# A Bimodal Diagnostic Cloud Fraction Parameterization. Part II: Evaluation and Resolution Sensitivity

Kwinten Van Weverberg*

*Met Office, Exeter, United Kingdom*

Cyril J. Morcrette

*Met Office, Exeter, United Kingdom*

Ian Boutle

*Met Office, Exeter, United Kingdom*

**Corresponding author address:* Atmospheric Processes and Parametrizations, Met Office, FitzRoy Road, Exeter, United Kingdom

E-mail: kwinten.vanweverberg@metoffice.gov.uk

# ABSTRACT

A wide range of approaches exists to account for subgrid cloud variability in regional simulations of the atmosphere. This paper addresses the following questions: (1) Is there still benefit in representing subgrid variability of cloud in convection-permitting simulations? (2) What is the sensitivity to the cloud fraction parameterization complexity? (3) Are current cloud fraction parameterizations scale-aware across convection-permitting resolutions? These questions are addressed for regional simulations of a six-week observation campaign in the US Southern Great Plains. Particular attention is given to a new diagnostic cloud fraction scheme with a bimodal subgrid saturation-departure PDF, described in Part I. The model evaluation is performed using ground-based remote sensing synergies, satellite-based retrievals and surface observations. It is shown that not using a cloud-fraction parameterization results in underestimated cloud frequency and water content, even for stratocumulus. The use of a cloud-fraction parameterization does not guarantee improved cloud property simulations, however. Diagnostic and prognostic cloud schemes with a symmetric subgrid saturation-departure PDF underestimate cloud fraction and cloud optical thickness, and hence overestimate surface shortwave radiation. These schemes require empirical bias-correction techniques to improve the cloud cover. The new cloud-fraction parameterization, introduced in Part I, improves cloud cover, liquid water content, cloud base height, optical thickness and surface radiation compared to schemes reliant on a symmetric PDF. Furthermore, cloud parameterizations using turbulence-based, rather than prescribed constant subgrid variances, are shown to be more scale-aware across convection-permitting resolutions.

## 1. Introduction

Over the past decades, different approaches have been proposed to represent subgrid cloud variability in numerical weather prediction (NWP) models. This occurred against a backdrop of ever increasing computing power and decreasing grid spacing, providing both opportunities and challenges for model development. On the one hand, the increased computing power allows for a more physical representation of processes that induce subgrid variability. On the other hand, the continuous increase in model resolution necessitates the development of scale-aware parameterizations that require less resolution-specific tuning.

The inception of subgrid cloud fraction (*CF*) parameterizations dates back to the 1970s with simple schemes, relating the *CF* to the grid-box mean relative humidity (*Sundqvist* 1978; *Del Genio* 1996). Later adjustments involved the assumption of an underlying probability density function (PDF) of the saturation departure (SD) (*Smith* 1990; *LeTreut and Li* 1991; *Lewellen and Yoh* 1993; *Lohmann et al.* 1999). The *CF* in these schemes is diagnosed by integration over the saturated part of the PDF, introducing sensitivity to the higher-order PDF moments. It is common practice to keep assumptions about the PDF moments fairly simple. The PDF variance is often kept constant by specifying a critical relative humidity (RHcrit) at which cloud just starts to form in a sub-saturated grid box, while higher-order moments (e.g. skewness) are ignored. A few notable exceptions to this simplified approach are *Ricard and Royer* (1993) and *Lohmann et al.* (1999), who obtain the PDF-variance from the turbulence parameterization, while still ignoring the PDF skewness.

Later advances in *CF* parameterization development abandoned the diagnostic approach and introduced prognostic equations for the time evolution of either the *CF* itself (*Tiedtke* 1993; *Wilson et al.* 2008), or the underlying uni-variate (saturation-departure or humidity) PDF moments

(*Tompkins* 2002). A further step-change in cloud scheme development involves the integration of cloud, turbulence and convection in so-called assumed-PDF schemes (*Golaz et al.* 2002; *Gerard* 2007; *Larson et al.* 2012). These parameterizations provide a fully self-consistent set of prognostic equations for all higher-order moments of multi-variate subgrid PDFs. These advanced schemes are mathematically elegant, but still require a host of closure assumptions, for instance in the specification of sometimes hard-to-observe source and sink terms for higher-order moments. Hence, while the time scales of certain cloud-related processes warrant a prognostic approach, these schemes have considerable added complexity, making them increasingly un-tractable for inevitable tuning in an operational context. Furthermore, the question remains whether the greater complexity is justified in an operational environment with fierce competition for computing power between more advanced physics and higher resolution.

This paper presents an in-depth evaluation of six *CF* parameterization approaches at convection-permitting scales, with particular attention for the new diagnostic bimodal *CF* scheme described in *Van Weverberg et al.* (2020), hereafter Part I. This new scheme allows for bimodal and skewed subgrid distributions within the entrainment zone. Hindcasts with a near-operational regional configuration of the U.K. Met Office Unified Model (UM) are performed for a 6-week observation campaign, the Midlatitude Continental Convective Clouds Experiment (MC3E; *Jensen et al.* (2016)) over the U.S. Southern Great Plains (SGP). These hindcasts are thoroughly evaluated using high-quality ground-based remote-sensing synergies, satellite retrievals and surface measurements.

More specifically, this paper aims to address the following questions: (1) Is there still benefit in the use of *CF* parameterizations at convection-permitting scales? (2) What is the sensitivity of convection-permitting simulations to a number *CF* parameterization approaches? How does the performance of the bimodal cloud scheme compare to conventional approaches with various complexity? (3) Are *CF* schemes tested here scale aware across a range of convection-permitting resolutions?

A description of the observations used in addition to those introduced in Part I is given in Section 2. An overview of the model configurations is provided in Section 3. The results section consists of three subsections. First, a single case study evaluation is shown, focusing on the detailed time-height evolution of the cloud properties. Second, a more general evaluation of the entire MC3E campaign is performed, focusing on vertical profiles and diurnal cycles of cloud properties. Third, scale awareness of the observed and simulated cloud properties is investigated. The main conclusions are summarized in Section 5.

## 2. Observations

### *a. Vertical Cloud Locations and Water Content*

The foremost source of information about cloud locations and water content in this study are the Active Remote Sensing of CLouds (ARSCL) and the Microbase ARM synergistic data products respectively. A detailed description of these products is provided in Part I. It is re-iterated that, based on observed wind speeds (*Toto and Jensen* 2016) and given the model grid length, a moving time window is applied on these cloud locations to establish the 'observed' cloud fraction. The same time window is used to establish the grid-box mean liquid water content.

Given the observational uncertainty, most of this paper focuses on non-precipitating, liquid clouds. We refer to Part I for details about the precipitation screening and the uncertainties associated with ARSCL and Microbase.

*b. Liquid Water Path*

Observations of the liquid water path (*LWP*) are obtained from the microwave radiometer (MWR) at the SGP site, which is also used as a constraint in the Microbase. The MWRRET ARM data product (*Turner et al.* 2007) Best-Estimate *LWP* is used, which is retrieved using an algorithm that combines information from the MWR brightness temperatures, surface-based meteorological data and radiosondes. Uncertainty estimates for the *LWP* are provided for individual measurements and are reported in the results section.

*c. Vertical Relative Humidity and Boundary Layer Height*

Radiosondes are routinely launched at the SGP site 4 times a day, and more frequently during the MC3E campaign. Relative humidity (RH) from the morning (1130 UTC, 0630 LT) and evening (2330 UTC, 1830 LT) radiosonde launches is used in this paper. These profiles are compared against simulated profiles nearest to the SGP and for the output time closest to the mean time between radiosonde launch and it reaching an altitude of 500 hPa. Uncertainty in the radiosonde RH is about 3%.

Information about the observed boundary-layer depth is obtained from the Planetary Boundary Layer (PBL) Height ARM product (*Sivaraman et al.* 2013). This product provides an estimate of the PBL heights for each available sounding, using four different methods. The *Heffter* (1980) and *Liu and Liang* (2010) methods use the potential-temperature profile and two additional estimates use the bulk Richardson Number. The average of these estimates is used as the observed PBL height, while the variability between these methods is shown as the observational uncertainty.

Last, this study uses the interpolated sounding ARM product (*Toto and Jensen* 2016), which interpolates observed soundings to a regular time-height grid with 332 levels and a 1-minute resolution. RH between the observed sounding launches is scaled using MWR observations.

*d. Surface Radiation*

Surface radiation observations are obtained from the ARM Best Estimate Cloud and Radiation Data product (*Xie et al.* 2010). Surface downwelling longwave (LW) and shortwave (SW) radiation from the SGP Central Facility radiometer were used, with a temporal resolution of 60 s. This product averages two out of three different co-located instruments measuring irradiances that agree best with each other (*Shi and Long* 2002). Uncertainties are of the order of 6 and 2.5% for SW and LW radiation respectively (*Stoffel* 2005).

*e. Satellite Cloud Optical Thickness and Water Path*

Data from the Moderate Resolution Imaging Spectroradiometer (MODIS), on-board the Aqua satellite were used as an additional independent observation of the cloud properties. Cloud Optical Thickness ($COT$) and Water path ($WP$) from the Collection 6 Level-2 Aqua-Modis cloud products (*Platnick et al.* 2017) was used from the mid-afternoon (1400 or 1500 LT) overpass over the SGP. Note that all pixels identified as 'overcast cloudy', 'partly cloudy' and 'cloud edge' were included in the analysis. The optical properties from MODIS are retrieved simultaneously using multispectral reflectances for the liquid and ice phase, using visible, infrared and thermal channels (*Platnick et al.* 2017). A simple regridding to the model grid was performed, since MODIS and the evaluated model configuration both have a resolution of 1 km. Model evaluation was performed in a model-to-observation approach, using the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP) (*Bodas-Salcedo et al.* 2011). This software uses simplified synthetic retrieval processes to mimic what the satellite observations would be, given the model's simulated cloud fields. Uncertainty in the MODIS $COT$ and $WP$ retrievals is provided on a per-pixel basis and is reported in the results section.

*f. Surface Precipitation*

The National Center for Environmental Prediction (NCEP) routinely produces the Stage IV radar-based, gauge-adjusted surface precipitation product with a 4-km spatial resolution and hourly sampling (*Lin et al.* 2005). Data covering the entire 1-km simulation domain are used in this paper. Uncertainties in the Stage IV rainfall estimates are generally within 25% (*Westcott et al.* 2008).

## 3. Model configurations

All simulations in this paper are integrated with the Met Office Unified Model (UM, vn11.4), using horizontal grid spacings of 4, 2, 1, and 0.5 km, nested within the GA6 configuration global model (*Walters et al.* 2017) at a resolution of N512 ($\simeq$ 30 km grid spacing near the SGP). Nesting of the successive domains was done one-way only, with no impact of the high-resolution domains onto the coarse-resolution domains.

We refer to Part I for more details about the model configurations and the scientific details other than the cloud schemes used. All three configurations from Part I, only varying in their cloud scheme settings, are used in the forthcoming analysis. These include the operational mid-latitude configuration (RA2-M, *Bush et al.* (2019)), using the diagnostic *Smith* (1990) cloud scheme and an empirical adjustment of the cloud cover (RA2M), a configuration using the Smith cloud scheme without the operational cloud cover adjustment (NOEACF), and a configuration with the new diagnostic bimodal cloud scheme (BM). More details about these configurations can be found in Part I. Three additional configurations are included in this paper. First, a version of the UM without a cloud fraction ($CF$) parameterization is used (NOCF), but with otherwise identical settings to the previous permutations.

A second additional configuration uses the Smith cloud scheme, but rather than using a prescribed and time-invariant profile of RHcrit, it diagnoses variances from the turbulence parameterization, following *Van Weverberg et al.* (2016) (SMITH-TKE). Subgrid saturation-departure variance $\sigma_s^2$ in this configuration is estimated as follows:

$$\sigma_s^2 = a_L^2\overline{q_T'^2} - 2a_Lb_L\overline{q_T'\theta_{liq}'} + b_L^2\overline{\theta_{liq}'^2} \tag{1}$$

where $a_L$ and $b_L$ account for latent heat release, $\overline{q_T'^2}$ is the variance of the total water ($q_T = q_v + q_{liq}$), $\overline{\theta_{liq}'^2}$ is the variance of the liquid potential temperature ($\theta_{liq}$) and $\overline{q_T'\theta_{liq}'}$ the co-variance between the two.

Following *Mellor and Yamada* (1982), (co-)variances are parameterized as:

$$\overline{\theta_{liq}'^2} = B_zS_h{l_{bl}}^2(\frac{\partial\theta_L}{\partial z})^2 \tag{2}$$

$$\overline{q_T'^2} = B_zS_h{l_{bl}}^2(\frac{\partial q_T}{\partial z})^2 \tag{3}$$

$$\overline{q_T'\theta_{liq}'} = B_zS_h{l_{bl}}^2\frac{\partial q_T}{\partial z}\frac{\partial\theta_L}{\partial z} \tag{4}$$

where $l_{bl}$ is the blended subgrid mixing length following *Boutle et al.* (2014), $B_z = 15$ as in *Nakanishi* (2001) and $S_h$ is the stability function. Note that these variances are subsequently translated into RHcrit as follows:

$$RHcrit = 1 - \frac{\sqrt{6}\sigma_s}{a_Lq_{sat}(\overline{T_L})} \tag{5}$$

We refer to *Van Weverberg et al.* (2016) for more information about these formulations.

A last configuration employs a more complicated prognostic *CF* scheme, the Prognostic Clouds and Condensate scheme (PC2), described in *Wilson et al.* (2008) (PC2-TKE). This cloud scheme calculates time-step tendencies of *CF* and liquid water content (*LWC*) from each parameterization that affects temperature or moisture and carries memory of the cloud state from previous time steps. This scheme uses the turbulence-based variances in its cloud initiation term (Equation 1), like SMITH-TKE, and is used in the global atmosphere configuration (GA7; *Walters et al.* (2019)), and in the tropical regional configuration (RA2-T, *Bush et al.* (2019)) at the Met Office. An overview of all model experiments is provided in Table 1.

## 4. Results

### *a. Case Study Evaluation*

Before providing a statistical analysis of the entire MC3E campaign, a detailed case study evaluation is shown here. Given the difficulty of the *Smith* (1990) scheme to produce full cloud cover in stratocumulus conditions (Part I), the case of 27 April 2011 was selected, when an extensive stratocumulus field moved over the SGP. Figure 1 shows the time-height cross sections of cloud fraction (*CF*) and liquid water content (*LWC*) at the ARM SGP site, as retrieved by ARSCL and Microbase and as simulated using all model configurations. Observations (Figures 1a and h) show an overcast stratocumulus deck, breaking up in the late afternoon. The afternoon cloud and PBL top are fairly stationary at about 2000 m and the *LWC* suggests occasional light rain not reaching the surface (e.g. at 1300 LT).

The NOCF (without a *CF* parameterization) displays an overcast cloud layer, although the cloud base is lower than observed (Figure 1b). Furthermore, clouds severely lack water, while breaking too late (Figure 1i).

