Model Improvement via Systematic Investigation of Physics Tendencies

MAY WONG, GLEN ROMINE, AND CHRIS SNYDER

National Center for Atmospheric Research, Boulder, Colorado

ABSTRACT

Deficiencies in forecast models commonly stem from inadequate representation of physical processes; yet, improvement to any single physics component within a model may lead to degradations in other physics components or the model as a whole. In this study, a systematic investigation of physics tendencies is demonstrated to help identify and correct compensating sources of model biases. The model improvement process is illustrated by addressing a commonly known issue in warm-season rainfall forecasts from parameterized convection models: the misrepresentation of the diurnal precipitation cycle over land, especially in its timing. Recent advances in closure assumptions in mass-flux cumulus schemes have made remarkable improvements in this respect. Here, we investigate these improvements in the representation of the diurnal precipitation cycle for a spring period over the United States, and how changes to the cumulus scheme impact the model climate and the behavior of other physics schemes. The modified cumulus scheme improves both the timing of the diurnal precipitation cycle and reduces midtropospheric temperature and moisture biases. However, larger temperature and moisture biases are found in the boundary layer as compared to a predecessor scheme, along with an overamplification of the diurnal precipitation cycle, relative to observations. Guided by a tendency analysis, we find that biases in the diurnal amplitude of the precipitation cycle in our simulations, along with temperature and moisture biases in the boundary layer, originate from the land surface model.

1. Introduction

The aim of model development is an accurate and realistic simulation of the true atmosphere, including its climate. This is essential in numerical weather prediction for skillful and reliable forecasts and also plays a vital role in data assimilation as observations are combined with short forecasts to construct improved estimates of the state. Systematic errors in the forecast system degrade those short forecasts and also violate assumptions common in data assimilation systems, thereby leading to inaccurate initial conditions. While forecast verification methods can help identify systematic model errors, few consider all of the model components and partitioning of the model tendencies, an approach that can highlight interactions among physics schemes and other potential sources of systematic model errors. Exceptions include studies that employed an initial tendency method, pioneered by Klinker and Sardeshmukh (1992) and applied successfully by Rodwell and Palmer (2007), Williams and Brooks (2008), and Cavallo et al. (2016) among others

in assessing climate model, clouds, and tropical cyclone forecast errors, respectively.

The initial tendency method utilizes the series of short forecasts produced by a cycling data-assimilation system. The difference between each short forecast and the next analysis reflects both random and systematic forecast errors, and averaging the differences over many cycles isolates the systematic error. If the tendencies from each physical process are accumulated in the forecast model and also averaged over many cycles, the systematic error can be decomposed across the responsible model processes, enabling the investigator to trace sources of systematic model error to specific model components and to assess how changes to a model component impact other physics components.

The initial tendency method is demonstrated in this study by addressing a common issue in most cumulus parameterizations in the representation of the diurnal precipitation cycle. The diurnal cycle of precipitation over the continental United States has been extensively studied by many (e.g., Wallace 1975; Dai et al. 1999; Carbone et al. 2002; Davis et al. 2003; Clark et al. 2007; Carbone and Tuttle 2008; Zhang and Klein 2010; Surcel et al. 2010; Dirmeyer et al. 2012; Mooney et al. 2017).

These studies have consistently found that the diurnal precipitation cycle in this region is well characterized by a late-afternoon maximum in the Southeast (near the Gulf Coast) and over the Rocky Mountains. East of the Rockies and into the Great Plains, precipitation maxima occur later at around local midnight, due to eastward propagation of the storms that often originate near the Rockies. Other contributing mechanisms to the diurnal precipitation pattern include moisture provided by the Great Plains low-level jet and the mountain–plain solenoid circulation due to differential heating (Carbone and Tuttle 2008).

Several studies have highlighted the deficiency in parameterized convection models in not being able to skillfully forecast warm-season rainfall as they lack an accurate representation of the diurnal precipitation cycle (e.g., Davis et al. 2003; Fritsch and Carbone 2004). Convection-permitting models were found to perform better at capturing the timing of the diurnal precipitation cycle (e.g., Ban et al. 2015; Wong and Skamarock 2016). The increase in model resolution has been instrumental in improving the quantitative precipitation forecast skill in numerical weather prediction and climate models (Bernardet et al. 2000; Done et al. 2004; Davis et al. 2006; Tang et al. 2013; Prein et al. 2015), although an over-amplification of the mean diurnal cycle in these higher resolution models remains an issue (e.g., Clark et al. 2007; Wong and Skamarock 2016; Weber and Mass 2019). Meanwhile, global data assimilation and ensemble and/or extended-range forecast systems (such as those for sub-seasonal and seasonal prediction systems) will continue to at least partially rely on parameterized-convection models (e.g., in variable-resolution models). In global climate models, because of the complexity in the coupling of Earth system components and the computational expense of running century-long simulations, the current state-of-the-art model resolution remains at a parameterized-convection scale. For example, the HighRes Model Intercomparison Project uses a model horizontal grid spacing of approximately 25 km (Haarsma et al. 2016).

The major shortcoming in the representation of the diurnal precipitation cycle motivated several studies to use it as a benchmark framework to assess the physical realism in parameterized convection models (e.g., Pritchard and Somerville 2009; Surcel et al. 2010). Errors in the diurnal precipitation cycle point to model errors in representing the interactions between the large-scale environment and the parameterized convective processes, including those related to clouds and precipitation. Common evaluation diagnostics include the diurnal cycle amplitude and phase (e.g., Dai et al. 1999), and rainfall intensity and frequency of occurrence (e.g., Dai et al. 1999, 2007; Mooney et al. 2017). Traditionally, cumulus parameterization schemes are also evaluated in terms of their heating and drying rates (e.g., Johnson 1984; Emanuel 1991; Grell et al. 1991; Bechtold et al. 2014; Xie et al. 2002).

Here, we employ the initial tendency technique and other diagnostics to analyze how a change in cumulus scheme affects the diurnal cycle of precipitation and model climate, as well as interactions with other physics schemes. In particular, this cumulus scheme is based on the Tiedtke (1989) cumulus parameterization with several modifications, including a closure devised by Bechtold et al. (2014) that incorporates the impact of boundary layer forcing, and allows the gradual buildup of convective available potential energy (CAPE) over the early daytime hours. The concept is based on the observation that CAPE buildup due to boundary layer thermodynamics may not necessarily be in balance with that from the free-tropospheric forcings (Zhang 2002), in contrast to the quasi-equilibrium assumption that is often made in cumulus schemes (e.g., Arakawa and Schubert 1974). Zhang and Wang (2017) further made the recent Bechtold et al. (2014) closure available to the research community through the Advanced Research Weather Research and Forecasting (WRF-ARW) Model (Skamarock et al. 2008; Powers et al. 2017), following the ECMWF Integrated Forecasting System Cy40r1 documentation (ECMWF 2014). Collectively, along with closures described in Bechtold et al. (2004), Bechtold et al. (2008) and Sundqvist (1978), the cumulus parameterization is available within WRF-ARW as the "New Tiedtke" scheme; we refer readers to Zhang and Wang (2017) and references within for a detailed description of the implementation.

The paper is organized as follows. Section 2 describes our model systems and methods of diagnosis, including the use of initial tendencies and semiprognostic tests. An assessment of the diurnal precipitation cycle in the model systems is presented in section 3. In section 4, improvements and degradations in the model climate using the New Tiedtke scheme, as compared to that of its predecessor, the "Tiedtke" scheme in WRF-ARW (Tiedtke 1989; Zhang et al. 2011) are shown and we further investigate the model bias differences using an initial tendency analysis. Biases related to the amplitudes of the diurnal precipitation cycles are investigated in section 5. Finally, a summary is provided in section 6.

2. Methodology

a. Analysis and forecast systems

To conduct our forecast evaluation, we use a continuously cycled data assimilation system based on the Data Assimilation Research Testbed (DART), which is a community facility for executing and developing

ensemble data assimilation systems (Anderson et al. 2009). A partial cycling system is a data assimilation system that periodically initializes with an external global analysis to introduce information obtained from a global data assimilation system. A continuously cycled data assimilation system, on the other hand, always uses its own analyses as initial conditions. Over an extended period, a continuously cycled data assimilation system is expected to evolve toward its model climate and permits diagnosis of systematic model biases associated with the model and physics configuration used to advance the state between analysis cycles. Here, "model climate" is used to refer to time-averaged characteristics of the model integrations. In our study, the model climate is investigated using 6-h forecasts and by examining characteristics of the systematic model errors, which are evident as model biases relative to observations.

Rodwell and Palmer (2007) and Rodwell and Jung (2008) highlighted the use of initial tendencies within continuously cycled analysis systems to assess the underlying physics in a global model, while Cavallo et al. (2016) demonstrated the use in a regional model setting. These studies tracked the time-averaged tendencies over 6-h forecast integrations to help diagnose the sources of systematic model biases. The use of native analyses (i.e., those generated by the same model as the forecast system) in these studies is particularly important, as external analyses may introduce additional biases that can strongly influence short forecast integrations (Klocke and Rodwell 2014; Cavallo et al. 2016). For example, Cavallo et al. (2016) examined the impact of initializing forecasts with nonnative analyses, and showed that tendencies from the early forecast hours are strongly sensitive to not only external analyses, but also analyses from the same model but with a different physics scheme. Extending the forecast period will help reduce the impact of using nonnative analyses, but will also allow time for the model components to interact with one another, complicating the tendency analysis. Therefore, a separate cycled analysis system is run for each model physics configuration examined in this study. It is assumed here that a 6-h forecast integration is sufficiently short to minimize the interactions among the model physical processes.

The configuration of each analysis system is based on the National Center for Atmospheric Research (NCAR) ensemble system, as described in Schwartz et al. (2015, 2019). Initializations to each set of forecasts are provided by DART analyses that use the same corresponding model physics to generate the background atmospheric states. Mesoscale analyses using these model systems are produced every 6 h at a 15-km horizontal grid spacing for the period from 1 May to 15 June 2017.

The same conventional observations are used in each data assimilation system, including rawinsondes, Aircraft Communication Addressing and Reporting System (ACARS), atmospheric motion vectors, global positioning system refractivity profiles, meteorological aerodrome reports (METAR), and marine surface observations. The analyzed variables include horizontal winds, potential temperature, water vapor mixing ratio, pressure, and geopotential height; microphysical variables are cycled.

Our model system is largely based on the physics suite used by the operational High-Resolution Rapid Refresh (HRRR) as implemented in version 3. The HRRR is the operational analysis-forecast system with partial hourly cycling on a 3-km grid over the continental United States (Benjamin et al. 2016). This version of the HRRR is based on WRF-ARW, version 3.8.1. Since the New Tiedtke scheme is only available starting from WRF-ARW, version 3.9.1.1, the scheme is ported to the HRRR code base. In this study, the same model code is run on a single mesoscale domain at 15-km horizontal grid spacing (Fig. 1). Due to the coarser model resolution, the model setup is more akin to the NOAA Rapid Refresh (RAP) (Benjamin et al. 2016), except for the choice of the cumulus scheme. Several model configurations are used to examine systematic model biases in selected physics suites (Table 1). In addition to the cumulus scheme, the land surface model is also different from the RAP system in the "Noah-MP" configuration, which uses the Noah-MP land surface model (Niu et al. 2011; Yang et al. 2011), instead of the Rapid Update Cycle land surface model (RUC-LSM; Smirnova et al. 1997, 2016). Relative to the NCAR ensemble system, which used the Tiedtke scheme, the modified physics components (relative to the Tiedtke configuration) include the RUC-LSM and the MYNN planetary boundary layer (PBL) scheme (Nakanishi and Niino 2006, 2009; Olson et al. 2019).

