Evaluation of Mesoscale Convective Systems in Climate Simulations: Methodological Development and Results from MPAS-CAM over the United States

ZHE FENG,[a] FENGFEI SONG,[a] KOICHI SAKAGUCHI,[a] AND L. RUBY LEUNG[a]

[a] *Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, Washington*

(Manuscript received 28 February 2020, in final form 12 August 2020)

ABSTRACT: A process-oriented approach is developed to evaluate warm-season mesoscale convective system (MCS) precipitation and their favorable large-scale meteorological patterns (FLSMPs) over the United States. This approach features a novel observation-driven MCS-tracking algorithm using infrared brightness temperature and precipitation features at 12-, 25-, and 50-km resolution and metrics to evaluate the model large-scale environment favorable for MCS initiation. The tracking algorithm successfully reproduces the observed MCS statistics from a reference 4-km radar MCS database. To demonstrate the utility of the new methodologies in evaluating MCS in climate simulations with mesoscale resolution, the process-oriented approach is applied to two climate simulations produced by the Variable-Resolution Model for Prediction Across Scales coupled to the Community Atmosphere Model physics, with refined horizontal grid spacing at 50 and 25 km over North America. With the tracking algorithm applied to simulations and observations at equivalent resolutions, the simulated number of MCS and associated precipitation amount, frequency, and intensity are found to be consistently underestimated in the central United States, particularly from May to August. The simulated MCS precipitation shows little diurnal variation and lasts too long, while the MCS precipitation area is too large and its intensity is too weak. The model is able to simulate four types of observed FLSMP associated with frontal systems and low-level jets (LLJ) in spring, but the frequencies are underestimated because of low-level dry bias and weaker LLJ. Precipitation simulated under different FLSMPs peak during the daytime, in contrast to the observed nocturnal peak. Implications of these findings for future model development and diagnostics are discussed.

KEYWORDS: Convective storms; Mesoscale systems; Storm environments; Cloud tracking/cloud motion winds; Climate models; Model evaluation/performance

1. Introduction

Realistic representation of the hydrologic cycle and related extremes in Earth system models has important societal benefits. As Earth continues to warm, hydrological cycle changes such as "the wet get wetter and the dry get drier" (Held and Soden 2006; Trenberth 2011) have significant implications for infrastructure planning and management of water resources. More importantly, the increased water vapor supply to storms under a warmer climate and the additional latent heat released can invigorate storms and enhance precipitation intensity (Trenberth et al. 2003), increasing the risk of flooding. Large disagreements in phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012) projections of future precipitation show that the current generation of global climate models (GCMs) have large uncertainties in simulating one of the key components of the water cycle. In contrast, high-resolution GCMs have been shown to provide improved, globally consistent information on hydrologic hazards and impacts compared to traditional coarse resolution models (Roberts et al. 2018).

In the midlatitude, intense precipitation events are commonly associated with synoptic-scale disturbances such as extratropical cyclones and mesoscale disturbances such as tropical cyclones. In the contiguous United States (CONUS), mesoscale convective systems (MCSs), the largest form of deep convective storms, produce over 50% of annual and seasonal rainfall over large regions east of the Continental Divide (ECONUS; Feng et al. 2019; Haberlie and Ashley 2019). MCSs consist of an assembly of cumulonimbus clouds on scales of 100 km or greater and produce mesoscale circulations (Houze 2004, 2018). They are also responsible for over half of the extreme daily rainfall events, trailed by synoptic and tropical systems in the ECONUS (Stevenson and Schumacher 2014). The warming climate in the past three decades has already resulted in an observed increase in the frequency and intensity of MCS precipitation during spring (Feng et al. 2016), and such increases are projected to further intensify under future warming (Prein et al. 2017).

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-20-0136.s1.

Corresponding author: Zhe Feng, zhe.feng@pnnl.gov

DOI: 10.1175/JCLI-D-20-0136.1

MCSs are notoriously difficult to simulate in traditional GCMs. This is partly due to the multiscale interactions between convective-scale dynamics and microphysics and the upscale feedbacks and interactions through latent heating (Feng et al. 2018; Yang et al. 2017), which are challenging for GCMs because the scale-separation assumptions used in most convection parameterizations are violated (Moncrieff 2010). Failure of traditional GCMs in simulating MCSs over ECONUS is manifested in the erroneous diurnal cycle of precipitation and large warm bias in the near-surface temperature (Lin et al. 2017). Various causes of the long-standing warm and dry biases in GCMs are documented in the Clouds Above the United States and Errors at the Surface (CAUSES) project (Morcrette et al. 2018), including biases in simulating different types of clouds (Van Weverberg et al. 2018) and errors in the surface energy budget (Ma et al. 2018).

