



RESEARCH ARTICLE

10.1002/2017JD027188

Key Points:

- Surface temperature biases in SGP are associated with large shortwave and longwave radiation biases in NWP and climate models
- Convective clouds contribute most to the radiation biases, although most models experience contributions from the surface albedo as well
- Models rain too much and too often in the afternoon, yet have convective clouds that are too transparent to SW radiation

Correspondence to:

K. Van Weverberg,
kwinten.vanweverberg@metoffice.gov.uk

Citation:

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>

Received 25 MAY 2017

Accepted 8 OCT 2017

Accepted article online 28 FEB 2018

CAUSES: Attribution of Surface Radiation Biases in NWP and Climate Models near the U.S. Southern Great Plains

K. Van Weverberg¹, C. J. Morcrette¹ , J. Petch¹ , S. A. Klein² , H.-Y. Ma² , C. Zhang² , S. Xie² , Q. Tang² , W. I. Gustafson Jr³ , Y. Qian³ , L. K. Berg³ , Y. Liu³ , M. Huang³ , M. Ahlgrimm⁴ , R. Forbes⁴ , E. Bazile⁵ , R. Roehrig⁵ , J. Cole⁶ , W. Merryfield⁶ , W.-S. Lee⁶ , F. Cheruy⁷ , L. Mellul⁷, Y.-C. Wang⁸ , K. Johnson⁹ , and M. M. Thieman¹⁰

¹Met Office, Exeter, UK, ²Lawrence Livermore National Laboratory, Livermore, CA, USA, ³Pacific Northwest National Laboratory, Richland, WA, USA, ⁴European Centre for Medium-Range Weather Forecasts, Reading, UK, ⁵CNRM, Meteo-France/CNRS, Toulouse, France, ⁶Environment and Climate Change Canada, Victoria, British Columbia, Canada,

⁷Laboratoire de Meteorologie Dynamique, Paris, France, ⁸Academia Sinica, Taipei, Taiwan, ⁹Brookhaven National Laboratory, Upton, NY, USA, ¹⁰Science Systems and Applications, Inc, Norfolk, VA, USA

Abstract Many Numerical Weather Prediction (NWP) and climate models exhibit too warm lower tropospheres near the midlatitude continents. The warm bias has been shown to coincide with important surface radiation biases that likely play a critical role in the inception or the growth of the warm bias. This paper presents an attribution study on the net radiation biases in nine model simulations, performed in the framework of the CAUSES project (Clouds Above the United States and Errors at the Surface). Contributions from deficiencies in the surface properties, clouds, water vapor, and aerosols are quantified, using an array of radiation measurement stations near the Atmospheric Radiation Measurement Southern Great Plains site. Furthermore, an in-depth analysis is shown to attribute the radiation errors to specific cloud regimes. The net surface shortwave radiation is overestimated in all models throughout most of the simulation period. Cloud errors are shown to contribute most to this overestimation, although nonnegligible contributions from the surface albedo exist in most models. Missing deep cloud events and/or simulating deep clouds with too weak cloud radiative effects dominate in the cloud-related radiation errors. Some models have compensating errors between excessive occurrence of deep cloud but largely underestimating their radiative effect, while other models miss deep cloud events altogether. Surprisingly, even the latter models tend to produce too much and too frequent afternoon surface precipitation. This suggests that rather than issues with the triggering of deep convection, cloud radiative deficiencies are related to too weak convective cloud detrainment and too large precipitation efficiencies.

1. Introduction

Despite continued improvements in the representation of atmospheric physics and the relentless increase in spatial resolution, many Numerical Weather Prediction (NWP) and climate models still suffer from persistent biases in a number of key variables. One particular bias that is shared between a vast number of NWP and climate models is a too warm lower troposphere near the midlatitude continents in summer (Ahlgrimm & Forbes, 2012; Brisson et al., 2016; Ma et al., 2014; Van Weverberg et al., 2015). Given the nature of this bias and the unequivocal importance of the surface temperature for both weather forecasting and climate applications, there is an urgent need to fully understand its origin.

Research so far has brought forward three hypotheses about why the midlatitude continents, and in particular the Southern Great Plains (SGP) in the American Midwest, tend to be too warm in many atmospheric models. Klein et al. (2006) postulated that mesoscale convective systems, initiated in the lee of the Rocky Mountains, are underrepresented in many of the global circulation models (GCMs) that they investigated. A lack of convective clouds directly leads to excessive downwelling shortwave (SW) radiances and hence enhanced heating of the surface and the boundary layer. Moreover, the lack of convective rainfall could result in soils being too dry and hence overestimated Bowen ratios, further exacerbating the excess heating of the surface. Other studies have hinted at deficiencies in the representation of shallow cumulus clouds near the SGP as the main

©2018 Crown copyright. This article is published with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

culprit for excessive surface SW radiation in many GCMs (Cheinet et al., 2005; Qian et al., 2012). In contrast, Ahlgrimm and Forbes (2012) could not confirm such a contribution from shallow clouds near the SGP in their NWP model. A third hypothesis puts the spotlight mainly on deficiencies in the land surface. Van Weverberg et al. (2015) found that in one of two models analyzed, the land surface was the predominant contributor to the temperature bias near the SGP. Koster et al. (2004) identified the SGP area as one of the major hot spots of land-atmosphere coupling, suggesting the local climate to be very sensitive to deficiencies in the land surface parameterization. A deficit in the evapotranspiration in the CMIP5 models over the midlatitude continents explained biases in the surface temperature very well, according to Mueller and Seneviratne (2014). Obviously, all three hypotheses are highly interconnected and might be at play at the same time to various degrees. Indeed, Cheruy et al. (2014) found correlations between both a deficit in the evaporative fraction and the cloud radiative effect, on the one hand, and a midlatitude continental warm bias in the CMIP5 models, on the other hand.

The CAUSES (Clouds Above the United States and Errors at the Surface) project was established in response to the inconclusive understanding of the midlatitude warm bias, under the auspices of GASS (Global Atmospheric System Studies, part of GEWEX—Global Energy and Water Exchanges—under the World Climate Research Programme). Its aim is to quantify how clouds, radiation, and surface processes contribute to the warm bias in a number of atmospheric models, with a particular focus on the SGP area. By means of short 5 day hindcasts, initialized from near-true state analysis fields, the models can be evaluated against observations as the temperature biases develop. Indeed, Xie et al. (2012) and Ma et al. (2014) showed that when running a number of GCMs in NWP mode, within just a few days, biases emerged with the same sign as biases that develop when running these models in climate mode.

Since 1993, the U.S. Department of Energy Atmospheric Radiation Measurement (ARM) Program has been collecting a wealth of observations at its Central Facility in Oklahoma. This supersite is conveniently located in the core of the area where the warm bias manifests itself in many models (Ma et al., 2014; Morcrette et al., 2017) and within fairly close proximity of the area with the most frequent mesoscale convective systems in the Great Plains (Geerts et al., 2017). Comparing the model hindcasts carried out within the CAUSES project against these observations should enhance our understanding of the mechanisms leading to the warm bias.

An overview of the CAUSES project and a description of the warm bias present in the atmospheric models that took part in CAUSES is presented in Morcrette et al. (2017). Nearly all of these 11 models (with grid spacings of 30–300 km) exhibited a warm bias persistent throughout the summer and present throughout most of the boundary layer.

It is clear from earlier research (Ahlgrimm & Forbes, 2012; Qian et al., 2012; Van Weverberg et al., 2015) that many models participating in CAUSES experience a bias in their surface radiation balance near the SGP. While the extent to which this bias feeds back in the surface energy balance and the screen level temperature is still uncertain, the radiation bias is very likely to play a critical role in the inception or growth of the warm bias.

The aim of this paper is to study the nature and origin of radiation biases in the models participating in CAUSES. The focus will be on the ARM Central Facility in the SGP, taking advantage of the high-quality and high-frequency observations at this site. In the framework of CAUSES, a number of global and limited area models have been run in hindcast mode between April and August 2011, a period overlapping with the Midlatitude Continental Convective Clouds (MC3E) experiment (Jensen et al., 2015). This intensive observation period (IOP) was specifically targeted at observing the full life cycle of deep convective storms, which are one of the possible regimes the warm bias is associated with (Klein et al., 2006).

More specifically, the goal of this paper is twofold. First, an attribution study is performed to quantify the origin of the net surface SW and longwave (LW) biases in nine atmospheric models that participated in CAUSES. Using an array of surface radiation measurement stations, it is assessed how much of the net surface radiation bias arises from deficiencies in the surface properties, clouds, column water vapor, or aerosols. Second, a more in-depth analysis is performed to further attribute the net radiation errors to specific cloud regimes, based on observed and simulated vertical profiles of cloudiness near the ARM SGP Central Facility. A unique feature of this analysis is its emphasis on the coevolution in space and time of cloud and radiation errors, rather than an evaluation of these two properties separately. A detailed analysis of the cloud and precipitation properties associated with the regimes that matter most from a radiation point of view is performed to help focus model parameterization efforts to most effectively reduce its radiation errors.

Observations and model data are described in section 2. The methods employed to perform the radiation attribution study and the cloud regime analysis are described in section 3. Section 4 provides an overview of the findings of this study and these are summarized in section 5.

2. Observation and Model Data

2.1. Description of Observations

2.1.1. Simulation Period and Mesoscale Continental Convective Cloud Experiment

Within the CAUSES project, models were run for the spring and summer 2011 (April–August). Typically, the warm bias over the SGP appears during this time of the year (Ma et al., 2014). Moreover, this particular period includes the MC3E field campaign (Jensen et al., 2015). From 22 April to 6 June, routine observations from the ARM program were complemented by data from this IOP. No data specifically collected during MC3E was used in this paper, although the liquid and ice water path retrievals were only available for the MC3E period at the time of writing. Furthermore, the inclusion of MC3E makes it possible for follow-up studies to explore the more MC3E-specific observations as well. It should be mentioned that while April and May 2011 were fairly normal in terms of temperature and precipitation, June, July, and August were exceptionally warm and June and July were exceptionally dry (McManus et al., 2011) compared to the long-term climatology. However, most of the MC3E period was still associated with fairly typical conditions, with deep convection occurring on 8 out of 45 days (Jensen et al., 2015).

2.1.2. ARM Best Estimate Data

Given the fairly coarse resolution of most of the participating models in CAUSES (Table 1), a direct comparison of model data with individual station data is not ideal. Therefore, the domain-averaged data contained in the ARM Best Estimate (ARMBE; Xie et al., 2010) two-dimensional gridded surface data value-added product (Tang & Xie, 2015) have been used in this study. This product merges key surface measurements at the ARM SGP site and interpolates the data to a regular 2-D grid to facilitate data application. It uses top of atmosphere (TOA) radiative fluxes obtained from the Geostationary Operational Environmental Satellite (GOES-11) and surface radiative fluxes obtained from the radiometer measurements at the SGP Central Facility and the surrounding Extended Facilities (RADFLUXANAL). Clear-sky radiative fluxes for these stations were obtained following the empirical fitting techniques described in Long and Ackerman (2000) and Long and Turner (2008). Vertically integrated water vapor (I WV) is retrieved from Microwave Radiometer stations (MWRRET) colocated with the radiometers at the various sites. More information about these data products and their uncertainties is provided in Table 2.

It should be emphasized that all data in the ARMBE are purely from the observations and hence are not influenced by any background model. Data were processed with a 1 h temporal resolution (similar to the radiation time step in most models under consideration) and averaged over a $2 \times 2^\circ$ grid surrounding the ARM SGP Central Facility, using the methods described in Zhang et al. (2001). This area includes measurements from the SGP Central Facility in Lamont, Oklahoma (E13) and three of the extended facilities (E9, E11, and E15). The location of these four sites is shown in Figure 1.

Surface precipitation data were also used from the ARMBE gridded 2-D product, which averages hourly precipitation data from the Oklahoma and Kansas Mesonet and the ARM standard rain gauge data to a $4 \times 4^\circ$ grid with a resolution of $0.25 \times 0.25^\circ$, surrounding the ARM SGP Central Facility.

2.1.3. Central Facility Data

For the cloud regime analysis, vertically distributed information about the cloud position is required. While satellites with passive instruments can provide this information to some extent, they cannot detect low clouds when a high cloud layer is present. This complicates the analysis, in particular, for convective clouds with high cloud tops, one of the main regimes of interest. While this could be addressed by using active remote sensing such as those on CloudSat, the orbit is such that it will only pass over a given location once every few days. Therefore, it was deemed more sensible to make use of the detailed cloud observations at the SGP ARM Central Facility (Table 2). The Active Remote Sensing of Clouds (ARSCL) ARM Value-Added Product (VAP) provides very detailed vertically distributed information of clouds, based on cloud radar, lidar, ceilometer, and radiometer data (Clothiaux et al., 2000).

By averaging the binary ARSCL cloud locations over a time window, “observed” cloud fractions can be obtained that can be compared against the model cloud fractions. To be consistent with the model grid resolution, this averaging time window is derived for each model separately by dividing the horizontal model grid

Table 1*Overview of All Models Included in the Attribution Study*

| Model name | Description of GCM code | Approximate dx (deg) | Vertical levels | Radiation dt (min) | Radiation parameterization | Convection parameterization |
|------------|--------------------------|----------------------|-----------------|--------------------|----------------------------|-----------------------------|
| CAM5 | Neale et al. (2012) | 1.0 | 30 | 60 | Iacono et al. (2008) | Zhang and McFarlane (1995) |
| WRFCLM | Ma et al. (2014) | 0.4 | 34 | 10 | Iacono et al. (2008) | Zhang and McFarlane (1995) |
| WRFNOAH | Ma et al. (2014) | 0.4 | 34 | 10 | Iacono et al. (2008) | Zhang and McFarlane (1995) |
| METUM | Walters et al. (2014) | 0.3 | 70 | 60 | Edwards and Slingo (1996) | Gregory and Rowntree (1990) |
| IFS | ECMWF (2015) | 0.4 | 91 | 60 | Morcrette et al. (2008) | Bechtold et al. (2014) |
| CNRM | Voldoire et al. (2013) | 1.4 | 31 | 180 | Fouquart and Bonnel (1980) | Bougeault (1985) |
| CANCM4 | von Salzen et al. (2013) | 2.8 | 35 | 60 | Li and Barker (2005) | Zhang and McFarlane (1995) |
| LMDZOR | Hourdin et al. (2013) | 2.5 | 39 | 60 | Morcrette (1991) | Emanuel (1993) |
| TAIESM | Neale et al. (2012) | 1.0 | 30 | 60 | Iacono et al. (2008) | Wang et al. (2014) |

Note. Provided are the main model reference, the number of vertical levels, the approximate grid spacing (dx), the radiation time step, and the references for the radiation and convection parameterizations

spacing by the observed wind speed, averaged between 1 and 10 km altitude. Using the actual wind speeds at each level (and hence applying a different time window for each model level) did not change the results much. Observed wind speeds are obtained from the Merged Sounding ARM product (Table 2). This approach has been widely used in model evaluation studies (Bouniol et al., 2010; Illingworth et al., 2007; Morcrette et al., 2012; Van Weverberg et al., 2015; Yang et al., 2006) and should give reliable results for lower-order moments and long time series (Gruetzun et al., 2013). It should be emphasized that this method yields slight differences in the observed cloud fractions that each of the models are compared against. Indeed, the smaller the model grid spacing, the more the inferred cloud fractions should tend to 0 or 1.

Wu et al. (2014) compared six different estimates of cloud fraction near the SGP for 10 years, including ARSCL and GOES-11. They found ARSCL-derived cloud fractions to be typically about 8% larger than GOES-11. Given that it is hard to judge which of the two observations should be considered more reliable, we assume an uncertainty of 8% for all cloud fractions in our analysis, but it should be stressed that it is hard to provide an objective estimate of the real uncertainty in cloudiness.

The goal of the cloud regime analysis in this study is to match cloud regimes with errors in the surface radiation balance. Hence, it is important that the cloud property observations are as consistent as possible with the surface radiation measurements. Therefore, only the radiometer data from the ARM Central Facility (E13) were used for the cloud regime attribution analysis, ignoring radiation measurements from the ARM extended facilities (E9, E11, and E15; Figure 1). For consistency with the cloud information, all radiation data were averaged over the same time window as the observed cloud fraction data. As for the ARMBE data described in the previous section, clear-sky radiative fluxes are obtained following the empirical fitting techniques described in Long and Ackerman (2000) and Long and Turner (2008). All-sky radiation uncertainties are taken from the ARM Solar Infrared Radiation Station Handbook (Stoffel, 2005) and represent inherent instrument calibration uncertainties. These are conservative estimates based on experiences with renewable energy applications. Clear-sky SW and LW irradiance uncertainties are determined by comparing 1 min clear-sky estimates against detected clear-sky data (Long & Ackerman, 2000) and a radiative transfer model (Long & Turner, 2008), respectively.

TOA fluxes were obtained from the Satellite Cloud Observations and Radiative Property retrieval System (SATCORPS; previously referred to as VISST) ARM Value-Added Product, which employs four visible and infrared channels from the GOES-11 geostationary satellite (Minnis & Smith, 1998; Minnis et al., 2011). This product is available at the native satellite resolution ("pixel-level," about 4 km) covering the region from 32°N–42°N and 105°W–91°W, surrounding the SGP ARM site, with a temporal resolution of 15 min. For the period from April to June, version 3 of this product was used, while version 4 was used for July and August. The changes between the two versions should be minor and include slight improvements to the ice cloud properties and cloud masks near twilight. To compare the TOA fluxes in a consistent way with the surface radiation measurements and the ARSCL-derived cloud properties, the VISST broadband albedo and broadband

Table 2
Overview of All Observations From the SGP Central Facility Used in This Study

| Variable | Shorthand | Data source | dt | dz | dx | Absolute uncertainty | Relative uncertainty | Data reference | Uncertainty reference |
|----------------------------------|-------------------|--------------|-------|----------|------|-------------------------|----------------------|------------------------------|-----------------------------|
| Surface downwelling shortwave | SW _D | RADFLUXANAL | 60 s | — | — | 10 W m ⁻² | 6% | Gaustad et al. (2016) | Stoffel (2005) |
| Surface downwelling longwave | LW _D | RADFLUXANAL | 60 s | — | — | 4 W m ⁻² | 2.5% | Gaustad et al. (2016) | Stoffel (2005) |
| Surface upwelling shortwave | SW _U | RADFLUXANAL | 60 s | — | — | 15 W m ⁻² | 6% | Gaustad et al. (2016) | Stoffel (2005) |
| Surface upwelling longwave | LW _U | RADFLUXANAL | 60 s | — | — | 4 W m ⁻² | 2.5% | Gaustad et al. (2016) | Stoffel (2005) |
| Surface down clear-sky shortwave | SWCS _D | RADFLUXANAL | 60 s | — | — | 20 W m ⁻² | 3% | Gaustad et al. (2016) | Long and Ackerman (2000) |
| Surface down clear-sky longwave | LWCS _D | RADFLUXANAL | 60 s | — | — | 4 W m ⁻² | 2.5% | Gaustad et al. (2016) | Long and Turner (2008) |
| 1–10 km wind speed | WSPD | MERGSOUND | 60 s | 20–200 m | — | — | — | Troyan et al. (2016) | — |
| Cloud fraction | CF | ARSL | 4 s | 30 m | — | 0.08 | — | Johnson et al. (2016) | Wu et al. (2014) |
| Cloud fraction | CF | GOES | 900 s | — | 4 km | 0.08 | — | Minnis et al. (2011) | Wu et al. (2014) |
| Integrated water vapor | IWV | MWRRET | 60 s | — | — | 0.6 kg m ⁻² | — | Gaustad and Riihimaki (2016) | Campos and Sisterson (2015) |
| TOA upwelling shortwave | TOA OSR | VISST | 900 s | — | 4 km | 23 W m ⁻² | 9% | Minnis et al. (2011) | Khaiyer et al. (2011) |
| TOA upwelling longwave | TOA OLR | VISST | 900 s | — | 4 km | 9 W m ⁻² | 3% | Minnis et al. (2011) | Khaiyer et al. (2011) |
| Liquid water content | LWC | MICROBASE-KA | 4 s | 30 m | — | — | 15% | Dunn et al. (2011) | Zhao et al. (2014) |
| Ice water content | IWC | MICROBASE-KA | 4 s | 30 m | — | — | 55% | Dunn et al. (2011) | Zhao et al. (2014) |
| Surface precipitation | RAIN | ARM BE | 1 h | — | — | 0.02 mm h ⁻¹ | — | Tang and Xie (2015) | Ritsche (2008) |
| 2 m temperature | T2M | ARM BE | 1 h | — | — | 0.4 K | — | Tang and Xie (2015) | Ritsche (2008) |

Note. Provided are the abbreviations used throughout the text, the ARM retrieval method or value-added product, the temporal sampling (dt), the vertical spacing (dz), horizontal spacing (dx), the relative and absolute uncertainty (as a standard deviation) and the main reference to each data product and its uncertainty. Note that for the radiation measurements, the uncertainty for any particular measurement is the absolute or relative uncertainty listed, whichever is larger.

