EF. However, the EF can be affected not only by precipitation but also by other land processes. Another related issue is which type of precipitating convection systems dominate the errors in the surface precipitation. As indicated in previous studies (e.g., Jiang et al., 2006), nearly half of the total summer mean precipitation over the U.S. Great Plains are those convective systems that are initiated from the lee of the Rockies and propagate toward the Great Plains.

**Table 3**

CMIP5/AMIP June–August  $T_{2m}$  Mean Bias (K), As Well As Predicted  $T_{2m}$  Bias (K) and Bias Contributions From Radiation (K) and Evaporative Fraction (K) Derived From Equation (1) at the ARM SGP

Model	$T_{2m}$ bias	Predicted $T_{2m}$ bias	Predicted $T_{2m}$ bias contributed from radiation	Predicted $T_{2m}$ bias contributed from EF
ACCESS1-0	4.68	3.80	1.12	2.68
ACCESS1-3	5.46	4.20	1.84	2.36
bcc-csm1-1-m	5.63	4.57	0.55	4.02
bcc-csm1-1	5.63	4.57	0.55	4.02
BNU-ESM	3.73	4.31	1.71	2.60
CCSM4	4.54	2.54	1.31	1.23
CESM1-CAM5	3.44	1.55	0.57	0.98
CSIRO-Mk3-6-0	10.41	7.49	2.54	4.95
CanAM4	7.44	4.87	1.33	3.54
FGOALS-g2	3.70	3.41	0.53	2.88
FGOALS-s2	8.26	6.63	1.43	5.20
GFDL-HIRAM-C360	5.10	3.51	0.99	2.52
GFDL-HIRAM-C180	3.58	2.10	0.57	1.53
GISS-E2-R	0.72	0.04	-0.33	0.37
HadGEM2-A	5.17	4.37	1.10	3.27
inmcm4	4.27	6.32	1.10	5.22
IPSL-CM5A-LR	6.44	4.85	0.82	4.03
IPSL-CM5A-MR	5.99	5.31	1.01	4.30
IPSL-CM5B-LR	3.44	4.35	0.93	3.42
MIROC5	7.20	4.38	1.68	2.70
MPI-ESM-LR	3.99	0.96	0.87	0.09
MPI-ESM-MR	3.74	0.70	0.78	-0.08
NorESM1-M	4.78	2.42	1.42	1.00

Note. The reference field is Atmospheric Radiation Measurement continuous forcing data set from 1999 to 2011. See Zhang et al. (2018) for more details.

Most contemporary climate models are still unable to capture such propagating systems because of the coarse horizontal resolutions. Therefore, quantification of the precipitation error from the propagating systems will further help quantify the relative importance of precipitation bias to  $T_{2m}$  bias.

Lastly, from the analysis between WRF-CLM and WRF-Noah, it is clear that land processes and land-atmosphere interactions can play an important role on the  $T_{2m}$  simulations through affecting soil moisture and surface turbulent fluxes. One of the goals for CAUSES is to better identify possible culprits for the  $T_{2m}$  bias and to provide possible solution and guidance for model development. As Van Weverberg et al. (2018) demonstrates the importance of model cloud parameterizations on radiation bias, it is also important to highlight possible pathways for  $T_{2m}$  bias improvement for the land model side. For example, Qian et al. (2013) show a better simulations of soil moisture and EF by implementing an irrigation scheme in the land model. Williams et al. (2016) also show some success in correcting EF and precipitation biases in the single-column model simulations by modifying leaf area index and bare soil resistance in the land model at the ARM SGP. Therefore, parameterizations in the land models that could affect surface EF and soil moisture simulations should also be further investigated and verified with available measurements to better constrain and improve their representations in order to achieve better  $T_{2m}$  simulations.

## Appendix A: Theoretical Interpretation of the Warm Bias

We now derive an approximate equation to quantify the magnitudes of radiation and evaporative fraction (EF) contributions to the warm bias. This equation is based upon the surface energy balance and expresses the physical mechanisms discussed in sections 3 and 4. We start with the time-averaged surface energy balance (the sign convention is defined near the start of section 3):

$$\text{SWNET} = \text{LWNET} + \text{SH} + \text{LH}, \quad (\text{A1})$$

where SWNET and LWNET are the surface absorbed shortwave and longwave radiation and SH and LH are the sensible and latent heat fluxes. Equation (A1) expresses the balance between the surface's gain of energy through radiation and its loss through the turbulent heat fluxes. With the long-term ( $>1$  week) average, we neglect the storage and ground-heat flux terms which are found to be small. For example, in both hindcast and AMIP models (Zhang et al., 2018), the radiative input exceeds the heat flux sum by  $10 \text{ W m}^{-2}$  on average for June–August at the SGP. For the observations, the imbalance is around  $-10 \text{ W m}^{-2}$  for 2011 and  $+5 \text{ W m}^{-2}$  for the multiyear climatology covering the period 1999–2011. These imbalances are typically smaller than the magnitude of the model errors in the shortwave radiation or heat fluxes.