Our region of interest is primarily over the Great Plains in the continental United States; areal averages are computed over the region within 30° to 45°N and −105° to −82°E (shaded region in Fig. 1). Some statistics will be computed over the full period 1 May–15 June 2017, but additional focus will be placed on the 2-week period from 1 to 15 June 2017, when there is weak synoptic forcing for convective organization, and convective evolution is more strongly tied to the diurnal heating cycle.

b. Tendencies

The time integration scheme in WRF-ARW is a third-order Runge–Kutta (RK) predictor-corrector scheme. The dynamics tendencies for each model time step are
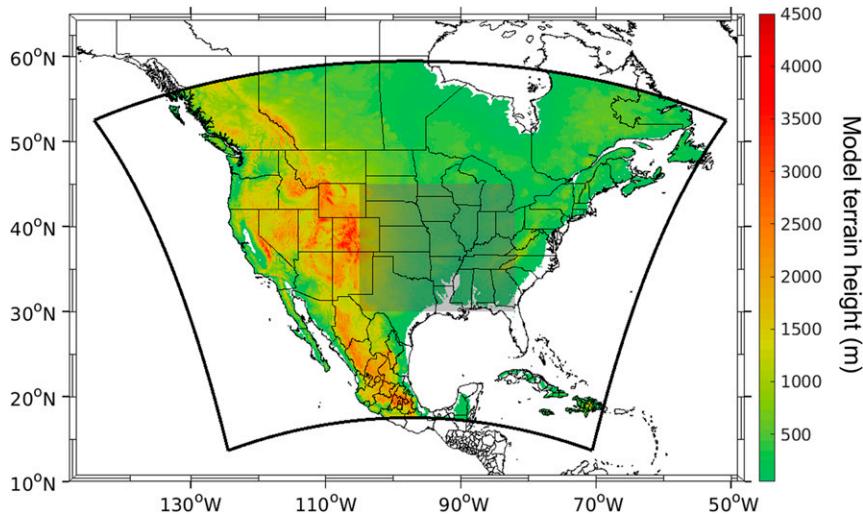
FIG. 1. Model domain configuration used in the analysis and forecast systems (solid line). Areal averages are computed over the central United States (shaded region). Filled contours show model terrain height greater than 50 m.

computed as a sum of the final predictor of the "large time-step dynamics" (i.e., over the third and final RK step) and the accumulated tendencies over the series of acoustic timesteps within that last RK step (Skamarock and Klemp 2008). The tendencies from explicit horizontal diffusion, PBL, radiation, and cumulus schemes at each model time step are evaluated at the beginning of the time step. The tendency due to the microphysics is computed at the end of the model time step as a saturation adjustment step (Klemp et al. 2007). The land surface model tendencies are provided as input into the PBL scheme as a surface boundary condition; therefore, their combined effect is reflected in the PBL tendencies.

We implemented tendency diagnostics for the atmospheric state variables: dry[1] potential temperature (θ), water vapor mass mixing ratio (q_v), and model wind components (u, v). The prognostic variables are cast in their flux form, and can be written as

$$(\rho_d \phi)^{t+\Delta t} = (\rho_d \phi)^t + \Delta t F_{\text{dynamics}}^{t*} + \Delta t F_{\text{phys/diffusion}}^t, \quad (1)$$

where ϕ is a state variable; ρ_d is the dry density; and F_{dynamics} and $F_{\text{phys/diffusion}}$ are the flux-form tendencies from the dynamics, and physics and explicit horizontal diffusion, respectively. The superscripts denote the time level at which the terms are evaluated (as described earlier), where Δt is the time step, t is the current time

level, and $t*$ is the last predictor–corrector RK step. Dividing Eq. (1) by ρ_d^t returns the expected ϕ tendencies from the physics and explicit horizontal diffusion schemes. Further adding and subtracting $\phi^{t+\Delta t}$ in Eq. (1) on the lhs and rearranging the terms, we get

$$\phi^{t+\Delta t} - \phi^t = \phi^{t+\Delta t}(1 - \rho_d^{t+\Delta t}/\rho_d^t)$$
$$+ \Delta t/\rho_d^t F_{\text{dynamics}}^{t+\Delta t} + \Delta t/\rho_d^t F_{\text{phys/diffusion}}^t, \quad (2)$$

where the appearance of the first term on the rhs is due to the reformulation of the flux-form advection term to its advective form. [Note that when no advection occurs over a time step (i.e., $\rho_d^{t+\Delta t}/\rho_d^t = 1$), the time tendency for ϕ will be solely from the physics and explicit diffusion.] Hereafter, this term will be included in the dynamics tendency term.

The tendency diagnostics are computed inline at every model time step as an accumulated sum. This eliminates the need to output and postprocess tendencies at every time step (as suggested in Cavallo et al. 2016). The output interval for the tendency diagnostics will then define the averaging interval for the tendencies. Hourly accumulated tendencies are output in this study.

c. Semiprognostic tests

To examine in detail the differing responses of a parameterization scheme to the large-scale environment, semiprognostic tests are conducted. These tests are single time-step ($\Delta t = 75$ s) experiments where instantaneous three-dimensional gridscale variables are provided as input to a parameterization scheme (Grell et al. 1991; Jakob and Siebesma 2003). This allows a

---

[1] The flux-form moist potential temperature [$\Theta_m = \Theta(1 + a'q_v)$ where $a' \equiv R_v/R_d \approx 1.61$ (see Klemp et al. 2007)] is an alternative prognostic variable, and its use will require the total tendency of water vapor mixing ratio in the computation of the θ budget.

TABLE 1. Model physics configurations used in the analysis and forecast systems.

	Tiedtke	New Tiedtke	Noah-MP
Cumulus	Tiedtke[a]	New Tiedtke[b]	New Tiedtke
Land surface	RUC-LSM[c]	RUC-LSM	Noah-MP[d]
Microphysics	Thompson[e]	Thompson	Thompson
Planetary boundary layer (PBL)	MYNN[f]	MYNN	MYNN
Surface layer	MYNN	MYNN	MYNN
TKE cycling	No	No	Yes
Radiation (longwave and shortwave)	RRTMG[g]	RRTMG	RRTMG

[a] Zhang et al. (2011).
[b] Zhang and Wang (2017).
[c] Smirnova et al. (1997, 2016).
[d] Niu et al. (2011); Yang et al. (2011).
[e] Thompson et al. (2008).
[f] Nakanishi and Niino (2006, 2009); Olson et al. (2019).
[g] Iacono et al. (2008).

close examination of the parameterization scheme behavior without any feedback from the rest of the model. Both the New Tiedtke and Tiedtke cumulus schemes depend on tendencies due to advection, the PBL and/or radiation schemes. Here, hourly averaged tendency output is provided as input where needed. We have examined the semiprognostic tendencies from both cumulus schemes using the actual instantaneous tendencies and the hourly averaged tendencies as input to the schemes. Differences between the semiprognostic cumulus tendencies and those computed online are small with a mean absolute difference of ~8% over the central United States and a vertical profile correlation coefficient of 0.99.

The purpose of the semiprognostic tests is to identify the leading parameterized processes in a particular physics scheme that drive the model behavior. In our study, we will focus on the cumulus tendencies and the impact of the trigger function and parameterized bulk updraft and downdraft processes. In both schemes, at most one type of convection (shallow, midlevel, and deep) can be activated in each grid column. A cumulus cloud depth is diagnosed as a part of the trigger function based on a simple "first-guess" entraining plume (with no downdrafts). In the Tiedtke scheme, if the cloud pressure depth meets a threshold of 150 hPa, deep convection is activated; otherwise, shallow convection is assumed for the cloudy grid column. A similar first-guess entraining plume model is used in the New Tiedtke scheme, but is based on a diagnostic updraft vertical velocity and includes simplified microphysics (Jakob and Siebesma 2003; Bechtold et al. 2004). The cloud depth threshold is 200 hPa for deep convection to be activated. We can therefore expect to see a correlation between the diagnosed cloud-top heights and the type of convection (e.g., Suhas and Zhang 2014). Decomposed

updraft and downdraft cumulus tendencies will also be shown to highlight the processes most impacted by the changes in closures.

One would expect the largest contribution from the cumulus schemes (i.e., tendency) to occur at peak convection, here defined as the time of diurnal maximum in the precipitation cycle. As will be shown, the timings of the diurnal cycle in the two schemes differ greatly. We therefore conduct the semiprognostic tests at a time step close to the predicted time of peak convection with respect to each cumulus scheme. Alternatively, one can use identical input fields generated from a model integration using one of the cumulus schemes; however, the other scheme's behavior can be sensitive to how the forecast environment is modulated by the choice of the first and may not reflect the actual behavior of the scheme.

3. Diurnal precipitation cycle

a. Timing and rainfall amounts

The strongest diurnal precipitation signal occurs during 1–15 June 2017 (Fig. 2). Precipitation from 48-h forecast integrations using the New Tiedtke, Tiedtke, and Noah-MP configurations and initialized with consistent analyses during this period are shown in Fig. 3a. Observed precipitation estimates from the Multi-Radar Multi-Sensor precipitation analyses (Zhang et al. 2016) show a diurnal cycle that peaks at approximately 2300 UTC. This is typical timing of the diurnal convection over central United States, and is consistent with previous studies (e.g., Carbone et al. 2002). We note that for this particular 2-week period, contrary to past climatological studies, there is little sign of nocturnal convection, which is typically observed to peak at approximately 0800 UTC. A Hövmoller diagram of the observed precipitation indicates that there are
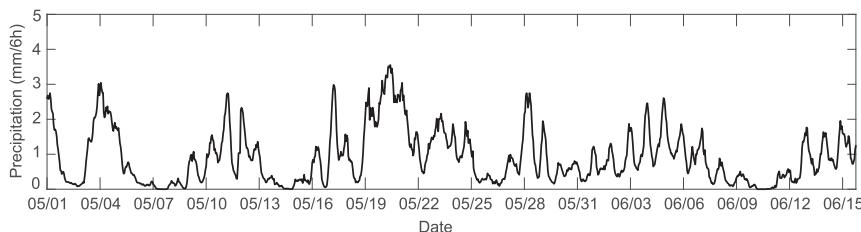
FIG. 2. Time series of observed precipitation rate (mm $6 h^{-1}$) based on MRMS precipitation estimates area-averaged over the central United States during 1 May–15 Jun 2017.

few eastward propagating systems during this period (not shown).

The New Tiedtke scheme is able to capture the timing of the diurnal cycle well, especially the timing of peak convection. Similar timing improvements have been documented in Bechtold et al. (2014), who devised the CAPE closure implemented in this scheme by Zhang and Wang (2017). The Tiedtke scheme, on the other hand, shows too early an onset of the peak convection, similar to past studies that noted this common deficiency in parameterized-convection models (e.g., Dai et al. 1999; Ban et al. 2015). Both models overpredict the magnitude of peak precipitation amounts, which we will return to in section 5 along with further discussions on the Noah-MP forecasts. Examining the 0–24 h and 24–48 h forecast periods, we see that the predicted precipitation amounts are very similar, indicating that little model spinup time is needed. This is one of the advantages of using consistent analyses in our forecast initializations; forecasts using nonnative analyses (e.g., using a model and/or model resolution different from the forecast model) typically take longer to spin up.