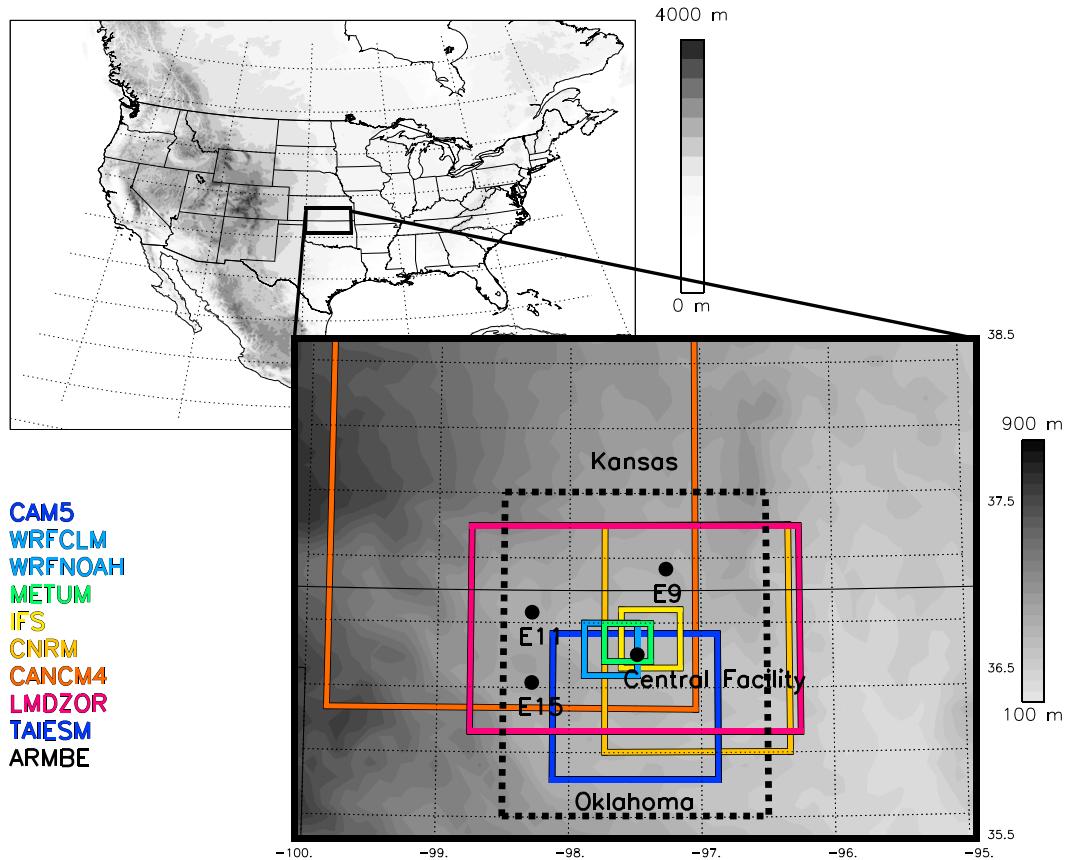


Figure 1. Domain overview of all models analyzed in this study. The footprints of the grid column nearest to the Southern Great Plains site is shown as a colored rectangle in the inset of the figure. The locations of the Central Facility and the three Extended Facilities are denoted by the black dot, and the area of data used to construct the gridded 2-D Atmospheric Radiation Measurement Best Estimate data set is shown by the black dashed rectangle. The orography of North America and the Southern Great Plains is shown by the shading.

outgoing longwave radiation (OLR) from the pixel closest to the SGP site was first selected. Subsequently, the same time window as for the surface radiation and cloud averaging was used to average over all satellite scenes available within that time window, applying a zenith angle adjustment to calculate the grid box average outgoing shortwave radiation (OSR). Spring and summer root-mean-square errors associated with the VISST retrievals are in the order of 5% for the OLR and 9% for the OSR (Khaiyer et al., 2011), based on comparison against CERES Aqua/Terra.

Last, a brief evaluation is performed of the liquid and ice water paths (LWP and IWP, respectively) associated with the cloud regimes. These are calculated from the liquid and ice water contents in the MICROBASE ARM Value-Added Product. This is an updated product from the version described in Dunn et al. (2011) and combines data from the ceilometer, micropulse lidar, microwave radiometer, merged sounding, and the vertically pointing Ka-band cloud radar (KAZR) at SGP. At the time of writing, this product was available with a 4 s temporal and 30 m vertical resolution for the period of the MC3E campaign. Hence, the analysis of the LWP and IWP in the models will be restricted to this period only (18 April to 6 June 2011). Similar to all above observations, the Microbase retrievals were time averaged using the 1–10 km averaged wind speeds from the merged sounding. Zhao et al. (2014) estimated the uncertainties in the retrievals of the liquid and ice water contents (LWC and IWC) to be in the order of 15 and 55%, respectively. The uncertainty in the IWC could be much higher than this value, however, given that uncertainties in the ice crystal habit and ice size distribution assumptions were not considered in these numbers. This needs to be borne in mind when interpreting the results shown here. Furthermore, the uncertainties mentioned in Zhao et al. (2014) are for an older version of this VAP, using the millimeter wavelength cloud radar, rather than the KAZR.

2.2. Description of Atmospheric Models

The models that took part in the CAUSES project vary in terms of their complexity, model resolution, and application. Some models, such as the IFS, are applied mostly in an NWP context, while others, such as the LMDZOR, CNRM, or CAM5 are used predominantly in climate applications. It is worth mentioning that most of the climate models used in this study consist of the atmospheric components of model versions intermediate between the Coupled Model Intercomparison Projects Phases 5 and 6 (CMIP5 and CMIP6; Eyring et al., 2016). Further differences exist in the domains these models are typically integrated over. The WRF models are limited area models (LAMs), while all other models are global models. Nevertheless, they all share a fairly distinct warm bias near the SGP, as outlined in Morcrette et al. (2017) An overview of the models that were analyzed in the context of this paper is provided in Table 1. Note that fewer models are analyzed here compared to Morcrette et al. (2017), since not all modeling centers provided all necessary fields to carry out the radiation and cloud regime analysis.

All models were run in hindcast mode, similar to the approach followed in Ma et al. (2014). ERA-Interim reanalyzes (Dee et al., 2011) were used for the initialization at 00 UTC each day between 1 April and 31 August 2011, and all models were run freely for 5 days. Hence, a comparison can be made against observations at different lead times, shedding light on the growth of the model biases over time. Some key differences between the nine models in this study are highlighted in Table 1. Grid spacings of the models vary from 30 km in the METUM to 240 km in the CANCM4, while the number of vertical levels varies from 30 in the CAM5 and TAIESM to 91 levels in the IFS model. All models under investigation have parameterized deep convection and prognostic cloud microphysics. Cloud cover is diagnosed in most models. The IFS and METUM, however, have prognostic cloud fraction formulations. More details about all the models are provided in Morcrette et al. (2017).

Given the focus on understanding the radiation biases in the models, data were analyzed at the model radiation time steps (when the radiative transfer is calculated). This time step is 1 h for most models under investigation (Table 1). The CNRM model has a longer radiation time step of 3 h and hence was analyzed at this coarser temporal resolution. The WRF models have a radiation time step of 10 min, but for practical reasons were only analyzed at hourly intervals.

While the METUM, IFS, and WRF simulations perform the radiative transfer prior to updating the cloud fields within a time step, all other models first update the cloud fields and subsequently perform the radiative transfer. This calling order matters when trying to trace the origin of biases in the surface radiation balance to the cloud properties in the models. Hence, the surface radiation fields in the IFS, METUM, and WRF models were compared against the cloud fields one physics time step prior to the radiation time step.

The analysis in the remainder of this paper focuses on the grid column closest to the ARM SGP Central Facility (36.60°N , -97.50°W). An overview of all model footprints of the grid column nearest to the SGP site is given in Figure 1. As can be seen on this figure, the location of the SGP site is fairly far from the center of the grid box nearest to it for many models. Hence, observed SW radiation has been adjusted using the cosines of the zenith angles of these two locations for each of the models. This adjustment is generally very small, but matters when comparing, for example, the observed and simulated diurnal cycles of radiation, mainly for the coarser models.

3. Methods

3.1. Radiation Attribution Methodology

The first goal of this paper consists of disentangling the net surface shortwave ($\overline{\Delta SW_{\text{net}}}$) and longwave ($\overline{\Delta LW_{\text{net}}}$) radiation biases into contributions from various processes as follows:

$$\overline{\Delta SW_{\text{net}}} = C_{\alpha} + C_{\text{cloud}} + C_{\text{iwv}} + C_{\text{res}} \quad (1)$$

$$\overline{\Delta LW_{\text{net}}} = C_{\text{surf}} + C_{\text{cloud}} + C_{\text{iwv}} + C_{\text{res}} \quad (2)$$

where C_{α} , C_{surf} , C_{cloud} , C_{iwv} , and C_{res} are contributions from the surface albedo, surface LW emission, cloud, integrated water vapor, and a residual term, respectively. Appendix A provides a detailed derivation of each of these terms, but equations for the most important terms (C_{α} , C_{surf} , and C_{cloud}) are as follows:

$$C_{\alpha} = \overline{(\alpha_M - \alpha_O)(SW_{\text{OD}})} \quad (3)$$

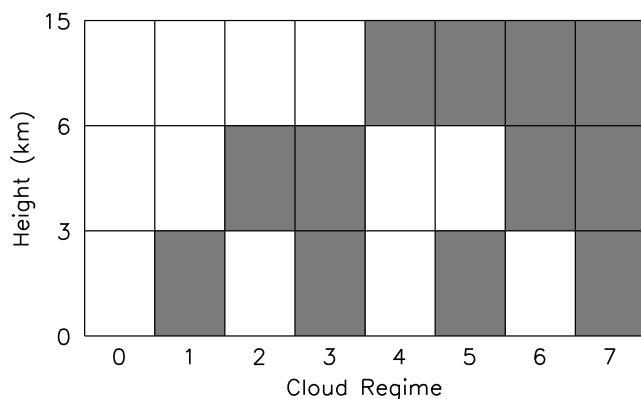


Figure 2. Overview of the cloud regime definitions. Regimes are based on any nonzero cloud fraction in the low (0–3 km), middle (3–6 km), and high levels (6–20 km) of the troposphere. Cloudy regions for each of the regimes are highlighted by gray shading.

where SW_{OD} is the observed downwelling SW radiation and α_M and α_O are the simulated and observed surface albedos, respectively.

$$C_{surf} = \overline{(LW_{MU} - LW_{OU})} \quad (4)$$

where LW_{MU} and LW_{OU} the modeled and observed upwelling LW radiation, respectively.

$$C_{cloud} = \overline{(1 - \alpha_O)((SW_{MD} - SW_{OD}) - (SWCS_{MD} - SWCS_{OD}))} \quad (5)$$

where $SWCS_{MD}$ and $SWCS_{OD}$ are the clear-sky SW downwelling radiation for the model and the observations, respectively.

The detailed derivations for each of these components are provided in Appendix A. The residual term includes contributions from errors in the aerosol optical depth and also numerical inaccuracies and issues with the clear-sky radiative transfer calculations in the models (Pincus et al., 2015).

Given that these contributions are not necessarily normally distributed, a Wilcoxon rank-sum test was applied on every contribution term to test whether the term was significantly different from zero at a 95% confidence interval.

3.2. Cloud Regime Analysis

A second goal of this paper is to further disentangle the contribution from clouds to the net radiation bias (equation (A5)) into contributions from different cloud regimes. Indeed, cloud-related radiation biases could predominantly originate from issues with low clouds, deep clouds, or even high clouds. As guidance to model development, it is instructive to know which of these regimes contribute most to the radiation biases.

Therefore, eight cloud regimes were defined, based on nonzero cloud occurrence in the lower (0–3 km altitude), middle (3–6 km altitude), and upper (6–15 km altitude) troposphere. To number the regimes, the absence of cloud in each layer is assigned a value of 0, while the presence of cloud in the lower, middle, and upper troposphere is given a value of 1, 2, and 4, respectively, similar to Van Weverberg et al. (2015), yielding a total of eight cloud regimes as shown in Figure 2.

It should be stressed that the regime definitions are based on any nonzero (>0.1%) cloud occurrence, and hence, clouds assigned to a certain regime could still span a wide range of cloud fractions. However, this regime classification was deemed the most sensible in terms of compromising between having a manageable number of regimes and being able to identify the most problematic cloud regime.

For each output time, the cloud contribution to the overall radiation bias C_{cloud} (equation (A5)) could be assigned to an observed-simulated cloud regime pair. It is worth highlighting here again that, in contrast to the radiation attribution methodology described in section 3.1, the cloud regime analysis only uses the surface radiation data and vertical cloud profiles from the ARM SGP Central Facility.

4. Results

4.1. Description of SW and LW Radiation Biases

Figures 3a and 3b show the monthly running mean biases of the net SW and LW radiation, respectively, averaged over all hindcast days ($T + 00$ to $T + 120$). The net biases averaged over the entire simulation period are shown in Table 3. The largest net SW biases occur in the WRFCLM, the CANCM4, and the LMDZOR simulations. Smaller biases are present in the IFS, the CNRM, and the CAM5 simulations, albeit even these models experience an important positive bias in the net SW. Net LW biases are particularly negative for the CANCM4 and LMDZOR models, lacking nearly 50 W m^{-2} of net surface LW radiation. Much smaller biases of about -10 to -15 W m^{-2} are present in most other models. Since all models were run in hindcast mode, it is instructive to look at the average growth of the biases over five hindcast days as well (Figures 4a and 4b). From this figure, there does not appear to be a very systematic growth over the 5 days in any of the models, apart from a decrease in the net LW bias in the LMDZOR and CANCM4 near the end of May and July, respectively. None of the growth rates shown in Figures 4a and 4b exceed the observational uncertainty (denoted by the gray shading). Overall, just 1 day into the hindcast, most models tend to have a fairly important radiation bias that persists throughout the remainder of the 5 day hindcast (not shown).

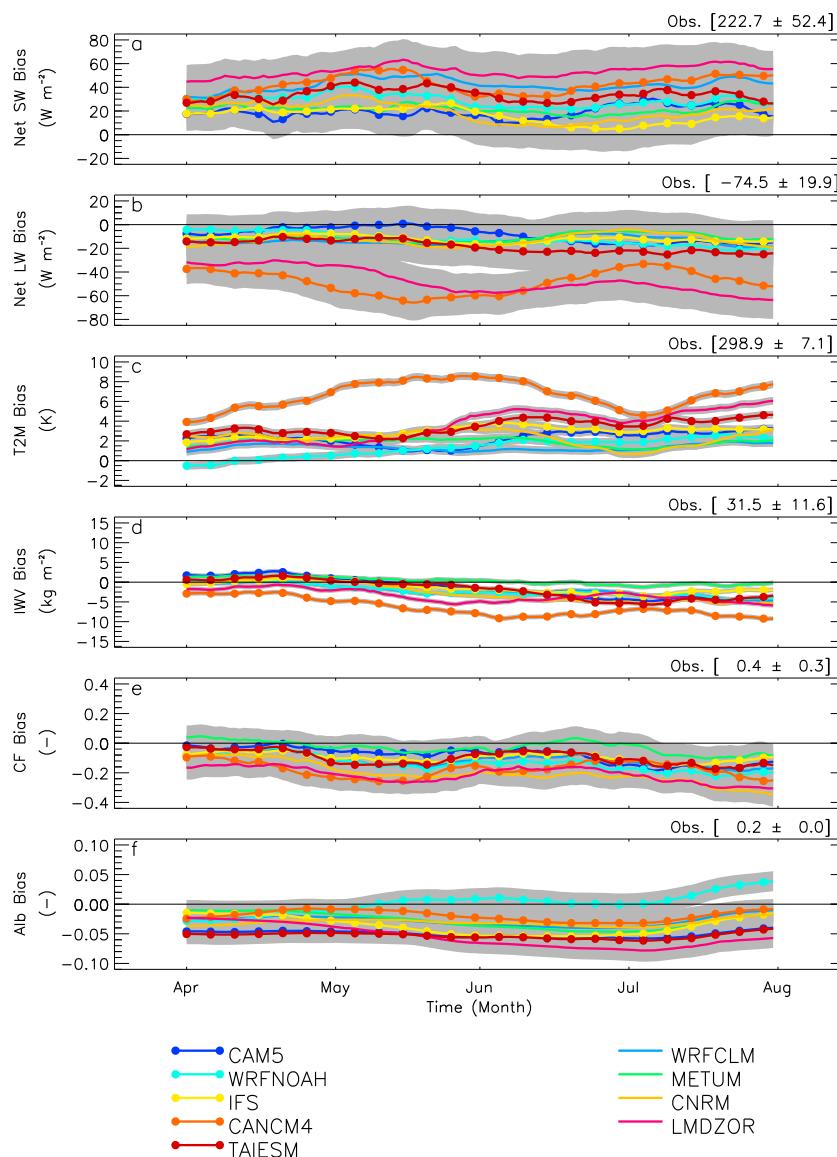


Figure 3. One-month running averages of the bias in (a) surface net shortwave radiation, (b) surface net longwave radiation, (c) 2 m temperature, (d) integrated water vapor, (e) total cloud fraction, and (f) surface albedo, for the grid column nearest to the Southern Great Plains (as shown in Figure 1) and averaged over all hindcast days ($T + 00$ to $T + 120$). The colors indicate different models as shown in the legend. Observations used to calculate the biases are obtained using the Atmospheric Radiation Measurement Best Estimate by Tang and Xie (2015), for a region of 2° by 2° surrounding the Southern Great Plains Central Facility, as shown in Figure 1. The gray shading indicates the uncertainty in the observations, as calculated using the values in Table 2 and applying the rules for error propagation where necessary. The numbers to the upper right of each panel indicate the mean and standard deviation of the variable denoted in the panel below for the full simulation period.