All three Smith-based configurations (NOEACF, RA2M and SMITH-TKE) struggle to maintain full cloud cover throughout the afternoon and underestimate the $LWC$ (Figure 1). As expected, the $CF$ is somewhat better captured by RA2M with occasionally overcast conditions (Figure 1d). However, these overcast moments remain interspersed with partially cloudy episodes, unlike in the observations (Figure 1a and d).

Interestingly, the advanced prognostic cloud scheme (PC2-TKE) produces too small $CF$ as well (1f), but improves the $LWC$ (Figure 1m) compared to the aforementioned configurations.

Persistent near-overcast conditions and larger $LWC$ in the afternoon are achieved using the new bimodal cloud scheme (BM, Figure 1g and n). Hence BM outperforms all other configurations for this stratocumulus case, although the cloud breaks up too late.

From Figure 1, all configurations reproduce the large-scale humidity (blue contours), temperature (not shown) and the PBL depth (red lines) fairly well, apart from a near-surface dry bias. It is remarkable that the similar humidity environments in all simulations are capable of producing fairly large differences in $CF$ and $LWC$, dependent on the cloud scheme used.

Given the uncertainties associated with the ground-based retrievals, Figure 2 provides an independent evaluation against MODIS observations. As mentioned in Section 2, the COSP software was used to provide synthetic observations from the model simulations.

Observed cloud optical thickness ($COT$), condensed water path ($WP$) and cloud top pressure ($CTP$) at 1500 LT are given in Figure 2a, h and o. The location of the SGP site, used for the cross sections in Figure 1, is marked by the yellow diamond on Figure 2 while the MODIS overpass time is indicated on the cross sections in Figure 1 with the vertical dashed gray line. Consistent with Figure 1a, MODIS displays an optically thick stratocumulus sheet covering most of the domain, with a few breaks near the Texas Panhandle and over Kansas (Figure 2a). The $WP$ is fairly high within the closed cells, although there is considerable variability (Figure 2h). Apart from a few high-level clouds to the northeast of the SGP site, clouds tops are generally fairly low (Figure 2o), confirming the principal contribution from the low-level stratocumulus deck to the $WP$ in Figure 2h.

All model configurations, apart from BM, fail to show a continuous, optically thick stratocumulus sheet with large $WP$ (Figure 2b-f and i-m), confirming their poor representation of cloud properties evident from Figure 1. Note that RA2M exhibits fairly similar cloud properties than NOEACF (Figure 2 c and d). Indeed, the operational bias adjustments in RA2M are limited to the $CF$, not affecting the $LWC$.

The clear improvement in $CF$ and $LWC$ in BM in Figure 1 is reflected in the COSP-diagnostics (Figure 2g and n). The $COT$ and $WP$ are substantially larger and more continuous in BM compared to the other configurations, bringing them closer to the observations.

Figure 3 shows the observed and simulated diurnal cycles of surface SW and LW downwelling radiation during 27 April and Table 2 shows evaluation statistics for each simulation. Apart from a brief gap in the clouds around 1000 LT, most of the daylight period remains overcast in the observations, leading to subdued downwelling SW and enhanced downwelling LW radiation. All simulations transmit too much SW radiation through the afternoon stratocumulus (Figure 3a), consistent with the underestimated $CF$ and/or $WP$ (Figure 1 and 2). The overcast stratocumulus sheet in NOCF (Figure 1b) fails to reflect sufficient downwelling SW (Figure 3a and Table 2), given its underestimated $LWC$ (Figure 1i). The PC2-TKE, having larger $LWC$, but much more broken clouds (Figure 1f and m), also has excessive surface SW (Figure 3a and b). It is only when both $CF$ and the $LWC$ are well-captured, as in BM, that the surface radiation statistics improve (Figure 3 and Table 2).

From the evidence presented so far, the bimodal cloud scheme is clearly beneficial for the 27 April stratocumulus case. One of the remarkable findings for the other configurations is the limited impact of using a more realistic saturation-departure variability in SMITH-TKE and PC2-TKE. Figure 4 provides a time-height cross section of the variable RHcrit and associated variances for SMITH-TKE on 27 April 2011. Since these variances are based on a combination of the turbulent kinetic energy (TKE) and the local thermodynamic gradients (section 3) the lowest RHcrit and largest variances manifest themselves near the PBL top. Most of the stratocumulus cloud (highlighted by the dotted area) resides well within this region of high variability. The RHcrit parameterization described in *Van Weverberg et al.* (2016) imposes a resolution-dependent minimum limit on the RHcrit, based on aircraft observations. From Figure 4, this minimum limit (about 86% for the 1 km grid spacing shown here) is reached throughout the entire cloudy region. The NOEACF has a fairly similar value of 80% at this altitude, which explains why SMITH-TKE does not yield vastly different cloud properties. Future research might revisit these resolution-dependent RHcrit limits in SMITH-TKE.

A similar analysis is provided for BM (Figure 5). As explained in Part I, the individual modes in the mixture of PDFs in the bimodal cloud scheme are symmetric and Gaussian, with variances based on an extension of the *Furtado et al.* (2016) scheme to all liquid clouds. These turbulence-based unimodal variances are provided in Figure 5a. In contrast to the turbulence-based variances in SMITH-TKE (Figure 4), the variances following *Furtado et al.* (2016) are not maximized near the PBL top. These variances are more uniquely related to the TKE and turbulent mixing length, and do not have a gradient-related term in their formulation like SMITH-TKE. Hence, variances here are larger within the PBL, where TKE is maximized.

A unique feature of the bimodal cloud scheme is the assignment of a bimodal subgrid SD PDF to each level encompassed by an entrainment zone (EZ, see Part I). This is done by combining the modes of variability from the bottom and the top of the EZ, using their respective mean and variance, and weighting them to conserve the SD at the level of interest. Any level outside the

EZ uses a unimodal distribution with the local turbulence-based variance. We refer to Part I for more details about the bimodal cloud scheme. Figure 5b shows the combined variance of the bimodal mixture distribution. The EZ is denoted by the gray contour and clearly, the variances of the mixture distribution within the EZ are much larger than the local turbulence-based variance. Hence, the mixture-distribution variance is maximized near the PBL top in Figure 5b, as typically observed (*Price* 2006; *Wood and Field* 2000; *Turner et al.* 2014; *Wulfmeyer et al.* 2016; *Osman et al.* 2018). Note that the mixture-distribution variances near the PBL top are larger than those obtained from the turbulence-based RHcrit formulation in *Van Weverberg et al.* (2016) (Figure 4b). A variance of $10^{-6} kg^2 kg^{-2}$ would roughly correspond to an RHcrit of about 40%. Further research will focus on the evaluation of the bimodal variances using aircraft and lidar data.

An implicit feature of the bimodal scheme is its ability to produce skewed mixture distributions of SD, by applying variable weights to each of the two PDFs. Observed humidity profiles from lidar and aircraft (*Turner et al.* 2014; *Wulfmeyer et al.* 2016; *Wood and Field* 2000) typically show negatively skewed distributions just below the PBL top. Figure 5c shows the skewness associated with the mixture distribution in the bimodal scheme for 27 April. Consistent with the aforementioned observations, the stratocumulus cloud just below the inversion resides in a broad region of negative skewness, which is responsible for the much larger $CF$ and $LWC$ for a given environmental RH as noted in Figure 1g and n.

### *b. Statistical Evaluation of MC3E*

#### 1) Vertical Profiles of Cloud Properties

The previous section demonstrated the benefit of the bimodal cloud scheme for a single stratocumulus case, focusing on the physical mechanism leading to improved $CF$ and $LWC$. This section explores whether these improvements can be confirmed for a wider range of conditions.

Figure 6 shows vertical profiles of observed and simulated non-precipitating, liquid cloud properties, averaged over the entire MC3E period. Evaluation statistics are provided in Table 3. All model configurations underestimate the MC3E-averaged vertically distributed *CF* (Figure 6a), although there is considerable variability in their frequency of occurrence (FOO; Figure 6b) and the amount of cloud when present (AWP; Figure 6c).

The NOCF underestimates the average *CF* (Figure 6a) to a similar degree than some configurations that do use a *CF* parameterization (Table 3). However, NOCF experiences a large compensating error between a too low FOO (Figure 6b), while by definition always having AWP= 1 (Figure 6c). The average *CF* in NOEACF is closer to the observations than NOCF, although exhibits an opposite compensating error of too large FOO and too small AWP. The bias-adjustment in RA2M leads to increased FOO and AWP compared to NOEACF (Figure 6b and c), yielding the best-captured average *CF* (Figure 6a). However, while the RA2M AWP is improved (yet still underestimated), the FOO becomes even more overestimated, particularly for very low clouds. The BM produces similar average *CF* than NOEACF (Figure 6a), but captures the observed FOO better than any other configuration (Figure 6b). The AWP in BM is comparable to RA2M (Figure 6c). Interestingly, the two configurations using the turbulence-based RHcrit (SMITH-TKE and PC2-TKE) produce similar average *CF* (Figure 6a), despite the greater complexity in the PC2-TKE cloud scheme. Their small average *CF* is mainly due to the underestimated FOO. The AWP in these two configurations is better captured than in all other configurations, particularly by SMITH-TKE. This is likely related to generally larger turbulence-based RHcrit in SMITH-TKE than e.g. the constant RHcrit in NOEACF.

Average *LWC* profiles are provided in Figure 6d. Similar to the *CF*, average *LWC* in all simulations is biased low, although there is considerable difference between the model configurations (Table 3). The average *LWC* in NOCF is the smallest of all configurations (Figure 6d). However, given that the *CF* is also significantly underestimated (Figure 6a), the in-cloud *LWC* is fairly well captured (Figure 6e). The NOEACF and RA2M exhibit similarly underestimated *LWC* (Figure 6d). However, given the larger *CF* in RA2M than in NOEACF (Figure 6a), the in-cloud *LWC* in RA2M is lower and significantly underestimated (Figure 6e). The BM average and in-cloud *LWC* are fairly well-captured (Figure 6d), which is an improvement compared to RA2M and NOEACF. Note that the microphysics and radiation schemes use in-cloud *LWC* rather than the grid-box mean *LWC*. The SMITH-TKE and PC2-TKE have too small average *LWC* (Figure 6d and Table 3), although the in-cloud *LWC* (Figure 6e) is overestimated in PC2-TKE.

Figure 6f shows the simulated and observed RH profiles for the morning (0630 LT; dashed lines) and evening (1830 LT; solid lines) radiosonde launches at the SGP, averaged over the MC3E period. The triangles denote the average PBL heights at the times of the soundings. The impact of the different cloud schemes on the RH profiles is surprisingly small, although note that these profiles include all MC3E days, many of which were non-cloudy. While RH is only slightly too dry in the morning, the simulated afternoon boundary-layer RH is much drier than observed. This could indicate excessive entrainment of dry free-tropospheric air into the boundary layer. Furthermore, the model dry bias might be a consequence of the lack of cloud, but could also be the origin of the cloud biases. The dry, warm bias in the U.S. Great Plains (*Morcrette et al.* 2018) has been shown before to be related more to land-surface and precipitation deficiencies, rather than to clouds (*Van Weverberg et al.* 2018). It is beyond the scope of this paper to investigate the complex interactions between land surface, precipitation and clouds and their role in the warm, dry bias over the SGP, but this bias should be borne in mind when analyzing the results shown here.

2) DIURNAL CYCLES OF CLOUD PROPERTIES

Figure 7 provides diurnal cycles of non-precipitating cloud properties, averaged over the MC3E period. Evaluation statistics are provided in Table 3. Simulated and observed vertically integrated cloud cover (*CC*) is derived assuming a maximum-random overlap. Note that, in contrast to the analysis so far, mixed-phase clouds are included here, since the following analysis does not rely on the uncertain Microbase *LWC*. The boundary between low- and mid-level cloud and mid- and high-level cloud is defined as 3000 and 6000 m altitude, respectively.

The observed diurnal cycle of low-level cloud exhibits a clear diurnal cycle, peaking near 1200 LT and reaching a minimum near local midnight (Figure 7a). All model configurations reproduce this diurnal cycle, although with an earlier-than-observed peak and too small low *CC* in general, most notably in the NOCF. The bias is particularly large in all simulations in the afternoon, with BM and RA2M outperforming the other configurations (Table 3).

The observed FOO of low *CC* peaks in the afternoon, while the AWP is fairly large throughout the diurnal cycle and drops to about 80% around noon (Figure 7b an c). Most experiments appear to have too frequent low cloud in the morning, in particular the RA2M and NOEACF, while the afternoon FOO is reasonably captured by most configurations (Figure 7b). The NOCF, however, only generates cloud half as frequently as observed throughout the entire diurnal cycle. All simulations, apart from the binary NOCF, underestimate the low *CC* AWP throughout the entire diurnal cycle (Figure 7c). In the afternoon, this is true in particular for PC2-TKE, while the RA2M and BM are slightly outperforming the other configurations.

While the focus of this paper in on low-level clouds, the surface radiation and satellite evaluation in the next sections is also affected by mid- and high-level clouds. Hence, Figure 7d-i shows the diurnal cycle of mid- and high-level *CC*. There is less mid-level than low-level observed cloud with limited diurnal variation (Figure 7d). The weak maxima in the morning and the evening coincide with maxima in the surface precipitation, shown later. The models underestimate the mid-level *CC* (Table 3), mainly during the observed maxima, and hardly show any diurnal variability (Figure 7d). The only exceptions are NOCF and PC2-TKE, showing enhanced mid-level *CC* in the afternoon. As for the low *CC*, the underestimation in the mid-level *CC* is almost entirely due to the very low AWP (Figure 7f), since the FOO is well-captured in all models (Figure 7e). The notable exception is PC2-TKE, showing a too large FOO of mid-level *CC* in the afternoon, causing the afternoon peak in average mid-level *CC* seen in Figure 7d.

The observed high-level *CC* again shows peaks in the morning and evening, coinciding with maxima in the surface precipitation as shown later (Figure 7g). High-level *CC* is well-captured by PC2-TKE and NOCF, but largely underestimated by all other configurations (Figure 7g and Table 3). Note that all diagnostic cloud schemes presented here calculate ice *CF* as a simple diagnostic function of ice water content, produced by the microphysics, following e.g. *Abel et al.* (2017). Apparently, this simple treatment leads to significantly underestimated AWP (Figure 7i), while the FOO is well captured (Figure 7h). For ice clouds, the more advanced PC2 scheme in PC2-TKE outperforms the other configurations.

Some additional diurnal cycle diagnostics are provided in Figure 8. The observed liquid water path ($LWP$) of non-precipitating clouds shows a very limited diurnal cycle (Figure 8a). All simulations produce a more distinct diurnal cycle than observed, underestimating $LWP$ at night, but better capturing or overestimating $LWP$ around noon. The smallest $LWP$ is produced by NOCF throughout most of the diurnal cycle, consistent with the analysis so far (Table 3). The BM enhances the $LWP$ compared to RA2M and NOEACF mainly in the afternoon, when the PBL and hence the EZ are deeper. Arguably, the late afternoon $LWP$ is somewhat better captured by BM than the other configurations.