Precipitation forecasts initialized every 6 h in the analysis systems (Fig. 3b) show resemblance to those in the 0–24- and 24–48-h forecasts. In particular, the diurnal cycle timing error in the Tiedtke forecasts is similar, and improvement of the diurnal cycle timing in the New Tiedtke forecasts relative to the Tiedtke forecasts in the 0–24- and 24–48-h forecasts is also seen in the 6-h forecasts. The amplitudes of the diurnal cycle in both sets of forecasts are slightly improved but still overestimated. The 1800 UTC initialized Tiedtke precipitation forecasts benefited the most from the improved initial conditions; however, they were insufficient to correct the timing error in the Tiedtke forecasts. The similarity between the 6- and 48-h forecast diurnal cycles is critical, as this indicates that model errors leading to timing and amplitude errors in the 48-h forecasts are likely already present in the short-term 6-h forecasts.

Interestingly, there is a spurious precipitation peak at the first hour in the New Tiedtke 6-h forecasts, most notably at 1900 UTC (Fig. 3b). The modified CAPE

closure in the New Tiedtke scheme is highly dependent on the tendencies from the PBL and radiation schemes, which require time to spin up from their own initialization procedures. For example, when total kinetic energy (TKE) is not provided in the initial conditions, the PBL scheme will carry out its own initialization procedure for the field, which may lead to spinup errors in the growth of the parameterized boundary layer height (J. B. Olson 2019, personal communication). Within a data assimilation framework, one way to reduce the PBL spinup time is to cycle the TKE. Figure 3b shows that model spinup in the first-hour precipitation is enhanced in a TKE-cycled system (Noah-MP) relative to the New Tiedtke run, where TKE was not cycled. Additional forecasts (not shown) with New Tiedtke and Noah-MP indicate that the first-hour precipitation bias reduces by about 40% when an input TKE field is provided (either recycled from the previous 6-h forecast valid at the initialization time or from a Noah-MP analysis). After the first hour, the New Tiedtke and Noah-MP forecasts,
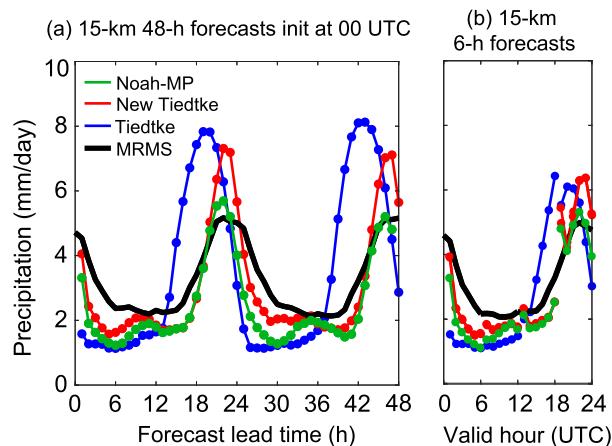
FIG. 3. Domain-averaged diurnal precipitation cycle from (a) 15-km 48-h forecasts, and (b) 15-km background (6-h) forecasts from the data assimilation system. "New Tiedtke" forecasts initialized with consistent DART analyses are shown in red. Similarly, those using the "Tiedtke" and "Noah-MP" configurations are in blue and green, respectively. MRMS observed precipitation estimates are in black. Forecasts shown are initialized for the period 1–15 Jun 2017.

regardless of any TKE input, converge before diverging depending on the land surface model used. Future potential considerations to further enhance model spinup include cycling of the higher-order moments in the PBL scheme (e.g., variances), more frequent calling of the radiation scheme (which is currently called at a 15-min interval), and/or cycling of the physics tendencies that are input to the New Tiedtke scheme. Reducing model spinup time will be important especially for data assimilation systems that run on a shorter (e.g., hourly) cycling window. Future investigations will examine the impact of using the New Tiedtke scheme in an hourly cycling analysis system.

b. Amount, frequency, and intensity

For measuring the skill in predicted diurnal precipitation cycles, in addition to rainfall amount, past studies have examined the frequency of occurrence of rainfall events and rainfall intensity (Dai et al. 1999, 2007; Evans and Westra 2012; Mooney et al. 2017). These additional metrics help quantify the nature of the precipitation event. For example, the same precipitation amount can be produced by light but very frequent precipitation events, or short episodes of heavy rainfall, which will have different impact on the surface hydrology. The occurrence of a rainfall event at a particular hour and location (in our case, a grid cell on our verification grid) is defined using a precipitation threshold, and the mean rainfall intensity at that location is the rainfall rate exceeding the given threshold averaged over the number of occurrences. The total amount of precipitation is equal to the product of the frequency of occurrence and the rainfall intensity. We note that, however, when we further spatially average these quantities, this no longer holds exactly, since generally $\overline{f \times I} \neq \overline{f} \times \overline{I}$, where f and I are the frequency of occurrence and rainfall intensity, respectively, and the overbar denotes spatial averaging. The exception is if all grid cells satisfy the imposed precipitation threshold (i.e., f is a constant and does not vary in space), or if the rainfall intensity happens to be spatially uniform (which is highly unlikely due to the convective nature of the precipitation in our period).

We conservatively remap (i.e., conserving the domain-average rainfall amount) the observed precipitation estimates and each hourly precipitation forecasts onto a regular $0.15° \times 0.15°$ ($\Delta x \approx 16$ km, $\Delta y \approx 13$ km) verification grid (shaded region in Fig. 1). The rainfall amount, frequency of occurrence and rainfall intensity are then calculated for each grid cell and forecast hour on the verification grid. We compute these metrics for various precipitation thresholds as shown in Fig. 4. The smallest measurable precipitation amount on the 0.01° MRMS grid is 0.1 mm h^{-1}. The smallest precipitation threshold of 4.44×10^{-4} mm h^{-1} is chosen based on the

minimum possible nonzero observed value on the 0.15° verification grid, which is approximately 0.1 mm h^{-1} divided by 225 grid cells. The other two precipitation thresholds help highlight how the frequency of occurrence and rainfall intensity vary with increasing precipitation intensity, as compared to observations.

Figure 4 shows the metrics calculated for the 15-km parameterized convection forecasts. At the two lowest thresholds, both parameterized-convection models tend to overestimate the frequency of occurrence, especially at peak precipitation. On the other hand, rainfall intensity is underestimated by both models, except that the Tiedtke model shows an overprediction of the rainfall intensity during peak precipitation. This is typical of most cumulus schemes and is consistent with results from past studies such as in Mooney et al. (2017) and Evans and Westra (2012). The timing (vertical dot–dashed line) of the peak observed precipitation (black dashed line) is more in phase with the timing of the frequency of occurrence than with the rainfall intensity variation, which was also found in past studies (e.g., Dai et al. 1999; Mooney et al. 2017).

As we increase the precipitation threshold to 0.1 mm h^{-1} (Figs. 4d–f), we see that the frequency of occurrence for the New Tiedtke forecasts drops dramatically relative to that at the lowest threshold of 4.44×10^{-4} mm h^{-1} (note the change in y axis), indicating that much of the precipitating grid cells are dominated by very low rainfall rates. The lack of change in the frequency × intensity (i.e., approximate rainfall amount) curves also indicates that these are grid cells with precipitation of very small magnitudes. There is still an overprediction of the frequency of occurrence especially at peak convection in the New Tiedtke forecasts. Although the rainfall intensity in the New Tiedtke forecasts is clearly underpredicted, we see that the diurnal variation corresponds well with that in the observations. The Tiedtke forecasts on the other hand shows a clear timing error in the peak intensity. For moderate to heavy rainfall events (precipitation threshold = 2.54 mm h^{-1}; Figs. 4g–i), the New Tiedtke forecasts show improvement in both its frequency of occurrence and rainfall intensity relative to the forecasts at lower thresholds. The New Tiedtke forecasts still have a slight overprediction of the peak precipitation occurrence. On the other hand, the Tiedtke scheme appears to strongly overpredict the frequency of occurrrence at peak convection. Both schemes underestimate the frequency and intensity during the nocturnal hours.

4. Impact on systematic biases

a. Verification against observations

To examine the impact the New Tiedtke scheme has on systematic biases, 6-h forecasts from the analysis
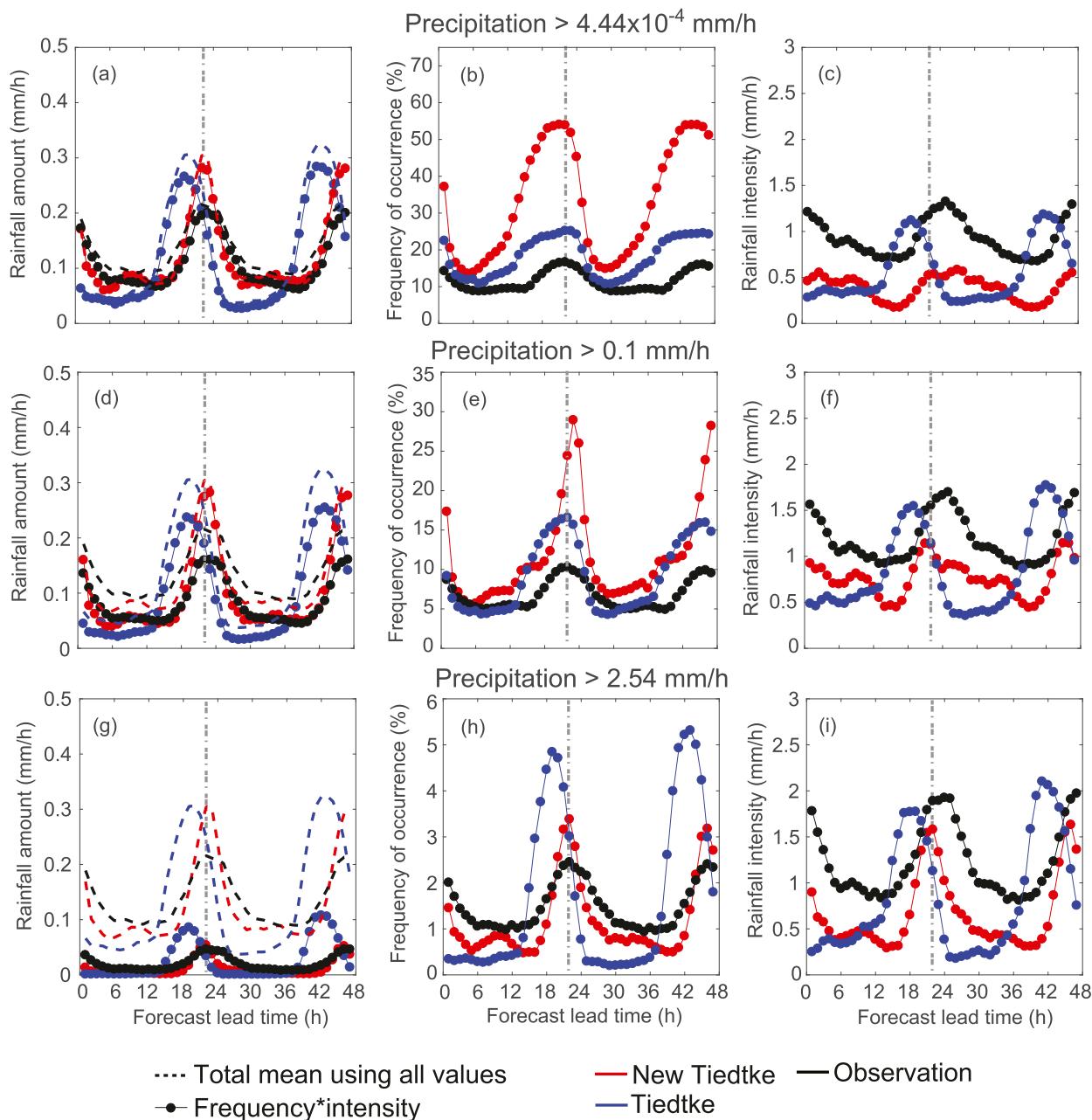
FIG. 4. (a),(d),(g) Area-averaged rainfall amount (mm h^{-1}), (b),(e),(h) frequency of occurrence (%), and (c),(f),(i) rainfall intensity (mm h^{-1}) for the 15-km New Tiedtke forecasts (red) and Tiedtke forecasts (blue) with the following precipitation thresholds applied: $P > 4.44 \times 10^{-4}$ mm h^{-1}, $P > 0.1$ mm h^{-1}, and $P > 2.54$ mm h^{-1}. Observed values based on MRMS precipitation estimates are plotted in black. Vertical gray dot–dashed line indicates approximately the timing of the observed peak total mean precipitation (black dashed line). Averages are computed over the verification grid.