Since the context of this radiation attribution study is the midlatitude warm bias, it is worth evaluating the temperature errors coinciding with the radiation biases in each of the models. The monthly running mean 2 m temperature errors in all models are depicted in Figure 3c, and the average temperature bias over the full simulation period is shown in Table 3. All models suffer from a persistent warm bias throughout most of the simulation period. The bias is particularly large in models with large positive biases in the net SW radiation (CanCM4, LMDZOR, and TAIESM; Figures 3a and 3b). Interestingly, some other models, such as the WRFCLM and WRFNOAH have fairly modest 2 m temperature biases, yet rather large errors in the net surface SW (Figures 3a and 3c).

Table 3

Biases of Net Shortwave (SW), Net Longwave (LW), 2 m Temperature (T2M), Integrated Water Vapor (I WV), Cloud Fraction (CF), and Albedo (α) for All Models

| Model | SW (W m ⁻²) | LW (W m ⁻²) | T2M (K) | I WV (kg m ⁻²) | CF (%) | α (%) |
|---------|----------------------------|----------------------------|------------|-------------------------------|-----------|-----------------|
| CAM5 | 19.0 | -9.2 | 2.3 | -1.4 | -8.0 | -4.9 |
| WRFCLM | 41.4 | -14.1 | 1.2 | -2.2 | -11.8 | -2.7 |
| WRFNOAH | 29.3 | -12.0 | 1.1 | -2.1 | -13.4 | 0.2 |
| METUM | 22.1 | -10.9 | 1.8 | 0.4 | -2.2 | -2.5 |
| IFS | 15.5 | -12.1 | 2.8 | -1.3 | -9.3 | -3.3 |
| CNRM | 20.4 | -14.7 | 2.4 | -2.7 | -20.4 | -3.1 |
| CANCM4 | 41.1 | -47.6 | 6.4 | -6.3 | -17.2 | -2.0 |
| LMDZOR | 51.8 | -46.2 | 3.3 | -3.4 | -21.6 | -5.1 |
| TAIESM | 32.7 | -18.5 | 3.4 | -1.7 | -9.6 | -5.2 |

Note. Data are obtained from the Atmospheric Radiation Measurement Best Estimate data (Tang & Xie, 2015) and include the entire simulation period (April–August 2011) and the full hindcast time ($T + 00$ to $T + 120$).

It is instructive to compare the radiation biases to the biases in a number of variables that are known to affect the radiative transfer. Figures 3d–3f show the monthly running mean relative biases of the I WV, cloud cover, and albedo for all the models. Observations are taken from the gridded ARMRE data as described in section 2. From Figure 3d, the I WV remains within about 10% of the observations throughout most of the simulation period, apart from the CANCM4 (and to a lesser extent the LMDZOR), which tends to exhibit a more prominent dry bias of up to 20%. Many models show a tendency of developing a dry bias over summer (Figure 3d), but again show fairly modest change of the bias over the hindcasts (Figure 4d). It is possible that the dry I WV bias, mainly later in summer, plays a role in the radiation biases seen in Figures 3a and 3b, given that dry atmospheres are more transparent to SW radiation and at the same time emit less LW radiation back to the surface. Further, all models tend to have insufficient cloudiness throughout most of the simulation period (Figure 3e and Table 3), with the smallest bias magnitudes in the METUM and the largest biases in the LMDZOR, CNRM, and CANCM4. Some models tend to get cloudier over the five hindcast days, in particular, the CAM5, TAIESM, and METUM around June, partly offsetting the negative bias in cloudiness (Figure 4e) later in the hindcasts. A lack of cloud would undoubtedly contribute to a positive bias in the downwelling SW and a negative bias in the downwelling LW radiation; hence, cloud cover biases seen in Figure 3e seem consistent with the radiation biases in Figures 3a and 3b.

A third possible contributor to SW radiation biases could be the surface albedo (defined as SW_U/SW_D), shown in Figure 3f and Table 3. Apart from the WRFNOAH, all model surface albedos are lower than observed. The albedo bias remains very constant throughout the hindcasts (Figure 4f), as would be expected given the typically slow changes in surface properties that affect the albedo. Again, a lower-than-observed albedo is consistent with a positive bias in the net SW radiation at the surface seen in Figure 3a.

It clearly is hard to identify one single culprit for the radiation biases present in all models, since many processes seem to push the SW and LW biases in the same direction throughout most of the simulation period. The next section will present a detailed attribution analysis as outlined in section 3 to quantify how much each of the above mentioned processes separately contributes to the overall radiation bias.

4.2. Attribution of Biases

4.2.1. Attribution of the Net Shortwave Bias

Since the growth rates of the biases in surface radiation over the hindcasts were found to be small (Figure 4), the remainder of this paper will focus on the entire simulation period from April to August 2011 and include the full hindcast lengths.

Given that biases in the surface albedo, clouds, and the I WV all tend to be consistent with a positive net SW bias and a negative net LW bias in most models (section 4.1), a more detailed and quantitative analysis is needed. Therefore, the bias attribution method laid out in section 3 has been applied to all model simulations. Figure 5 shows the contributions from albedo, clouds, I WV, and the residual contribution to the net SW radiation bias, as defined in equations (A4), (A5), (A7), and (A8), respectively. Also shown in this figure

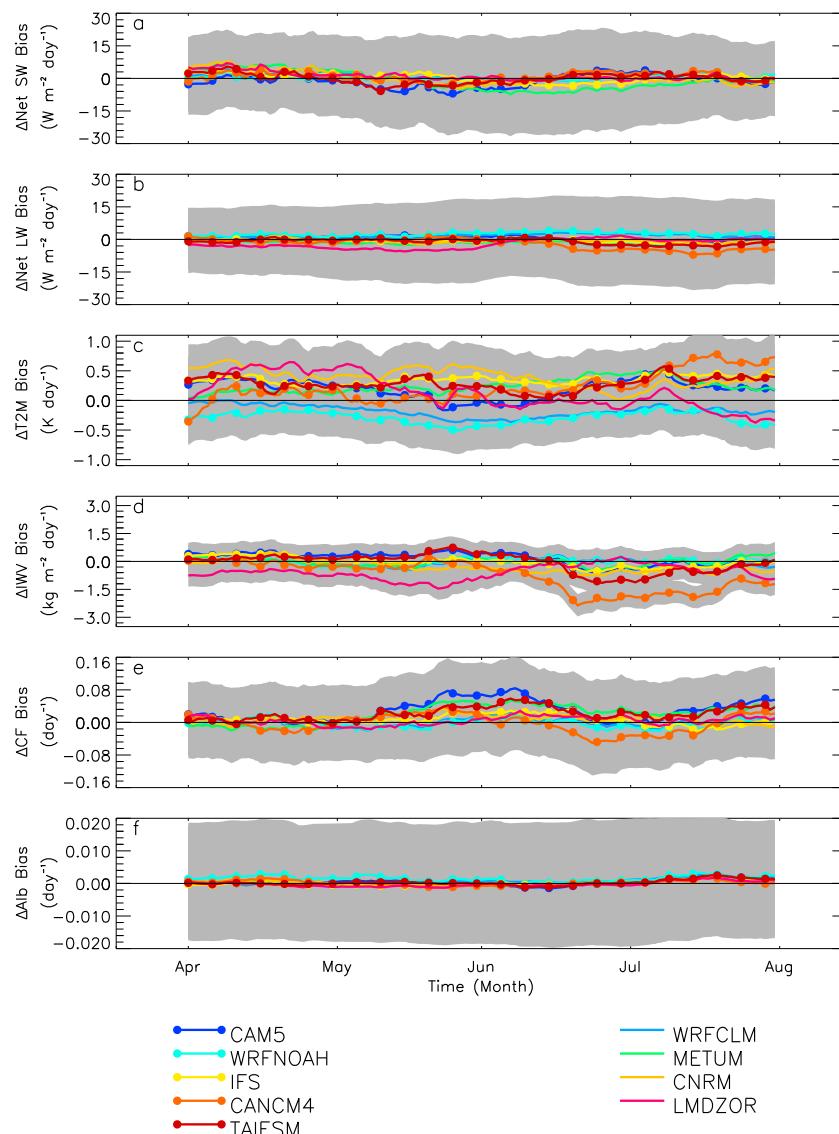


Figure 4. As in Figure 3 but for the average growth of the bias per hindcast day.

is the distribution around the mean contributions (as boxes and whiskers). The edges of the boxes are highlighted if the contribution is significant, based on a Wilcoxon rank-sum test.

Consistent with Figure 3f, the too low surface albedo in most models has a significant impact on the net SW radiation balance. The largest contribution of the surface albedo to the net SW radiation bias is found in the CAM5, the TAIESM, and the LMDZOR models, exceeding 15 W m^{-2} , which is an important portion of the total SW radiation bias. It is interesting to note that both the CAM5 and TAIESM share the same land surface scheme. The WRFNOAH model has a significantly negative contribution from the surface albedo to the net SW bias (Figure 5c), consistent with the overestimated albedo throughout most of the simulation period in Figure 3f.

In all models, apart from the CAM5, clouds appear to be the dominant source of the net SW radiation bias. The radiatively most important cloud biases are found in the WRF models, the CANCM4, and the LMDZOR (Figures 5b, 5c, 5g, and 5h), which is consistent with the large cloud fraction biases in these models in Figure 3e. It should be highlighted that other than errors in the cloud cover (as shown in Figure 3e), errors can be present in the cloud optical properties even when the correct cloud cover is predicted. Indeed, the METUM has a fairly unbiased cloud cover on average (Table 3 and Figure 3e), yet still has a dominant contribution from cloud errors to the net SW radiation bias. Similar or slightly larger cloud-related radiation biases are found in the IFS, CNRM, and TAIESM (Figures 5e, 5f, and 5i), despite a more pronounced lack of cloud cover in these models.

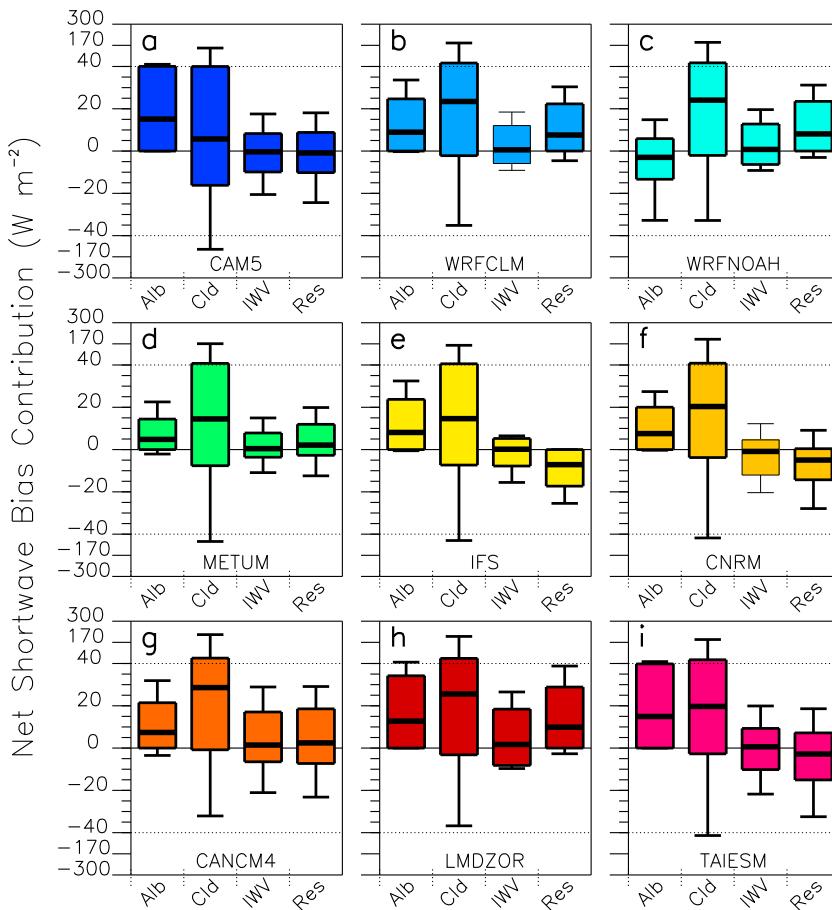


Figure 5. Net shortwave radiation bias attribution in all models. Contributions from albedo (Alb), clouds (Cld), integrated water vapor, and residual processes (Res) are provided and calculated from equations (A4) to (A8). Boxes contain the 15th–85th percentiles of the distribution containing the full simulation period (April to August 2011, including day and night) and the five hindcast days. The mean contribution is indicated by the bold horizontal line, and whiskers indicate the 5th and 95th percentiles. Contributions that are significantly different from zero, based on a Wilcoxon Rank-Sum test (with 5% confidence interval) are highlighted by bold box edges and darker shading. Observational data used in this analysis obtained are from the Atmospheric Radiation Measurement Best Estimate data described in Tang and Xie (2015). Note that the Y axis is stretched below -40 W m^{-2} and above 40 W m^{-2} (indicated by the horizontal dotted line) to include all data on the panels.

compared to the METUM (Figure 3e). The smallest, although still significant contribution of cloud errors to the overall net SW bias is found in the CAM5. Indeed, the cloud-related SW bias is smaller than the surface albedo-related bias (Figure 5a) for this model. This finding is consistent with Van Weverberg et al. (2015).

From Figure 3d, it became clear that the atmosphere in most models is too dry over the SGP during the simulation period. However, SW radiation bias contributions from the biased Iwv composites are very small for all models (Figure 5). Somewhat surprisingly, the small Iwv contributions are still significant for most models, apart from the WRFCCLM and the CNRM.

The residual bias combines effects from the aerosol optical depth and inaccuracies in the clear-sky radiative transfer calculations. While generally small, this term still entails a significant contribution to the overall net SW radiation bias in all models. The sign of this bias varies among the different models in question, however. The LMDZOR model appears to have the largest positive radiation bias associated with the residual term. It is unclear whether this model underestimates the aerosol loading in the SGP or whether possible inaccuracies in the radiative transfer could explain this positive contribution. Interestingly, the IFS and CNRM have a fairly large, but negative, impact of the residual term on the overall net SW bias (Figure 5e), partially offsetting positive contributions from the surface albedo and clouds (Figure 5e).

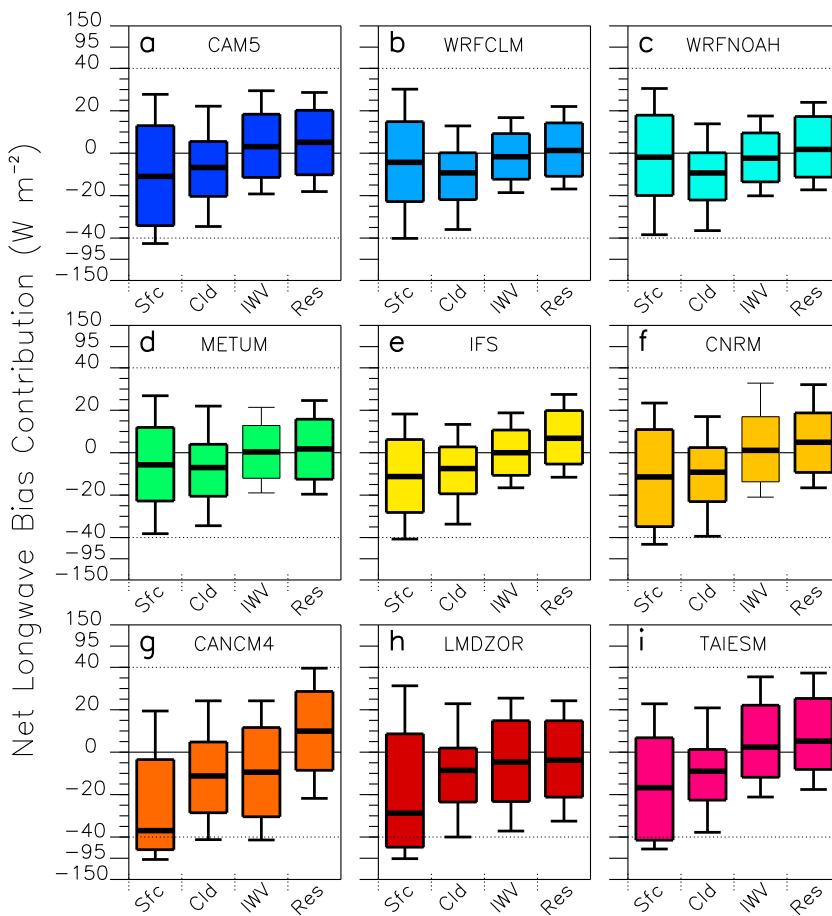


Figure 6. As in Figure 5 but for the net longwave radiation bias.

In summary, this analysis has shown that cloud issues are the dominant source of the radiation biases in all models apart from the CAM5. The largest cloud radiative problems occur in the WRF models, the CANCM4, and the LMDZOR. A positive, but mostly secondary effect originates from albedo errors in most models, apart from the WRFNOAH.

4.2.2. Attribution of the Net LW Bias

Figure 6 provides an overview of the contributions from each of the processes outlined in section 3 to the net LW radiation bias. Given that all models have too warm near-surface temperatures (Figure 3c), it is unsurprising that the surface contribution to the net LW bias is significantly negative in all models. Indeed, following the Stefan-Boltzmann law, warmer surfaces ought to have a larger upwelling LW radiation. The largest warm biases are found in the CANCM4 and LMDZOR models (Figure 3c), which also have the largest contributions from the surface to their overall net LW radiation bias. The surface temperatures in the WRFNOAH were comparable to the observed temperatures (Figure 3c) and hence the contribution of the surface to the overall net LW radiation bias is small (albeit still significant).

As for the net SW radiation bias, errors in the cloud properties contribute very significantly to the overall net LW radiation bias in all models (second bar and whiskers in each panel of Figure 6). It is interesting to note that the variability among the different models in terms of the cloud contribution to the net LW radiation bias is smaller than for the net SW bias. Most models exhibit net LW biases of about -12 to -8 W m^{-2} due to errors in the cloud fractions or radiative properties. Also note that the negative LW cloud contribution generally does not offset the positive SW cloud contribution.

The dry Iwv bias (Figure 3d) appears to have a significant, but generally small impact on the net LW radiation in all models, apart from the METUM and the CNRM. A clearly negative contribution can be found for those models with the largest overall dry bias, such as the CANCM4 and the LMDZOR (Figures 6g and 6h).

Interestingly, while still having a dry bias in the atmosphere on average (Figure 3d), some models, such as the CAM5 and the TAIESM, have a positive contribution from I WV to the LW radiation bias (Figures 6a and 6i). A mechanism explaining this apparent contradiction is provided in Appendix B.