In order to isolate the role of EF, we use the definition of  $\text{EF} = \text{LH}/(\text{SH} + \text{LH})$  to eliminate LH from (A1):

$$\text{SWNET} = \text{LWNET} + \gamma \text{SH}, \quad (\text{A2})$$

where we have defined  $\gamma = 1/(1 - \text{EF})$ . If (A2) holds for both models and the observations, the model errors would also satisfy (A2)

$$\text{SWNET}' = \text{LWNET}' + (\gamma \text{SH})', \quad (\text{A3})$$

where  $(\cdot)' = \text{O}_{\text{mod}} - \text{O}_{\text{obs}}$ , is the error term defined as model minus observation value. We now also separate the upward longwave (LWUP) error from the downward longwave (LWDN) error and after expansion of the product and rearrangement we arrive with

$$(\text{SWNET}' + \text{LWDN}') - \text{SH}_{\text{mod}}\gamma' = \text{LWUP}' + \gamma_{\text{obs}} \text{SH}', \quad (\text{A4})$$

Equation (A4) expresses the physical balance between on the one hand errors in radiative input (first term on left-hand side) and EF (second term on left-hand side) (represented by  $\gamma'$ ) which drive errors in surface temperature and on the other hand the compensation by the errors in LWUP and SH, which we now take as being proportional to errors in  $T_{2m}$ :

$$(\text{SWNET}' + \text{LWDN}') - \text{SH}_{\text{mod}}\gamma' = \left( \frac{\partial \text{LWUP}}{\partial T_{2m}} + \gamma_{\text{obs}} \frac{\partial \text{SH}}{\partial T_{2m}} \right) T_{2m}', \quad (\text{A5})$$

With straightforward algebraic manipulation, we can rewrite errors in  $\gamma$  in terms of errors in  $\text{EF}'$ :

$$(\text{SWNET}' + \text{LWDN}') - \text{EF}'\gamma_{\text{obs}}(\text{SH}_{\text{mod}} + \text{LH}_{\text{mod}}) = \left( \frac{\partial \text{LWUP}}{\partial T_{2m}} + \gamma_{\text{obs}} \frac{\partial \text{SH}}{\partial T_{2m}} \right) T_{2m}', \quad (\text{A6})$$

Rearrangement of (A6) yields our rough approximate equation for  $T'_{2m}$  errors:

$$T'_{2m} = \frac{(SWNET' + LWDN')}{\frac{\partial LWUP}{\partial T_{2m}} + \gamma_{obs} \frac{\partial SH}{\partial T_{2m}}} - \frac{EF' \gamma_{obs} (SH_{mod} + LH_{mod})}{\frac{\partial LWUP}{\partial T_{2m}} + \gamma_{obs} \frac{\partial SH}{\partial T_{2m}}}, \quad (A7)$$

Equation (A7) quantifies two separate components of  $T'_{2m}$  arising from separate errors in the radiative input and EF. To test whether equation (A7) accurately relates errors in radiative input and EF to  $T'_{2m}$ , we must supply values of  $\gamma_{obs}$  and the derivatives in the denominator.