systems are verified against observations. Forecast biases averaged over the entire period (1 May–15 June 2017) based on rawinsondes and aircraft data from ACARS are shown in Fig. 5. The verification is done using all available observations (including those not assimilated to ensure a consistent set of observations for verification across all experiments). Midtropospheric

temperatures in the New Tiedtke forecasts are consistently warmer than in the Tiedtke forecasts, leading to a statistically significant difference in the temperature bias (Figs. 5a,b). At those levels, the New Tiedtke forecasts also show a reduced moisture bias as compared to that in the Tiedtke forecasts (Fig. 5c). Both the New Tiedtke and Tiedtke forecasts show a strong cold and moist bias

in the boundary layer (Figs. 5a–c), with the New Tiedtke scheme showing more negative temperature and more positive moisture biases. It is important to note that temperature observations from ACARS have different observational measurement errors from those in rawinsondes, where the former exhibits a greater variance with aircraft type, flight altitude, and phase of flight (Ballish and Kumar 2008). Consistent with their findings, the average temperature profiles from ACARS are warmer than those from radiosondes, leading to a more negative forecast bias than those compared against radiosondes. There are some small differences in the wind biases but not statistically significantly so (Figs. 5d,e); therefore, for the remainder of the paper, we will focus on the moisture and heat budgets.

b. Link between model biases and tendencies

Figure 6 shows the total heating and moistening tendencies averaged in 6-h windows (0000–0600, 0600–1200, 1200–1800, and 1800–0000 UTC) from the New Tiedtke analysis system (Figs. 6a,c), and their differences from those in the Tiedtke analysis system (Figs. 6b,d). The mean tendency and differences averaged over all four windows are also shown (in gray), and the vertical profiles are dominated by the mean differences between 1800 and 0000 UTC. This is not surprising as that is when convection is expected to be the most active in the diurnal cycle over this region. We also see that the vertical profiles of the accumulated tendency differences during 1800–0000 UTC are very similar (of the same order of magnitude) to those of the bias differences in Fig. 5.

We further decompose the 1800–0000 UTC heating and moistening tendencies by model components (Fig. 7). In general, the signs of the individual physics and dynamics tendencies are similar between the New Tiedtke and Tiedtke forecasts. A notable difference is that the New Tiedtke cumulus scheme allows for deeper net tropospheric heating and drying than the Tiedtke cumulus scheme. In the upper troposphere, it is also evident that the cumulus schemes interact differently with the microphysics scheme, which provides more net heating there in the Tiedtke forecasts (Fig. 7c). Examining the cumulus and microphysics schemes together (magenta dashed line) however suggests that the cumulus scheme is the main driver of the net tendency differences. We also see that the Tiedtke scheme is more active at vertically mixing moisture within the boundary layer (Fig. 7f). This difference in the boundary layer mixing is partially compensated by the PBL scheme, but the

New Tiedtke forecasts are still generally more moist near the surface and drier at the top of the boundary layer relative to the Tiedtke forecasts (black line in Fig. 7f).

Comparing Figs. 7c and 7f to the biases shown in Figs. 5a–c, we see that the differences in model behavior align with the bias differences. In particular, the stronger midtropospheric drying and heating by the New Tiedtke scheme led to reduced temperature and moisture biases in the midtroposphere, and less mixing in the boundary layer led to forecasts with colder and more moist (worse) biases.

c. Parameterized processes controlling model climate

Each scheme's behavior at peak convection, which is when most of the convective heating and drying of the atmosphere occurs, likely drives the differences evident in the mean profiles between 1800 and 0000 UTC (Fig. 6). Instantaneous model output fields at 2000 UTC (2300 UTC) from 6-h forecasts initialized with consistent 1800 UTC DART analyses are provided as input to the semiprognostic tests with the Tiedtke (New Tiedtke) schemes. These times are selected based on the approximate peak convection time by each scheme. Figure 8a shows the domain-averaged total differences (black dotted lines) in the heating and moistening tendencies from the semiprognostic tests. The vertical profiles are very similar to the 6-h average profiles presented in Figs. 7c and 7f, indicating that the 1800–0000 UTC tendency differences are indeed dominated by differences at peak convection for this period. The triggering of convection and decomposed cumulus tendencies from these semiprognostic tests are examined to identify the leading parameterized processes that drive the tendency differences.

1) TRIGGER FUNCTION

The trigger function in a cumulus scheme is the procedure that determines the occurrence of convection and its type (e.g., Bechtold et al. 2004). Past studies have shown that the trigger function in cumulus schemes can have a strong impact on the development of convective activities in parameterized convection models. For example, Kain and Fritsch (1992) discussed how trigger functions that depend on resolvable-scale tendencies may be more susceptible to forecast errors, in situations where the role of large-scale forcing is small in the convective activity. Xie and Zhang (2000) demonstrated how a reformulation of the trigger function significantly improved model biases in their single-column experiments. Suhas and Zhang (2014) recently evaluated various commonly used trigger functions for
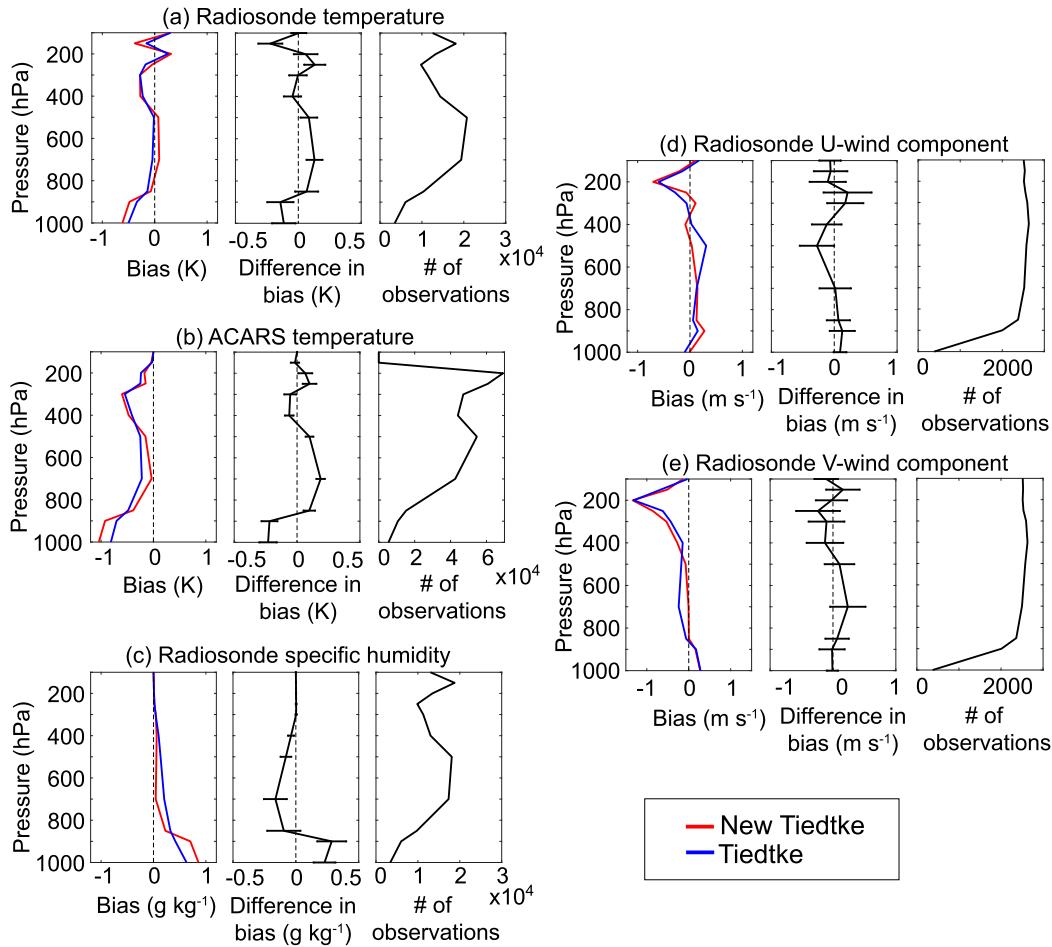
FIG. 5. Average 6-h forecast bias (forecast − observation) compared against (a) radiosonde and (b) ACARS for temperature (K), and radiosonde for (c) specific humidity, (d) *u*-wind component, and (e) *v*-wind component. Biases are averaged over the central United States and for the entire period from 1 May to 15 Jun 2017. 98% confidence intervals are shown for the mean bias differences.

deep convection over a point location based on frequency of occurrence over several periods of interest. In their study, the authors defined an observed deep convective event as a precipitation event exceeding a given threshold. Similar to their study, we will first compare the difference in frequency of triggering of deep, shallow, and midlevel convection between the two cumulus schemes.

Figure 9 shows the percentage of grid points averaged over the central U.S. domain (darker markers) and the whole 15-km domain (fainter markers), where deep, shallow, or midlevel convection is active, or convection is inactive. Further analysis shows that cumulus heating and moistening tendencies over grid columns with activated midlevel convection were insignificant (not shown), and therefore will not be further described. Based on Fig. 9, it is evident that the New Tiedtke scheme more frequently triggers deep convection than

the Tiedtke scheme, whereas the Tiedtke scheme preferentially triggers shallow convection covering almost 50% of the central U.S. domain even at peak convection. The variability over time of the deep convection trigger in the New Tiedtke scheme is large over the central United States (i.e., the scheme is fairly responsive to the different large-scale forcing over the period). The shallow convection trigger on the other hand is less sensitive and variable.