From Figure 6a and b, RA2M and NOEACF showed a significant overestimation of the *CF* at very low levels, below 1000 m. This seems to be related to their too low cloud bases (Figure 8b), mainly at night and in the evening. Cloud bases are better represented by the other configurations (Table 3).

Interestingly, the low *CC* shown in Figure 7a displays a smaller difference between RA2M and BM than the vertical *CF* profile in the previous section (Figure 6a). The physical low-cloud depth explains some of these discrepancies (Figure 8c and Table 3). Indeed, throughout the entire day, clouds in the RA2M and NOEACF are deeper than in all other configurations and the observations. Hence, the reason the vertical cloud fraction profiles in RA2M (Figure 6a) are better captured than in all other configurations, is mainly due to a compensation between excessively deep clouds, whilst having a too small AWP. Cloud depth is well captured in all configurations with turbulence-based variances.

### 3) Diurnal Cycle of Surface Precipitation

The parameterization of *CF* matters for the simulated radiative transfer, but also for the precipitation microphysics. While not the principal focus of this paper, it is interesting to cast a glance at the diurnal cycle of precipitation to verify its behavior in the various cloud configurations. From Figure 9, the observed domain-average diurnal cycle of precipitation shows the often reported double peak in the morning and evening (*Klein et al.* 2006; *Van Weverberg et al.* 2018). Consistent with earlier studies, all configurations struggle to simulate this double peak and produce a single mid-afternoon peak, with very limited impact of the *CF* parameterization. However, the domain-average precipitation is reduced and closer to the observations in the BM (Table 4 and Figure 9).

4) SATELLITE-BASED EVALUATION

Figure 10 shows histograms of $COT$ and $WP$, including all MODIS overpasses over the SGP (between 1400 and 1600 LT). Note that all diurnal cycle figures in this paper denote this MODIS overpass window using gray shading. While MODIS only constitutes a snapshot in the diurnal cycle, it provides additional evidence for model biases shown in the previous sections. Uncertainties in the MODIS $COT$ and $WP$ are provided on a per-pixel basis and are shown here by assessing the (Gaussian) probability that any individual pixel belongs to a particular bin, given the pixel-level mean and standard error, and are indicated with the gray-shaded areas in Figure 10.

Consistent with the previous sections, the $COT$ and $WP$ in NOCF is underestimated throughout the entire range of values (Figure 10a and b). A smaller underestimation across the range of $COT$ and $WP$ values is present in RA2M and NOEACF. Note that the COSP diagnostics shown here are fairly similar for RA2M and NOEACF, given that the $LWC$ is unaffected by the operational cloud adjustments in RA2M. Consistent with the ground-based retrievals in Figure 8a at the MODIS overpass time, BM most accurately reproduces the observed $COT$ and $WP$ histograms (Figure 10). PC2-TKE has too frequent small and too infrequent large values of $COT$ and $WP$. Note that PC2-TKE has the largest low- and mid-level $CC$ frequency (Figure 8b and e) at the MODIS overpass time. $COT$ and $WP$ are underestimated for the entire value range in SMITH-TKE.

Simulated and observed $COT$ and $WP$ statistics are provided in Table 5 for the full domain (including clear-sky), and for liquid-phase and ice-phase grid points only. Frequencies are based on any grid point with COT exceeding 0.3, which is typically considered the satellite's lower detection limit.

For $COT$ both the arithmetic ($\overline{arCOT}$) and the extinction-weighted mean ($\overline{exCOT}$) are shown. The latter is more relevant for surface radiation. Transmission-weighted mean $COT$ is calculated as follows:

$$\overline{exCOT} = -\mu log(\frac{1}{N}\sum exp(\frac{-\tau_i}{\mu_i})) \tag{6}$$

where $\mu$ is the arithmetic mean of the solar zenith angle cosines, $\tau_i$ is the $COT$ for grid point $i$, $\mu_i$ is the solar zenith angle cosine of grid point $i$ and $N$ is the number of grid points.

The liquid-cloud frequencies shown in Table 5 correlate well with the ground-based low and mid-cloud frequencies at the MODIS overpass time (Table 5 and Figure 7b and e), with a severe lack of cloud in NOCF, too frequent cloud in PC2-TKE and fairly well-captured frequencies in RA2M and BM. Continuing to focus on the liquid phase $LWP$ and $COT$, note that statistics in Table 5 are for grid points that at are least partially cloudy. As all cloudy grid points in NOCF are completely overcast, their average $LWP$ and $COT$ are larger than for any other configuration and the observations. While the liquid-phase $WP$ and $\overline{arCOT}$ are too small in PC2-TKE, its liquid $\overline{exCOT}$ is much better captured (Table 5). This hints at persistent substantial broken cloud, always reflecting much of the incoming SW, while overcast conditions or very small cloud cover occur rather infrequently. The liquid-phase $WP$ and $COT$ are very similar for the other configurations, with a tendency to underestimate the average $WP$ and $\overline{exCOT}$.

All configurations overestimate the frequency, $WP$ and $COT$ of ice-phase clouds (Table 5), consistent with the mid- and high-level clouds in Figures 7e and h, at the MODIS overpass time (note the dip in the observed $CC$ at this time). The ice-phase $\overline{arCOT}$ is largest in RA2M, PC2-TKE and BM, while $\overline{exCOT}$ is smaller in RA2M and BM than in PC2-TKE. This is consistent with the relatively larger influence of more broken cloud on $\overline{exCOT}$ than on $\overline{arCOT}$ for a given water content. Indeed, recall that PC2-TKE captured the large AWP for high clouds considerably better than the diagnostic cloud scheme configurations. The focus of this paper is on low-level cloud, but this finding suggests ample room for improving ice *CF* parameterizations.

Combining liquid-phase, ice-phase and clear-sky grid points, $\overline{arCOT}$ and $WP$ at the MODIS overpass time are overestimated in all configurations (Table 5). In contrast, $\overline{exCOT}$ is underestimated in all configurations, apart from PC2-TKE. The discrepancy between underestimated $\overline{exCOT}$ and overestimated $\overline{arCOT}$ reflects the typically too small AWP in most cloud configurations. If the excessive total water contents in all configurations (Table 5) are associated with partial cloudiness, rather than overcast conditions, their impact on the radiative transfer will be comparatively limited.

#### 5) Surface Radiation

One of the principal advantages *CF* parameterizations in coarser-scale models is the benefit for surface radiation statistics. Figure 11 shows simulated and observed diurnal cycles of downwelling SW and LW radiation averaged for the MC3E period. Table 6 provides surface radiation statistics.

Note that the radiation data were not screened for the occurrence of precipitation. Furthermore, the radiation scheme assumes some in-cloud heterogeneity for liquid clouds (*Cahalan et al.* 1994) that is not seen by the COSP diagnostics. This somewhat complicates a direct one-to-one comparison with the cloud properties in Figure 7.

Nevertheless, in combination with the ground-based and satellite-based analysis so far, a consistent picture emerges that there is considerable benefit in using a *CF* parameterization at convection-permitting scales. The NOCF, significantly lacking cloud and liquid water, unsurprisingly overestimates the downwelling SW (Figure 11a and Table 6) and performs worse than other configurations using a *CF* parameterization. All other configurations also experience too large downwelling SW, consistent with their lack of cloud. Unsurprisingly, schemes performing better in terms of cloud cover, also exhibit better radiation characteristics. The RA2M has the largest cloud cover, and hence smallest SW bias in the morning (Figure 11a). The BM performs best for low $CF$ and $LWP$, and hence SW radiation in the afternoon (Figure 11a). These two configurations overall show the smallest SW bias of all simulations (Table 6)

A negative bias in the downwelling LW emerges in the afternoon and at night in all simulations (Figure 11b), consistent with lack of cloud at these times (Figure 7). Again, RA2M and BM perform better than the other configurations for the LW bias (Table 6).

### *c. Sensitivity to Horizontal and Vertical Resolution*

All analysis so far has been concerned with the 1 km grid-spacing simulations. As models are run at increasingly fine resolution, the need for scale-aware cloud parameterizations becomes ever more pressing (e.g. *Tompkins* (2003)). Moreover, operationally at the Met Office, convection-permitting ensembles use similar physics as the deterministic regional simulations, but with a slightly larger grid spacing of about 2 km (*Bush et al.* 2019; *Hagelin et al.* 2017). It is important that no systematic biases are introduced by lowering the resolution.

This section explores the scale-awareness of all model configurations, using the domains with 0.5, 1, 2 and 4 km grid spacing. The same time- and space invariant vertical profile of RHcrit was applied across these resolutions for RA2M and NOEACF. The BM, SMITH-TKE and PC2-TKE all use variances in their subgrid saturation-departure distributions that are linked with scale-aware turbulence diagnostics (see section 3).

Figure 12 provides vertical profiles of average $CF$, FOO and AWP, averaged over the entire MC3E period for all non-precipitating, liquid clouds. ARSCL observations have been regridded to assumed grid lengths of 0.5, 1, 2 and 4 km. While observed average $CF$ is insensitive to the resolution (as expected), the FOO becomes larger and the AWP smaller as the grid spacing increases (Figure 12a, d and g). Hence, the observed *CF* exhibits more binary behavior as the grid spacing decreases. However, even at 500m grid spacing, considerable subgrid variability is observed.

All model configurations using a constant variability profile (NOCF, RA2M and NOEACF) have similar average *CF* profiles across the different resolutions (Figure 12b). However, the observed tendency towards more binary *CF* with decreasing grid spacing is absent in these configurations (evident from the FOO and AWP in Figure 12e and h). As such, without re-tuning of RHcrit for each grid spacing, their bias in the FOO and AWP will be resolution-dependent.

Model configurations with scale-aware variability in the *CF* parameterization also show limited resolution sensitivity in the average *CF* (Figure 12c). However, these configurations do exhibit considerable variability in terms of their FOO and AWP with varying resolution (Figure 12f and i). BM scales too much with resolution, showing more variability between the 0.5 and 2 km FOO and AWP than the observations. Moreover, the 4 km simulation appears to have reduced FOO and enhanced AWP compared to the 2 km simulation, unlike the observations. The PC2-TKE on the other hand barely scales with resolution, while SMITH-TKE scales much more analogous to the observations.

A caveat to be made with the scale-awareness analysis shown here, is that the UM is a rather diffusive model. Indeed, *Klaver et al.* (2019) have shown that at least for the global version of the UM, the effective resolution is close to four times the grid spacing. For regional simulations the effective resolution might be even lower, as suggested by the power spectra in *Boutle et al.* (2014), showing poorly represented energy cascades for scales smaller than about 10 times the grid spacing. This might explain some of the poor performance of the simulations in terms of

AWP in this paper, although there is a large discrepancy even between the 2 km observed and the 0.5 km simulated FOO and AWP (Figure 12) for the scale-aware simulations.

To better understand the origin of the resolution-dependent FOO and AWP in BM, vertical profiles of several scale-aware diagnostics for the different resolutions are provided in Figure 13. The turbulence-based unimodal saturation-departure variance, calculated following Equation 13 in Part I, relies on two scale-aware diagnostics, the mixing length ($l_{bl}$) and the vertical velocity variance (${\sigma_w}^2$). Recall that $l_{bl}$ is the blended mixing length (*Boutle et al.* 2014), combining the mixing length from a 1D boundary-layer scheme ($l_{1D}$, *Lock et al.* (2000)) and a 3D Smagorinsky scheme ($l_{smag}$, *Smagorinsky* (1963)). The ${\sigma_w}^2$ is parameterized following *Walters et al.* (2019) as ${\sigma_w}^2 = K_m {\tau_{turb}}^{-1}$ where $\tau_{turb}$ is a turbulence time scale and $K_m$ is the eddy diffusivity for momentum.

The scale-awareness of the low-cloud FOO (Figure 12e) predominantly carries the footprint of the $l_{bl}$ scaling with resolution (Figure 13a). Of the two mixing lengths that are blended in the *Boutle et al.* (2014) parameterization, it is the $l_{smag}$ that scales with resolution, growing larger with increasing grid spacing. At the same time, the blending of $l_{smag}$ and $l_{1D}$ is itself resolution-dependent within the PBL. For grid spacings larger than 1 km, $l_{smag}$ becomes larger than $l_{1D}$, but at the same time the weight of $l_{smag}$ becomes smaller. Hence, the 2 km $l_{bl}$ is smaller than the 1 and 0.5 km $l_{bl}$ within the boundary layer. This effect becomes even more obvious for the 4 km $l_{bl}$, which is smaller than any other $l_{bl}$ below 1.5 km altitude. This is the principal reason for the slightly odd behavior of the 4 km simulation in terms of FOO and AWP. Hence, this is not an inherent characteristic of the bimodal cloud scheme, but rather of the mixing length blending.

The ${\sigma_w}^2$ scales more uniformly across the vertical profile and increases with increasing grid length (Figure 13b), even for the 4 km simulation.

As the $l_{bl}$ increases with height, and the ${\sigma_w}^2$ is maximized near the surface, the unimodal saturation-departure variance profile in Figure 13c emerges (solid lines). The subgrid variance increases as the grid spacing increases, as expected, except for the 4 km simulation, due to the inverse scaling of $l_{bl}$ for this resolution. However, the variance maximum also moves higher in the atmosphere as the resolution decreases, also reflected in the low-cloud FOO scaling (Figure 12e). This is probably not desirable and suggests room for improvement in the way the $l_{bl}$ scales with resolution, at least for relatively coarse grid spacing. Note that the mixing-length blending provides large improvements across the turbulent gray-zone (100 m - 1 km) (*Boutle et al.* 2014).

For grid points encompassed by an EZ, the bimodal scheme diagnoses cloud by combining modes from the bottom and the top of the EZ. Hence, the variance of the mixture of these two PDFs can be calculated (Equation 30 in Part I). As expected, this mixture variance tends to be larger, and scales similarly compared to the local unimodal variances of the individual modes (Figure 13c).

Furthermore, skewness of the mixture distribution in the bimodal scheme can be calculated from Equation 31 in Part I and is shown for the different resolutions in Figure 13d. Skewness remains fairly constant with changing resolution, which is desirable, since the depth of the EZ and the presence of a mixture distribution should be independent of the horizontal resolution, at least as long as the grid spacing is larger than the length scale of free-tropospheric intrusions into the mixed layer ($< 100m$ as evident from lidar observations, e.g. *Wulfmeyer et al.* (2016)).

Many of the processes relevant to cloud processes, such as dry air entrainment at the boundary-layer top, are sensitive to the vertical resolution as well. Hence, Figure 14 shows the observed and simulated profiles of average cloud fraction, FOO and AWP, assuming two vertical level sets, and using a horizontal grid spacing of 1 km. The L70 level set has been used for all analysis discussed so far, and has a vertical grid spacing decreasing from 20m near the surface to about 200 m near 3 km altitude. The L140 level set has twice as many vertical levels and hence double the resolution of the L70 level set.

Only the RA2M and BM simulations are shown in this figure. The vertical resolution at L70 is clearly high enough to capture most subgrid variability relevant for cloud formation, as the sensitivity of RA2M and BM to the vertical resolution is very limited, consistent with observations. This also highlights that cloud deficiencies in all configurations, i.e. the general lack of cloudiness, are not primarily a vertical resolution problem, but point to more fundamental issues of the cloud and boundary-layer parameterizations.