Errors in LWUP are theoretically expected to be given by  $LWUP' = 4\varepsilon\sigma T_s^3 T'_s$ , where  $\varepsilon$  is the surface longwave emissivity,  $\sigma$  is the Stefan-Boltzmann constant, and  $T'_s$  is the surface skin temperature error. With a definition of  $\kappa$  as the ratio of  $T'_{2m}$  to  $T'_s$ , one has  $LWUP' = 4\varepsilon\sigma T_s^3 T'_{2m}/\kappa$ . We do realize that model error in surface emissivity are neglected in this study, and this could be a factor for biases in the longwave fluxes. In practice, we find that the theoretical expectation matches the model data extremely well. For the nine hindcast models in the present study, the intermodel correlation coefficient between  $T'_{2m}$  and LWUP' is 0.94 and the linear regression slope is  $7.3 \text{ W m}^{-2} \text{ K}^{-1}$ . For the 23 AMIP models, the correlation coefficient and linear regression slope are 0.93 and  $7.1 \text{ W m}^{-2} \text{ K}^{-1}$ . This is not far off for the theoretical expectation; given a  $T_s = 300 \text{ K}$ ,  $4\varepsilon\sigma T_s^3$  is  $5.8 \text{ W m}^{-2} \text{ K}^{-1}$ , implying  $\kappa \sim 0.8$ . A value of  $\kappa < 1$  is reasonable given that  $T'_{2m}$  should be smaller than  $T'_s$ . As an aside, we note that while one would expect errors in LWDN to be correlated with  $T'_{2m}$  through the longwave emission by air in the boundary layer, we find that LWDN errors have weak and statistically insignificant intermodel correlations with  $T'_{2m}$ ; the correlation coefficients between LWDN' and  $T'_{2m}$  are 0.13 and 0.26 for the hindcast and AMIP model ensembles, respectively. LWDN errors do not vary systematically with temperature errors because of a partial compensation between the extra LWDN from an overly warm boundary layer with the reduced LWDN from a lack of clouds and water vapor under a biased warm state (Zhang et al. (2018)). For this reason, we neglect this dependence in this derivation and treat LWDN as a (typically small) error in radiative input to the surface.

Errors in SH are theoretically expected to be given by  $\rho c_p C_{dh} V (T'_s - T'_{2m}) = \rho c_p C_{dh} V (1 - \kappa) T'_{2m}/\kappa$ , where  $\rho$  is the air density,  $c_p$  is the specific heat of air at constant pressure,  $C_{dh}$  is the heat exchange coefficient, and  $V$  is the surface wind speed. Empirically, we find that the relationship between SH' and  $T'_{2m}$  to not be as good as those for LWUP; for the hindcast models, the correlation coefficient is +0.76 and the regression slope is  $11.0 \text{ W m}^{-2} \text{ K}^{-1}$ , whereas for the AMIP models the correlation coefficient is +0.65 and the regression slope is  $6.0 \text{ W m}^{-2} \text{ K}^{-1}$ . The greater error in this relationship and the more widely divergent regression slopes may reflect the neglect of model errors in the heat exchange coefficient and surface wind speed, to list just a few possibilities. With characteristic values of  $\rho = 1 \text{ kg m}^{-3}$ ,  $c_p = 1,004 \text{ J kg}^{-1} \text{ K}^{-1}$ ,  $C_{dh} = 3 \times 10^{-3}$ , and  $V = 7 \text{ m s}^{-1}$ , the values of  $\kappa$  implied by the linear regression slopes are 0.7 for the hindcast models and 0.8 for the AMIP models; this is in agreement with the value of  $\kappa$  implied by the errors in LWUP.

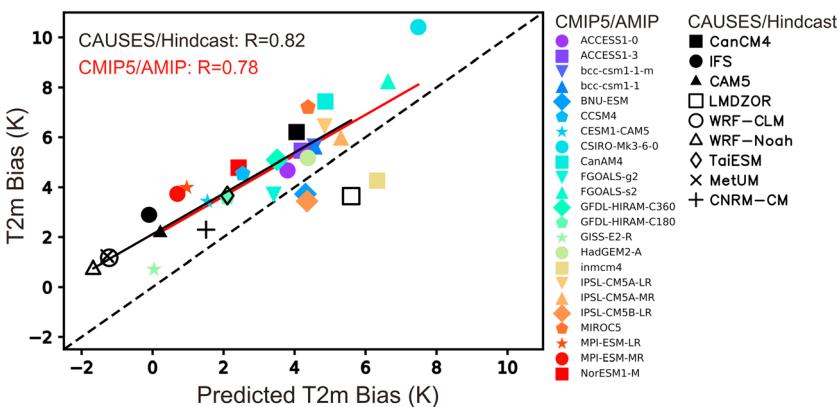
Recognizing that this derivation is only approximate, we now take characteristic values for the terms in (A7). Specifically, we assign values of  $7 \text{ W m}^{-2} \text{ K}^{-1}$  and  $9 \text{ W m}^{-2} \text{ K}^{-1}$  to  $\frac{\partial LWUP}{\partial T_{2m}}$  and  $\frac{\partial SH}{\partial T_{2m}}$ , respectively. We assign  $\gamma_{obs} = 2.5$ , using an  $EF_{obs} = 0.6$  derived considering that the June-July-August average  $EF_{obs}$  was 0.55 for 2011 (the year of the hindcasts) and 0.66 for the 1999–2011 climatology. With these values, we have our approximate relationship between errors in surface air temperature and errors in radiative input and evaporative fraction:

$$T'_{2m} = \frac{(SWNET' + LWDN')}{30 \text{ W m}^{-2} \text{ K}^{-1}} - \frac{2.5 EF' (SH_{mod} + LH_{mod})}{30 \text{ W m}^{-2} \text{ K}^{-1}}. \quad (A8)$$

Equation (A8) makes a prediction of a model's  $T'_{2m}$  from given its errors in radiative input and evaporative fraction as well as the model's value for the sum of the sensible and latent fluxes. We now test the value of (A8) through a scatterplot of this predicted  $T'_{2m}$  against the actual  $T'_{2m}$  for the AMIP (data source from Zhang et al., 2018) and hindcast models (Figures A1). Considering all the approximations made in the derivation of (A8), the predicted  $T'_{2m}$  agrees reasonably well the actual  $T'_{2m}$ . The correlation coefficients of +0.78 across the 23 AMIP models and +0.82 across the nine hindcast models are equal to or larger in magnitude than any of the individual correlations with radiative input error or EF errors. This suggests that (A8) captures some of

## Acknowledgments

We thank three anonymous reviewers for their helpful comments and we also thank Yunyan Zhang and Thomas Phillips for discussion of this work. The CAUSES project is endorsed and sponsored by the Global Energy and Water cycle Exchanges (GEWEX)/Global Atmospheric System Studies (GASS) and the U.S. Department of Energy's Regional and Global Climate Modeling (RGCM) and Atmospheric System Research (ASR) Programs. Data have been used from the DOE Atmospheric Radiation Measurement (ARM) Climate Research User Facility (<https://www.arm.gov/data>), and we are grateful to the ECMWF for making their ERA-Interim reanalysis available (<https://www.ecmwf.int/en/research/climate-reanalysis/era-interim>). Computing resources and CAUSES simulations archive were provided from the Livermore Computing Center at LLNL and the National Energy Research Scientific Computing Center (NERSC), contract DE-AC02-05CH11231. H. M., S. K., S. X., C. Z., S. T., and Q. T. were funded by the RGCM, ASR, and ARM Programs of the U.S. DOE, and their work was performed under the auspices of the U.S. DOE by LLNL under contract DE-AC52-07NA27344. C. M. and K. V. W. at UKMO received support from ASR via grant DE-SC0014122 and M. A. at ECMWF via DE-SC0005259. L. K., W. G., M. H., Y. L., and Y. Q. were supported by the U.S. DOE ASR program and Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the U.S. DOE under contract DE-AC05-76RL01830, and a portion of the computing was done using PNLL Institutional Computing. The IDRIS computational facilities (Institut du Développement et des Ressources en Informatique Scientifique, CNRS, France) were used to perform all the LMDZOR/CAUSES computation (grant DARI-0292). R. R. acknowledges the support from the DEPHY2 project, funded by the French national program LEFE/INSU. The work of TaiESM development is supported by Ministry of Science and Technology, Taiwan, R.O.C. under grants NSC100-2119-M-001-029-MY5 and MOST105-2119-M-001-018. The simulations will be available online through the NERSC Science Gateways (details provided on <https://www.nersc.gov/users/sciencegateways/>). Due to the large volume of data sets and limited disk space, data will be shared online upon request.



**Figure A1.** Scatterplot of the predicted June–August mean error in surface air temperature from equation (A8) (abscissa) versus the actual surface air temperature error (ordinate) from 9 CAUSES hindcast models and 23 CMIP5/AMIP models. See Tables 2 and 3 for a summary of the values. The critical values for correlation coefficients with 9 and 23 models are  $\pm 0.67$  and  $\pm 0.41$ , respectively, with probability  $p = 0.05$ .

the separate influence of radiation and evaporative fraction errors on  $T'_{2m}$ . However (A8) is not an unbiased prediction of  $T'_{2m}$ ; for example, the data and regression line in Figure A1 indicate that models have a positive surface air temperature bias of +2 K when (A8) predicts zero bias. The warm bias in the ERA-Interim initial conditions (Morcrette et al., 2018) may partially explain why hindcast models have a greater temperature bias than that predicted by (A8) but cannot explain why AMIP models lie above the 1:1 line.