To evaluate the realism of the convective cloud distribution determined as part of the trigger function, we compare the estimated cloud-top pressure with those retrieved from satellite-based observations over the central U.S. domain. The satellite-based cloud-top pressure is retrieved from *GOES-13* available through the Satellite Clouds and Radiation Property Retrieval System (NASA Langley Research Center 2017). In that system, daytime retrievals use the iterative Visible
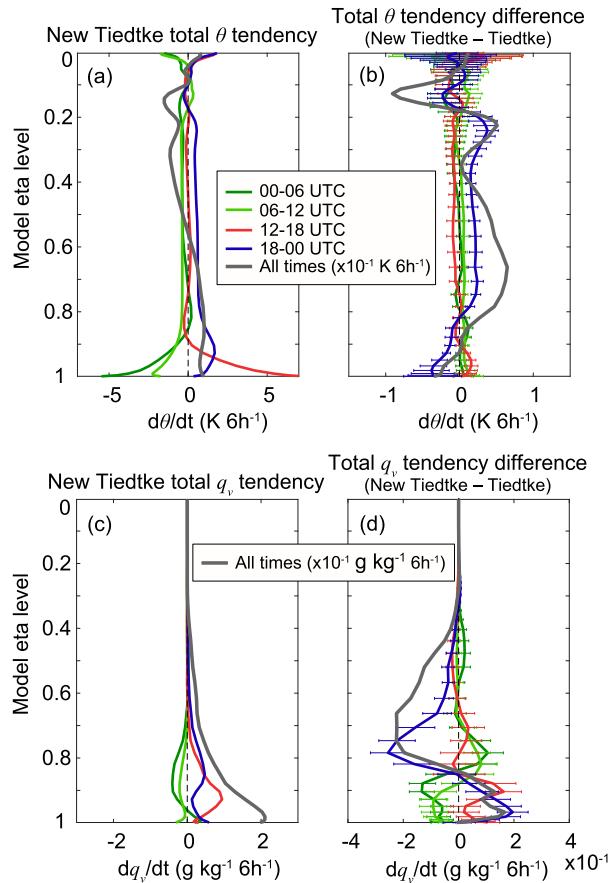
FIG. 6. Total 6-h model tendency for (a) potential temperature and (c) water vapor mixing ratio from the New Tiedtke 15-km background averaged over the full period 1 May to 15 Jun 2017 (dark gray), 0000–0600 (dark green), 0600–1200 (light green), 1200–1800 (red), and 1800–0000 UTC (blue). The total tendency and differences averaged over the entire period are multiplied by a factor of 10 for plotting purposes. Differences between the New Tiedtke and Tiedtke model tendencies are shown for (b) potential temperature and (d) water vapor, with the 98% confidence interval of the mean difference shown. Differences are statistically significant where the confidence interval does not include zero.

Infrared Shortwave-infrared Split-window Technique (VISST; Minnis et al. 2008, 2011, 2016) to obtain best-matched cloud properties relative to radiative models. Cloud properties such as optical depth, cloud effective radii, and cloud effective temperature, are retrieved from top-of-atmosphere calibrated reflectances and brightness temperatures, as well as ancillary information including atmospheric variables from the MERRA reanalyses (Rienecker et al. 2011). The cloud products are available half-hourly at an approximately 4 km resolution.

Figure 10 shows the frequency distribution of cloud-top pressure averaged over the 15 days at 2000 UTC for Tiedtke and 2300 UTC for New Tiedtke, and from

GOES-13 aggregated over the two closest retrieval times. Bin intervals of 25 hPa are used, and the percentage of grid points is calculated over all grid points in the central U.S. domain (including clear-sky areas). Broadly speaking, low cloud-top pressures (between 100 and 400 hPa) represent high cirrus clouds and deeper convective clouds, whereas high cloud-top pressures (>600 hPa) represent shallow clouds. The bimodal nature of the histograms of SatCORPS cloud-top pressure is also found in cases examined by Jones et al. (2018). To distinguish deeper cumulus clouds with high cloud tops from high cirrus clouds, the frequency distribution of cloud-top pressure for pixels with an optical depth $\tau \geq 3$ is also plotted (dashed lines in Fig. 10).

We found a similar bimodal distribution of the diagnosed cloud-top pressures from the two cumulus schemes. The Tiedtke scheme tends to overpredict the frequency and cloud-top pressures of the shallow clouds as compared to the satellite-derived observations across the corresponding range of lower cloud tops. This is consistent with our finding that the Tiedtke scheme is active in triggering shallow convection. Differences in the distributions of the satellite-derived low cloud-top pressures at the two times of day indicate that the observed higher clouds are more prevalent at 2300 UTC than at 2000 UTC, where the former is the typical peak convection time observed over the central United States. The Tiedtke scheme at its peak convection underpredicts the frequency of the deep convective clouds, even when compared to a less convective observed period at 2000 UTC. The New Tiedtke scheme, on average, is better able to capture the distribution of the high cloud tops ($\tau \geq 3$), although with a slight overprediction. This indicates that there is some physical evidence in the satellite-derived observations for the more active triggering of deep convection in the New Tiedtke scheme and relatively less frequent triggering of shallow convection.

2) CONTRIBUTIONS TO THE LARGE-SCALE HEAT AND MOISTURE BUDGETS

We now investigate how differences between the two cumulus schemes lead to the total tendency differences (Figs. 7c,f) and to the bias differences (Fig. 5). To do this, we turn to the individual processes that contribute to the large-scale heat and moisture budgets in each scheme. The New Tiedtke and Tiedtke cumulus schemes both include updraft and downdraft convective transport of heat and moisture, local compensating vertical motion, condensation/sublimation in updrafts, evaporation of cloud-liquid water in downdrafts and of precipitation in the unsaturated subcloud layer, and latent heat release
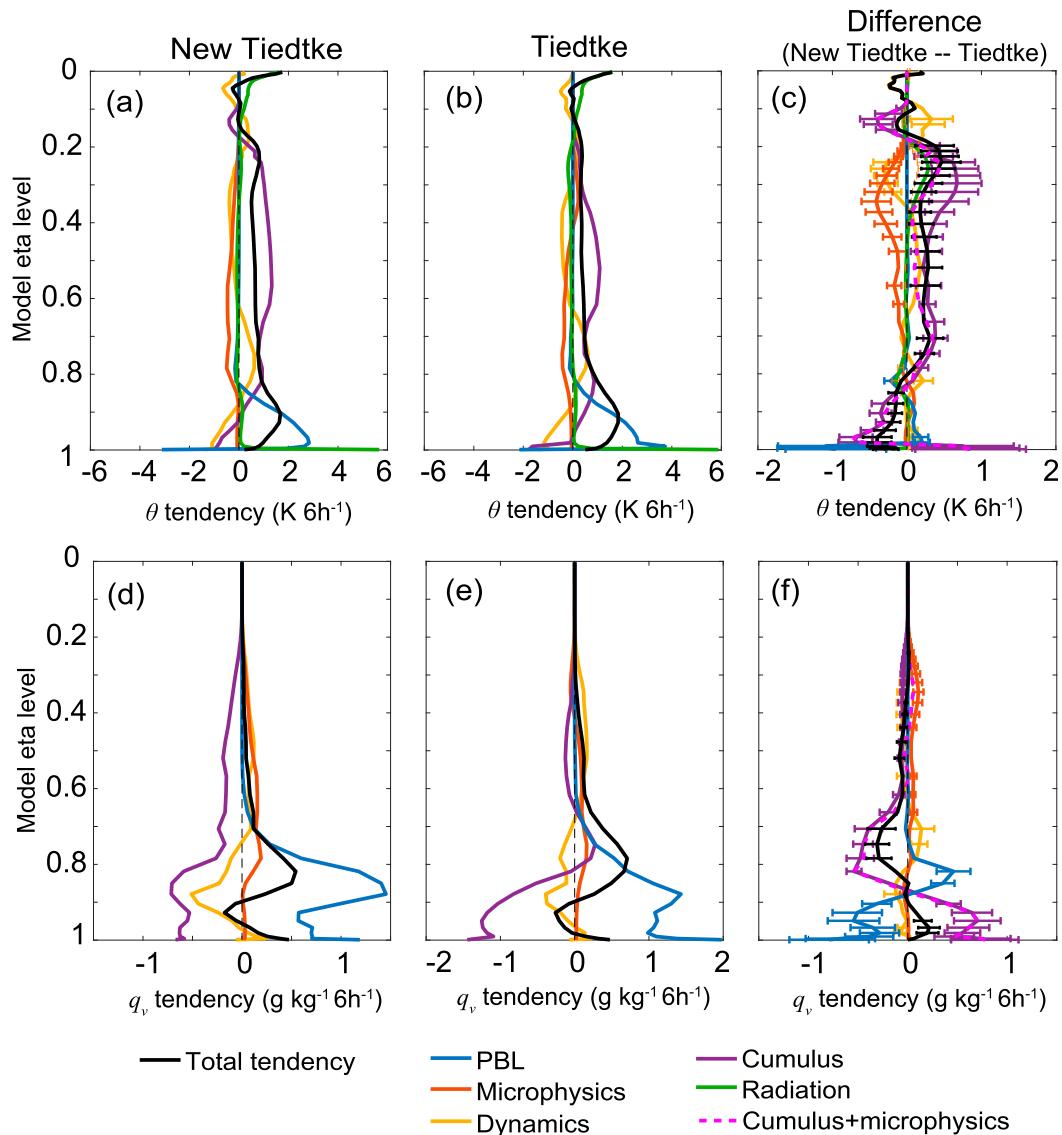
FIG. 7. New Tiedtke and Tiedtke accumulated 6-h tendencies from 1800 to 0000 UTC and their mean differences (98% confidence interval shown).

from the conversion between snow and rain (i.e., the freezing/melting of precipitation). In addition, the New Tiedtke scheme also parameterizes the process of freezing of condensates (i.e., cloud-liquid water to cloud-ice) in updrafts.

We further stratify the differences in the semi-prognostic cumulus tendencies by domain-averaging them according to convection type. Figures 8b and 8c show the domain-averaged differences over deep convection and shallow convection columns, respectively. To reduce the influence of the triggering frequency on the tendency differences, cumulus tendencies are also averaged over grid columns with the same convection type (not shown). These tendency differences are nearly opposites