## 5. Discussion and Conclusions

This second of two papers presents an in-depth evaluation of different approaches to represent subgrid cloud variability in numerical weather prediction models at convection-permitting scales. A wide range of observations, including ground-based remote sensing, satellite-based retrievals and surface observations are used in this evaluation, gathered during the Midlatitude Continental Convective Clouds Experiment (MC3E) at the US Southern Great Plains.

Simulations are performed using an operational regional model configuration, with 6 permutations to its cloud fraction (*CF*) parameterization and with a range of horizontal and vertical grid spacings. A first set of simulations ignores any subgrid cloud variability. Three configurations use the *Smith* (1990) diagnostic cloud scheme, each with variations to its subgrid saturation-departure PDF: using a constant variability profile, using the same constant variability profile, but additionally using operationally-used bias-adjustment techniques, and using more realistic turbulence-based and scale-aware subgrid variability. A more advanced prognostic *CF* scheme is used in a fifth configuration, and a sixth configuration uses a newly developed diagnostic cloud scheme, introduced in Part I. This schemes assumes a mixture of PDFs in the entrainment zone, a dry mode from the free troposphere and a moist mixed-layer mode.

Revisiting the first of three research questions asked in the introduction, it is shown that low-cloud simulations still benefit from the use of a $CF$ parameterization at convection-permitting scales. The omission of subgrid cloud variability leads to less than half the observed frequency of low cloud and correspondingly small domain-average $WP$. Consequently, cloud optical thickness is largely underestimated and the surface shortwave radiation overestimated. Even for a stratocumulus case, the omission of a $CF$ scheme leads to underestimated water contents and excessive surface shortwave radiation, despite maintaining full cloud cover.

However, including a diagnostic or prognostic $CF$ scheme does not automatically lead to large improvements and the simulations proved to be sensitive to the choice of the specific $CF$ scheme. A diagnostic $CF$ scheme with prescribed, constant variance profiles (through a critical relative humidity) only manages to produce good cloud cover when operational bias-adjustment techniques are applied. However, this configuration is only able to do so through a compensating error between too large cloud frequency and too deep clouds, but a too small cloud amount when present. The combination of too frequent clouds with too small $WP$ results in well-captured surface radiation and optical thickness.

A diagnostic $CF$ scheme that assumes a turbulence-based, but symmetric subgrid variability PDF, does not perform well for most metrics shown. This configuration exhibits negative biases in its low cloud cover, liquid water path and liquid cloud optical thickness.

The new diagnostic bimodal $CF$ parameterization, introduced in Part I, outperforms the other configurations in this paper for a number of metrics. The scheme produces the largest cloud cover in the afternoon, with well-captured frequency and larger cloud amount when present than the other $CF$ parameterizations. The bimodal scheme exhibits the best liquid cloud optical thickness and water path and its surface radiation biases smaller than in any other configuration, barring the operational configuration. The improvements manifest themselves via clouds near the boundary-layer top being able to experience negatively-skewed saturation-departure distributions. This brings a larger portion of the grid box in a supersaturated state than an un-skewed distribution with identical variance and mean conditions.

The simulation with a prognostic $CF$ scheme produces frequent low- and mid-level clouds in the afternoon, but underestimates the cloud amount when present. This scheme produces the largest, and overestimated, in-cloud water contents, while still overestimating the downwelling shortwave radiation. This advanced scheme does not outperform simpler schemes for the diagnostics shown here. Despite the complexity and memory of the cloud state in this scheme, there is still an implicit assumption of symmetric subgrid variability. Hence, the inclusion of time-variable higher-order moments of the distribution such as skewness, appears to matter more for the model performance than whether the cloud scheme is prognostic or diagnostic. It will be further investigated whether some of the bimodal cloud scheme concepts can be used in a prognostic framework as well.

The cloud frequency and the cloud amount when present is insensitive to the horizontal grid spacing in configurations relying on a constant variance profile, such as the operational configuration. Observed clouds, in contrast, clearly become more binary as the assumed grid spacing decreases. The schemes using scale-aware variance, linked to the turbulence scheme, display the observed increase in cloud frequency and decrease in cloud amount when present with increasing grid spacing. However, the cloud frequency in the bimodal scheme seems to vary more with resolution than the observed frequency. This appears to be related to the large variability of the blended turbulence mixing length, parameterized for gray-zone turbulence.

The analysis in this paper was not restricted to low clouds alone and revealed some interesting model deficiencies in high-cloud cover that are worth investigating in more detail in future studies.

All simulations with a diagnostic $CF$ parameterization underestimate the average high-level $CF$, as their high-level clouds are often broken when overcast conditions are observed. At the same time all configurations have excessive ice water contents in the afternoon, regardless of their $CF$ bias. The former issue is indicative of excessive variability imposed on high clouds in all diagnostic $CF$ parameterizations. The excessive ice water content on the other hand is likely related to the parameterization of ice microphysics.

Furthermore, all simulations produce too little cloud and excessive surface SW radiation, while the boundary layer tended to be too dry. It is hard to tell which of these two issues is the cause or the effect, but the area of the U.S. Great Plains has been known for its difficulties in terms of the simulated surface-energy balance (*Koster et al.* 2004; *Klein et al.* 2006; *Ma et al.* 2018; *Van Weverberg et al.* 2018). The wealth of observations at the SGP ARM site still warrants its use for in-depth model evaluation, but it would be worth repeating the analysis in other regions with better captured climatology. Another reason for repeating this exercise for very different environments is inspired by the need for different cloud configurations over the tropics and mid-latitudes for regional forecasts at the Met Office (*Bush et al.* 2019). As such, the bimodal cloud scheme could be a step towards unification of these different configurations.

Last, while a broad range of observations was used to evaluate the cloud schemes in this paper, an evaluation of the higher-order moments of the subgrid saturation-departure distribution and the assumptions of the new bimodal scheme is desirable. This will be the focus of future research, using large eddy simulations, aircraft and lidar observations.

*Acknowledgments.* The authors are grateful to Adrian Lock and Vince Larson for stimulating discussions and to Alejandro Bodas-Salcedo for providing the COSP software. The ground-based observational data for this article were gratefully obtained from the US Department of Energy

ARM data archive (http://www.archive.arm.gov/armlogin/login.jsp), sponsored by the DOE Office of Science, Office of Biological and Environmental Research, Environmental Science Division. The Aqua MODIS Cloud Properties L2 were acquired from the Level 1 and Atmosphere Archive Distribution System (LAADS) Distributed Active Archive Center (DAAC), located in the Goddard Space Flight Center in Greenbelt, Maryland (https://ladsweb.modaps.eosdis.nasa.gov/). The work of K. Van Weverberg was supported by the Met Office Weather and Climate Science for Service Partnership (WCSSP) Southeast Asia as part of the Newton Fund.

## References


Abel, S.J., Boutle I.A., Waite, K., Fox, S., Brown, P.R.A., Cotton, R. (2017). The role of precipitation in controlling the transition from stratocumulus to cumulus clouds in a northern hemisphere cold-air outbreak. *Journal of the Atmospheric Sciences*, 74, 2293-2314.

Bodas-Salcedo, A., and co-authors (2011). COSP: Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, 92, 1023-1043.

Boutle, I.A., Eyre, J.E.J, Lock, A.P. (2014). Seamless stratocumulus simulations across the turbulent gray zone. *Monthly Weather Review*, 142, 1655-1668.

Bush, M. and coauthors. (2019). The Met Office Regional Atmosphere and Land configurations (RAL)- Version 1. *Geoscientific Model Development*, in review.

Cahalan, R.F, Ridgway, W., Wiscombe, W.J., Bell, T.L. (1994). The albedo of stratocumulus clouds. *Journal of the Atmospheric Sciences*, 51, 2434-2455.

Del Genio, A.D. (1996). A prognostic cloud water parameterization for global climate models. *Journal of Climate*, 9, 270-304.

Field, P.R., Hill, A.A, Furtado, K., Korolev, A. (2014). Mixed-phase clouds in a turbulent environment. Part II: Analytic treatment. *Quarterly Journal of the Royal Meteorological Society*, 140, 870-880.

Furtado, K., Field, P.R., Boutle, I.A., Morcrette, C.J., Wilkinson, J.M. (2016). A physically based subgrid parameterization for the production and maintenance of mixed-phase clouds in a general circulation model. *Journal of the Atmospheric Sciences*, 73, 279-291.

Gerard, L. (2007). An integrated package for subgrid convection, clouds and precipitation compatible with meso-gamma scales. *Quarterly Journal of the Royal Meteorological Society*, 133, 711-730.

Golaz, J.-C., Larson, V.E., Cotton, W.R. (2002). A pdf-based model for boundary layer clouds. Part I: Method and model description. *Journal of the Atmospheric Science*, 39, 3540-3551.

Hagelin, S., Son, J., Swinbank, R., McCabe, A., Roberts, N., Tennant, W. (2017). The Met Office convective-scale ensemble, MOGREPS-UK. *Quarterly Journal of the Royal Meteorological Society*, 143, 2846-2861.

Hefter, J.L. (1980). Transport layer depth calculations. *Second Joint Confer ence on Applications of Air Pollution Meteorology*, New Orleans, Louisiana.

Jensen, M. P., and co-authors (2016). The midlatitude continental convective clouds experiment (MC3E). *Bulletin of the American Meteorological Society*, 97, 1667-1686.

Klaver, R., Haarsma, R., Vidale, P.L., Hazeleger, W. (2019). Effective resolution in high resolution global atmospheric models for climate studies. *Atmospheric Science Letters*, 21, e952.

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

Koster, R. D., Dirmeyer, P. A., Guo, Z. C., Bonan, G., Chan, E., Cox, P., Gordon, C. T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C. H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., Sud, Y. C. Taylor, C. M., Verseghy, D., Vasic, R., Xue, Y. K., Yamada, T. (2004). Regions of strong coupling betweeen soil moisture and precipitation. *Science* 305, 1138-1140.

Larson, V.E., Schanen, D.P., Wang, M., Ovchinnikov, M., Ghan, S., (2012). pdf Parameterization of Boundary Layer Clouds in Models with Horizontal Grid Spacings from 2 to 16 km. *Monthly Weather Review*, 140, 285-306.

LeTreut, H., Li, Z.X. (1991). Sensitivity of an atmospheric general circulation model to prescribed SST changes: Feedback effects associated with the simulation of cloud optical properties. *Climate Dynamics*, 5, 175-187.

Lewellen, W.S, Yoh, S. (1993). Binormal model of ensemble partial cloudiness. *Journal of the Atmospheric Sciences*, 50, 1228-1237.

Lin, Y., Mitchell, K.E. (2005). The NCEP stage II/IV hourly precipitation analyses: development and applications. *19th Conf.Hydrology*, available at: https://ams.confex.com/ams/pdfpapers/83847.pdf.

Liu, S, Liang, X.Z., (2010). Observed diurnal cycle climatology of planetary boundary layer height. *Journal of Climate*, 23, 5790-2807.

Lock, A.P, Brown, A.R., Bush, M.R., Martin, G.M., Smith, R.N.B. (2000). A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests. *Monthly Weather Review*, 128, 3187-3199.

Lohmann, U., McFarlane, N., Levkov, L., Abdella, K., Albers, F. (1999). Comparing different cloud schemes of a single column model by using mesoscale forcing and nudging technique. *Journal of Climate*, 12, 438-461.

Ma, H.-Y. and coauthors. (2018). CAUSES: On the role of surface energy budget errors to the warm surface air temperature error over the Central United States. *Journal of Geophysical Research*, 123, 2888-2909.

Mellor, G.L., Yamada, T. (1982). Development of a turbulence closure model for geophysical fluid problems. *Reviews Geophysics and Space Physics*, 20 , 851-875.

Morcrette, C.J. and coauthors (2018). Introduction to CAUSES: Description of weather and climate models and their near-surface temperature errors in 5 day hindcasts near the Southern Great Plains. *Journal of Geophysical Research*, 123, 2655-2683.

Nakanishi M. (2001). Improvement of the Mellor-Yamada turbulence closure model based on large-eddy simulation data. *Boundary Layer Meteorol.*, 99, 349-378.

Osman, M.K., Turner, D.D., Heus, T., Newsom, R. (2018). Characteristics of water vapor turbulence profiles in convective boundary layers during the dry and wet seasons over Darwin. *Journal of Geophysical Research*, 123, 4818-1836.

Platnick, S., Meyer, K. G., King, M. D., Wind, G., Amarasinghe, N., Marchant, B., Arnold, G. T., Zhang, Z., Hubanks, P. A., Holz, R. E., Yang, P., Ridgway, W. L., Riedi, J. (2017). The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples From Terra and Aqua. *IEEE Transactions on Geoscience and Remote Sensing*, 55, 502-525, 2017.

Price, J.D. (2006). A study of probability distributions of boundary-layer humidity and associated errors in parametrized cloud-fraction. *Quarterly Journal of the Royal Meteorological Society*, 127, 739-758.

Ricard, J.L., J.F. Royer. (1993). a statistical cloud scheme for use in an AGCM. *Ann. Geophyscicae*, 11, 1095-1115.

Shi, Y., Long, C.N. (2002). Best estimate radiation flux value added product: Algorithm operational details and explanations. *Technical Report DOE/SC-ARM/TR-008*. US Department of Energy. Available online at: http://www.arm.gov/publications/tech reports/doe-sc-arm-tr-008.pdf.

Sivaraman, C., McFaralane, S., Chapman, E., Jensen, M.P., Toto, T. (2013). Planetary Boundary Layer (PBL) Height Value Added Product (VAP): Radiosonde Retrievals *Technical Report DOE/SC-ARM/TR-132*. US Department of Energy. Available online at: http://www.arm.gov/publications/tech reports/doe-sc-arm-tr-132.pdf.

Smagorinsky, J. (1963). General circulation experiments with the primitive equations. Part I: The basic experiment. *Monthly Weather Review*, 91, 99-164.

Smith, R.N.B.. (1990). A scheme for predicting layer clouds and their water content in a general circulation model. *Quarterly Journal of the Royal Meteorological Society*, 116, 435-460.

Stoffel, T. (2005). Solar Infrared Radiation Station (SIRS) Handbook. *Technical Report DOE/SC-ARM/TR-025*. US Department of Energy. Available online at: https://www.arm.gov/publications/tech reports/handbooks/sirs handbook.pdf?id=68.

Sundqvist, H. (1978). A parameterization scheme for non-convective condensation including prediction of cloud water content. *Quarterly Journal of the Royal Meteorological Society*, 104, 677-690.

Tiedtke, M. (1993). Representation of clouds in large-scale models. *Monthly Weather Review*, 121, 3040-3061.

Tompkins, A.M. (2002). A prognostic parameterization for the subgrid-scale variability of water vapor and clouds in large-scale models and its use to diagnose cloud cover. *Journal of the Atmospheric Sciences*, 59, 1917-1942.

Tompkins, A.M. (2003). Impact of temperature and humidity variability on cloud cover assessed using aircraft data. *Quarterly Journal of the Royal Meteorological Society*, 129, 2151-2170.