## References

- 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. *Meteorological Monographs*, 57, 23.1–23.17. <https://doi.org/10.1175/AMSMONOGRAPH-D-15-0044.1>
- Bond, D. (2005). *Soil Water and Temperature System (SWATS) Handbook* (ARM Technical Report TR-063). Washington, DC: U.S. Department of Energy.
- Brown, A., Milton, S., Cullen, M., Golding, B., Mitchell, J., & Shelly, A. (2012). Unified modeling and prediction of weather and climate. *Bulletin of the American Meteorological Society*, 93(12), 1865–1877. <https://doi.org/10.1175/BAMS-D-12-00018.1>
- Carbone, R. E., Tuttle, J. D., Ahijevych, D. A., & Trier, S. B. (2002). Inferences of predictability associated with warm season precipitation episodes. *Journal of the Atmospheric Sciences*, 59(13), 2033–2056. [https://doi.org/10.1175/1520-0469\(2002\)059%3C2033:IOPAWW%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059%3C2033:IOPAWW%3E2.0.CO;2)
- 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>
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597. <https://doi.org/10.1002/qj.828>
- Doelling, D. R., Loeb, N. G., Keyes, D. F., Nordean, M. L., Morstad, D., Nguyen, C., et al. (2013). Geostationary enhanced temporal interpolation for CERES flux products. *Journal of Atmospheric and Oceanic Technology*, 30(6), 1072–1090. <https://doi.org/10.1175/JTECH-D-12-00136.1>
- Hourdin, F., Grandpeix, J.-Y., Rio, C., Bony, S., Jam, A., Cheruy, F., et al. (2012). LMDZ5B: The atmospheric component of the IPSL climate model with revised parameterizations for clouds and convection. *Climate Dynamics*, 40(9–10), 2193–2222. <https://doi.org/10.1007/s00382-012-1343-y>
- 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 (MCCE). *Bulletin of the American Meteorological Society*, 97(9), 1667–1686. <https://doi.org/10.1175/BAMS-D-14-00228.1>
- Jiang, X., Lau, N.-C., & Klein, S. A. (2006). Role of eastward propagating convection systems in the diurnal cycle and seasonal mean of summertime rainfall over the U.S. Great Plains. *Geophysical Research Letters*, 33, L19809. <https://doi.org/10.1029/2006GL027022>
- Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., et al. (2011). Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research*, 116, G00J07. <https://doi.org/10.1029/2010JG001566>
- Kato, S., Loeb, N. G., Fred, 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>
- Lee, W.-L., Liou, K. N., & Wang, C.-C. (2013). Impact of 3-D topography on surface radiation budget over the Tibetan plateau. *Theoretical and Applied Climatology*, 113(1–2), 95–103. <https://doi.org/10.1007/s00704-012-0767-y>