of the domain averages (Figs. 8b,c). In particular, the former show that deep convective transport and the associated condensational heating in New Tiedtke is weaker than those in the Tiedtke scheme, while the shallow convective mixing of moisture is stronger. This indicates that the stronger warming in the mid- to upper levels of the model in the New Tiedtke scheme is dominantly due to more grid columns activated with deep convection, which led to more parameterized condensational heating aloft. As previously indicated, the more active triggering of deep convection is supported by satellite-derived observations, and has led to an overall improvement in the midtropospheric temperature bias (Figs. 5a,b). The more frequent and deeper convective updraft transport of moisture to the
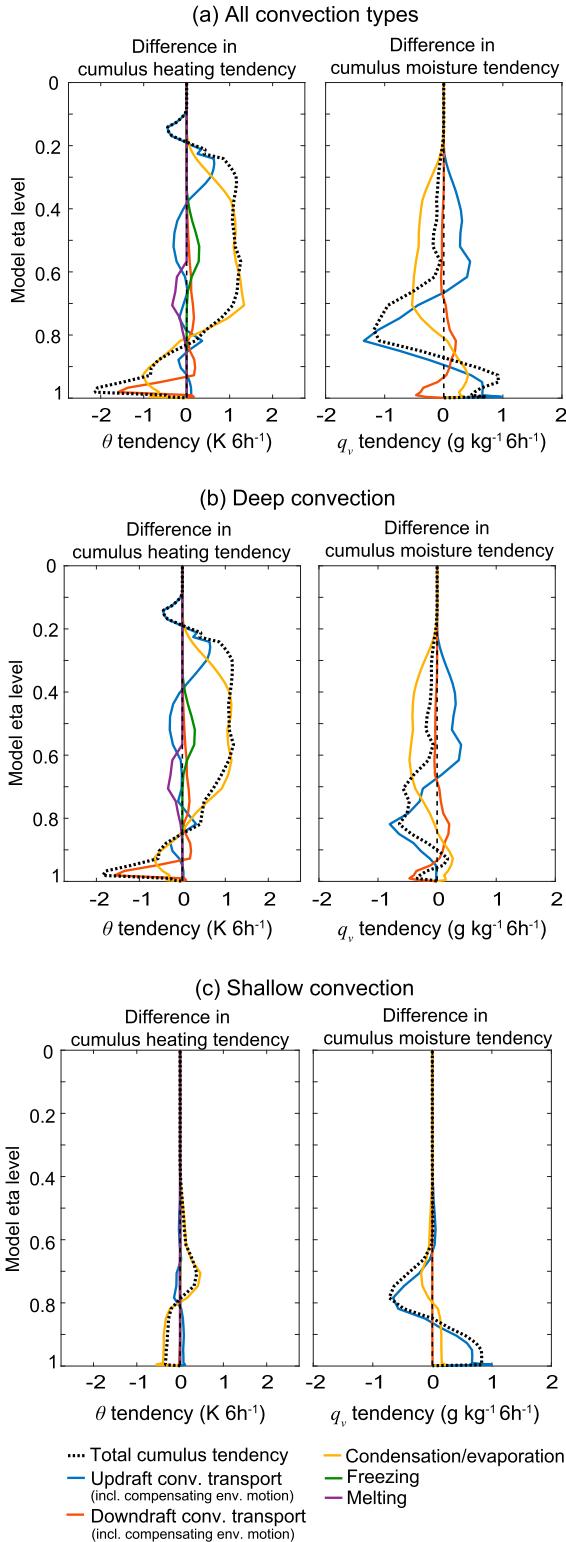
FIG. 8. (a) Domain-averaged vertical profiles of differences (New Tiedtke − Tiedtke) in the cumulus heat and moisture budgets at all convection-triggered grid points at peak convection time for each cumulus scheme. (b) As in (a), but over only "deep convective" columns. (c) As in (a), but over only "shallow convective" columns.
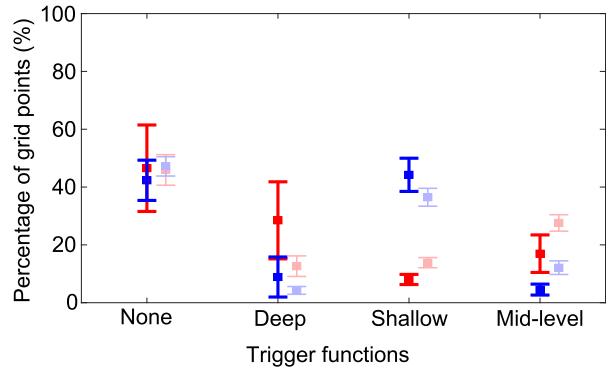
FIG. 9. Average percentage of grid points where trigger functions are activated for deep convection, shallow convection, and mid-level convection by the New Tiedtke scheme (red and light red) and the Tiedtke scheme (blue and light blue) based on the semi-prognostic tests. Darker colors represent percentages over central United States; light colors represent percentages over the whole 15-km domain. Error bars denote one standard deviation. Averages are taken over the period 1–15 Jun 2017 at 2300 UTC for New Tiedtke and 2000 UTC for Tiedtke.

mid- to upper model levels (Fig. 8b) led to a net drying tendency near the top of the boundary layer. The net moistening effect aloft, however, appears small due to a compensating drying tendency by condensation.

Consistent with the total model tendency and temperature biases seen earlier, the New Tiedtke scheme in these semiprognostic tests shows greater cooling tendencies near the surface, related to a more negative cold bias than the Tiedtke scheme in the boundary layer. This is partly due to more evaporative cooling and downdraft convective transport (Fig. 8a). Relative to the Tiedtke scheme, the New Tiedtke scheme generates more parameterized rainfall, which can contribute to more surface evaporation, all else equal. The resulting evaporative cooling can also be exacerbated by positive moisture biases in other model components, which will be shown in the next section to have a strong control on the parameterized precipitation amount and the amplitude of the diurnal precipitation cycle.

As shown earlier, the Tiedtke scheme is very active in triggering shallow convection. The more active shallow mixing leads to more boundary layer vertical mixing of the water vapor (Fig. 8c). This results in the signature we see in the difference in moisture biases between the two forecasts (Fig. 5c). The smaller boundary layer moist bias in the Tiedtke scheme (as compared to the New Tiedtke scheme) is therefore an artifact of excessive vertical mixing of moisture away from the surface. This overactive shallow mixing also contributes to a greater moist bias above the boundary layer in the Tiedtke forecasts. Conversely, the less active shallow convection
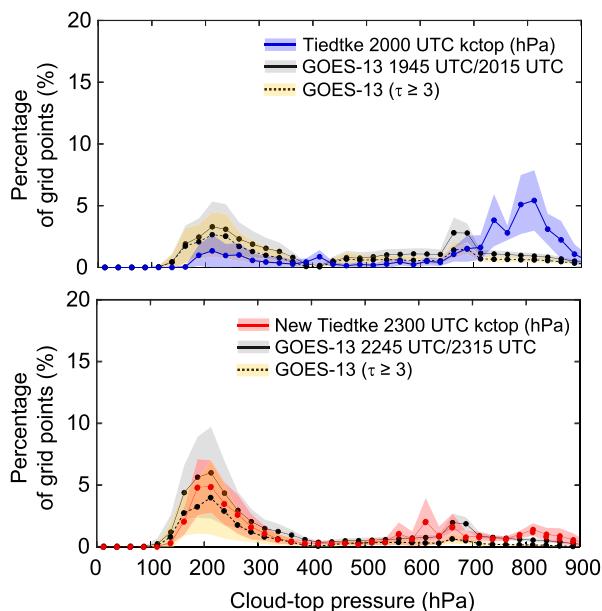
FIG. 10. (top) Frequency distribution of cloud-top pressures at 2000 UTC from the Tiedtke scheme (blue) and those at 1945 UTC/2015 UTC aggregated from all *GOES-13* pixels (only *GOES-13* pixels with optical depth $\tau \geq 3$) in the central U.S. domain are shown in black solid (dashed) lines. (bottom) As in top, but those at 2245 UTC/2315 UTC from *GOES-13* are aggregated (black solid and dashed) and at 2300 UTC from the New Tiedtke scheme (red). Solid lines with markers represent the mean frequency distribution averaged over 1–15 Jun 2017; shadings denote one standard deviation.

in the New Tiedtke scheme results in greater accumulation of boundary layer moisture and increased positive moisture bias in the boundary layer.

5. Overprediction of diurnal amplitude

To diagnose the potential cause of the diurnal precipitation cycle overamplication, the water vapor budget is examined. Figure 11 shows the domain-integrated observed and forecast precipitation, and volume-integrated water vapor mixing ratio tendencies from the dynamics and physics tendencies. Volume-integrated quantities are integrals in the horizontal and vertical dimensions over the central United States. The model tendencies and precipitation are from 48-h New Tiedtke and Noah-MP forecasts initialized with consistent DART analyses. Since model column physics do not allow lateral exchange of prognostic variables from one grid column to another, horizontal moisture convergence/divergence over our domain of interest is represented by the volume-integrated dynamics tendency. As the PBL scheme only allows vertical mixing, the volume-integrated PBL tendency represents the net source of moisture parameterized by the land

surface model, which per design exhibits a strong diurnal cycle associated with the incident solar radiation (Smirnova et al. 1997, 2016).

Similar to Fig. 3, the diurnal amplitude of the 6-h precipitation is overpredicted in the New Tiedtke forecasts (dotted lines in Fig. 11). The cumulus scheme is the main driver in converting the available moisture into precipitation. Aside from adjustment of the large-scale temperature and humidity, the cumulus scheme also detrains cloud condensates from the convective updrafts to the gridscale column. These adjustments will have an impact on the microphysical process rates, which, as found here, have a net tendency to replenish moisture through evaporation. Net horizontal moisture convergence over the region is small relative to the surface moisture fluxes from the land surface model (labeled as "PBL").

To investigate whether the diurnal amplitude of our precipitation forecasts is sensitive to the land surface model, we switched to the Noah-MP land surface model (Niu et al. 2011). The resulting land surface model moisture tendencies (solid blue line in Fig. 11) are drastically reduced during the daytime hours. Similar differences in behavior in the LSM surface moisture fluxes were also found in Duda et al. (2017), who conducted a multiphysics ensemble experiment by perturbing the LSM and found that the RUC-LSM tends to generate an anomalously large surface moisture flux. The reduction in the surface moisture flux from changing the LSM to Noah-MP systematically affects the amount of water vapor that the cumulus scheme has available to convert to clouds and precipitation, and leads to a diurnal amplitude in better agreement with that from MRMS. In addition to a better amplitude of the diurnal precipitation cycle (Figs. 3 and 11), the 6-h forecasts in the Noah-MP analysis system also indicate an elimination of the moist bias in the boundary layer and a reduced temperature bias (albeit with an opposite sign; Fig. 12). We note that similar sensitivity of the amplitude of the diurnal precipitation cycle to the land surface model is also found in downscaled 3-km convection-permitting forecasts run with the same model systems but without the cumulus scheme (not shown).

6. Summary

Error sources associated with inadequate representation of the physical processes are difficult to disentangle. Improvement to any single physics component may lead to degradations in other physics components and overall deterioration in forecast skill. Traditional forecast verification methods that examine skill scores,
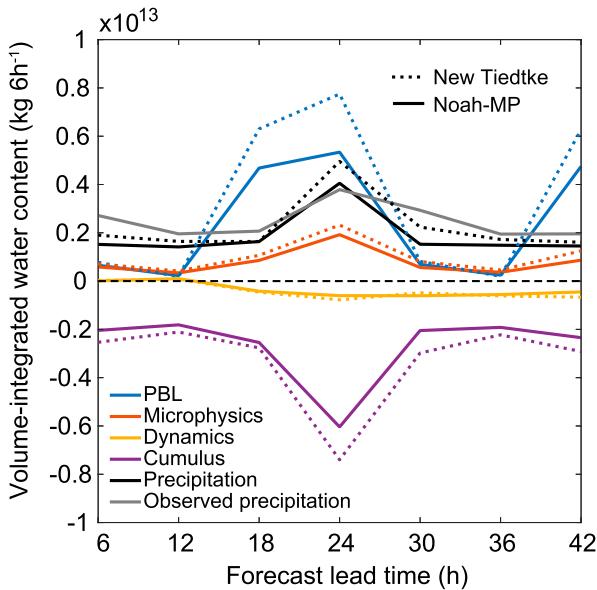
FIG. 11. Volume-integrated water vapor mixing ratio tendencies from the dynamics (yellow) and physics (PBL in blue, cumulus in purple, and microphysics in red), and domain-integrated modeled and observed precipitation in black and gray, respectively. Dotted lines are the New Tiedtke forecasts and solid lines are the Noah-MP forecasts.

such as temperature biases, are useful in assessing systematic model errors, but less informative in disentangling their sources within a forecast model. A systematic investigation of physics tendencies within a data assimilation system has been demonstrated here to help identify and correct compensating model biases. Short 6-h forecast tendencies are used to guide the investigation of systematic model behaviors in longer range (1–2 day) forecasts. These 6-h forecasts were generated as a part of a continuously cycled data assimilation system. The initialization with consistent DART analyses helps reduce the impact of external model biases in our analysis and interpretation. Moreover, the data assimilation framework facilitates the computation of biases compared against conventional observations and serves as a verification platform. The implementation of model tendencies within the data assimilation system allowed a process-based perspective on understanding model biases.