Toto, T., Jensen, M.P. (2016). Interpolated Sounding and Gridded Sounding Value - Added Products. *Technical Report DOE/SC-ARM/TR-183*. US Department of Energy. Available online at: http://www.arm.gov/publications/tech reports/doe-sc-arm-tr-183.pdf.

Turner D.D., Clough, S.A., Liljegren, J.C., Clothiaux, E.E., Cady-Pereira, K.E., Gaustad, K.L. (2007). Retrieving liquid water path and precipitable water vapor from the Atmospheric Radiation Measurement (ARM) microwave radiometers. *IEEE: Transactions on Geoscience and Remote Sensing*, 45, 3680-3690.

Turner D.D., Wulfmeyer V., Berg L.K., Schween J.H. (2014). Water vapor turbulence profiles in stationary continental convective mixed layers. *Journal of Geophysical Research*, 119, 11151-11165.

Van Weverberg, K., Boutle, I.A., Morcrette, C.J., Newsom, R.K. (2016). Towards retrieving critical relative humidity from ground-based remote-sensing observations. *Quarterly Journal of the Royal Meteorological Society*, 142, 2867-2881.

Van Weverberg, K., and coauthors. (2018). CAUSES: Attribution of surface radiation biases in NWP and climate models near the U.S. Southern Great Plains. *Journal of Geophysical Research*, 123, 3612-3644.

Van Weverberg, K., Morcrette, C.J., Boutle, I.A. (2020). Bi-Modal Diagnostic Cloud Fraction Parameterization. Part I: Motivating Analysis and Scheme Description. *Monthly Weather Review*, submitted.

Walters, D. and coauthors (2017). The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations. *Geoscientific Model Development*, 10, 1487-1520.

Walters, D. and coauthors (2019). The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations. *Geoscientific Model Development*, 12, 1909-1963

Westcott, N.E., Knapp, H.V., Hilberg, S.D. (2008). Comparison of gage and multi-sensor precipitation estimates over a range of spatial and temporal scales in the Midwestern United States. *Journal of Hydrology*, 351, 1-12.

Wilson, D.R., Bushell, A.C., Kerr-Munslow, A.M., Price, J.D., Morcrette, C.J. (2008). PC2: A prognostic cloud fraction and condensation scheme. I: Scheme description. *Quarterly Journal of the Royal Meteorological Society*, 134, 2093-2107.

Wood, R., Field, P.R. (2000). Relationships between total water, condensed water, and cloud fraction in stratiform clouds examined using aircraft data. *Journal of the Atmospheric Sciences*, 57, 1888-1905.

Wulfmeyer, V., Muppa, S.K., Behrendt, A., Hammann, E., Spaeth, F., Sorbjan, Z., Turner, D.D., Hardesty, R.M. (2016). Determination of Convective Boundary Layer Entrainment Fluxes, Dissipation Rates, and the Molecular Destruction of Variances: Theoretical Description and a Strategy for Its Confirmation with a Novel Lidar System Synergy. *Journal of the Atmospheric Sciences*, 73, 667-692.

Xie, S., and coauthors. (2010). ARM climate modeling best estimate data: A new data product for climate studies. *Bulletin of the American Meteorological Society*, 91, 13-21.

# LIST OF TABLES

TABLE 1. Experiment overview. Apart from the cloud scheme configuration, these experiments have identical settings as indicated in the text.

| Experiment | Cloud Scheme Configuration |
|---|---|
| NOCF | No cloud fraction scheme |
| NOEACF | Smith cloud fraction scheme |
| RA2M | Smith cloud fraction scheme + operational adjustments (EACF + ACF) |
| SMITH-TKE | Smith cloud fraction scheme + turbulence-based variances |
| PC2-TKE | PC2 cloud fraction scheme + turbulence-based variances |
| BM | Bimodal cloud fraction scheme |

TABLE 2. Surface radiation statistics averaged over the 27 April case in all 6 model configurations. Shown are the Root-mean Squared Error (RMSE) and bias for the downwelling shortwave and downwelling longwave radiation. The best and worst statistic values for each variable are highlighted in boldface and italic respectively.

| Experiment | SW Bias W m$^{-2}$ | SW RMSE W m$^{-2}$ | LW Bias W m$^{-2}$ | LW RMSE W m$^{-2}$ |
|---|---|---|---|---|
| NOCF | *96.4* | 196.0 | -21.0 | *38.7* |
| NOEACF | 87.4 | *212.6* | -17.3 | 25.1 |
| RA2M | 57.9 | 192.3 | -13.0 | 20.5 |
| SMITH-TKE | 79.8 | 194.2 | -17.5 | 26.6 |
| PC2-TKE | 95.6 | 205.4 | *-23.1* | 33.5 |
| BM | **16.9** | **13.2** | **-8.8** | **19.5** |

TABLE 3. Cloud evaluation statistics for the entire MC3E period and for all 6 model configurations. Shown are the Root-mean Squared Error (RMSE) and bias for the 3D cloud fraction (CF) and liquid water content (LWC), 2D low- (LOW CLD), mid- (MID CLD) and high-level (HIGH CLD) cloud cover, total cloud cover (TOTAL CLD), liquid water path (LWP), cloud base height (CLD BASE), and low cloud depth (CLD DEPTH). Note that 3D cloud fraction statistics include all grid points that are cloudy in any of the 6 configurations or the observations. All data were filtered for the occurrence of precipitation in any of the simulations or the observations. The best and worst statistic values for each variable are highlighted in boldface and italic respectively.

| Experiment | NOCF | NOEACF | RA2M | SMITH-TKE | PC2-TKE | BM |
|---|---|---|---|---|---|---|
| CF BIAS (%) | *-6.8* | -4.5 | **-2.8** | -5.7 | -6.5 | -4.3 |
| CF RMSE (%) | *28.9* | **27.0** | 27.4 | 27.5 | 27.4 | 27.9 |
| LWC BIAS ($10^{-6} kgkg^{-1}$) | *-10.2* | -3.0 | -3.1 | -7.6 | -7.7 | **-0.1** |
| LWC RMSE ($10^{-6} kgkg^{-1}$) | 70.0 | 69.8 | 67.9 | **65.3** | 66.0 | *81.4* |
| LOW CLD BIAS (%) | *-9.6* | -6.7 | -3.8 | -7.8 | -8.4 | **-3.3** |
| LOW CLD RMSE (%) | *38.1* | **33.2** | 33.3 | 34.7 | 34.5 | 33.3 |
| MID CLD BIAS (%) | **1.7** | *-5.1* | -4.5 | -4.0 | -1.9 | -3.3 |
| MID CLD RMSE (%) | *34.6* | **28.9** | 29.0 | 30.5 | 30.5 | 29.3 |
| HIGH CLD BIAS (%) | 3.0 | *-12.4* | -10.7 | -8.3 | **-2.0** | -10.3 |
| HIGH CLD RMSE (%) | *46.4* | 42.3 | 41.9 | 42.6 | 43.1 | **41.6** |
| TOTAL CLD BIAS (%) | **-5.9** | *-15.9* | -12.3 | -14.3 | -9.0 | -12.3 |
| TOTAL CLD RMSE (%) | 44.8 | 45.0 | 44.0 | *45.8* | 45.7 | **44.0** |
| LWP BIAS ($10^{-3} kgm^{-2}$) | *-16.9* | -5.3 | -5.1 | -13.6 | -12.4 | **-0.7** |
| LWP RMSE ($10^{-3} kgm^{-2}$) | *105.1* | 72.2 | 74.9 | 72.5 | **70.3** | 90.8 |
| CLD BASE BIAS (m) | 149.9 | -389.1 | *-422.8* | -151.5 | **-109.3** | -297.9 |
| CLD BASE RMSE (m) | **590.3** | 762.0 | *792.1* | 664.4 | 627.9 | 712.9 |
| CLD DEPTH BIAS (m) | -489.5 | 397.1 | *466.9* | -181.2 | -169.3 | **-50.2** |
| CLD DEPTH RMSE (m) | 768.4 | 885.7 | *942.7* | 639.3 | **624.3** | 683.8 |

TABLE 4. Domain-average diurnal precipitation for the 1 km domain for NCEP Stage IV observations and the 6 model configurations for the entire simulation period.

| Experiment | Daily Precipitation (mm $day^{-1}$) |
|---|---|
| OBS | 3.36 |
| NOCF | 3.69 |
| NOEACF | 3.82 |
| RA2M | 3.76 |
| SMITH-TKE | 3.70 |
| PC2-TKE | 3.84 |
| BM | 3.38 |

TABLE 5. Statistics of the Cloud Optical Thickness ($COT$) and Water Path ($WP$) for MODIS and the 6 model configurations for each of the MODIS overpasses during the 6-week MC3E period. Data for the full domain was included and simulated $COT$ and $WP$ were obtained through the COSP diagnostics. Shown are the frequency of occurrence (FOO) of all cloud ($COT > 0.3$), liquid-phase cloud and ice-phase cloud, the arithmetic domain-averaged $COT$, liquid-phase $COT$ and ice-phase $COT$, the extinction-weighted domain averaged $COT$, liquid-phase $COT$ and ice-phase $COT$ and the domain-averaged $WP$, liquid-phase $WP$ and ice-phase $WP$. The largest and smallest values for each variable is highlighted in boldface and italic respectively. Observational uncertainty is provided for the $COT$ and $WP$.

| Experiment | FOO all % | FOO liq % | FOO ice % | $WP$ all kg m$^{-2}$ | $WP$ liq kg m$^{-2}$ | $WP$ ice kg m$^{-2}$ |
|---|---|---|---|---|---|---|
| observations | 46.1 | 28.3 | *17.8* | *0.11* ±0.04 | 0.12 ±0.03 | *0.36* ±0.17 |
| NOCF | *31.8* | *8.6* | 23.2 | 0.18 | **0.13** | 0.74 |
| NOEACF | 49.1 | 25.8 | 23.3 | 0.20 | 0.09 | 0.75 |
| RA2M | 49.2 | 27.1 | 22.1 | 0.20 | 0.09 | 0.80 |
| SMITH-TKE | 43.1 | 18.7 | **24.4** | 0.19 | 0.08 | 0.71 |
| PC2-TKE | **58.0** | **36.7** | 21.3 | 0.19 | *0.06* | 0.77 |
| BM | 50.2 | 27.5 | 22.7 | **0.22** | 0.11 | **1.02** |

| Experiment | $\overline{arCOT}$ all | $\overline{arCOT}$ liq | $\overline{arCOT}$ ice | $\overline{exCOT}$ all | $\overline{exCOT}$ liq | $\overline{exCOT}$ ice |
|---|---|---|---|---|---|---|
| observations | *10.2* ±2.1 | 18.6 ±2.3 | *23.1* ±7.3 | 0.53 | 1.99 | 1.91 |
| NOCF | 10.3 | **25.6** | 34.2 | *0.33* | **3.68** | 2.11 |
| NOEACF | 13.1 | 17.9 | 35.9 | 0.48 | 1.67 | *1.53* |
| RA2M | **13.6** | 18.8 | **38.1** | 0.48 | *1.58* | 1.68 |
| SMITH-TKE | 11.6 | 17.5 | 34.2 | 0.41 | 1.65 | 1.66 |
| PC2-TKE | 12.2 | *11.7* | 37.5 | **0.63** | 1.71 | **2.18** |
| BM | 13.0 | 17.4 | 36.3 | 0.50 | 1.72 | 1.70 |

TABLE 6. Surface radiation statistics averaged over the entire 6 week MC3E period in all 6 model configurations. Shown are the Root-mean Squared Error (RMSE) and bias for the downwelling shortwave and downwelling longwave radiation. The best and worst statistic values for each variable are highlighted in boldface and italic respectively.

| Experiment | SW Bias W $m^{-2}$ | SW RMSE W $m^{-2}$ | LW Bias W $m^{-2}$ | LW RMSE W $m^{-2}$ |
|---|---|---|---|---|
| NOCF | *24.4* | *137.0* | *-7.6* | *22.2* |
| NOEACF | 12.4 | 126.3 | -5.4 | **20.5** |
| RA2M | **5.2** | **126.1** | **-3.2** | 21.4 |
| SMITH-TKE | 16.1 | 127.9 | -5.7 | 21.2 |
| PC2-TKE | 15.6 | 126.8 | -6.4 | 21.5 |
| BM | 6.9 | 127.6 | -3.8 | 20.8 |

# LIST OF FIGURES

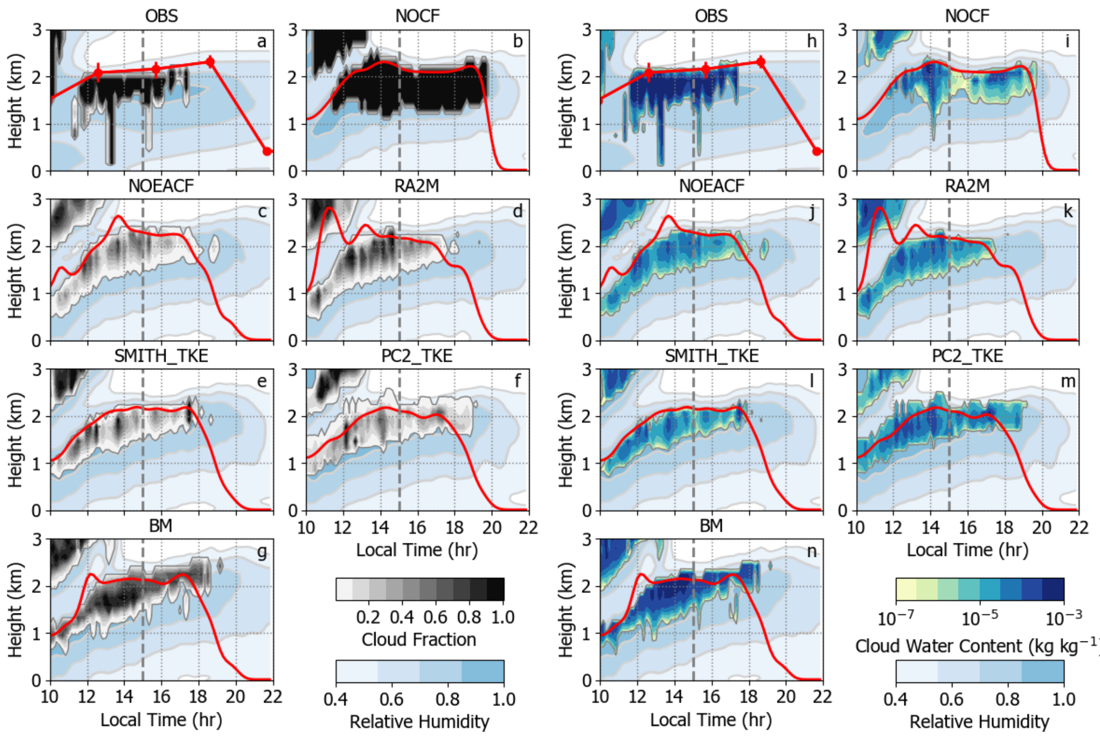



FIG. 1. Time-height cross sections of *CF* (left) and water content (right) as observed (a and h) and simulated using the NOCF (b and i), NOEACF (c and j), RA2M (d and k), SMITH-TKE (e and l), PC2-TKE (f and m) and BM (g and n) configurations for the stratocumulus case of 27 April 2011 at the location of the Southern Great Plains Central Facility in Oklahoma. Also plotted are the observed (from the interpolated soundings) and simulated relative humidity using blue shading in the background. The red lines provide the boundary-layer height as observed (average of four methods for the observations, with variability between the methods provided as error bars) and as simulated by all experiments. The gray dashed vertical line indicates the time of the MODIS overpass shown in Figure 2. The *CF* in panel a is derived from ARSCL, and the *LWC* in panel h is obtained from Microbase.