- 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(4), 1810–1827. <https://doi.org/10.1002/2015MS000490>
- Ma, H.-Y., Xie, S., Boyle, J. S., Klein, S. A., & Zhang, Y. (2013). Metrics and diagnostics for precipitation-related processes in climate model short-range hindcasts. *Journal of Climate*, 26(5), 1516–1534. <https://doi.org/10.1175/JCLI-D-12-00235.1>
- 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>
- Ma, P.-L., Rasch, P. J., Fast, J. D., Easter, R. C., Gustafson, W. I. Jr., Liu, X., et al. (2014). Assessing the CAM5 physics suite in the WRF-Chem model: Implementation, resolution sensitivity, and a first evaluation for a regional case study. *Geoscientific Model Development*, 7(3), 755–778. <https://doi.org/10.5194/gmd-7-755-2014>
- Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., et al. (2013). The surfexv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. *Geoscientific Model Development*, 6, 929–960. <https://doi.org/10.5194/gmd-6-929-2013>
- Merryfield, W. J., Lee, W.-S., Boer, G. J., Kharin, V. V., Scinocca, J. F., Flato, G. M., et al. (2013). The Canadian Seasonal to Interannual Prediction System. Part I: Models and initialization. *Monthly Weather Review*, 141(8), 2910–2945. <https://doi.org/10.1175/MWR-D-12-00216.1>
- 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>
- Neale, R., J. H. Richter, A. J. Conley, S. Park, P. H. Lauritzen, A. Gettelman, D. L. Williamson, S.-J. Lin (2012). Description of the NCAR Community Atmosphere Model (CAM5.0) (NCAR Tech. Note NCAR-TN-486+STR, 274 pp.).
- 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>
- 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>
- 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>
- Qian, Y., Long, C. N., Wang, H., Comstock, J. M., McFarlane, S. A., & Xie, S. (2012). Evaluation of cloud fraction and its radiative effect simulated by IPCC AR4 global models against ARM surface observations. *Atmospheric Chemistry and Physics*, 12(4), 1785–1810. <https://doi.org/10.5194/acp-12-1785-2012>
- Santanello, J. A. Jr., Roundy, J., & Dirmeyer, P. A. (2015). Quantifying the land-atmosphere coupling behavior in modern reanalysis products over the U.S. Southern Great Plains. *Journal of Climate*, 28(14), 5813–5829. <https://doi.org/10.1175/JCLI-D-14-00680.1>
- Schneider, J. M., Fisher, D. K., Elliott, R. L., Brown, G. O., & Bahrmann, C. P. (2003). Spatiotemporal variation in soil water: First results from the ARM SGP CART network. *Journal of Hydrometeorology*, 4(1), 106–120. [https://doi.org/10.1175/1525-7541\(2003\)004%3C0106:SVISWF%3E2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004%3C0106:SVISWF%3E2.0.CO;2)
- Tang, Q., & Xie, S. (2015). Station-based surface data value-added product, DOE/SC-ARM/TR-151, ARM Climate Research Facility. <https://doi.org/10.5439/1178332>
- 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. <https://doi.org/10.1175/BAMS-D-11-00094.1>
- 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>
- 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: Atmospheres*, 123. <https://doi.org/10.1002/2017JD027188>
- Voldoire, A., Sanchez-Gomez, E., Salas y Melia, D., Decharme, B., Cassou, C., Sénési, S., et al. (2013). The CNRM-CM5.1 global climate model: Description and basic evaluation. *Climate Dynamics*, 40, 2091–2121. <https://doi.org/10.1007/s00382-011-1259-y>
- Walters, D. N., Williams, K. D., Boulton, I. A., Bushell, A. C., Edwards, J. M., Field, P. R., et al. (2014). The Met Office Unified Model Global Atmosphere 4.0 and JULES Global Land 4.0 configurations. *Geoscientific Model Development*, 7(1), 361–386. <https://doi.org/10.5194/gmd-7-361-2014>
- Wang, Y.-C., Pan, H.-L., & Hsu, H.-H. (2015). Impacts of the triggering function of cumulus parameterization on warm-season diurnal rainfall cycles at the Atmospheric Radiation Measurement Southern Great Plains site. *Journal of Geophysical Research: Atmospheres*, 120, 10,681–10,702. <https://doi.org/10.1002/2015JD023337>
- 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 U.S. Southern Great Plains. *Journal of Geophysical Research: Atmospheres*, 121, 12,125–12,144. <https://doi.org/10.1002/2016JD025223>
- Williams, K. D., Bodas-Salcedo, A., Déqué, M., Fermepin, S., Medeiros, B., Watanabe, M., et al. (2013). The transpose-AMIP II experiment and its application to the understanding of Southern Ocean cloud biases in climate models. *Journal of Climate*, 26(10), 3258–3274. <https://doi.org/10.1175/JCLI-D-12-00429.1>
- Xie, S., Cederwall, R. T., & Zhang, M. H. (2004). Developing long-term single-column model/cloud system-resolving model forcing 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>
- Xie, S., Ma, H.-Y., Boyle, J. S., Klein, S. A., & Zhang, Y. (2012). On the correspondence between short- and long-timescale systematic errors in CAM4/CAM5 for the years 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., Clothiaux, E. E., & Gaustad, K. L. (2010). ARM climate modeling best estimate data. *Bulletin of the American Meteorological Society*, 91, 13–20. <https://doi.org/10.1175/2009BAMS2891>
- Zhang, C., Xie, S., Klein, S. A., Ma, H.-Y., Tang, S., Van Weverberg, K., et al. (2018). CAUSES: Diagnosis of the Summertime Warm Bias in CMIP5 Climate Models at the ARM Southern Great Plains Site. *Journal of Geophysical Research: Atmospheres*, 123. <https://doi.org/10.1002/2017JD027200>