The residual component (last box plot in each panel on Figure 6) is generally fairly small, but significant for all models. Fairly large positive contributions to the overall net LW radiation are found in the CANCM4 and the TAIESM (Figures 6g and 6i). The LMDZOR model is the only model with a large negative contribution from the residual term to the net LW radiation bias (Figure 6h). As for the net SW radiation bias (Figure 5), there is a need to further investigate what causes these small but significant contributions from the residual term in each of the models. This could be related to aerosol properties and also to issues with the radiative transfer code itself, such as the definitions of the solar and infrared spectra in each of the models (e.g., Li et al., 2010). Indeed, for many models the SW and LW biases in the residual term seem to balance each other (Figures 5 and 6). It should be mentioned that from Morcrette et al. (2017), the warm bias extends fairly deep into the troposphere. Hence, it can be expected that many models suffer from an excess in downwelling LW radiation associated with the warmer atmosphere. Ideally, this effect should end up in the residual term in Figure 6, but it should be borne in mind that temperature and water vapor content of the atmosphere are highly correlated. Hence, the contribution from a too warm troposphere will be hard to disentangle between the I WV and residual components in our analysis.

4.3. Cloud Contribution to Radiation Bias

4.3.1. Cloud Property Analysis

Given that clouds cause a significant portion of the net SW and LW radiation biases in many of the models in this study, a more detailed evaluation of the cloud fields is desirable. As a first step, this section deals with the ability of the models to replicate the vertical and diurnal structure of observed cloudiness. The next section will then establish the link between the cloud and radiation deficiencies, using the cloud regime definitions outlined in section 3.

Figure 7 shows the diurnal cycle of the average cloud fraction profile at the SGP in the observations (Figure 7a) and in all of the nine models under investigation. For this particular period in 2011, observed cloud fractions showed a maximum in the midlevels in the morning and evening, and a minimum in the afternoon. In the lower troposphere, a weak maximum in cloud fraction is apparent around noon (Figure 7a). This diurnal cycle of the cloud fraction profiles compares well with Zhang and Klein (2010), who looked at 11 years of cloud data over the SGP. Similar to Morcrette et al. (2012), we acknowledge here that errors in the average cloud fraction (AVG) entail a combination of errors in the frequency of occurrence (FOO) and the amount when present (AWP). Potential compensating errors between the FOO and the AWP matter in the light of radiation errors given the nonlinear impact of clouds on the radiative transfer. Furthermore, revealing these errors can also help unveil the origins of the cloud deficiencies in the various models under investigation. The observed FOO (Figure 8a) shows how cloud frequently builds up in the low levels around local noon, then deepens throughout the afternoon, culminating in an abundance of high clouds in the late evening. However, there is also a second, weaker peak in the FOO in the early morning. This peak appears to be more restricted to the middle and high levels, possibly associated with the nocturnal elevated convective storms that frequently occur in the SGP (Geerts et al., 2017). The observed AWP (Figure 9a) is smallest in the afternoon, mainly in the low levels of the atmosphere, indicating that the large FOO (Figure 8a) in the afternoon is associated with partly cloudy scenes, typical for cumulus and congestus-type cloud.

From Figures 7b–7d, the CAM5 and the WRF models have similar diurnal cycles in their cloud fraction profiles. This is somewhat surprising given their very different cloud-related SW and LW radiation errors (panels (a)–(c) of Figures 5 and 6). However, recall that these models all share the same set of atmospheric physical parameterizations (Morcrette et al., 2017). These three models tend to have a very distinct maximum in the cloudiness in the late afternoon, at all levels of the atmosphere (Figures 7b–7d). The fairly well captured afternoon AVG cloud cover masks large compensating overestimations in the FOO (Figures 8b–8d) and underestimations in the AWP (Figures 9b–9d), however. Possibly, this hints at the CAM5 and the WRF models triggering deep (convective) cloud too easily, but systematically underestimating the cloud fraction associated with these clouds. This would be consistent with earlier findings by, for example, Wang et al. (2014). The nocturnal AVG cloud cover, and in particular the early morning observed maximum, is largely underestimated in these three models (Figure 7), due to too infrequent cloud (Figure 8) and too small AWP (Figure 9).

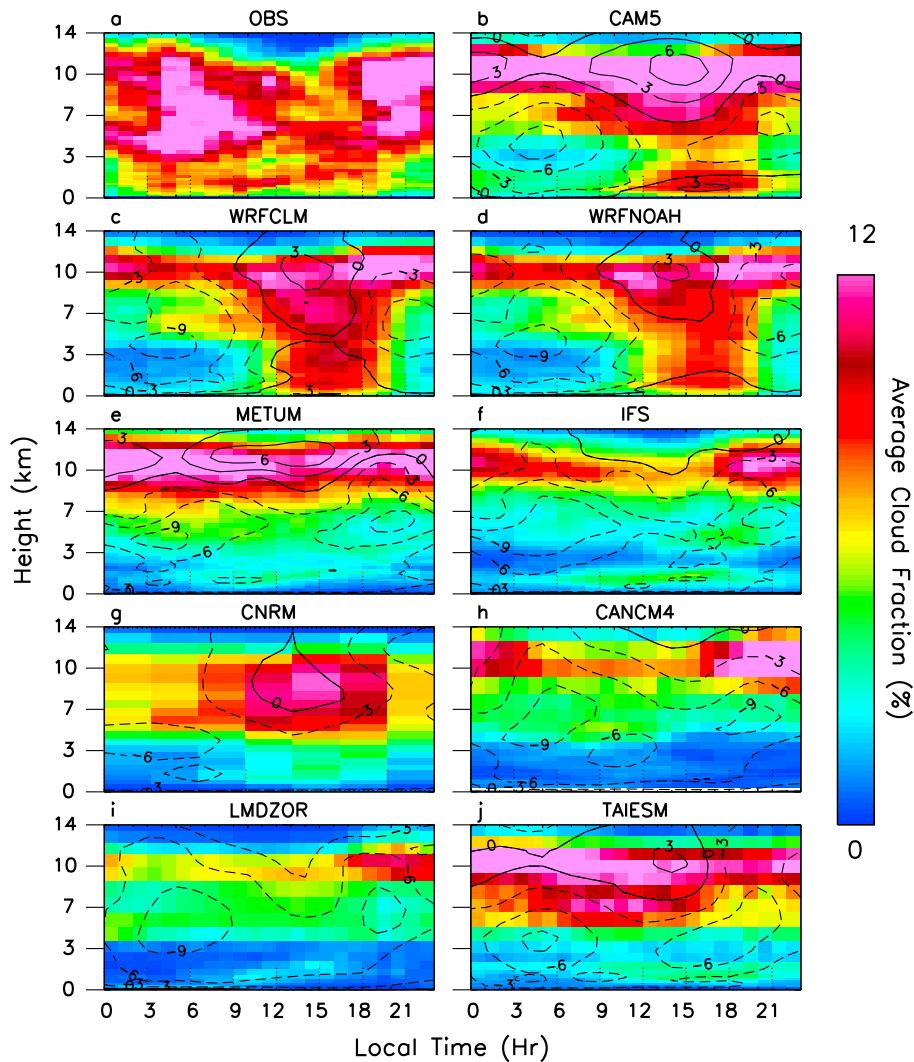


Figure 7. Diurnal cycle of the average cloud fraction profile in the observations (a) and each of the models (b–j). The observed cloud fraction shown in panel (a) is the fraction regressed to the IFS grid definition, since this model has the most detailed vertical resolution. Observations regressed to each of the respective model grids were used to calculate the biases shown by the black contours. The bias for each of the models is shown as solid (positive) and dashed (negative) lines on top of the color shading. Contour intervals are 3%. All panels include data from the full simulation period (April–August 2011) and all hindcast days ($T + 00$ to $T + 120$) for the respective models, including cloudy and noncloudy data points

The diurnal evolution of the AVG cloud fraction profile is fairly similar between the METUM and the IFS (Figures 7e and 7f). In contrast to the CAM5 and WRF models, the METUM and IFS lack the prominent maximum in the AVG cloud cover in the afternoon. Both models exhibit fairly weak diurnal cycles in their AVG cloud cover, with persistent overestimations of the high cloud cover and underestimations of the middle and low cloud cover. While the AVG cloud fraction profiles in these two models look similar, their FOO (Figures 8e and 8f) and AWP (Figures 9e and 9f) are very different. The METUM seems to have cloud too infrequently in the low level and midlevels (Figure 8e), but captures the AWP fairly well. Despite its lack of midlevel AVG cloud, the IFS largely overestimates the midlevel and high-level cloud frequency (FOO; Figure 8f), mainly in the early evening hours. However, apart from within the boundary layer, the AWP in the IFS is dramatically underestimated compared to the observations.

A more pronounced diurnal cycle in cloudiness is present in the CNRM (Figure 7g), compared to the IFS and METUM. However, rather than the observed maximum in midlevel and upper-level cloudiness in the morning and late evening, the CNRM has its maximum in AVG cloud fraction in the afternoon. At this time,

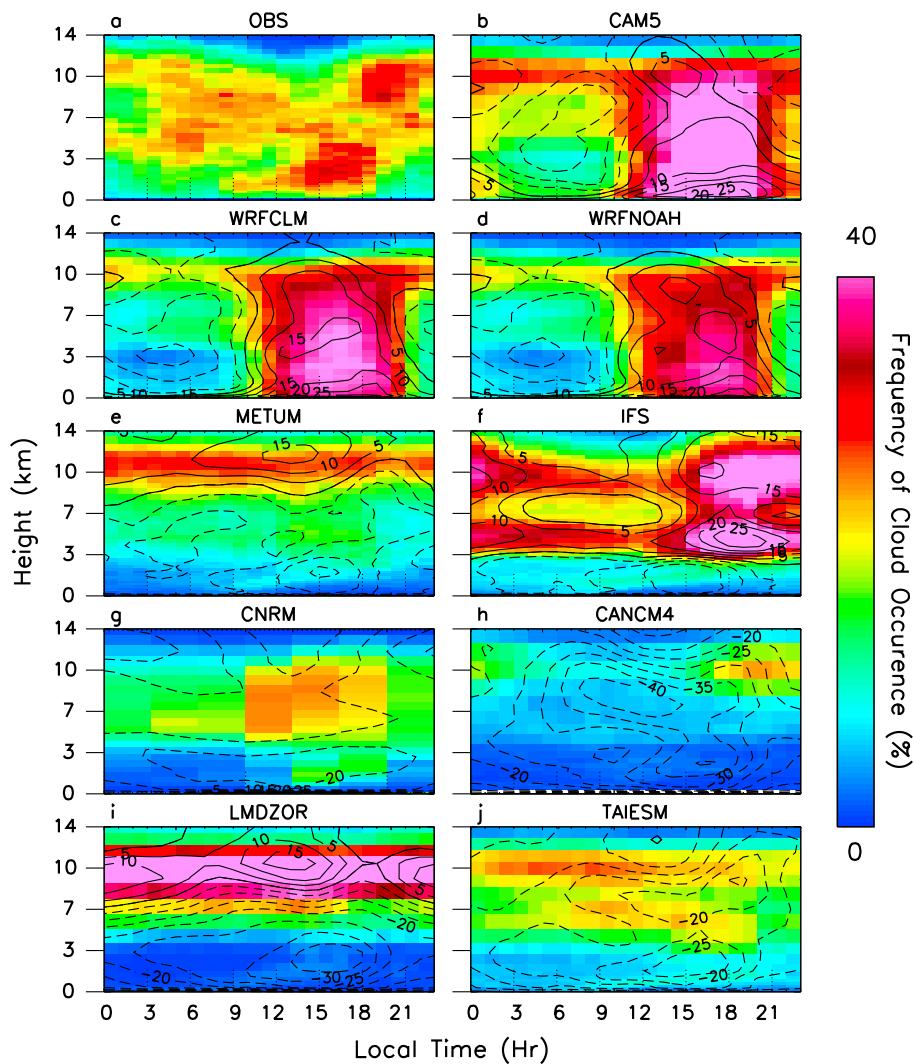


Figure 8. As in Figure 7 but for the frequency of occurrence of any nonzero cloud fraction. Contour intervals are 5%.

the upper-level AVG cloud fraction is overestimated, while the low-level cloudiness remains underestimated throughout the entire diurnal cycle. The latter is mainly due to underestimating the FOO (Figure 8g), while the AWP is fairly well captured, or even slightly overestimated in the CNRM (Figure 9g).

The CANCM4, LMDZOR, and TAIESM all share a deficiency in their low-level and midlevel cloud throughout the entire day (Figures 7h–7j). In the CANCM4 and LMDZOR this is mainly due to very infrequent cloud at these levels of the atmosphere (Figures 8h and 8i). The TAIESM (Figure 7j) has better captured low-level cloud statistics than the LMDZOR and the CANCM4, but has dramatically less cloud in the afternoon than the CAM5 (Figure 7b). This is remarkable, given that TAIESM and CAM5 share most of their dynamics and physics. The main difference between the CAM5 and TAIESM resides in the cloud macrophysics and convective triggering, which clearly has an important impact. The updated convective triggering in the TAIESM, described in Wang et al. (2014), acts mainly to suppress the convective initiation, which has reduced the overestimated FOO in the afternoon compared to the CAM5. However, given that the AWP in the TAIESM (Figure 9j) is relatively similar to the CAM5 and still underestimated at midlevels, the net result is now an underestimation of the AVG cloud amount. The AVG high-level cloud in the TAIESM is better captured than in the CANCM4 and LMDZOR (and the CAM5), both in terms of the FOO and the AWP (Figures 8 and 9). The CANCM4 and LMDZOR have fairly similar and somewhat underestimated AVG cloud amounts in the upper levels of the troposphere (Figures 7h and 7i), but this masks very different behavior in terms of the FOO and the AWP.

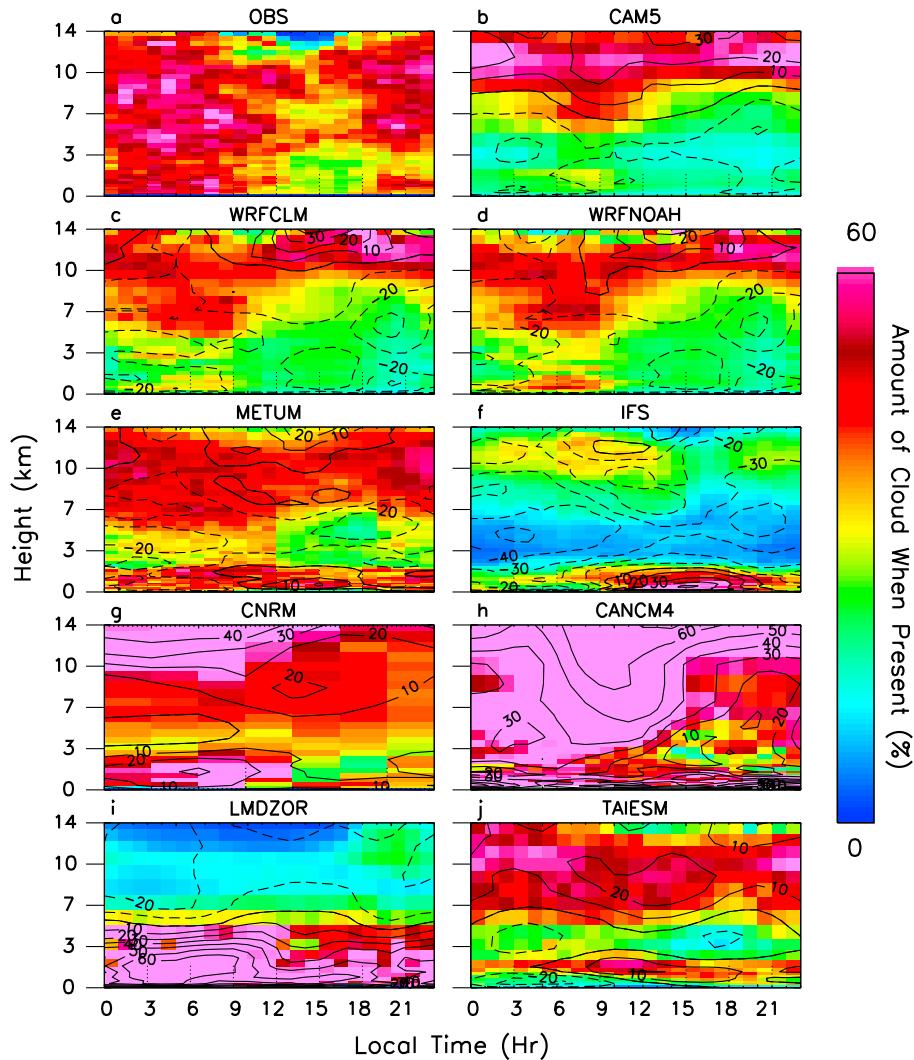


Figure 9. As in Figure 7 but for the amount when present of any nonzero cloud fraction. Contour intervals are 10%.

Indeed, while the CANCM4 exhibits a compensation between far too small FOO and too large AWP (Figures 8 and 9h), the LMDZOR compensates a far too large frequency with much too small AWP (Figures 8 and 9i).

4.3.2. SW Radiative Effect of Cloud Regimes

It is clear from the previous section that all models struggle to replicate the observed cloud fraction and frequency, although each of the models appears to have very model-specific issues. To further tease out the specific issues in each of the models, it is necessary to perform a more detailed analysis. While the average diurnal cycles of the cloud fraction profiles in Figure 7 shed some light on the vertical structure of the clouds, they do not provide information on how the model biases in the vertical model levels are vertically overlapped. Indeed, a bias in the average high- and low-level cloud fraction could be the result of unconnected issues with high-level clouds during a particular period and low-level clouds during a different period, but they could also originate from a bias in deep cloud properties, including low and high-level cloud at the same time. The regime classification proposed in section 3 enables a simple analysis of the vertical cloud structure, while keeping the tie between cloud issues that are vertically overlapped.