.

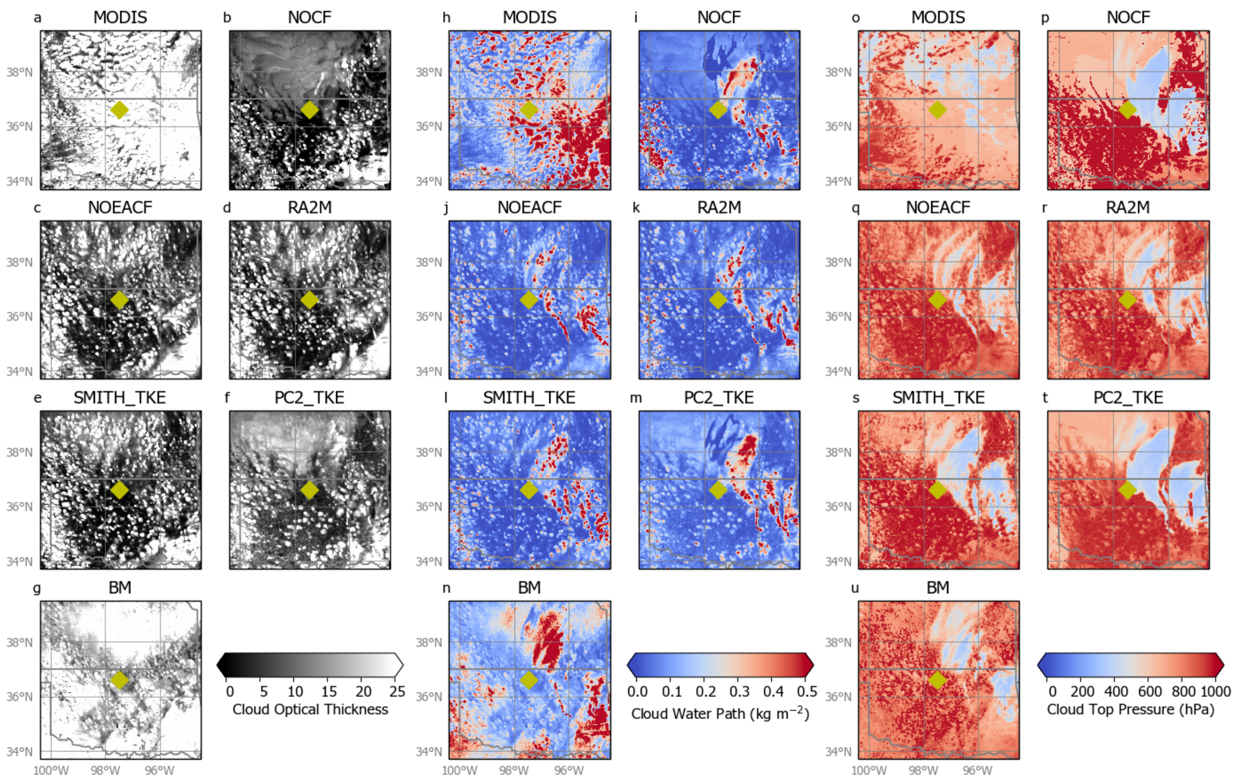



FIG. 2. Cloud optical thickness (left), cloud water path (middle) and cloud top pressure (right) at 1500 LT as observed by Aqua MODIS (a, h and o) and as a simulated using the NOCF (b, i and p), NOEACF (c, j and q), RA2M (d, k and r), SMITH-TKE (e, l and s), PC2-TKE (f, m and t) and BM (g, n and u) configurations for the stratocumulus case of 27 April 2011 for the entire 1-km grid-spacing domain. State boundaries are denoted by gray solid lines, and the location of the Southern Great Plains Central Facility is highlighted with the yellow diamond. Simulations have been run through COSP to provide synthetic optical thickness and water paths as would be observed by MODIS.

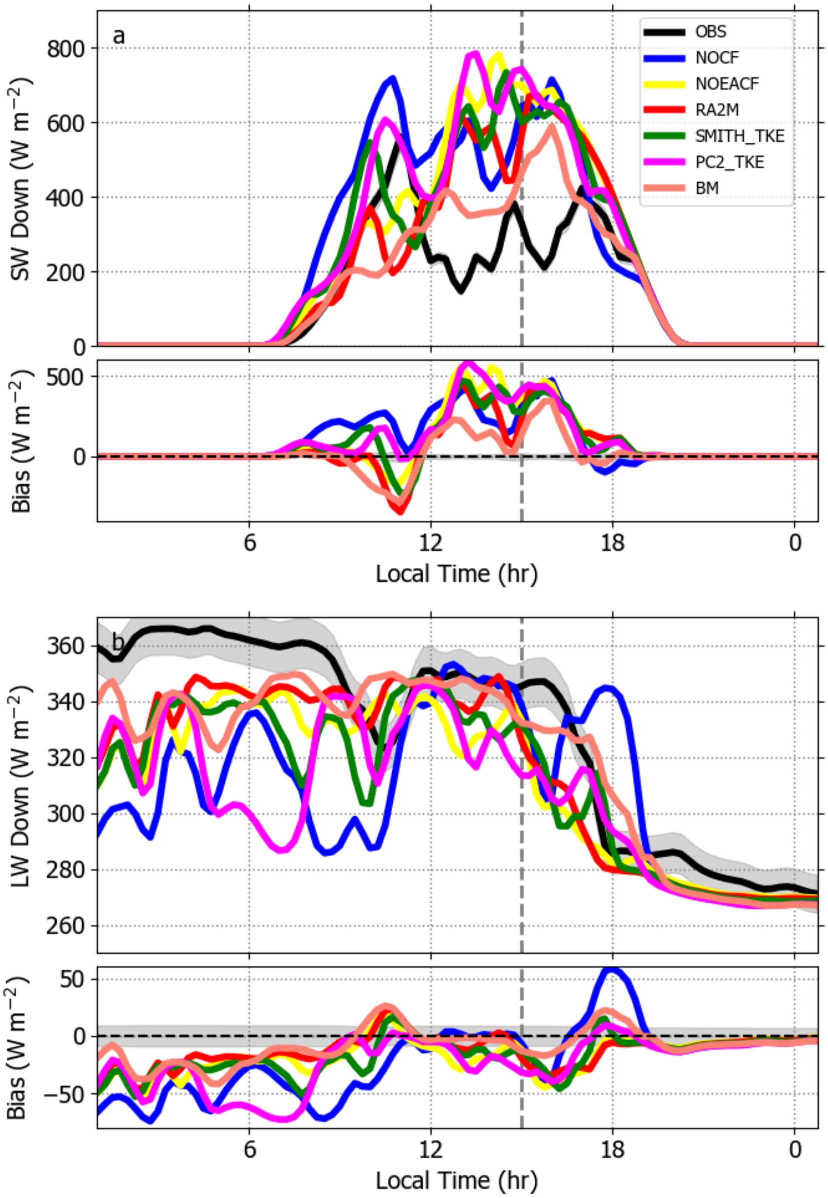
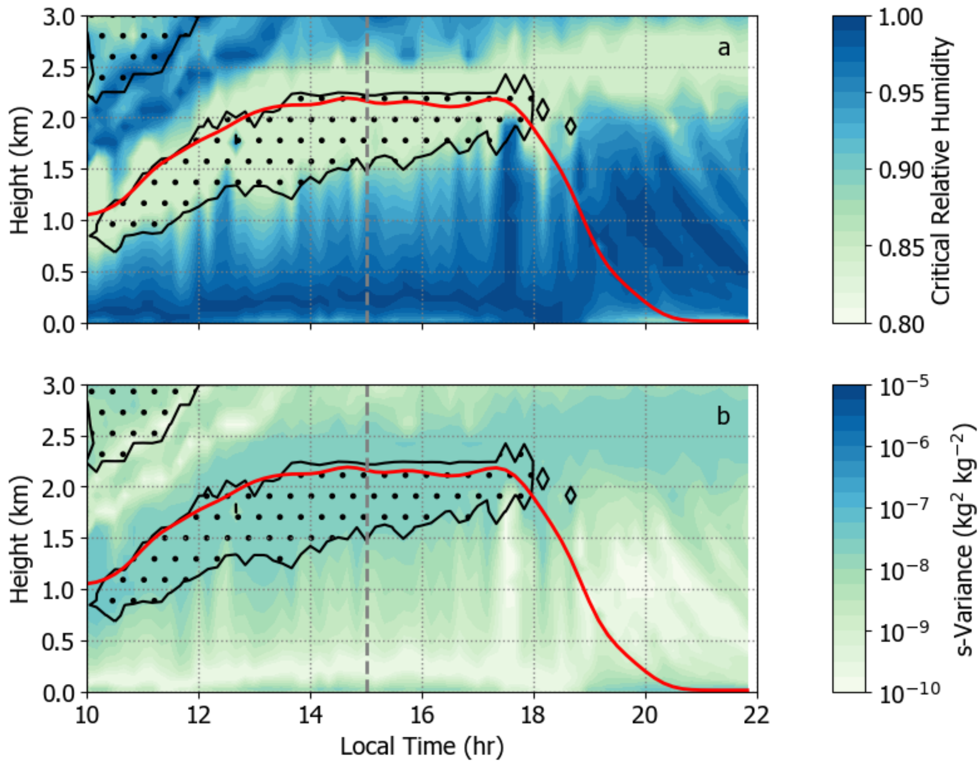



FIG. 3. Diurnal cycle of observed (ARM Best-Estimate, black) and simulated downwelling shortwave (a) and longwave (b) surface radiation for the stratocumulus case of 27 April 2011 at the Southern Great Plains Central Facility. Also shown is the absolute bias of the downwelling radiation against the observations. The gray shading denotes the uncertainty in the observations. Observed and simulated data have a 15 min frequency and a Gaussian filter was applied to the observations and the simulations to filter out small-scale noise. The gray dashed vertical line indicates the time of the MODIS overpass shown in Figure 2.




1011 FIG. 4. Time-height cross section of the critical relative humidity (a) and s-variance (b) for the SMITH-TKE configuration for the stratocumulus case of 27 April 2011. The dotted shaded area denotes *CF* larger than 1% and red solid line indicates the top of the mixed-layer. The gray dashed vertical line indicates the time of the MODIS overpass shown in Figure 2.

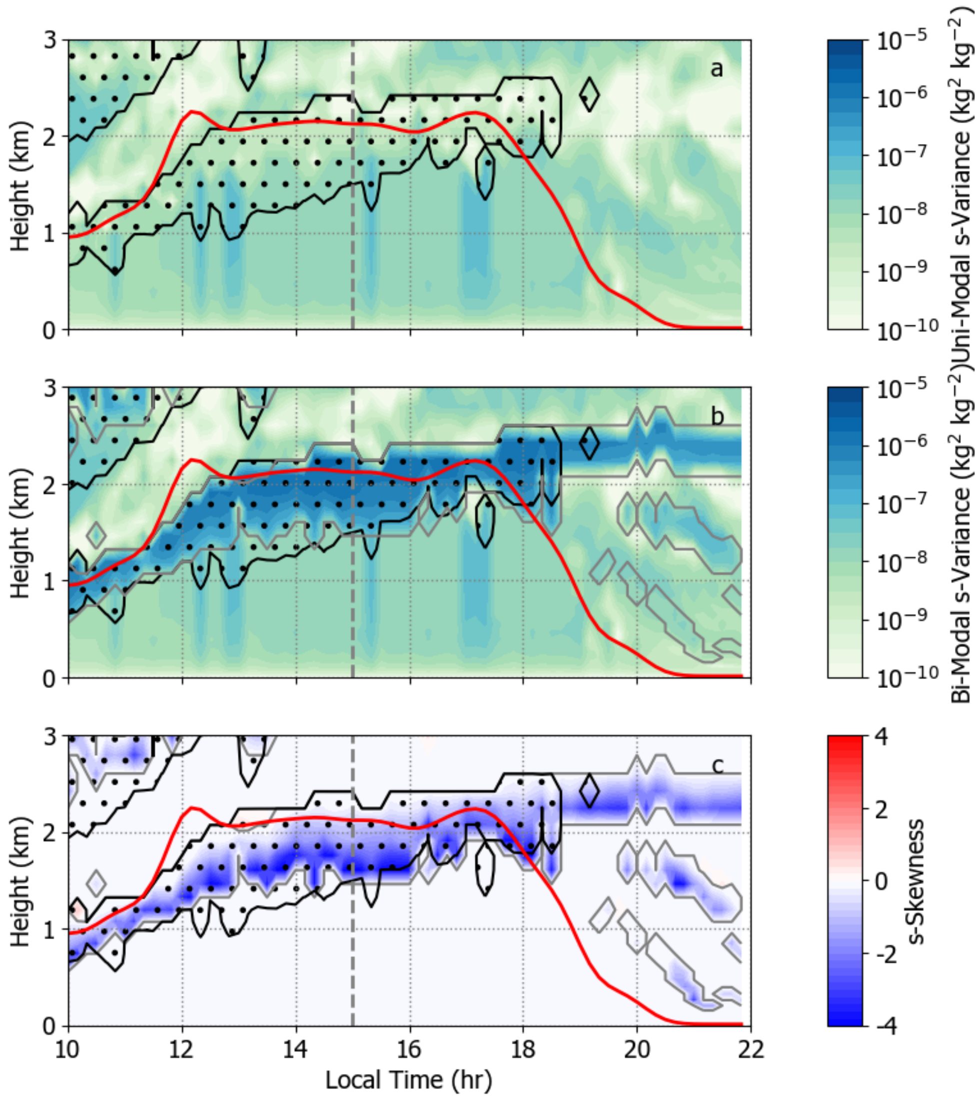



FIG. 5. Time-height cross section of the local turbulence-based unimodal variance (a), the bimodal mixture variance (b) and bimodal mixture skewness (c) for the BM configuration for the stratocumulus case of 27 April 2011. The solid gray contour in panels b and c shows the diagnosed entrainment zone, while the dotted shaded areas denote *CF* larger than 1%. The red solid line indicates the top of the mixed-layer and the gray dashed vertical line indicates the time of the MODIS overpass shown in Figure 2.