To illustrate the model improvement process, we evaluated an improved cumulus scheme ("New Tiedtke") over a warm-season rainfall period in the central United States, and examined the impact on the model climate and the interactive roles of other model components. Although improvements were found in the midtropospheric temperature and moisture biases, in the boundary layer, the New Tiedtke forecasts showed colder and more moist biases. Model tendencies and semiprognostic tests

were used to better understand and disentangle model processes related to the bias changes. As expected, the cumulus tendencies were found to be the dominant driver of the forecast differences. The 6-h tendencies, however, did show interactions of the cumulus scheme with other physics schemes. For example, the cumulus schemes interacted differently with the microphysics scheme, where the impact was mostly found in the upper-troposphere latent heating/cooling. Also, differences in the strength of boundary layer mixing as parameterized by the cumulus scheme were partially compensated by interactions with the PBL scheme. Despite these interactions, the cumulus scheme appears to be the leading factor in the forecast bias changes.

Further investigation with the moisture tendency budget revealed that the apparent deterioration in the near-surface moisture bias in the New Tiedtke forecasts was dominated by large daytime surface moisture fluxes parameterized by the land surface model. Based on our semiprognostic tests, the New Tiedtke scheme was found to be more (less) active in triggering deep (shallow) convection than the Tiedtke scheme, where the former appears to be better supported by satellite-retrieved cloud products. In other words, the drier (less positive) near-surface moisture bias in the Tiedtke forecasts was a result of the overactive boundary layer mixing by the cumulus scheme, an error compensating for the large surface moisture flux. Our results also indicate that the amplitude of the diurnal precipitation cycle was sensitive to the daytime surface moisture fluxes. With the Noah-MP LSM, the daytime surface moisture fluxes are reduced and the strong moist and cold biases in the boundary layer are much improved. The impact of the land surface model on the boundary layer biases warrants future investigation.

The methodology of using tendencies to diagnose sources of model error, as demonstrated here and in past studies, is not limited to the WRF-ARW Model. In practice, any forecast model within a stable and effective continuously cycling data assimilation framework is a candidate for this model improvement framework. For example, within the community facility DART, other model systems such as the Community Atmosphere Model (CAM) and the Model for Prediction Across Scales (MPAS) can also potentially be investigated. Similarly, the approach would also be viable for the National Oceanic and Atmospheric Administration (NOAA)'s new operational weather forecast model when run in data assimilation mode. Plans are underway to implement tendency diagnostics in the new NOAA forecast model to facilitate this type of process-based model error analysis. These models also follow the practice of using physics suites, which will especially
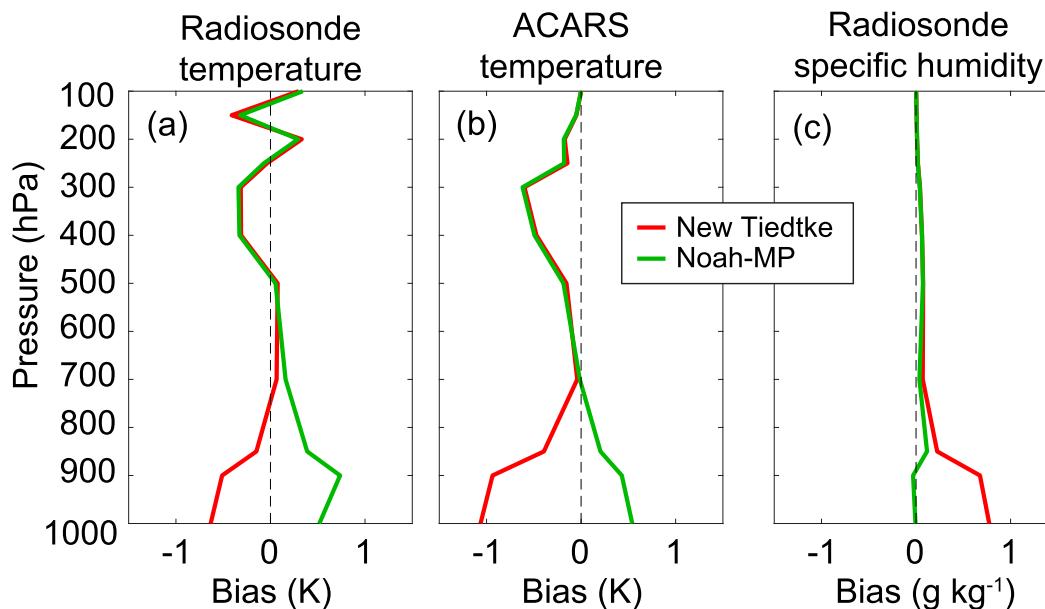
FIG. 12. Average 6-h forecast bias (forecast − observation) compared against (a) rawinsondes and (b) ACARS for temperature (K), and rawinsondes for (c) specific humidity (g kg^{-1}). Biases are averaged over the central United States and for the entire period from 1 May to 15 Jun 2017.

benefit from frameworks such as this that focuses on a holistic approach to model improvement. Model biases are clearly sensitive to model resolution; therefore, individual cycling systems are needed for different model-resolution systems. Future diagnosis of systematic model biases in other model systems and forecasts will be carried out.

REFERENCES

Anderson, J., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Arellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, https://doi.org/10.1175/2009BAMS2618.1.

Arakawa, A., and W. H. Schubert, 1974: Interaction of a cumulus cloud ensemble with large-scale environment, Part I. *J. Atmos. Sci.*, **31**, 674–701, https://doi.org/10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2.

Ballish, B. A., and V. K. Kumar, 2008: Systematic differences in aircraft and radiosonde temperatures: Implications for NWP

and climate studies. *Bull. Amer. Meteor. Soc.*, **89**, 1689–1708, https://doi.org/10.1175/2008BAMS2332.1.

Ban, N., J. Schmidli, and C. Schaer, 2015: Heavy precipitation in a changing climate: Does short-term summer precipitation increase faster? *Geophys. Res. Lett.*, **42**, 1165–1172, https://doi.org/10.1002/2014GL062588.

Bechtold, P., J. P. Chaboureau, A. Beljaars, A. K. Betts, M. Kohler, M. Miller, and J. L. Redelsperger, 2004: The simulation of the diurnal cycle of convective precipitation over land in a global model. *Quart. J. Roy. Meteor. Soc.*, **130**, 3119–3137, https://doi.org/10.1256/qj.03.103.

——, M. Koehler, T. Jung, F. Doblas-Reyes, M. Leutbecher, M. J. Rodwell, F. Vitart, and G. Balsamo, 2008: Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Quart. J. Roy. Meteor. Soc.*, **134**, 1337–1351, https://doi.org/10.1002/qj.289.

——, N. Semane, P. Lopez, J.-P. Chaboureau, A. Beljaars, and N. Bormann, 2014: Representing equilibrium and nonequilibrium convection in large-scale models. *J. Atmos. Sci.*, **71**, 734–753, https://doi.org/10.1175/JAS-D-13-0163.1.

Benjamin, S. G., and Coauthors, 2016: A North American Hourly Assimilation and Model Forecast Cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Bernardet, L. R., L. D. Grasso, J. E. Nachamkin, C. A. Finley, and W. R. Cotton, 2000: Simulating convective events using a high-resolution mesoscale model. *J. Geophys. Res.*, **105**, 14 963–14 982, https://doi.org/10.1029/2000JD900100.

Carbone, R. E., and J. D. Tuttle, 2008: Rainfall occurrence in the U.S. warm season: The diurnal cycle. *J. Climate*, **21**, 4132–4146, https://doi.org/10.1175/2008JCLI2275.1.

——, ——, D. Ahijevych, and S. Trier, 2002: Inferences of predictability associated with warm season precipitation episodes. *J. Atmos. Sci.*, **59**, 2033–2056, https://doi.org/10.1175/1520-0469(2002)059<2033:IOPAWW>2.0.CO;2.

Cavallo, S. M., J. Berner, and C. Snyder, 2016: Diagnosing model errors from time-averaged tendencies in the Weather Research and Forecasting (WRF) Model. *Mon. Wea. Rev.*, **144**, 759–779, https://doi.org/10.1175/MWR-D-15-0120.1.

Clark, A. J., W. A. J. Gallus, and T.-C. Chen, 2007: Comparison of the diurnal precipitation cycle in convection-resolving and non-convection-resolving mesoscale models. *Mon. Wea. Rev.*, **135**, 3456–3473, https://doi.org/10.1175/MWR3467.1.

Computational and Information Systems Laboratory, 2017: Cheyenne: HPE/SGI ICE XA System (NCAR Community Computing). National Center for Atmospheric Research, Boulder, CO, accessed 1 July 2019, https://doi.org/10.5065/D6RX99HX.

Dai, A., F. Giorgi, and K. E. Trenberth, 1999: Observed and model-simulated diurnal cycles of precipitation over the contiguous United States. *J. Geophys. Res.*, **104**, 6377–6402, https://doi.org/10.1029/98JD02720.

——, X. Lin, and K.-L. Hsu, 2007: The frequency, intensity, and diurnal cycle of precipitation in surface and satellite observations over low- and mid-latitudes. *Climate Dyn.*, **29**, 727–744, https://doi.org/10.1007/s00382-007-0260-y.

Davis, C. A., K. W. Manning, R. E. Carbone, S. B. Trier, and J. D. Tuttle, 2003: Coherence of warm-season continental rainfall in numerical weather prediction models. *Mon. Wea. Rev.*, **131**, 2667–2679, https://doi.org/10.1175/1520-0493(2003)131<2667: COWCRI>2.0.CO;2.

——, B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part II: Application to convective rain systems. *Mon. Wea. Rev.*, **134**, 1785–1795, https://doi.org/10.1175/MWR3146.1.

Dirmeyer, P. A., and Coauthors, 2012: Simulating the diurnal cycle of rainfall in global climate models: Resolution versus parameterization. *Climate Dyn.*, **39**, 399–418, https://doi.org/10.1007/s00382-011-1127-9.

Done, J., C. A. Davis, and M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the weather research and forecasting (WRF) model. *Atmos. Sci. Lett.*, **5**, 110–117, https://doi.org/10.1002/asl.72.

Duda, J. D., X. Wang, and M. Xue, 2017: Sensitivity of convection-allowing forecasts to land surface model perturbations and implications for ensemble design. *Mon. Wea. Rev.*, **145**, 2001–2025, https://doi.org/10.1175/MWR-D-16-0349.1.

ECMWF, 2014:Part IV: Physical processes. *IFS Documentation CY40R1*, Vol. 4, ECMWF, https://www.ecmwf.int/node/9204.

Emanuel, K. A., 1991: A scheme for representing cumulus convection in large-scale models. *J. Atmos. Sci.*, **48**, 2313–2329, https://doi.org/10.1175/1520-0469(1991)048<2313:ASFRCC>2.0.CO;2.

Evans, J. P., and S. Westra, 2012: Investigating the mechanisms of diurnal rainfall variability using a regional climate model. *J. Climate*, **25**, 7232–7247, https://doi.org/10.1175/JCLI-D-11-00616.1.