The ultimate goal of this paper is to find how errors in the cloud properties propagate into biases in the surface radiation. To this end, surface radiation biases can be assigned to their concurrent cloud regimes, allowing an unambiguous identification of the cloud regimes that matter most from a radiation error point of view. For the sake of brevity, the remainder of this analysis will focus on the SW radiation, although it should be mentioned that similar findings were made for the LW radiation errors in most models. To provide a way to evaluate

Table 4

Simulated and Observed Shortwave (SW) Total Cloud Radiative Effect (tCRE) Broken Down by Cloud Regime and Aggregated for All Cloud Regimes Combined

| – | 1 (W m ⁻²) | 2 (W m ⁻²) | 3 (W m ⁻²) | 4 (W m ⁻²) | 5 (W m ⁻²) | 6 (W m ⁻²) | 7 (W m ⁻²) | Total (W m ⁻²) |
|-------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|-------------------------------|
| CAM5 mod | 4.5 | 0.6 | 2.4 | 0.3 | 5.0 | 1.7 | 25.0 | 39.4 |
| CAM5 obs | 4.0 | 0.7 | 5.9 | 2.5 | 4.5 | 5.4 | 26.8 | 50.3 |
| WRFCLM mod | 2.9 | 1.1 | 2.2 | 0.4 | 3.5 | 1.5 | 11.1 | 22.7 |
| WRFCLM obs | 4.7 | 2.5 | 7.3 | 3.3 | 5.6 | 5.5 | 20.7 | 50.5 |
| WRFNOAH mod | 3.9 | 1.0 | 2.1 | 0.5 | 4.0 | 1.6 | 8.7 | 21.9 |
| WRFNOAH obs | 4.7 | 2.5 | 7.3 | 3.2 | 5.6 | 5.4 | 20.6 | 50.4 |
| METUM mod | 4.1 | 1.3 | 1.7 | 1.6 | 4.1 | 5.2 | 13.1 | 31.3 |
| METUM obs | 5.0 | 1.9 | 6.3 | 3.9 | 5.6 | 6.2 | 20.8 | 50.9 |
| IFS mod | 2.6 | 1.1 | 2.7 | 0.9 | 3.0 | 7.9 | 13.9 | 32.2 |
| IFS obs | 4.5 | 1.1 | 6.5 | 3.4 | 6.1 | 6.6 | 21.6 | 50.7 |
| CNRM mod | 3.2 | 0.1 | 0.7 | 0.2 | 1.3 | 3.8 | 15.3 | 24.6 |
| CNRM obs | 2.3 | 1.1 | 5.4 | 2.6 | 4.3 | 5.4 | 26.5 | 47.9 |
| CANCM4 mod | 1.7 | 1.5 | 1.9 | 3.1 | 1.3 | 5.3 | 1.5 | 16.2 |
| CANCM4 obs | 1.9 | 0.8 | 5.6 | 2.4 | 3.8 | 5.8 | 28.9 | 49.5 |
| LMDZOR mod | 1.8 | 0.5 | 0.6 | 1.4 | 1.3 | 7.7 | 5.8 | 19.0 |
| LMDZOR obs | 2.9 | 0.7 | 5.9 | 2.7 | 4.3 | 5.0 | 28.2 | 49.9 |
| TAIESM mod | 2.6 | 1.3 | 1.9 | 0.7 | 3.4 | 4.3 | 11.4 | 25.7 |
| TAIESM obs | 4.0 | 0.6 | 5.9 | 2.7 | 4.7 | 5.3 | 26.7 | 50.3 |

Note. Shown are the low cloud regime (1), mid cloud (2), mid/low cloud (3), high cloud (4), low/high cloud (5), mid/high cloud (6), and the deep cloud regimes (7). The regime with the largest absolute CRE error is highlighted in bold for each model. Radiation data are from the radiometers at the SGP Central Facility, and the Cloud information is from ARSCL. Data include the entire simulation period (April–August 2011) and the full hindcast time ($T + 00$ to $T + 120$) and are averaged over the entire diurnal cycle (day and night). Note that the observations are regressed to each of the model grids and hence vary slightly for each of the models.

the coevolution of cloud and radiation errors, the total shortwave cloud radiative effect (tCRE) is calculated for each of the models and the observations as follows:

$$\text{tCRE}_i = \overline{(\text{SW}_{\text{Di}} - \text{SWCS}_{\text{Di}})} \cdot \text{freq}_i = \text{aCRE}_i \cdot \text{freq}_i \quad (6)$$

where the index i denotes the regime (from 0 to 7), SW_{Di} and SWCS_{Di} are the average all-sky and clear-sky downwelling radiation associated with regime i , and freq_i is the frequency of regime i (as a fraction). Note that the CRE as defined here only includes downwelling radiation near the surface, as opposed to the more conventional definition using the net surface radiation.

The tCRE for each of the simulated and observed regimes is provided in Table 4. Furthermore, Figure 10 visualizes the two components of equation (6) for each of the models and the observations. The average shortwave cloud radiative effect (aCRE_i) of each regime ($\text{SW}_{\text{Di}} - \text{SWCS}_{\text{Di}}$) is shown as the height of each bar in Figure 10, while the FOO freq_i is shown as the width of each bar. Hence, following equation (6), the surface area of each bar (width \times height) is the SW tCRE of the regime in question, also listed in Table 4. For each model under investigation, the simulated regime properties are given above the zero line in the panel, while the regressed observed regime properties are given below the zero line in the panel. A flawless model would hence appear as a perfect reflection below the zero line of the regime properties above the zero line. This allows the SW tCRE of each of the models to be evaluated, while looking at the contributions from each of the eight cloud regimes, each of which is a product of two factors. If a particular regime has a bias in its tCRE, this could be due to the model not producing this regime frequently enough (width of the bars; freq_i in equation (6)) and/or the model producing the wrong radiative properties associated with the regime (height of the bars; aCRE_i in equation (6)). The C_{cloud} defined in equation (A5) is hence the difference in the simulated and observed tCRE, summed over all regimes (multiplied by the one minus observed albedo).

One of the more puzzling findings of the previous sections is the large similarity in the diurnal cycles of the vertical cloud properties between the CAM5, on the one hand, and the WRF models, on the other hand (panels (b)–(d) of Figures 7 to 9), while their cloud-related surface radiation errors are very different (Figure 5).

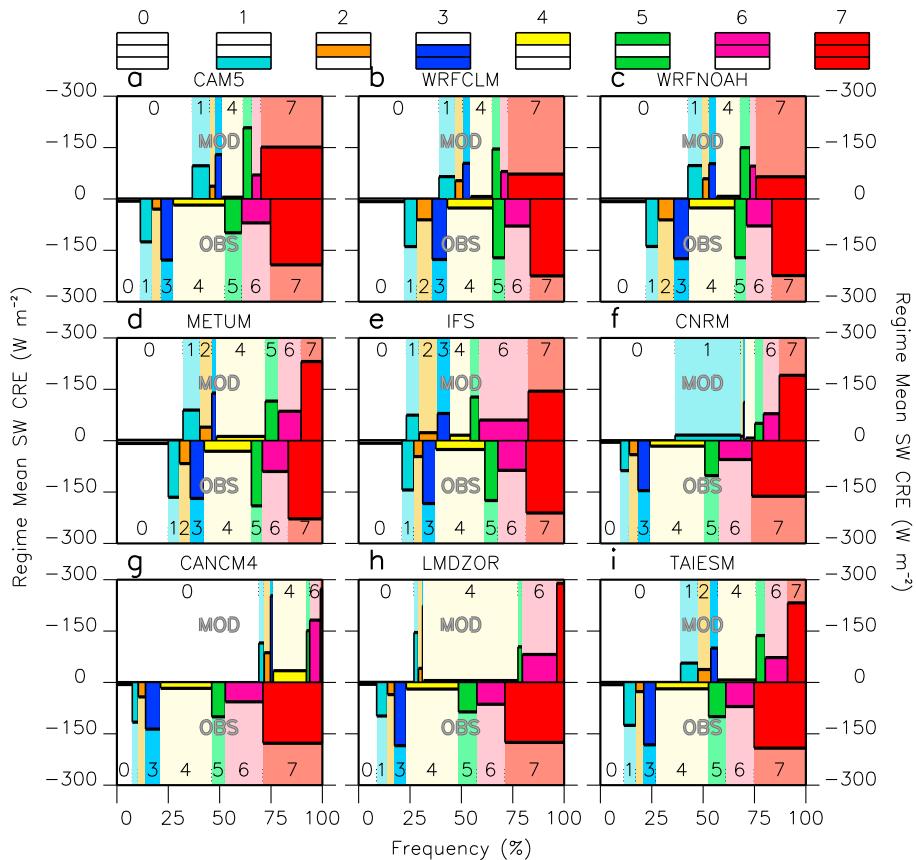


Figure 10. Statistics of shortwave cloud radiative effect (CRE) for each of the regimes in all models. Each of the colored bars represents one of the cloud regimes defined in Figure 2. In each panel, the bars above the zero line represent the model, while the bars below the zero line denote the observations. Watermark colors in the background have been added for each regime for clarity. The width of each bar indicates the relative frequency of the regime (and hence sums up to 1 for all regimes combined). The height of the bar represents the average CRE (aCRE) when the regime is present. The total CRE of the regime is defined as the product of the regime's frequency and its aCRE and is represented by the surface area of each of the bars. Data in this figure are from the full simulation period (April to August 2011, including daytime only) and the five hindcast days. A perfect model would hence imply a perfectly mirrored picture above and below the zero line. Regime numbers are indicated at the top of the figure as guidance, only if its frequency is larger than 0.07. Regime definitions are shown at the top of the figure as a reminder, with regime colors matching those in the panels.

Figure 10 can help explain this apparent contradiction. From this figure, the frequencies of all regimes (width of the bars) are very similar between the CAM5 (Figure 10a) and the WRF models (Figures 10b and 10c), consistent with the diurnal cycles of the cloud properties in the previous section (Figures 7–9). Indeed, these three models tend to have clear skies too often (regime 0). The frequency of the deep cloud regime (regime 7) is slightly overestimated in the CAM5 and largely overestimated in the WRF models. It should be noted that the frequency of the deep cloud regime in the afternoon in the CAM5 is much more overestimated than in the morning (not shown). This confirms that their large overestimation in the FOO (Figures 9b to 9d) at all levels of the atmosphere in the afternoon is largely due to the deep cloud regime. Interestingly, the average cloud radiative effects (aCRE) of the regimes exhibit greater differences between the CAM5 and the WRF models (height of the bars in Figures 10a to 10c). Given its large frequency as well as its large aCRE, the deep cloud regime (regime 7) dominates the overall cloud radiative effect (Table 4 and Figure 10). The tCRE of this regime is fairly well captured in the CAM5 (Table 4 and surface area of the red bar in Figure 10a), due to compensating errors between a too large frequency and a too small aCRE. Conversely, the tCRE of regime 7 in the WRF models is largely underestimated (Table 4 and Figures 10b and 10c), due to a very small aCRE for this regime. This analysis highlights that the too large downwelling surface SW radiation in the WRF models does not originate from missing deep cloud events, but rather from misrepresenting their CRE when they are present. The following section will further delve into the origins of the different CRE of regime 7 between the CAM5 and the WRF models. Given that these models share the same atmospheric physics parameterizations,

it would be interesting to find out whether this difference is predominantly a resolution effect or due to the different dynamical cores in these models. It is worth noting that many other regimes contribute small amounts to the overall bias in the CRE as well as in all three models, specifically the low/mid (3) and mid/high (6) regimes (Table 4). Indeed, mainly these latter regimes are responsible for the (small) contribution of clouds to the overall surface radiation errors in the CAM5 (Figure 5).

While the AVG cloud fraction profiles were shown to be similar between the METUM and the IFS (Figures 7e and 7f), their cloud regime statistics differ somewhat (Figure 10). The METUM appears to have too much clear sky at the expense of mainly the deep cloud regime (width of the red bars in Figure 10d and Table 4). The IFS captures the frequency of the deep cloud regime well, but tends to overestimate the occurrence of the mid/high cloud (regime 6) at the expense of high cloud (regime 4; Figure 10e). Focusing on the deep cloud regime, it appears from the surface area of the red bars in Figures 10d and 10e, and from Table 4, that the tCRE of this regime is very similar between the IFS and the METUM, and generally underestimated by nearly 40%. This is mainly a consequence of not having this regime often enough in the METUM, while the aCRE is well captured (Figure 10d). However, the inverse is true for the IFS. The latter model indeed produces regime 7 often enough, but typically with too low aCRE (Figure 10e). This is consistent with the very different profiles of the FOO and the AWP between these two models (panels (e) and (f) of Figures 8 and 9). While regime 7 is the main culprit for cloud-related radiation errors in both models (responsible for about 40% of the bias), other regimes do contribute as well. The high cloud regime (4) contributes in both models due to too weak aCRE in the METUM and too small frequency in the IFS. The low/mid cloud regime behaves similarly to the deep cloud regime and it is its frequency that is underestimated in the METUM (with a fair aCRE), while the aCRE is too small in the IFS (but with fair frequency). The aCRE of the low cloud regime (1) is too small in both models, but this is compensated by too large frequency of this regime in the METUM. The too small aCREs in the IFS associated with most regimes are consistent with, for example, Ahlgrimm and Forbes (2012), who found the IFS to be associated with positive biases in the net SW radiation, even when low-level, deep-level, or midlevel regimes were correctly simulated. It is worth mentioning that the METUM and IFS overall appear to best reproduce the regime frequencies. Interestingly, these are the only models that have a prognostic cloud fraction scheme as well.

Similar to the METUM, the CNRM misses a large portion of the tCRE associated with the deep cloud regime due to underestimating its frequency, while capturing its aCRE quite well (Figure 10f and Table 4). However, the main characteristic of the CNRM is its spurious large frequency of low-level cloud (regime 1), mainly in the afternoon. In Figure 8g there appears to be a maximum in the FOO in the lower levels, but this seems to be still smaller than observed and never more than 15%. One possible explanation is that this model consistently produces excessive shallow cloud in a single model layer, for example, closest to the top of the boundary layer. Since this level probably varies from day to day, the very large frequency of low cloud in Figure 10f could be smeared out over several levels in Figure 8g. This might suggest that the CNRM triggers its shallow convection scheme too readily, but limits these clouds to a single model layer, which could partly explain the very small aCRE. Further, the CNRM largely underestimates the abundance of regimes 4 and 6, while it has too much clear sky (regime 0), leading to an important total bias in the CRE of all regimes combined (Table 4), consistent with Figure 5f.

Both the CANCM4 and the LMDZOR dramatically underestimate the frequency of all cloud types that include low-level cloud (regimes 1, 3, 5, and 7; Figures 10g and 10h), leading to much too small tCREs for all these cloud types (Table 4). The very different FOO and AWP in the upper levels of the troposphere between these two models (panels (h) and (i) of Figures 8 and 9) are reflected in the very different frequency and aCRE of the high cloud regime (4; Figures 10g and 10h and Table 4). The CANCM4 misses much of the high cloud, but overdoes its aCRE, while the LMDZOR has more than double the frequency of high cloud, but with very limited aCRE. It should be highlighted here that the CANCM4 and LMDZOR are the two models with the largest overall bias in the total CRE, which appears consistent with the fact that these models also seem to have the largest 2 m temperature bias (Figure 3c).

The change in the convective triggering in the TAIESM compared to the CAM5 has had a dramatic impact on the cloud properties, as shown in the previous section. This also has important consequences from a radiation point of view. As shown in Figure 10h and Table 4, while the aCRE associated with regime 7 has been largely improved, now the frequency of deep cloud is severely underestimated in the TAIESM, compared to the overestimation in the CAM5. This is partly compensated for by an increase in the frequencies of the low, high,

and mid/high cloud regimes (1, 4, and 6), but overall this has led to a large increase in the bias in the total CRE for all regimes combined (Table 4).

While the analysis from Figure 10 helps understand which deficiencies in the cloud regime statistics are most relevant from a radiation point of view, it is still unclear what role is played by timing errors in the regime occurrences. Indeed, when a certain regime has a too low frequency, it would be helpful to understand which regime is typically simulated instead. For instance, is there a tendency of missing deep cloud events altogether, or of producing cloud that is not deep enough? Even when we get the right frequency of a particular regime over a diurnal cycle, errors in the radiative properties might exist. In this case, the question is whether we get biases in the radiative effect because the regimes occur at the wrong time of day, or whether the regime consistently has wrong radiative properties, even when its timing is correct. The latter is important in light of previous studies repeatedly finding that models with parameterized convection fail to reproduce the observed diurnal cycle of cloudiness and precipitation (Clark et al., 2007; Gustafson et al., 2015; Walther et al., 2013; Yang & Slingo, 2001).

To help unravel these issues, Figure 11 shows the diurnal cycles of the cloud-related SW radiation biases (as defined in equation (A5), subsampled according to the concurrent observed-simulated cloud regime pairs. The bold black lines in this figure represent the C_{cloud} for the respective time of day, while the height of the stacked bars indicates the SW radiation error associated with each of the observed-simulated regime pairs. Cloud regime pairs that are—on average—associated with a positive bias in the SW_D are shown above the zero line and regime pairs that contribute negatively to the SW_D are shown below the zero line. The colors of the bars denote the simulated cloud regime, while the hatching pattern indicates the observed cloud regime, as indicated in the legend on the right-hand side of the figure. The definitions of the cloud regimes are provided at the top of the figure as a reminder.

While the cloud fraction statistics and the regime frequencies in the CAM5 and the WRF models are very similar (Figures 7 to 10), the observed-simulated regime pairs contributing most to the SW bias are very different between these models (Figures 11a–11c). While the overall cloud contribution to the SW bias remains positive throughout the day (bold black lines) in the CAM5, there are often negative contributions when the model simulates deep cloud (regime 7; red bars below the zero line in Figure 11a) when other regimes are observed. This is consistent with deep clouds being triggered too frequently in the CAM5, often when no (0) or just high clouds (4) should be present. However, when the deep cloud regime is correctly simulated, it is associated with a positive SW bias (red bars with dense crossed hatching in Figure 11a), consistent with the too small aCRE of this regime in Figure 10a. Furthermore, simulating clear skies when other regimes (mainly 3 and 7) should be present contributes a fair amount to the SW bias as well (white bars above the zero line in Figure 11a). This analysis confirms that the CAM5 has compensating errors between too frequent deep cloud, but with too weak aCRE. This also means that if the aCRE of deep cloud would be improved in this model, it is likely that this model would develop an overall negative SW bias associated with clouds, given its too frequent deep cloud triggering. In contrast to the CAM5, virtually no cloud regime pairs contribute negatively to the overall SW bias in the WRF models (no bars below the zero line in Figures 11b and 11c). Even occasions of triggering deep cloud (7) instead of any other regime, coincide with a positive SW bias in these models (red bars with "M" symbol hatching above the zero line in Figures 11b and 11c). A very large contribution to the SW bias appears when the model correctly simulates deep cloud (red bars with dense crossed hatching in Figures 11b and 11c). The main difference between the CAM5 and the WRF models is the even smaller aCRE for deep clouds in the WRF models, while their cloud regime statistics are very similar (Figure 10).

The METUM and IFS had very similar AVG cloud fraction profiles (Figures 7e and 7f), but somewhat different regime statistics, with the high cloud and the mid/high cloud regimes simulated too frequently in the METUM and IFS, respectively. Simulating clear skies (white bars) or high cloud (yellow bars) instead of deep (densely crossed hatching) cloud events hence is a main contribution to the SW bias in the METUM (Figure 11d). These errors contribute fairly little in the IFS, for which the SW bias mainly originates from producing mid/high cloud (pink bars) when deep cloud should be present (Figure 11e).