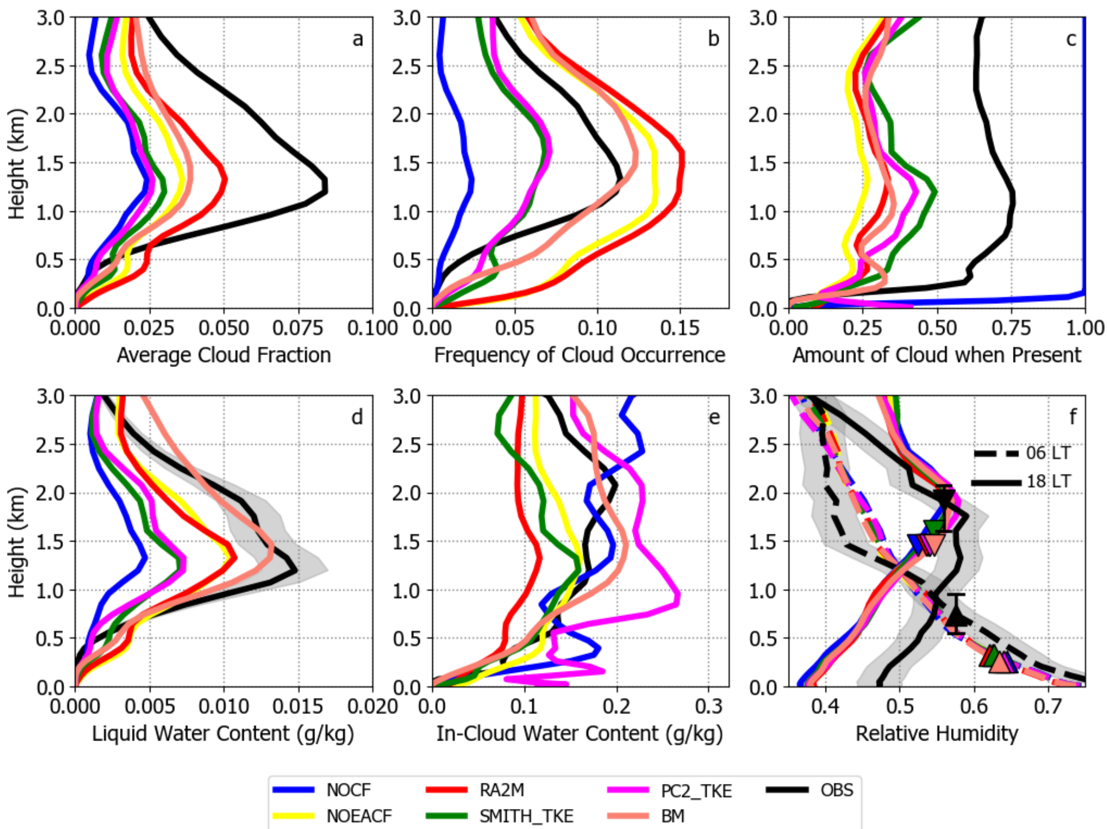
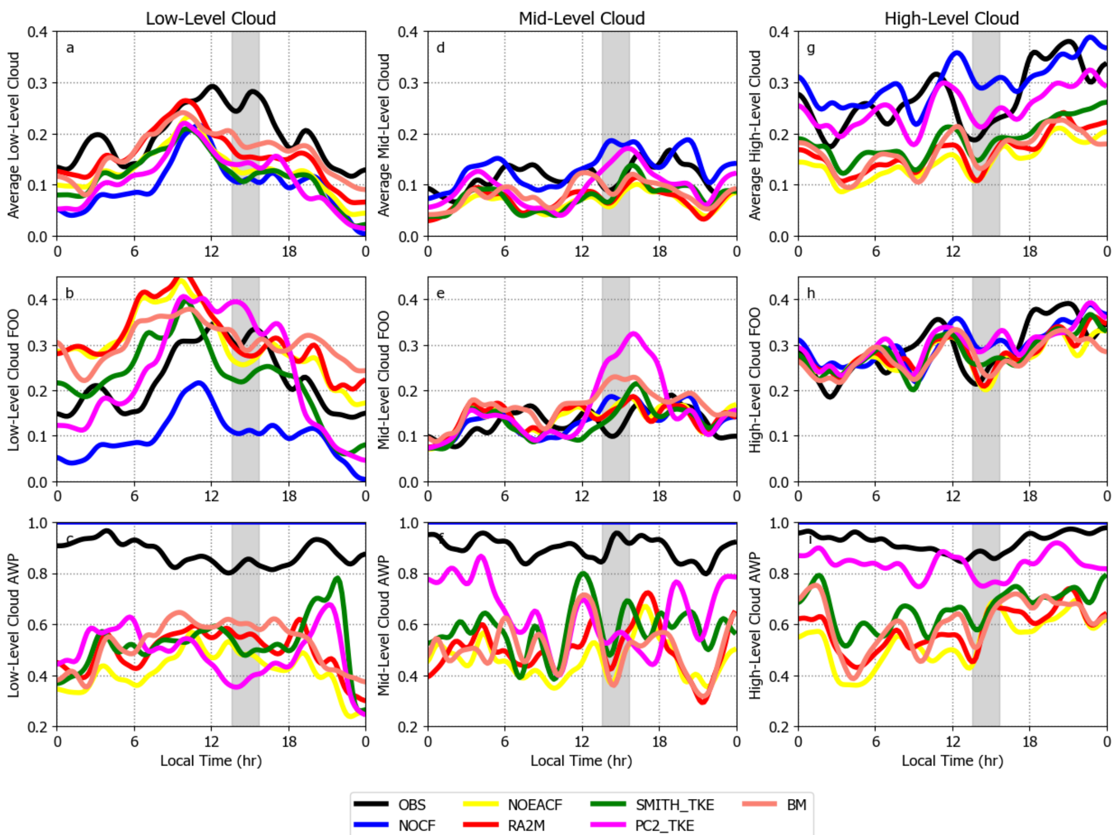



FIG. 6. Average vertical profiles of *CF* (a), frequency of cloud occurrence (FOO; b), amount of cloud when present (AWP; c), liquid water content (d) and in-cloud liquid water content (e) averaged over the entire 6 weeks of the MC3E campaign for rain-free times at the location of the Southern Great Plains Central Facility, as retrieved from ARSCL and Microbase (black lines) and as simulated by all model experiments. FOO and AWP are calculated based on any non-zero cloud occurrence. Panel (f) denotes vertical profiles of total relative humidity around 0630 LT (dashed lines) and 1830 LT (solid lines) as observed from soundings and as simulated, averaged over all balloon launch times during the MC3E campaign. Downward and upward triangles denote the average boundary layer height for the 1830 and 0630 LT profile respectively. The gray shading in the water content and the relative humidity profiles represents the observational uncertainty.




1029 FIG. 7. Diurnal cycles of average (top), frequency (FOO; middle) and amount when present (AWP; bottom) of low cloud (left), mid-level cloud (middle) and high-level cloud (right) averaged over all rain-free times during the 6 weeks of the MC3E campaign for the observations and the 6 model configurations. Cloud cover is calculated from ARSCL cloud locations regridded to the model levels for the observations and from the bulk *CF* in the simulations, assuming random-maximum overlap. The boundary between low and mid-level cloud is 3000 m and the boundary between mid- and high-level cloud is 6000 m. FOO and AWP are calculated based on any non-zero cloud occurrence. Observations and simulations were available with a 10 min frequency for the entire 6 week period. The gray shaded vertical band in each plot denotes the time-range of the MODIS overpasses. A Gaussian filter was applied to the observations and the simulations to filter out small-scale noise.

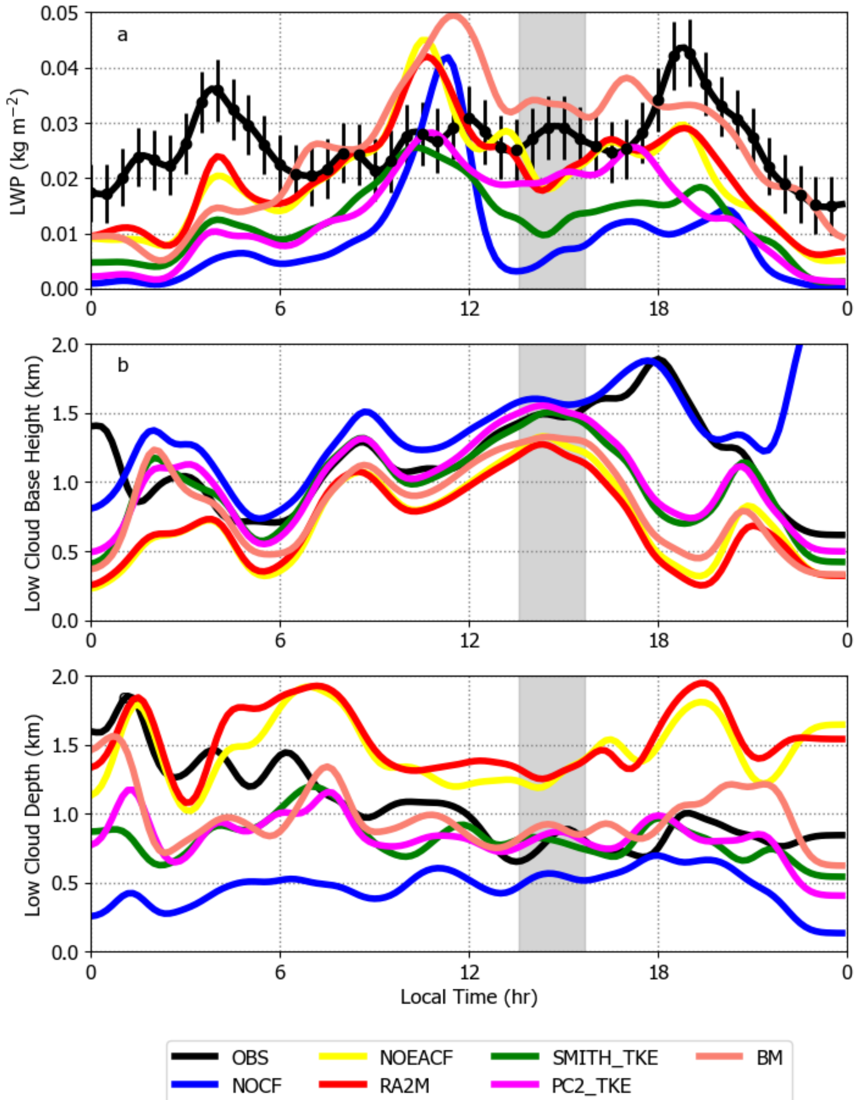



FIG. 8. Diurnal cycles of liquid water path (*LWP*; a), low cloud base height (b), and low cloud depth (c), averaged over all rain-free times during the 6 weeks of the MC3E campaign for the observations and the 6 model configurations. Cloud base and and cloud depth are based on any non-zero cloud occurrence and are averaged for clouds with a cloud base lower than 3000 m only. The *LWP* is obtained from the Microwave Radiometer (MWR) and vertical error bars denote the observational uncertainty. Cloud base height and cloud depth are obtained from ARSCL cloud locations. Observations and simulations were available with a 10 min frequency for the entire 6 week period. The gray shaded vertical band in each plot denotes the time-range of the MODIS overpasses. A Gaussian filter was applied to the observations and the simulations to filter out small-scale noise.

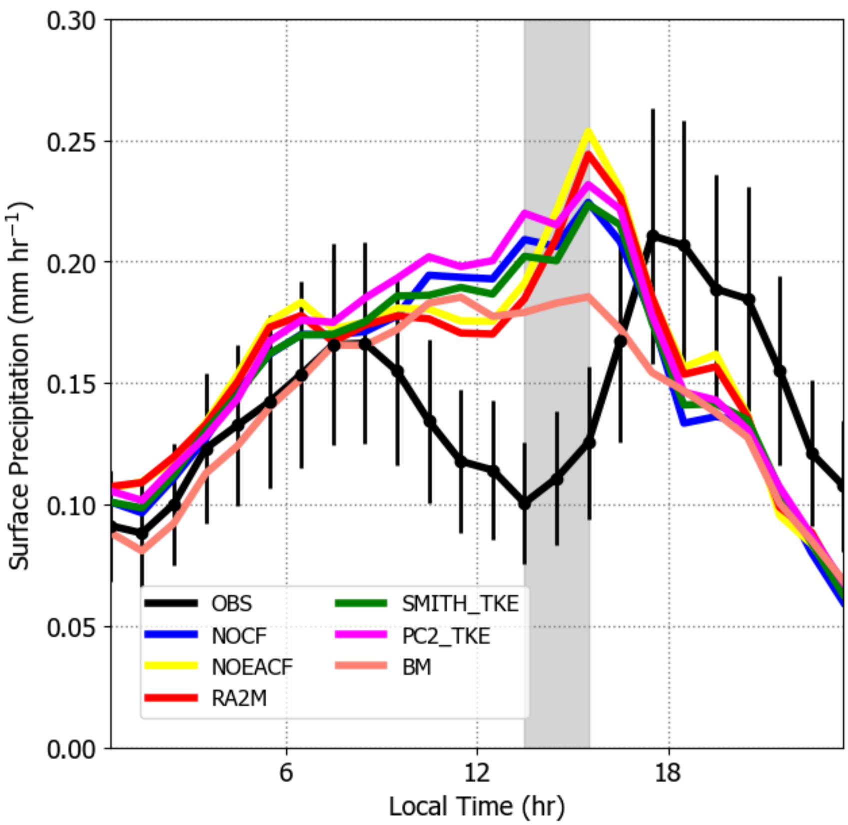



FIG. 9. Diurnal cycles of domain-average surface precipitation as observed from the NCEP Stage IV rain gauge radar merged product, and as simulated by the 6 model configurations, averaged over the 6 weeks of the MC3E campaign. Vertical error bars on the observed diurnal cycle denote the observational uncertainty. Observations and simulations were available with a 10 min frequency for the entire 6 week period. The gray shaded vertical band in each plot denotes the time-range of the MODIS overpasses.