Fritsch, J. M., and R. E. Carbone, 2004: Improving quantitative precipitation forecasts in the warm season: A USWRP research and development strategy. *Bull. Amer. Meteor. Soc.*, **85**, 955–966, https://doi.org/10.1175/BAMS-85-7-955.

Grell, G. A., Y. H. Kuo, and R. J. Pasch, 1991: Semiprognostic tests of cumulus parameterization schemes in the middle latitudes. *Mon. Wea. Rev.*, **119**, 5–31, https://doi.org/10.1175/1520-0493(1991)119<0005:STOCPS>2.0.CO;2.

Haarsma, R. J., and Coauthors, 2016: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geosci. Model Dev.*, **9**, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, https://doi.org/10.1029/2008JD009944.

Jakob, C., and A. P. Siebesma, 2003: A new subcloud model for mass-flux convection schemes: Influence on triggering, updraft properties, and model climate. *Mon. Wea. Rev.*, **131**, 2765–2778, https://doi.org/10.1175/1520-0493(2003)131<2765:ANSMFM>2.0.CO;2.

Johnson, R. H., 1984: Partitioning tropical heat and moisture budgets into cumulus and mesoscale components—Implications for cumulus parameterization. *Mon. Wea. Rev.*, **112**, 1590–1601, https://doi.org/10.1175/1520-0493(1984)112<1590:PTHAMB>2.0.CO;2.

Jones, T. A., P. Skinner, K. Knopfmeier, E. Mansell, P. Minnis, R. Palikonda, and W. Smith Jr., 2018: Comparison of cloud microphysics schemes in a Warn-On-Forecast system using synthetic satellite objects. *Wea. Forecasting*, **33**, 1681–1708, https://doi.org/10.1175/WAF-D-18-0112.1.

Kain, J. S., and J. M. Fritsch, 1992: The role of the convective trigger function in numerical forecasts of mesoscale convective systems. *Meteor. Atmos. Phys.*, **49**, 93–106, https://doi.org/10.1007/BF01025402.

Klemp, J., W. Skamarock, and J. Dudhia, 2007: Conservative split-explicit time integration methods for the compressible nonhydrostatic equations. *Mon. Wea. Rev.*, **135**, 2897–2913, https://doi.org/10.1175/MWR3440.1.

Klinker, E., and P. D. Sardeshmukh, 1992: The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sci.*, **49**, 608–627, https://doi.org/10.1175/1520-0469(1992)049<0608:TDOMDI>2.0.CO;2.

Klocke, D., and M. J. Rodwell, 2014: A comparison of two numerical weather prediction methods for diagnosing fast-physics errors in climate models. *Quart. J. Roy. Meteor. Soc.*, **140**, 517–524, https://doi.org/10.1002/qj.2172.

Minnis, P., and Coauthors, 2008: Cloud detection in nonpolar regions for CERES using TRMM VIRS and *Terra* and *Aqua* MODIS data. *IEEE Trans. Geosci. Remote Sens.*, **46**, 3857–3884, https://doi.org/10.1109/TGRS.2008.2001351.

——, and Coauthors, 2011: CERES edition-2 cloud property retrievals using TRMM VIRS and *Terra* and *Aqua* MODIS Data—Part I: Algorithms. *IEEE Trans. Geosci. Remote Sens.*, **49**, 4374–4400, https://doi.org/10.1109/TGRS.2011.2144601.

——, K. Bedka, Q. Trepte, C. R. Yost, S. T. Bedka, B. Scarino, K. Khlopenkov, and M. M. Khaiyer, 2016: A consistent long-term cloud and clear-sky radiation property dataset from the Advanced Very High Resolution Radiometer (AVHRR). CDR Program Doc. CDRP-ATBD-0826, 159 pp.

Mooney, P. A., C. Broderick, C. L. Bruyere, F. J. Mulligan, and A. F. Prein, 2017: Clustering of observed diurnal cycles of precipitation over the United States for evaluation of a WRF multiphysics regional climate ensemble. *J. Climate*, **30**, 9267–9286, https://doi.org/10.1175/JCLI-D-16-0851.1.

Nakanishi, M., and H. Niino, 2006: An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397–407, https://doi.org/10.1007/s10546-005-9030-8.

——, and ——, 2009: Development of an improved turbulence closure model for the atmospheric boundary layer. *J. Meteor. Soc. Japan*, **87**, 895–912, https://doi.org/10.2151/jmsj.87.895.

NASA Langley Research Center, 2017:Satellite clouds and radiation property retrieval system. NASA, accessed 1 July 2019, https://satcorps.larc.nasa.gov.

Niu, G.-Y., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1.

Model description and evaluation with local-scale measurements. *J. Geophys. Res.*, **116**, D12109, https://doi.org/10.1029/2010JD015139.

Olson, J. B., J. S. Kenyon, W. A. Angevine, J. M. Brown, M. Pagowski, and K. Suselj, 2019: A description of the MYNN-EDMF scheme and the coupling to other components in WRF-ARW. NOAA Tech. Memo. OAR GSD-61, 42 pp., https://doi.org/10.25923/n9wm-be49.

Powers, J. G., and Coauthors, 2017: The Weather and Research Forecasting Model: Overview, system efforts, and future directions. *Bull. Amer. Meteor. Soc.*, **98**, 1717–1737, https://doi.org/10.1175/BAMS-D-15-00308.1.

Prein, A. F., and Coauthors, 2015: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Rev. Geophys.*, **53**, 323–361, https://doi.org/10.1002/2014RG000475.

Pritchard, M. S., and R. C. J. Somerville, 2009: Assessing the diurnal cycle of precipitation in a multi-scale climate model. *J. Adv. Model. Earth Syst.*, **1**, https://doi.org/10.3894/JAMES.2009.1.12.

Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, https://doi.org/10.1175/JCLI-D-11-00015.1.

Rodwell, M. J., and T. N. Palmer, 2007: Using numerical weather prediction to assess climate models. *Quart. J. Roy. Meteor. Soc.*, **133**, 129–146, https://doi.org/10.1002/qj.23.

——, and T. Jung, 2008: Understanding the local and global impacts of model physics changes: An aerosol example. *Quart. J. Roy. Meteor. Soc.*, **134**, 1479–1497, https://doi.org/10.1002/qj.298.

Schwartz, C. S., G. S. Romine, M. L. Weisman, R. A. Sobash, K. R. Fossell, K. W. Manning, and S. B. Trier, 2015: A real-time convection-allowing ensemble prediction system initialized by mesoscale ensemble Kalman filter analyses. *Wea. Forecasting*, **30**, 1158–1181, https://doi.org/10.1175/WAF-D-15-0013.1.

——, ——, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2019: NCAR's real-time convection-allowing ensemble project. *Bull. Amer. Meteor. Soc.*, **100**, 321–343, https://doi.org/10.1175/BAMS-D-17-0297.1.

Skamarock, W. C., and J. B. Klemp, 2008: A time-split non-hydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.*, **227**, 3465–3485, https://doi.org/10.1016/j.jcp.2007.01.037.

Skamarock, W., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.

Smirnova, T. G., J. M. Brown, and S. G. Benjamin, 1997: Performance of different soil model configurations in simulating ground surface temperature and surface fluxes. *Mon. Wea. Rev.*, **125**, 1870–1884, https://doi.org/10.1175/1520-0493(1997)125<1870:PODSMC>2.0.CO;2.

——, ——, ——, and J. S. Kenyon, 2016: Modifications to the Rapid Update Cycle Land Surface Model (RUC LSM) available in the Weather Research and Forecasting (WRF) Model. *Mon. Wea. Rev.*, **144**, 1851–1865, https://doi.org/10.1175/MWR-D-15-0198.1.

Suhas, E., and G. J. Zhang, 2014: Evaluation of trigger functions for convective parameterization schemes using observations. *J. Climate*, **27**, 7647–7666, https://doi.org/10.1175/JCLI-D-13-00718.1.

Sundqvist, H., 1978: Parameterization scheme for non-convective condensation including prediction of cloud water-content. *Quart. J. Roy. Meteor. Soc.*, **104**, 677–690, https://doi.org/10.1002/qj.49710444110.

Surcel, M., M. Berenguer, and I. Zawadzki, 2010: The diurnal cycle of precipitation from continental radar mosaics and numerical weather prediction models. Part I: Methodology and seasonal comparison. *Mon. Wea. Rev.*, **138**, 3084–3106, https://doi.org/10.1175/2010MWR3125.1.

Tang, Y., H. W. Lean, and J. Bornemann, 2013: The benefits of the Met Office variable resolution NWP model for forecasting convection. *Meteor. Appl.*, **20**, 417–426, https://doi.org/10.1002/met.1300.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, https://doi.org/10.1175/2008MWR2387.1.

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779–1800, https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2.

Wallace, J., 1975: Diurnal variations in precipitation and thunderstorm frequency over the conterminous United States. *Mon. Wea. Rev.*, **103**, 406–419, https://doi.org/10.1175/1520-0493(1975)103<0406:DVIPAT>2.0.CO;2.

Weber, N., and C. Mass, 2019: Subseasonal weather prediction in a global convection-permitting model. *Bull. Amer. Meteor. Soc.*, **100**, 1079–1089, https://doi.org/10.1175/BAMS-D-18-0210.1.

Williams, K. D., and M. E. Brooks, 2008: Initial tendencies of cloud regimes in the Met Office Unified Model. *J. Climate*, **21**, 833–840, https://doi.org/10.1175/2007JCLI1900.1.

Wong, M., and W. C. Skamarock, 2016: Spectral characteristics of convective-scale precipitation observations and forecasts. *Mon. Wea. Rev.*, **144**, 4183–4196, https://doi.org/10.1175/MWR-D-16-0183.1.

Xie, S. C., and M. H. Zhang, 2000: Impact of the convection triggering function on single-column model simulations. *J. Geophys. Res.*, **105**, 14 983–14 996, https://doi.org/10.1029/2000JD900170.

——, and Coauthors, 2002: Intercomparison and evaluation of cumulus parametrizations under summertime midlatitude continental conditions. *Quart. J. Roy. Meteor. Soc.*, **128**, 1095–1135, https://doi.org/10.1256/003590002320373229.

Yang, Z.-L., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. *J. Geophys. Res.*, **116**, D12110, https://doi.org/10.1029/2010JD015140.

Zhang, C., and Y. Wang, 2017: Projected future changes of tropical cyclone activity over the western North and South Pacific in a 20-km-mesh regional climate model. *J. Climate*, **30**, 5923–5941, https://doi.org/10.1175/JCLI-D-16-0597.1.

——, ——, and K. Hamilton, 2011: Improved representation of boundary layer clouds over the southeast Pacific in ARW-WRF using a modified Tiedtke cumulus parameterization scheme. *Mon. Wea. Rev.*, **139**, 3489–3513, https://doi.org/10.1175/MWR-D-10-05091.1.

Zhang, G. J., 2002: Convective quasi-equilibrium in midlatitude continental environment and its effect on convective parameterization. *J. Geophys. Res.*, **107**, 4220, https://doi.org/10.1029/2001JD001005.

Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 621–638, https://doi.org/10.1175/BAMS-D-14-00174.1.

Zhang, Y., and S. A. Klein, 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. *J. Atmos. Sci.*, **67**, 2943–2959, https://doi.org/10.1175/2010JAS3366.1.