Radiation is called only once every 3 h in the CNRM, hence the coarser temporal resolution in its diurnal cycle of the cloud regime analysis in Figure 11f. The excessive frequency of the low cloud regime draws most of the attention in this model. The main regime that contributes to the overall SW radiation bias is producing just low-level cloud (blue bar) when deep cloud is observed (dense crossed hatching). This might hint

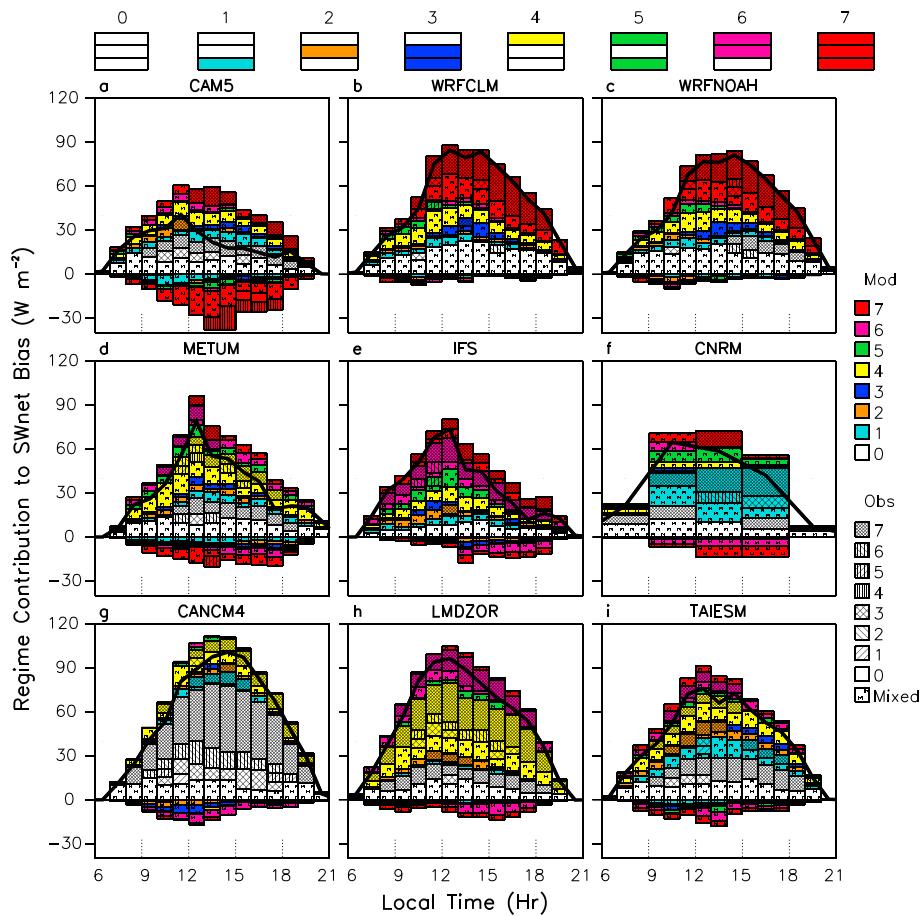


Figure 11. Decomposition of the diurnal cycle of the cloud-related net SW radiation bias (as in equation (A5) into the observed-simulated regime combination. All panels include data from the full simulation period (April–August 2011) and all hindcast days ($T + 00$ to $T + 120$) for the respective models. The height of each stacked colored bar denotes the average contribution of the respective observed-simulated regime pair to the overall C_{cloud} at that time of the day. Note that the legend is constructed so that the color indicates the simulated regime and the hatching style matches the observed regime, as indicated in the legend to the right of the panels. The temporal intervals correspond to the radiation time step in each model. Observed-simulated regime pairs that contribute less than 5 W m^{-2} are combined in the “mixed” category. Observed cloud fractions and radiation are obtained from ARSCL and RADFLUXANAL respectively. The black line overlayed on top of the bars denotes the overall C_{cloud} for each time in the respective model.

at a problem with the convective diagnosis in this model. Missing deep clouds (regime 0 instead of 7 altogether or correctly simulating deep clouds (7-7) also contributes a fair amount to the radiation bias.

The largest portion of the SW bias in the CANCM4 and LMDZOR models originates from missing deep cloud events (regime 7, dense hatching in Figures 11g and 11h). While the CANCM4 mainly simulates clear skies instead (white bars in Figure 11g), the LMDZOR mainly does so at the expense of high and mid/high clouds (yellow and pink bars in Figure 11h).

The partitioning of the cloud-related SW bias into observed-simulated regime pairs in the TAIIESM looks very different from the CAM5 (Figure 11i versus Figure 11a). Given that the deep cloud regime is severely underrepresented in the TAIIESM, rather than overestimated in the CAM5, there are hardly any negative contributions to the overall SW bias in the TAIIESM. Moreover, most of the cloud-related SW bias finds its origin in missing deep clouds (densely hatched bars) and simulating clear skies, mid/high, high, midlevel or even low clouds instead. Interestingly, there also does not appear to be a significant SW bias associated with correctly simulating the deep cloud regime in the TAIIESM, indicating the well-captured aCRE of this regime.

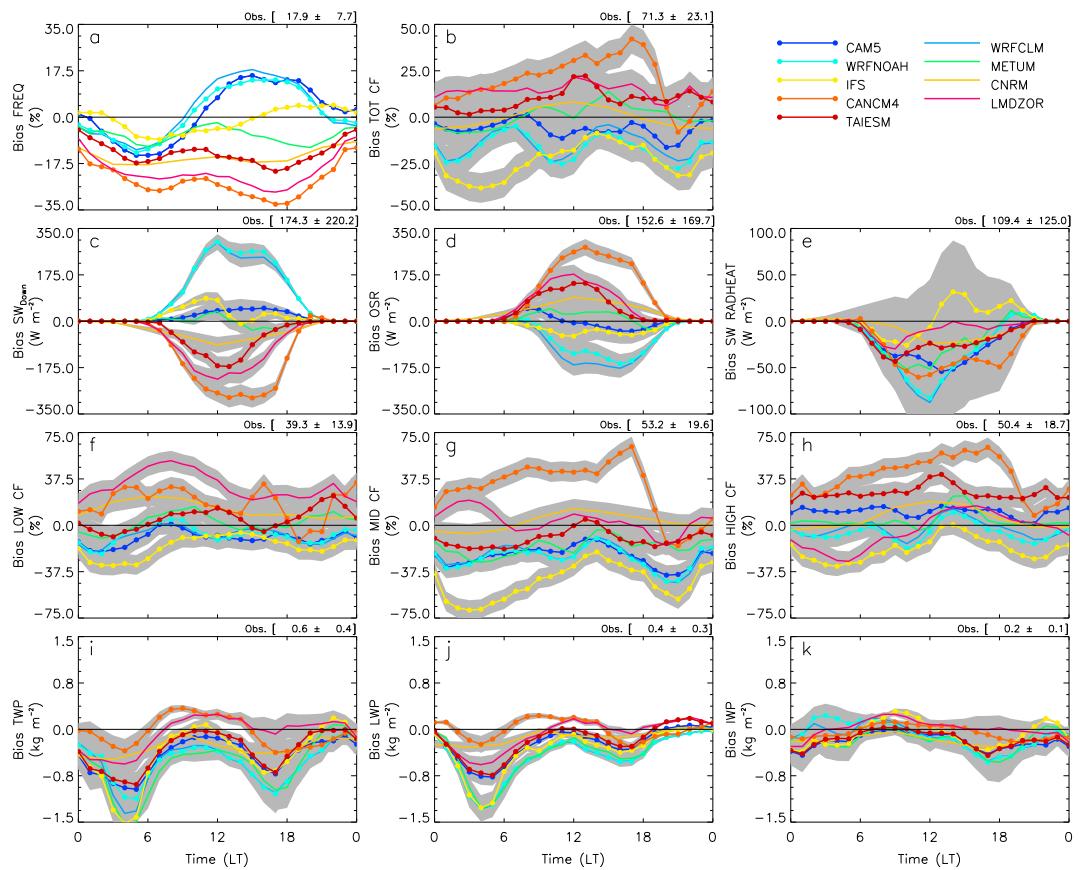


Figure 12. Absolute biases in the diurnal cycles of the cloud properties associated with regime 7 in all models. Shown are the biases in the frequency of the regime (a), the total cloud fraction based on a random maximum overlap (b), the surface downwelling shortwave (SW) (c), the top-of-atmosphere outgoing SW radiation (d), the column SW radiative heating as defined in equation (7) (e), low, middle, and high cloud fraction based on a random maximum overlap (f, g, and h), total, liquid, and ice water path (i, j, and k). Apart from the water paths, all data are for the full simulation period (April–August 2011) and all hindcast days ($T + 00$ to $T + 120$). The water paths are for the MC3E period only (18 April to 6 June 2011). Observations are for the Central Facility, as indicated in Table 2 and have been regressed to each model as outlined in the text. Gray shading indicates the observational uncertainty, based on the numbers in Table 2 and taking into account the rules for error propagation where necessary. The numbers to the upper right of each panel indicate the mean and standard deviation of the variable denoted in the panel below, averaged for the full simulation period and over all model grid lengths.

4.3.3. Cloud Property Evaluation of Deep Cloud Regime

While the scope of this paper is not to give a full diagnosis and explanation of the radiation errors in each of the models participating in the CAUSES project, a more detailed analysis of the cloud properties of the deep cloud regime is desirable, given its importance in the SW (and LW) radiative errors in all models.

The diurnal cycles in the absolute biases of a number of cloud and radiation properties associated with regime 7 are shown in Figure 12. Again, all observed properties have been regressed to their respective model grids before the bias was calculated. The numbers indicated above each panel denote the mean and standard deviation of the observed value of the variable shown in the panel, averaged over the diurnal cycle and all model grid lengths. It should be mentioned that the analysis period for the liquid, ice and total water path in this figure (Figures 12i–12k) is restricted to the MC3E IOP, given that these observations were only available during this period at the time of writing. The regime statistics for all models were very similar during this 6 week period compared to the full 5 month simulation period, however (not shown).

Consistent with the previous analysis, the frequency bias of the deep cloud regime varies widely from one model to another (Figure 12a). All models tend to underestimate the frequency of this regime in the early morning. This underestimation persists throughout the entire diurnal cycle in the TAIESM and CNRM, and even more in the CANCM4 and LMDZOR. The METUM continues to underestimate the frequency of regime

7 in the afternoon and evening, but captures its frequency reasonably well around noon. Distinct afternoon and evening overestimations of the deep cloud regime are present in the WRF and the CAM5, and IFS models, respectively.

Figure 12c shows the downwelling surface SW bias associated with regime 7. The bias is largest in the two WRF models, consistent with the shallow height of the red bars in Figures 10b and 10c. Magnitudes for the CAM5 and IFS are similar in the afternoon, experiencing slightly too large downwelling SW radiation for regime 7. The METUM has fairly well captured downwelling SW associated with regime 7 throughout the diurnal cycle, while the surface in the CNRM, TAIESM, CANCM4, and LMDZOR receives far too little SW when deep clouds are present.

From Figure 12d, the TOA OSR errors correlate reasonably well with the bias in the surface SW for most models (Figure 12c). This indicates that most of the surface radiation biases in the deep cloud regime can be explained by errors in the cloud top reflectivity. From the net TOA and surface (SFC) SW radiative fluxes, the SW column radiative heating can be estimated as follows:

$$\text{SWHEAT} = (\text{TOA SW}_D - \text{TOA SW}_U) - (\text{SFC SW}_D - \text{SFC SW}_U) \quad (7)$$

where the suffix *D* and *U* stand for downwelling and upwelling, respectively. Figure 12e provides the biases in the SW radiative heating for all models. While the uncertainties in the observed radiative heating are fairly large, most models, apart from the IFS in the afternoon, seem to underestimate the amount of SW heating during the daytime when deep clouds are present. For the WRF and CAM5, these biases in the radiative heating are consistent with the lack of cloud fraction in regime 7 (leading to subdued tropospheric SW absorption). For the other models (except the IFS), the lack of SW absorption seems contradictory with the lack of SW radiation reaching the surface during regime 7 (Figure 12c). This does suggest that the deep clouds in these models are very reflective, leaving less radiation to be absorbed both at the surface and within the troposphere.

Errors in the surface and TOA radiative properties of the deep cloud regime could have number of underlying causes. It could be that cloud fractions detrained from parameterized convective cores are too small, the water paths in these clouds could be underestimated, or effective radii could be poorly represented in the radiation schemes. Diurnal cycles of the total, low, mid, and high cloud fractions are given in Figures 12b and 12f–12h. From these panels, it is clear that the models allowing too much SW to reach the surface in the deep cloud regime (i.e., CAM5, WRF, and IFS), also have too small low-level and midlevel cloud fractions for this regime (Figures 12f and 12g). The large difference between the surface SW bias in the CAM5 and the WRF models is not due to very different biases in the cloud fractions between these models. Indeed, low, mid, and high cloud cover are fairly similar between these models in the afternoon. The IFS suffers from a persistent lack of cloud associated with regime 7 during the entire diurnal cycle. The cloud fractions in the CNRM and METUM are reasonably well captured. The large overestimations in cloudiness during the deep cloud regime in the LMDZOR, CANCM4, and the TAIESM are mainly due to large overestimations in the low, mid/high, and high cloud cover, respectively (Figures 12f–12h).

The total (TWP), liquid (LWP), and ice water path (IWP) biases associated with regime 7 are shown in Figures 12i to 12k. It should be mentioned here that the simulated water paths only include hydrometeor species that are seen by the radiation scheme, while the MICROBASE-derived water paths will include all condensed water in the tropospheric column. In many models, precipitating hydrometeors such as snow, rain, or graupel are ignored by the radiative transfer. While this complicates the evaluation of the water paths, our focus on explaining biases in the radiative transfer justifies the approach used here. Indeed, given that this analysis tries to uncover how cloud errors trickle down to radiation errors, it is the cloud properties seen by the radiation that are mainly of interest here. It should be kept in mind, however, that biases in the water paths listed here could either point to issues with the cloud properties or to the omission of some water species in the radiative transfer.

Again, the water path biases explain some, but not all of the biases seen in the surface and TOA radiation characteristics. The excessive afternoon downwelling SW at the surface and negative TOA OSR bias in some models (i.e., CAM5, WRF and IFS) correspond well with an important bias in TWP (Figure 12i), including both the LWP (Figure 12j) and the IWP (Figure 12k). Most other models also suffer from negative biases in their early morning and afternoon TWP. Water paths are better captured in the late morning in most models, and overestimated in the CANCM4 and the LMDZOR, which is mainly due to too large LWP (Figure 12j). Interestingly,

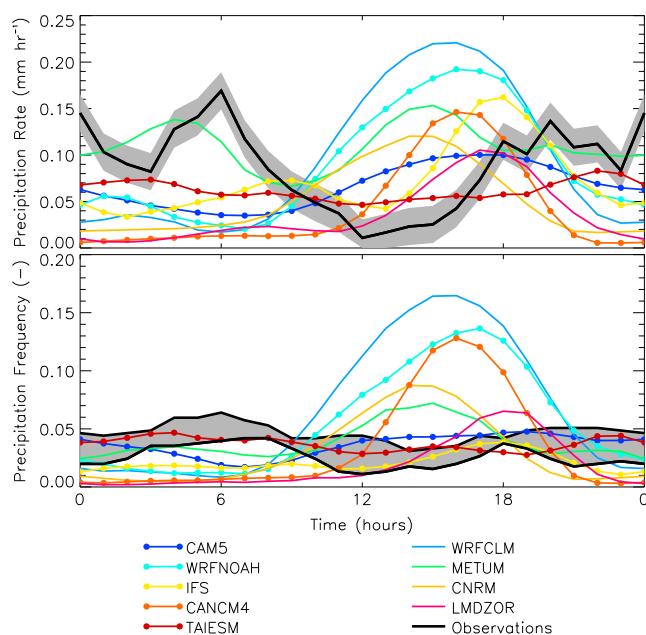


Figure 13. Average diurnal cycle of precipitation intensity (top) and frequency (bottom). Data in this figure are from the full simulation period (April to August 2011) and the five hindcast days. Observations are from the Atmospheric Radiation Measurement Best Estimate near the Central Facility as indicated in Table 2. Observed frequencies are calculated based both on a $0.25 \times 0.25^\circ$ and a $2.0 \times 2.0^\circ$ region. The gray-shaded area in the top panel denotes the observational uncertainty, while the gray shading in the bottom panel delineates the area covered by observed frequencies using the $0.25 \times 0.25^\circ$ and a $2.0 \times 2.0^\circ$ regions. The frequency includes precipitation rates above 0.5 mm h^{-1} only.

and as observed by the ARM-BE rainfall product, all averaged over the entire simulation period from April to August 2011. Note that the precipitation frequency is expected to change as the averaging area increases. Therefore, the observed frequencies are calculated both for $0.25 \times 0.25^\circ$ and $2.0 \times 2.0^\circ$ averaging domains, roughly covering the grid spacings of the models investigated. The gray-shaded area in Figure 13 hence delineates the minimum and maximum observed precipitation frequencies, considering varying averaging domains. Table 5 provides the total rainfall amounts as observed and simulated by all models. The observed precipitation rates and frequencies exhibit a distinctive peak in the evening and a second peak in the early morning, both coinciding with episodes of enhanced observed cloudiness (Figure 7a). Apart from the METUM, none of the models seem to capture the early morning peak in the precipitation rates (Figure 13, top). The afternoon peak in mean precipitation rates is largely overestimated in the WRF models, the IFS, the METUM, and the CANCM4 and its timing is too early in all models, although the timing is slightly better captured in the IFS and the LMDZOR (Figure 13, top). The latter is consistent with, for example, Rio et al. (2009), who showed that shallow cumulus in the LMDZOR played a key role in preconditioning the atmosphere for deep convection, leading to a delay in the onset of the rainfall maximum from midday to late afternoon. Despite missing the nocturnal precipitation peak, many models compensate for this error by having too much precipitation in the afternoon, so that their total precipitation amount is fairly well captured or even overestimated (Table 5). It is worth noting that the afternoon rainfall is too heavy even in the models that underestimate the frequency of the deep cloud regimes at that time, such as the METUM, the CNRM, the CANCM4,

the METUM has a fairly negative TWP bias throughout the day in the deep cloud regime, both due to too little liquid and ice. This suggests that the well-captured surface SW and TOA OSR in the METUM (Figures 12c and 12d) hides a compensating error between slightly too large cloud fractions and too small water paths (Figures 12b and 12i). Conversely, the IFS has slightly better captured TWP (Figure 12i), but its cloud fraction is severely underestimated (Figure 12b). The large difference between the WRF and CAM5 surface SW and TOA OSR biases seems to correspond with smaller LWP in the WRF models (Figure 12j), despite their identical physics packages. The IWP and cloud fractions are indeed very similar between these three models. It needs to be further investigated whether this difference is predominantly an effect of the different dynamics between the CAM5 and the WRF models, or an impact of their different resolutions.

4.3.4. Evaluation of Surface Precipitation

The various issues associated with the characteristics of the deep cloud regime all hint at important deficiencies in the triggering of deep convection in many of the models. For some models, such as the CANCM4, the dry IWV bias suggests that the large-scale dynamics might not be captured well enough to trigger convection. For the other models, it should be further understood whether the deep convective parameterization is not capable of triggering local diurnal-heating-induced convection, or whether the propagating systems over the Great Plains, initiated in the lee of the Rockies (Klein et al., 2006) are poorly represented. Apart from detrain cloud fraction and condensate, the convection parameterization also produces precipitation (and changes in the atmospheric heating). A closer look at the diurnal precipitation characteristics might hint at answers to some of the above questions.

Figure 13 shows the diurnal cycle of the mean hourly precipitation rates and the frequency of rainfall rates larger than 0.5 mm h^{-1} for the models

Table 5

Total Simulated and Observed Accumulated Surface Precipitation (mm) Over the Full Simulation Period (April–August 2011) for the Grid Point Nearest to the SGP Site

| OBS | CAM5 | WRFCLM | WRFNOAH | METUM | IFS | CNRM | CANCM4 | LMDZOR | TAIESM |
|-----|------|--------|---------|-------|-----|------|--------|--------|--------|
| 323 | 242 | 340 | 328 | 402 | 263 | 187 | 152 | 131 | 227 |

Note. Model data have been averaged over the five hindcast days.

and the LMDZOR (Figures 10d and 10f–10h). Conversely, while the CAM5 is one of the models with the largest frequency of the deep cloud regime (Figure 10a), it is among the models raining the least frequently and the least intensely around that time (Figures 13, top and 13, bottom). Interestingly, models that have a similar convection parameterization, such as CAM5, the WRF models and the CANCM4, behave fairly similarly in terms of the timing of the precipitation, but exhibit large differences in terms of the precipitation amounts. Indeed, these four models experience the largest rain rates in the afternoon and miss the nocturnal precipitation peak, but the afternoon rainfall in the WRF models exceeds the CANCM4 rainfall by far.