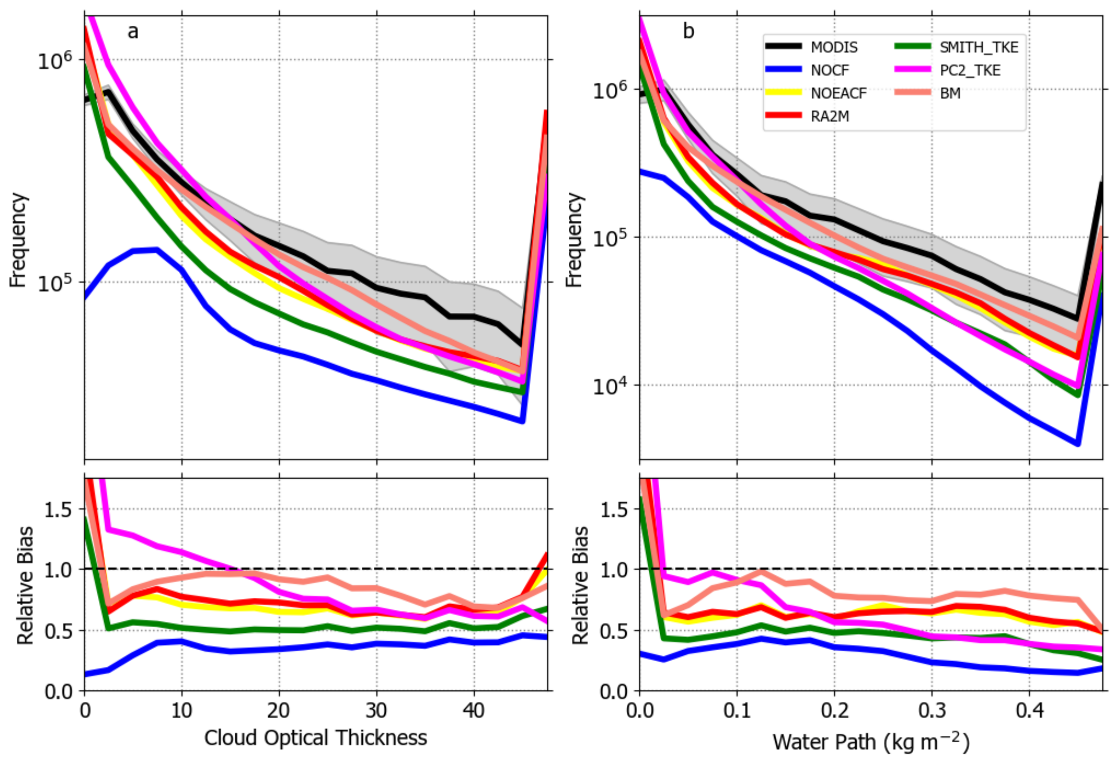



FIG. 10. Histograms of the liquid cloud optical thickness (a) and liquid water path (b) for the entire 1-km domain from MODIS (regridded to the model grid) and the 6 model configurations. Simulated values are obtained using the COSP algorithm to simulate how the model fields would be perceived from MODIS. Histograms only include cloud areas identified as liquid cloud phase by the MODIS/COSP retrieval algorithm and contain all MODIS overpasses over the SGP that occurred during the 6-weeks of the MC3E campaign, once a day, typically between 1400 and 1500 LT. Shown are the absolute frequencies of the histograms (top) as well as the relative bias of the simulations compared to the observations (bottom). The gray shaded area denotes the observational uncertainty, obtained as explained in the text.

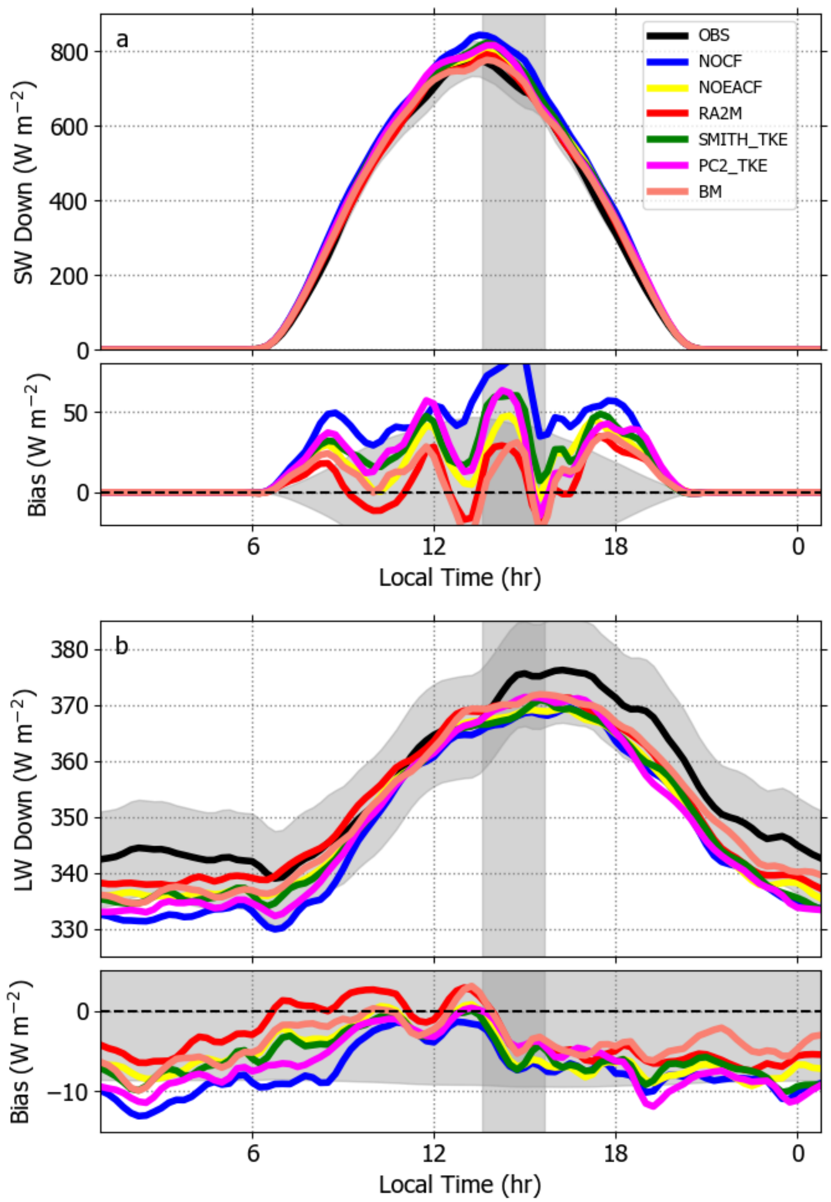



FIG. 11. Diurnal cycle of observed (ARM Best-Estimate; black) and simulated downwelling shortwave (a) and longwave (b) surface radiation averaged over the 6 weeks of the MC3E campaign at the SGP Central Facility. Also shown is the absolute bias of the downwelling radiation against the observations. The gray shading denotes the uncertainty in the observations. Observed and simulated data have a 15 min frequency and a Gaussian filter was applied to the observations and the simulations to filter out small-scale noise. The gray shaded vertical band in each plot denotes the time-range of the MODIS overpasses.

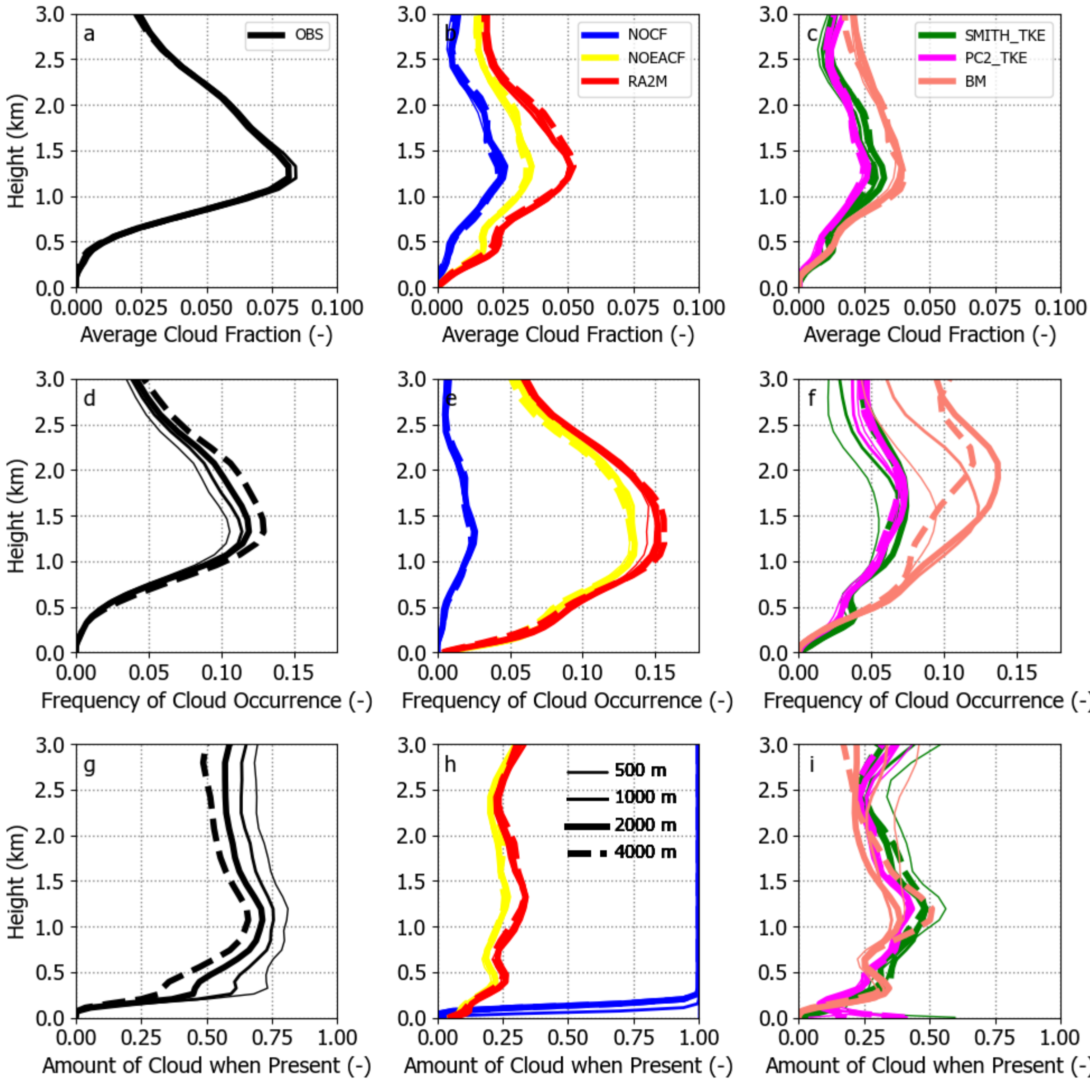



FIG. 12. Averaged vertical profiles of *CF* (top), frequency of cloud occurrence (middle) and amount of cloud when present (bottom), as observed (derived from ARSCL; left), as simulated by the NOCF, NOEACF and RA2M configurations (middle) and as simulated by the SMITH-TKE, PC2-TKE and BM configurations (right). Profiles are averaged over all non-rainy output times (10 min frequency) for the entire 6 weeks of the MC3E campaign at the location of the SGP Central Facility. Profiles are shown for the 4 km (thick dashed line), 2 km (thick solid line), 1 km (intermediate thick line) and 0.5 km (thin line) grid spacings.

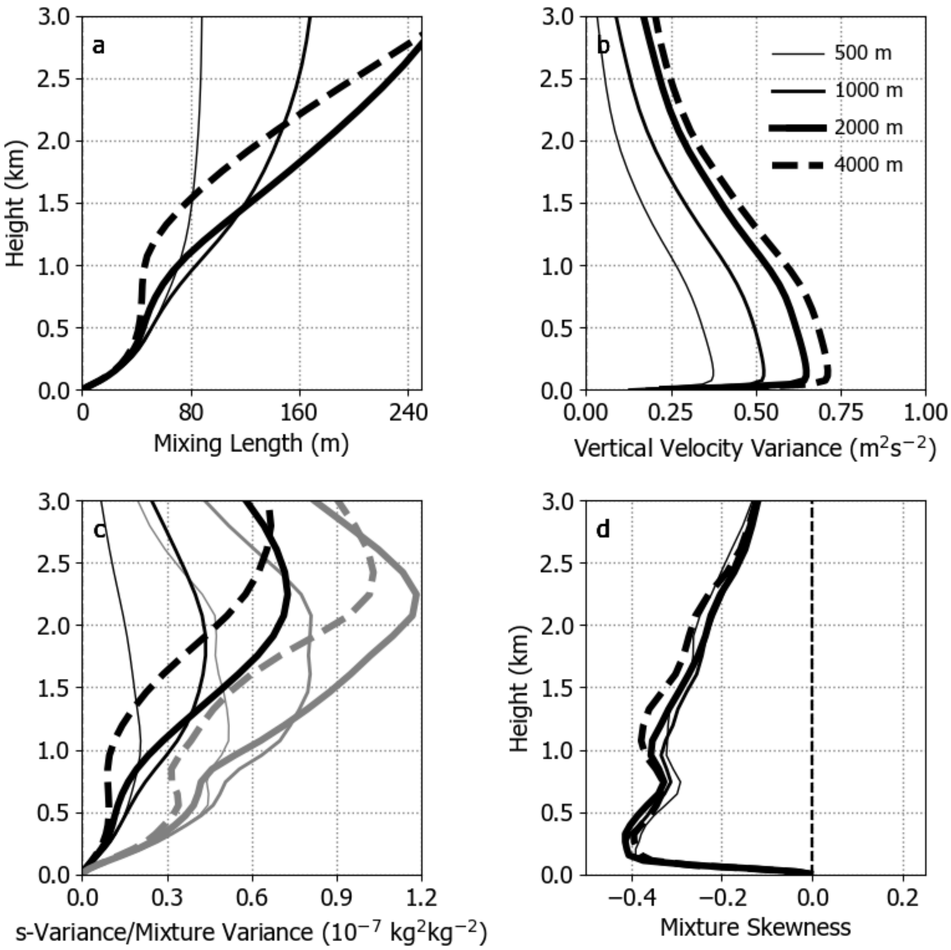
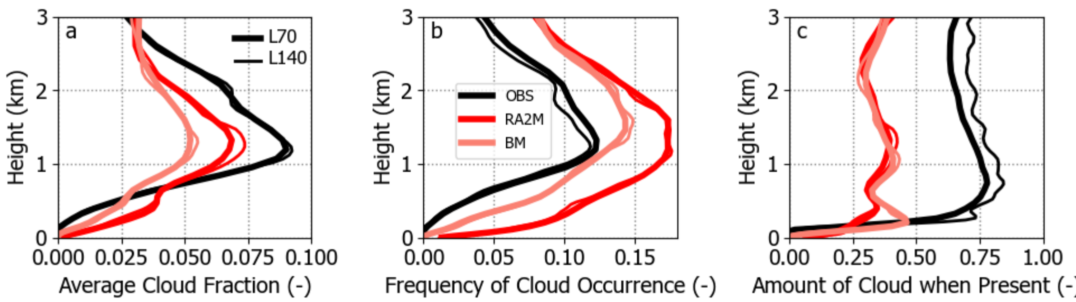



FIG. 13. Averaged vertical profiles of the blended mixing length (a), turbulent kinetic energy (b), local turbulence-based unimodal variance (black lines; c) and the bimodal mixture variance (gray lines; c) and bimodal mixture skewness (d) as simulated in the BM configuration. Profiles are averaged over all output times (10 min frequency) for the entire 6 weeks of the MC3E campaign at the location of the SGP Central Facility. Profiles are shown for the 4 km (thick dashed line), 2 km (thick solid line), 1 km (intermediate thick line) and 0.5 km (thin line) grid spacings.




1077 FIG. 14. Averaged vertical profiles of *CF* (left), frequency of cloud occurrence (middle) and amount of cloud
1078 when present (right), as observed (derived from ARSCL; black) and as simulated by the RA2M and BM. Profiles
1079 are averaged over all non-rainy output times (10 min frequency) for the entire 6 weeks of the MC3E campaign
1080 at the location of the SGP Central Facility. Profiles are shown for the L140 (thin line), and the L70 (thick line)
1081 vertical level sets. Horizontal grid spacing for all profiles is 1 km.