While far from conclusive, this analysis does suggest that most of the models trigger precipitation more than enough in the afternoon, but often lack the deep clouds that are observed with this precipitation. Nearly all models struggle to capture the nocturnal precipitation peak, which is likely related to propagating convective systems (Surcel et al., 2010; Trenberth et al., 2010), driven by cold pool dynamics and low-level jets. These mechanisms are known to be poorly represented in deep convective parameterizations, which likely explains their underrepresentation. The afternoon peak is more likely locally developed convection, associated with the diurnal heating of the surface (Zhang & Klein, 2010). Many models overestimate the intensity and have the timing of this peak too early. Possibly, this issue is related to the excess SW radiation reaching the surface in all models, heating the surface more than observed, and hence setting off convection too readily, but this needs further investigation. The fact that many models rain too heavily in the afternoon, while not detraining enough cloud from their convection schemes, suggest that precipitation efficiencies might be too high in these models over the SGP. It needs further study to confirm this hypothesis, but increasing the detrainment of cloud condensate in the convective environment and reducing the amount of condensate raining out to the surface could possibly improve many of the radiation issues highlighted. However, if not combined with an improved representation of the nocturnal precipitation peak, this might lead to dry biases in many models and errors in the surface energy balance. This indicates that solving one issue in the models, but not tackling a compensating error at the same time, could lead to deteriorations of the overall model performance.

5. Discussion and Conclusions

This paper constitutes the second part of a four-part series on the midlatitude continental warm bias seen in many NWP and climate models, with a particular focus on the U.S. SGP. In the framework of the CAUSES project, the first part of this series has documented the temporal and spatial characteristics of the warm bias in 11 GCMs and limited area models. Each of these models was initialized daily from ERA-Interim reanalysis between April and August 2011 and run freely in hindcast mode for 5 days, allowing for a detailed evaluation against the wealth of observations collected by the U.S. Department of Energy ARM Program near the Great Plains.

This second paper takes a closer look at the radiation biases that manifest themselves in the models taking part in the CAUSES project. More specifically, the two main objectives of this study are to perform (i) a systematic attribution study of radiation biases to errors in the surface properties, cloud characteristics, atmospheric water vapor, and aerosol and (ii) a more detailed analysis of the role of clouds and cloud regimes in the creation of surface radiation biases.

Figure 14 provides an overview table of the key findings for all models. The large positive net SW biases found in all models are predominantly caused by cloud and surface albedo deficiencies. From the first two columns in Figure 14, cloud issues tend to dominate over albedo issues in all models, apart from the CAM5. Most of the cloud contribution to the net SW radiation bias originates from issues with the deep cloud regime, defined as concurrently having cloud in the low, middle, and high troposphere. This regime makes up 40–80 % of the total cloud contribution in all models, apart from the CAM5 (third column in Figure 14). The reason why this regime contributes so much varies between the models however. Some models, like the WRF simulations or the IFS, produce this regime enough or even more frequently than observed ($FREQ_7$), but with very weak cloud radiative effects (CRE_7). Other models, including the METUM, CNRM, and TAIESM capture the CRE_7 , but underestimate the frequency of the deep cloud regime ($FREQ_7$). The CANCM4 and LMDZOR miss most of the deep cloud events ($FREQ_7$), but largely overdo their CRE_7 if they are present. The contribution of the deep cloud regime to radiation errors in the CAM5 is small due to a compensating error between too large $FREQ_7$ and too small CRE_7 (Figure 14).

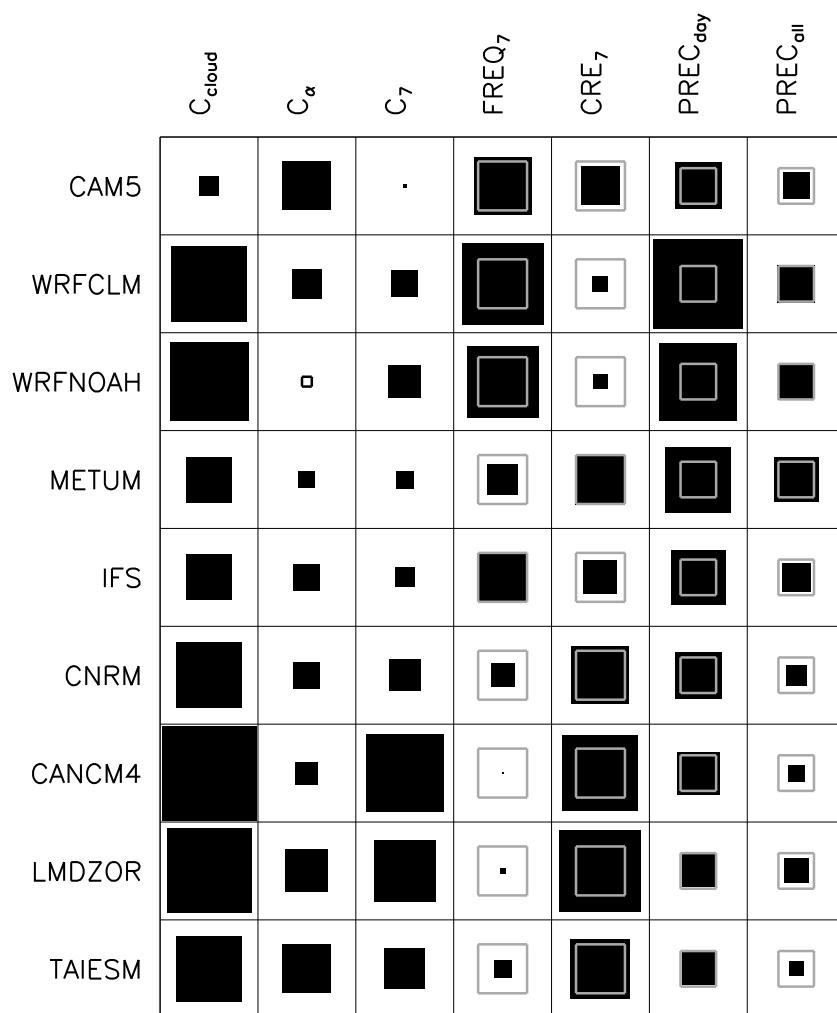


Figure 14. Qualitative summary table for all models. The first three columns provide the contribution to the net SW radiation bias from clouds in general (C_{cloud} ; first column), surface albedo (C_α ; second column), and the deep cloud regime (C_7 ; third column). The width and height of the squares scale with the magnitude of these contributions and are filled when the contribution is positive and empty when the contribution is negative. The last four columns represent the biases in a few key properties. Black-filled squares that are larger than the gray outline denote overestimations, while black squares that are smaller than the gray outline denote underestimations. Provided are the frequency bias of the deep cloud regime (FREQ₇; fourth column), the cloud radiative effect bias of the deep cloud regime (CRE₇; fifth column), the daytime precipitation bias (PREC_{day}; sixth column), and the overall precipitation bias (PREC_{all}; seventh column).

Interestingly, this does not imply that all models miss deep convection during the daytime. Indeed, a surprising finding of this study is that many models rain too much and too often during the afternoon (PREC_{day}, Figure 14), while at the same time having too much absorbed SW radiation at the surface. This might point to issues of excessive convective precipitation efficiencies in many of the models and too weak detrainment of convective cloud.

It should also be mentioned that the finding of sufficient daytime rainfall, but with weak cloud radiative impact, does not rule out that there are still underlying issues with the surface evaporative fraction. Indeed, while the daytime precipitation tends to be too large, many models miss the large observed nocturnal precipitation peak, leading to too small total rainfall amounts in all but three models (PREC_{all}, Figure 14) and possible impacts on the soil. Moreover, the excess SW radiation during the daytime in all models, be it due to surface albedo or cloud deficiencies, will also likely impact the surface energy balance. These issues will be further explored in the third part of this paper series (Ma et al., 2017).

As with most model evaluation studies, there are substantial uncertainties associated with many of the observations used here, in particular, associated with overlapping the ARSCL value-added product with radiometer measurements and satellite products (Qian et al., 2012). However, many of the discrepancies between the models and the observations and between the different models are much larger than these uncertainties and appear robust to slight changes to the assumptions in our methods.

While this study has been able to pinpoint some first-order mechanisms leading to severe biases in the surface radiation balance in all models participating in CAUSES, there are many outstanding issues yet to be resolved. In particular, it should be understood what is causing the weak deep cloud CRE in many models. For some models, this appears to be partly a precipitation efficiency issue (too heavy afternoon rain, but too little cloud), while for others the role of the large-scale environment might play a significant role as well. Moreover, in some models there might be room for improvement by including precipitating hydrometeors and more realistic particle properties in the radiative transfer. It should be teased out whether the excessive afternoon rainfall in many models is induced by this excessive surface heating or whether issues in the convection parameterizations are to blame for these excessive rainfall rates. A strategy to understand the role of convective triggering on the cloud and radiation biases should include the analysis of convection-permitting versions of those models presented here that can be run in that manner. Sensitivity studies focused on the assumptions in the convection parameterization, and the coupling of convection and other cloud physics are highly needed as well, leveraging the wealth of observations from the MC3E campaign. Furthermore, additional analysis of the propagation of convective systems initiated in the lee of the Rockies is needed, both with convection-permitting and nonconvection-permitting models.

Future studies, focused on specific problems in each of the models, should bring more clarity and provide possible ways forward in resolving some of the issues that this paper has highlighted. Although more work is required to actually fix the biases seen in these models, this paper has been able to help identify the root causes of the radiation errors in all models concerned, which helps focus model development and is a first step toward improving the representation of clouds and radiation over the midlatitudes in these models. In the future, the methodologies developed in this paper could be easily applied to improved versions of these models or other models that experience similar issues with clouds and radiation.

Appendix A: Surface Radiation Attribution

This appendix documents the radiation attribution analysis used in this paper, aimed at disentangling the net SW and LW radiation errors into contributions from the surface properties, clouds, integrated water vapor, and aerosol. The method described here is similar to the one used in Zhang et al. (2017).

The net SW radiation bias can be written as

$$\overline{\Delta SW_{\text{net}}} = \overline{(SW_{\text{MD}} - SW_{\text{OD}})} - (\overline{SW_{\text{MU}}} - \overline{SW_{\text{OU}}}) \quad (\text{A1})$$

where SW_{MD} and SW_{OD} are the modeled and observed downwelling SW radiation, respectively, and SW_{MU} and SW_{OU} are the modeled and observed upwelling SW radiation, respectively. The overbar indicates that these are averages over the entire simulation period.

Given the definition of the surface albedo $\alpha = \frac{SW_U}{SW_D}$, equation (A1) can be rewritten as

$$\overline{\Delta SW_{\text{net}}} = \overline{(1 - \alpha_M)(SW_{\text{MD}})} - \overline{(1 - \alpha_O)(SW_{\text{OD}})} \quad (\text{A2})$$

where α_M and α_O are the simulated and observed albedos, respectively. Note that we use the all-sky radiative fluxes for the albedo calculations. Trying to separate out the albedo effects in equation (A2) from the downwelling radiation effects, equation (A2) can be rearranged to give

$$\overline{\Delta SW_{\text{net}}} = \overline{(1 - \alpha_O)(SW_{\text{MD}} - SW_{\text{OD}})} + \overline{(\alpha_M - \alpha_O)(SW_{\text{OD}})} + \overline{(\alpha_O - \alpha_M)(SW_{\text{MD}} - SW_{\text{OD}})} \quad (\text{A3})$$

In this equation, the first term contains all factors contributing to the downwelling SW radiation bias, such as clouds, integrated water vapor (IWP), and aerosol, multiplied by the observed albedo (which is identical for all models in the intercomparison). The second term is the contribution to the ΔSW_{net} from errors in the surface albedo, and the third term is a covariance term between the contribution from surface albedo biases and the nonalbedo-related biases. It should be mentioned that the covariance term is generally very small ($< 1 \text{ W m}^{-2}$ in all models in this study) and will be ignored in the remainder of this paper.

Hence, the contribution of surface albedo errors to the overall net SW radiation bias can be written as

$$C_\alpha = \overline{(\alpha_M - \alpha_O)(SW_{OD})} \quad (A4)$$

Given the estimates of observed clear-sky SW radiation by Long and Ackerman (2000), the contribution of biases in the cloud properties to the overall net SW radiation biases can be written as

$$C_{\text{cloud}} = \overline{(1 - \alpha_O)(SW_{MD} - SW_{OD}) - (SWCS_{MD} - SWCS_{OD})} \quad (A5)$$

where $SWCS_{MD}$ and $SWCS_{OD}$ are the clear-sky SW downwelling radiation for the model and the observations, respectively. Note the multiplication with $(1 - \alpha_O)$ in this equation (as inherited from equation (A3)).

The remaining component now will be the clear-sky SW downwelling radiation bias:

$$C_{\text{clear}} = \overline{(1 - \alpha_O)(SWCS_{MD} - SWCS_{OD})} \quad (A6)$$

This component will contain contributions from biases in the vertically I WV, aerosol, and inaccuracies in the clear-sky radiative transfer. Focusing on the I WV first, it is instructive to compare the bias in the clear-sky radiation (equation (A6)) with the concurrent bias in I WV. For every output time (every hour in most models), the clear-sky net SW radiation bias can be assigned to one of two composites. A first composite contains all analysis times when the I WV is biased (negatively or positively) and a second composite encompasses analysis times with unbiased I WV. The threshold between the two composites has been set here to a bias of -5% or $+5\%$ of the observed I WV value, respectively. Although this is a fairly arbitrary threshold, variations between 1 and 10% did not change our conclusions.

Assuming that the bias in I WV is independent from any remaining biases associated with the clear-sky SW radiation (e.g., aerosol), the contribution from a biased I WV can be estimated as follows:

$$C_{\text{iwv}} = (C_{\text{clear,biasiwv}} - C_{\text{clear,nobiasiwv}}) \text{freq}_{\text{biasiwv}} \quad (A7)$$

where $C_{\text{clear,biasiwv}}$ and $C_{\text{clear,nobiasiwv}}$ are the clear-sky contributions for the biased and unbiased I WV composites, respectively, calculated following equation (A6). The parameter $\text{freq}_{\text{biasiwv}}$ is the frequency (as a fraction) of the biased composite.

All remaining potential contributions to the overall net SW radiation bias are contained in the residual contribution:

$$C_{\text{res}} = C_{\text{clear,nobiasiwv}} \quad (A8)$$

It should be mentioned that this last component could be further disentangled if aerosol optical depths had been available for the models, which unfortunately is not the case. Hence, this last residual component contains any contributions from aerosol, as well as potential inaccuracies in the radiation codes (e.g., Pincus et al., 2015) and inaccuracies from misassignments of clear and cloud skies in the observations.

The total $\overline{\Delta SW_{\text{net}}}$ can now be written as follows:

$$\overline{\Delta SW_{\text{net}}} = C_\alpha + C_{\text{cloud}} + C_{\text{iwv}} + C_{\text{res}} \quad (A9)$$

where the individual contributions are described in equations (A4), (A5), (A7), and (A8).

A similar approach can be followed for the net LW radiation bias, so that

$$\overline{\Delta LW_{\text{net}}} = C_{\text{surf}} + C_{\text{cloud}} + C_{\text{iwv}} + C_{\text{res}} \quad (A10)$$

where the C_{surf} is

$$C_{\text{surf}} = \overline{(LW_{MU} - LW_{OU})} \quad (A11)$$

The terms in equation (A10) no longer contain the albedo factor as for the SW contributions and the observed clear-sky downwelling longwave ($LWCS_{OD}$) is estimated following Long and Turner (2008).

Appendix B: Attribution of Net LW Bias to I WV

In section 4.2.2, it was found that some models experience a positive contribution from I WV to the net LW bias, despite these models being drier than observed. To understand this apparently contradicting behavior

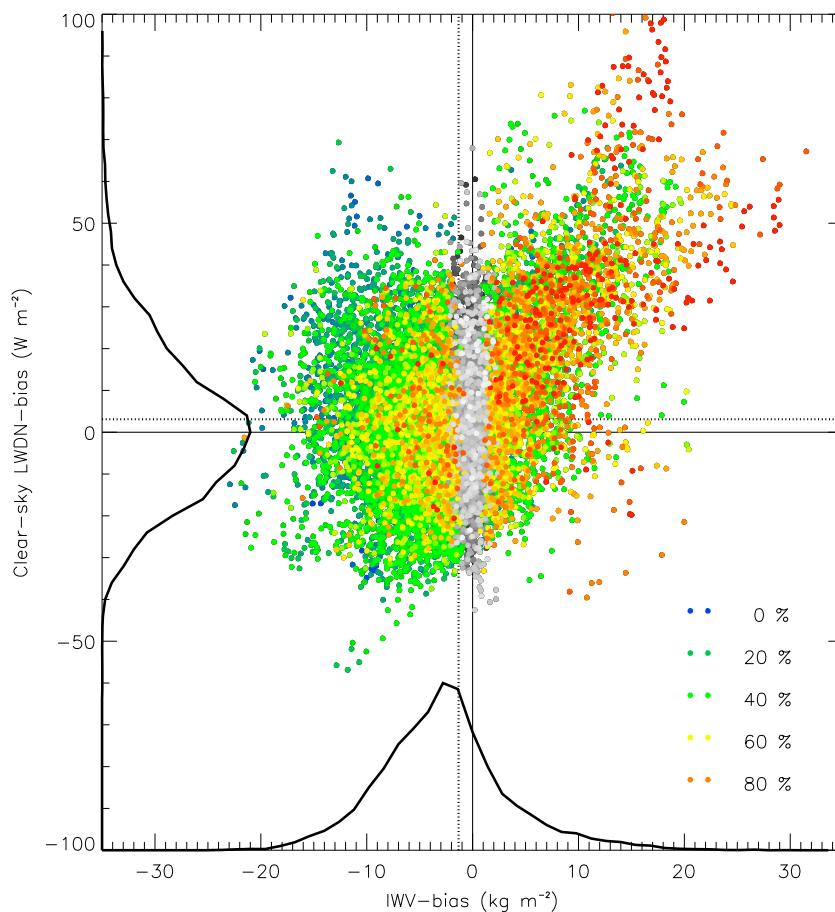


Figure B1. Scatter plot of the bias in integrated water vapor (IWV) versus the bias in clear-sky downwelling longwave radiation for the CAM5 against the Atmospheric Radiation Measurement Best Estimate data (Tang & Xie, 2015). Data points are from the entire simulation period (April–August 2011, with hourly intervals) and the five hindcast days ($T + 00$ to $T + 120$). Colors indicate the simulated near-surface (2 m) relative humidity for each respective data point. Only the data points with an IWV bias of more than $\pm 5\%$ have been colored. The shape of the distribution and the average of the IWV bias and the clear-sky downwelling longwave radiation bias are indicated by the black thick solid lines and the thin vertical and horizontal dotted lines, respectively.

in these models, it is instructive to have a closer look at the scatter plots shown in Figure B1. This figure shows the correlation between the IWV bias and the clear-sky downwelling LW radiation bias throughout the entire simulation period (April–August 2011) and the full 5 day hindcast ($T + 00$ to $T + 120$) for the CAM5. The colors in this figure represent the simulated near-surface relative humidity (RH) associated with each data point. From this figure, it appears that when the IWV is biased dry (data points to the left of the zero-IWV-bias line), the clear-sky downwelling LW radiation bias is not much reduced compared to when the IWV is well captured (gray dots). However, when the IWV is biased wet (data points to the right of the zero-IWV-bias line), the clear-sky downwelling LW radiation bias sharply increases. Moreover, these latter data points also seem to be associated with systematically larger relative humidity (as shown by the red colors for these data points in Figure B1). This finding is relevant since hygroscopic particles start to accumulate liquid water and a haze starts to form as soon as the relative humidity reaches about 75% (also bear in mind that 60% of the downwelling LW radiative impact of the atmospheric WV originates from the bottom 100 m of the atmosphere in clear-sky conditions McFarlane et al., 2013). Although still considered a clear-sky environment, this haze has nonnegligible effects on the downwelling LW radiation (Long & Turner, 2008). A dry bias in IWV in the CAM5 is typically associated with near-surface RH values closer to 50–60% and under these circumstances, slight differences in the water vapor content of the atmosphere will matter less than when the near-surface RH approaches 70–80%.

Hence, there appears to be a very nonlinear effect of the IWV-bias on the overall net LW bias. Despite the too dry atmospheres on average in the CAM5 (and many other GCMs considered here), the occasional wet IWV bias has a far larger impact on the downwelling LW, given its concurrent high near-surface RH.

Acknowledgments

This research was supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research and used data from the Atmospheric Radiation Measurement Climate Research Facility. The work of Morcrette and Van Weverberg was supported by DOE grant DE-SC001422. The Pacific Northwest National Laboratory (PNNL) is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830. A portion of the computational research was performed using PNNL Institutional Computing at Pacific Northwest National Laboratory and at the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-05CH11231. The work at Lawrence Livermore National Laboratory (LLNL) was supported by the Regional and Global Climate Modeling and Atmospheric System Research programs (H. Y. M., S. K., C. Z., and S. X.) as well as the Atmospheric Radiation Measurement (ARM) program (S. X., Q. T., and C. Z.) of the Biological and Environmental Research Division in the Office of Sciences of the U.S. Department of Energy (DOE) under the auspices of the U.S. Department of Energy by LLNL under contract DE-AC52-07NA27344. 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]. Roehrig acknowledges the support from the DEPHY2 project, funded by the French national program LEFE/INSU. The contribution from ECMWF to this study was supported by DOE grant DE-SC0005259. The simulations will be available online through the NERSC Science Gateways (details provided on <https://www.nerc.gov/users/sciencegateways/>). Due to the large volume of data sets and limited disk space, data will be shared online upon request. Observational data were obtained from the Atmospheric Radiation Measurement (ARM) Climate Research Facility, a U.S. Department of Energy Office of Science user facility sponsored by the Office of Biological and Environmental Research.

References

- Ahlgrimm, M., & Forbes, R. (2012). The impact of low clouds on surface shortwave radiation in the ECMWF model. *Monthly Weather Review*, 140, 3783–3794.
- Bechtold, P., Semane, N., Lopez, P., Chaboureau, J.-P., Beljaars, A., & Bormann, N. (2014). Representing equilibrium and non-equilibrium convection in large-scale models. *Journal of the Atmospheric Sciences*, 71, 734–753.
- Bougeault, P. (1985). A simple parameterization of the large-scale effects of cumulus convection. *Monthly Weather Review*, 113, 2108–2121.
- Bouniol, D., Protat, A., Delanoe, J., Pelon, J., Pirou, J.-M., Bouyssel, F., et al. (2010). Using continuous ground-based radar and lidar measurements for evaluating the representation of clouds in four operational models. *Journal of Applied Meteorology and Climatology*, 49, 1971–1991.
- Brisson, E., Van Weverberg, K., Demuzere, M., Devis, A., Saeed, S., Stengel, M., et al. (2016). How well can a convection-permitting climate model reproduce decadal statistics of precipitation, temperature and cloud characteristics? *Climate Dynamics*, 9, 3043–3061.
- Campos, E., & Sisterson, D. L. (2015). A unified approach for reporting ARM measurement uncertainties technical report (Tech. Rep. DOE/SC-ARM/TR-170). US Department of Energy. Argonne, IL: Argonne National Laboratory. Retrieved from <https://www.arm.gov/publications/techreports/doe-sc-arm-tr-170.pdf> (accessed 21 August 2017)
- Cheinet, S., Beljaars, A., Morcrette, J.-J., & Viterbo, P. (2005). Assessing physical processes in the ECMWF model forecasts using the ARM SGP observations (ECMWF ARM Rep. 1). Reading, UK: European Centre for Medium-range Weather Forecasts.
- Cheruy, F., Dufresne, J.-L., Hourdin, F., & Dutcher, 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>
- Clark, A. J., Gallus, W., & Chen, T.-C. (2007). Comparison of the diurnal precipitation cycle in convection-resolving and non-convection-resolving mesoscale models. *Monthly Weather Review*, 135, 3456–3473.
- Clothiaux, E. E., Ackerman, T. P., Mace, G. G., Moran, K. P., Marchand, R. T., Miller, M. A., et al. (2000). Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites. *Journal of Applied Meteorology*, 39, 645–665.
- 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, 553–597.
- Dunn, M., Johnson, K. L., & Jensen, M. P. (2011). The microbase value-added product: A baseline retrieval of cloud microphysical properties (Tech. Rep. DOE/SC-ARM/TR-095). US Department of Energy. Upton, New York: Brookhaven National Laboratory. Retrieved from <http://www.arm.gov/publications/techreports/doe-sc-arm-tr-095.pdf> (accessed 10 July 2015)
- European Centre for Medium-Range Weather Forecasts (ECMWF) (2015). IFS Documentation-Cy41r1. Operational implementation (Tech. Rep.). Reading, UK: European Centre for Medium-Range Weather Forecasts. Retrieved from <http://www.ecmwf.int/en/forecasts/documentation-an-support/changes-ecmwf-model/IFS-documentation> (last access: 19 January 2017)
- 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, 689–719.
- Emanuel, K. A. (1993). A cumulus representation based on the episodic mixing model: The importance of mixing and microphysics in predicting humidity. *AMS Meteorological Monographs*, 24, 185–192.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., et al. (2016). Overview of the Coupled Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9, 1937–1958.
- Fouquart, Y., & Bonnel, B. (1980). Computations of solar heating of the Earth's atmosphere: A new parameterization. *Beitrage zur Physik der Atmosphare*, 53, 35–62.
- Gaustad, K. L., & Riihimaki, L. (2016). Atmospheric Radiation Measurement (ARM) Climate Research Facility. 1996, updated hourly. MWR Retrievals (MWRRET1LILJCLOU). 2011-03-31 to 2011-09-01, 36.605 N 97.485 W: Southern Great Plains (SGP) Central Facility, Lamont, OK (C1), Atmospheric Radiation Measurement (ARM) Climate Research Facility Data Archive, Oak Ridge, Tennessee, USA. Data set accessed 2016-08-19 at <https://doi.org/10.5439/1027369>
- Gaustad, K. L., Riihimaki, L., & Long, C. (2016). Atmospheric Radiation Measurement (ARM) Climate Research Facility. 1994, updated hourly. Radiative Flux Analysis (RADFLUX1LONG). 2011-03-30 to 2011-09-01, 36.605 N 97.485 W: Southern Great Plains (SGP) Central Facility, Lamont, OK (C1), Atmospheric Radiation Measurement (ARM) Climate Research Facility Data Archive, Oak Ridge, Tennessee, USA. Data set accessed 2016-09-14 at <https://doi.org/10.5439/1179822>
- Geerts, B., Parsons, D., Ziegler, C. L., Weckwerth, T. M., Turner, D. D., Wurman, J., et al. (2017). The 2015 Plains Elevated Convection At Night (PECAN) field project. *Bulletin of the American Meteorological Society*, 98, 767–786.
- Gregory, D., & Rowntree, P. R. (1990). A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Monthly Weather Review*, 118, 1483–1506.
- Gruetzen, V., Quaas, J., Morcrette, C. J., & Ament, F. (2013). Evaluating statistical cloud schemes: What can we gain from ground-based remote sensing? *Journal of Geophysical Research: Atmospheres*, 118, 10,507–10,517. <https://doi.org/10.1002/jgrd.50813>
- Gustafson, W. I. Jr., Ma, P.-L., & Singh, B. (2015). Precipitation characteristics of CAM5 physics at mesoscale resolution during MC3E and the impact of convective timescale choice. *Journal of Advances in Modeling Earth Systems*, 6, 1271–1287. <https://doi.org/10.1002/2014MS000334>
- 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, 2193–2222.
- Iacono, M. J., Delamere, J. S., Mlawer, J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, 113, D13103. <https://doi.org/10.1029/2008JD009944>
- Illingworth, A. J., Hogan, R. J., O'Connor, E. J., Bouniol, D., Brooks, M. E., Delanoe, J., et al. (2007). CloudNet—Continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bulletin of the American Meteorological Society*, 88, 883–898.
- Jensen, M. P., Toto, T., Troyan, D., Ciesielski, P. E., Holdridge, D., Kyrouac, J., et al. (2015). The midlatitude continental convective clouds experiment (MC3E) sounding network: Operations, processing and analysis. *Atmospheric Measurement Techniques*, 8, 421–434.

- Johnson, K., Toto, T., & Giangrande, S. (2016). Atmospheric Radiation Measurement (ARM) Climate Research Facility. 2011, updated hourly. Active Remote Sensing of Clouds (ARSCL) product using Ka-band ARM Zenith Radars (ARSLKAZRBND1KOLLIAS). 2011-03-31 to 2011-08-31, 36.605 N 97.485 W: Southern Great Plains (SGP) Central Facility, Lamont, OK (C1), Atmospheric Radiation Measurement (ARM) Climate Research Facility Data Archive, Oak Ridge, Tennessee, USA. Data set accessed 2017-01-20 at <https://doi.org/10.1002/2017JD027188>
- Khaiyer, M. M., Nordeen, M. L., Palikonda, R., Yi, Y., & Rutan, D. A. (2011). Improved TOA broadband shortwave and longwave fluxes derived for SGF using GOES-11, Poster presented during the Second Science Team Meeting of the Atmospheric System Research Program. San Antonio, Texas. Retrieved from <https://asr.science.energy.gov/meetings/stm/posters/posterpdf/2011/P000311.pdf>
- 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>
- Koster, R. D., Dirmeyer, P. A., Guo, Z. C., Bonan, G., Chan, E., Cox, P., et al. (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, 305, 1138–1140.
- Li, J., & Barker, H. W. (2005). A radiation algorithm with correlated k-distribution. Part I: Local thermal equilibrium. *Journal of the Atmospheric Sciences*, 62, 286309
- Li, J., Curry, C. L., Sun, Z., & Zhang, F. (2010). Overlap of solar and infrared spectra and the shortwave radiative effect of methane. *Journal of the Atmospheric Sciences*, 67, 2372–2389.
- 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, 15,609–15,626.
- 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., Klein, S. A., Xie, S., Zhang, C., Tang, S., Tang, Q., et al. (2017). 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, 2888–2909. <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, 1781–1798.
- Ma, P.-L., Rasch, P. J., Fast, J. D., Easter, R. C., Gustafson, W. I., 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, 755–778.
- McFarlane, S. A., Long, C. N., & Flaherty, J. (2013). A climatology of surface cloud radiative effects at the ARM Tropical Western Pacific sites. *Journal of Applied Meteorology and Climatology*, 52, 996–1013.
- McManus, G. D., Shafer, M. A., & Johnson, H. (2011). Oklahoma monthly climate summary August 2011. Oklahoma climatological survey, University of Oklahoma
- Minnis, P., & Smith, W. L. (1998). Cloud and radiative fields derived from GOES-8 during SUCCESS and the ARM-UAV spring 1996 flight series. *Geophysical Research Letters*, 25, 1113–1116.
- Minnis, P., Sun-Mack, S., Young, D. F., Heck, P. W., Garber, D. P., Chen, Y., et al. (2011). CERES Edition-2 cloud property retrievals using TRMM VIRS and Terra and Aqua MODIS data, Part I: Algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 49, 4374–4400.
- Morcrette, C. J., Klein, S. A., Ma, H.-Y., Van Weverberg, K., Zhang, C., Petch, J., et al. (2017). CAUSES: Near-surface temperature errors in NWP and climate model 5-day hindcasts near the Southern Great Plains. *Journal of Geophysical Research*, 123. <https://doi.org/10.1002/2017JD027199>
- Morcrette, C. J., O'Connor, E. J., & Petch, J. C. (2012). Evaluation of two cloud parametrization schemes using ARM and Cloud-Net observations. *Quarterly Journal of the Royal Meteorological Society*, 138, 964–979.
- Morcrette, J.-J. (1991). Radiation and cloud radiative properties in the European Centre for Medium Range Weather Forecasts forecasting system. *Journal of Geophysical Research*, 96, 9121–9132.
- Morcrette, J.-J., Barker, H. W., Cole, J. N. S., Iacono, M. J., & Pincus, R. (2008). Impact of a new radiation package, McRad, in the ECMWF integrated forecasting system. *Monthly Weather Review*, 136, 4773–4798.
- 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., Gettelman, A., Park, S., Chen, C.-C., Lauritzen, P. H., Williamson, D. L., et al. (2012). Description of the NCAR community atmosphere model (CAM 5.0) (*Technical Note NCAR-TN-486+STR*). Boulder, CO: NCAR
- Pincus, R., Mlawer, E. J., Oropoulos, 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>
- Rio, C., Hourdin, F., Grandpeix, J.-Y., & Lafore, J.-P. (2009). Shifting the diurnal cycle of parameterized deep convection over land. *Geophysical Research Letters*, 36, L07809. <https://doi.org/10.1029/2008GL036779>
- Ritsche, M. T. (2008). Surface Meteorological Observations System (SMOS) handbook Tech. Rep. DOE/SC-ARM/TR-031). US Department of Energy. Argonne, IL: Argonne National Laboratory. Retrieved from <https://www.arm.gov/publications/techreports/handbooks/smshandbook.pdf> (accessed 25 August 2017)
- Stoffel, T. (2005). Solar Infrared Radiation Station (SIRS) handbook (Tech. Rep. DOE/SC-ARM/TR-025). US Department of Energy. Golden, CO: National Renewable Energy Laboratory. Retrieved from https://www.arm.gov/publications/techreports/handbooks/sirs_handbook.pdf?id=68 (accessed 21 August 2017)
- Surcel, M., Berenguer, M., & Zawadzki, I. (2010). The diurnal cycle of precipitation from continental radar mosaics and numerical weather prediction models. Part I: Methodology and seasonal comparison. *Monthly Weather Review*, 138, 3084–3106.
- Tang, Q., & Xie, S. (2015). 2D gridded surface data value-added product (Tech. Rep. DOE/SC-ARM/TR-152). US Department of Energy. Livermore, CA: Lawrence Livermore National Laboratory. Retrieved from <http://www.arm.gov/publications/techreports/doe-sc-arm-tr-152.pdf> (accessed 19 April 2017)
- Trenberth, K. E., Dai, A., Rasmusson, R. M., & Parsons, D. B. (2010). The changing character of precipitation. *Bulletin of the American Meteorological Society*, 91, 1205–1215.
- Troyan, D., Toto, T., & Giangrande, S. (2016). Atmospheric Radiation Measurement (ARM) Climate Research Facility. 1996, updated hourly. Merged Sounding (MERGESONDE1MACE). 2011-03-31 to 2011-08-31, 36.605 N 97.485 W: Southern Great Plains (SGP) Central Facility, Lamont, OK (C1), Atmospheric Radiation Measurement (ARM) Climate Research Facility Data Archive, Oak Ridge, Tennessee, USA. Data set accessed 2017-01-23 at <https://doi.org/10.5439/1034922>
- 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, 1785–1810.
- 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, 3190–3206.

- Volodire, A., Sanchez-Gomez, E., Salas y Melia, D., Decharme, B., Cassou, C., Senesi, S., et al. (2013). The CNRM-CM5.1 global climate model: Description and basic evaluation. *Climate Dynamics*, 40, 2091–2121.
- von Salzen, K., Scinocca, J. F., McFarlane, N. A., Li, J., Cole, J. N. S., Plummer, D., et al. (2013). The Canadian Fourth Generation Atmospheric Global Climate Model (canAM4). Part I: Representation of physical processes. *Atmosphere-Ocean*, 51, 104–125.
- Walters, D. N., Williams, K. D., Boulle, 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, 361–386.
- Walther, A., Jeong, J.-H., Nikulin, G., Jones, C., & Chen, D. (2013). Evaluation of the warm season diurnal cycle of precipitation over Sweden simulated by the Rossby Centre regional climate model RCA3. *Atmospheric Research*, 119, 131–139.
- Wang, Y.-C., Pan, H.-L., & Hsu, H.-H. (2014). 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, 10,684–10,702. <https://doi.org/10.1002/2015JD023337>
- Williams, K. D., Bodas-Salcedo, A., Deque, M., Fermeipin, S., Medeiros, B., Wanatabe, M., et al. (2013). The transpose-AMIP II experiment and its application to the understanding of the Southern Ocean cloud biases in climate models. *Journal of Climate*, 26, 3258–3274.
- Wu, W., Liu, Y., Jensen, M. P., Toto, T., Foster, M. J., & Long, C. N. (2014). A comparison of multiscale variations of decade-long cloud fractions from six different platforms over the Southern Great Plains in the United States. *Journal of Geophysical Research: Atmospheres*, 119, 3438–3459. <https://doi.org/10.1002/2013JD019813>
- Xie, S., Ma, H.-Y., Boyle, J., 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, 7937–7955.
- Xie, S., McCoy, R., Kein, S., & Cederwall, R. (2010). Clouds and more: ARM climate modeling best estimate data: A new data product for climate studies. *Bulletin of the American Meteorological Society*, 91, 13–20.
- Yang, F., Pan, H. L., Krueger, S. K., Moorthi, S., & Lord, S. J. (2006). Evaluation of the NCEP global forecast system at the ARM SGP site. *Monthly Weather Review*, 134, 3668–3690.
- Yang, G. Y., & Slingo, J. (2001). The diurnal cycle in the Tropics. *Monthly Weather Review*, 129, 784–801.
- Zhang, C., Xie, S., Klein, S. A., Ma, H.-Y., Tang, S., Van Weverberg, K., et al. (2017). 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>
- Zhang, G. J., & McFarlane, N. A. (1995). Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centre general circulation model. *Atmosphere-Ocean*, 33, 407–446.
- Zhang, M. H., Lin, J. L., Cederwall, R., Yio, J., & Xie, S. C. (2001). Objective analysis of ARM IOP data: Method, features and sensitivity. *Monthly Weather Review*, 129, 295–311.
- Zhang, Y. Y., & Klein, S. A. (2010). Mechanisms affecting the transition from shallow to deep convection over land: Inferences from observations of the diurnal cycle collected at the ARM Southern Great Plains site. *Journal of the Atmospheric Sciences*, 67, 2943–2959.
- Zhao, C., Xie, S., Chen, X., Jensen, M. P., & Dunn, M. (2014). Quantifying uncertainties of cloud microphysical property retrievals with a perturbation method. *Journal of Geophysical Research: Atmospheres*, 119, 5375–5385. <https://doi.org/10.1002/2013JD021112>