Besides important influence of physical processes, MCSs in ECONUS are closely related to the large-scale environments (Feng et al. 2019; Song et al. 2019). Long-term observations of MCSs in the United States show distinctly different large-scale environments associated with the initiation and structure of MCSs in spring versus summer (Feng et al. 2019; Song et al. 2019). During spring, baroclinic waves and frontal systems provide strong lifting mechanisms and the Great Plains low-level jet (LLJ) provides anomalous moisture for favorable dynamical and thermodynamical environments for MCS development. In contrast, during summer, favorable environments featuring significantly weaker baroclinic lifting and thermodynamic instability suggest much lower predictability of MCSs compared to spring (Song et al. 2019). Besides limitations in physics parameterizations, it is unclear if GCMs are able to simulate different types of observed large-scale environments favorable for MCS development (Song et al. 2019).

Several different approaches to higher resolution modeling have been used to improve simulation of MCSs. The multiscale modeling framework (MMF), in which a cloud-resolving model is embedded in each GCM grid column to replace the cumulus parameterizations, have shown some advantages in simulating MCS-like features over ECONUS (Kooperman et al. 2013; Lin et al. 2019; Pritchard and Somerville 2009a,b). Convection-permitting regional models (grid spacing ≤ 4 km) can produce significantly more realistic warm season MCSs comparable to observations (Feng et al. 2018; Prein et al. 2020), but they are still prohibitively expensive for long-term climate simulations. Global models with grid spacing finer than 100 km have the potential for substantial improvements at synoptic-scale processes (Roberts et al. 2018), which can result in more intense and less frequent precipitation that agrees better with observations (Chen and Dai 2019; Wehner et al. 2014). More recently, the variable-resolution (VR) approach, in which higher resolution over a region of interest and a gradual transition to coarser resolution elsewhere, has been explored for studying regional processes at a lower computational cost compared to global high-resolution models. Previous works showed that VR models are able to produce climate statistics such as precipitation in the refinement region comparable to their global uniform high-resolution counterparts (Gettelman et al. 2018; Sakaguchi et al. 2015). However, it is unclear if the increase in intense precipitation events with higher resolution reflects a correct representation of the underlying processes, which is critically important for providing reliable future projections of changes in the hydrological cycle based on models. Better simulations of synoptic-scale processes and intense precipitation by models with grid spacing finer than 100 km (Chen and Dai 2019; Demory et al. 2014; Roberts et al. 2018) may have important impacts on modeling MCS and their characteristics in the United States, but such effects have not been investigated.

Evaluating high-resolution simulations of MCS is particularly timely as a set of climate experiments with horizontal resolution of at least 50 km is becoming available as part of phase 6 of the Coupled Model Intercomparison Project (CMIP6; Eyring et al. 2016) High Resolution Model Intercomparison Project (HighResMIP; Haarsma et al. 2016) to systematically assess the robust benefits of increased model resolution based on a multimodel ensemble. Previous high-resolution climate model evaluation efforts have not focused on directly evaluating the ability of models to simulate MCS and non-MCS precipitation against observations. Therefore, new methods to evaluate MCS precipitation characteristics as well as their favorable large-scale environments simulated by climate models across different resolutions are needed.

In this study, we develop a process-oriented approach to systematically evaluate the performance of climate models at mesoscale resolutions (grid spacing 10–50 km) in simulating warm-season MCS-like precipitation features and their favorable large-scale environments over the United States. Climate simulations at mesoscale resolutions are becoming more commonly available (e.g., Caldwell et al. 2019; Gutjahr et al. 2019; Roberts et al. 2019), motivating the need to develop evaluation and diagnostic methods to probe their ability in capturing mesoscale features such as MCS, which has been a challenge for models with parameterized convection (e.g., Davis et al. 2003). To reduce the computational cost of running global simulations at mesoscale resolutions, the VR approach is taken by coupling the dynamical core of the Model for Prediction Across Scales (VR-MPAS; Skamarock et al. 2012) with VR capability to the widely used Community Atmosphere Model version 5 (CAM5) physics (Sakaguchi et al. 2015; Zhao et al. 2016). Simulations from this model with regional refinement at mesoscale resolution over North America are analyzed as a test case to demonstrate the utility of the new diagnostic approach developed in this study for model evaluation and guiding the directions for future development.

At the core of the developed framework is a novel MCS tracking algorithm for mesoscale resolution models based on high-resolution observations. To facilitate understanding of the model biases in simulating MCS, a metric is also developed to evaluate the ability of the model in reproducing the observed favorable large-scale environments for MCS development. The paper is organized as follows: section 2 describes the model simulation setup and observational datasets; section 3 details the methodological developments for tracking MCS and identifying their large-scale environments; application of the MCS evaluation methodology to two sets of climate simulations is presented in section 4; section 5 evaluates the

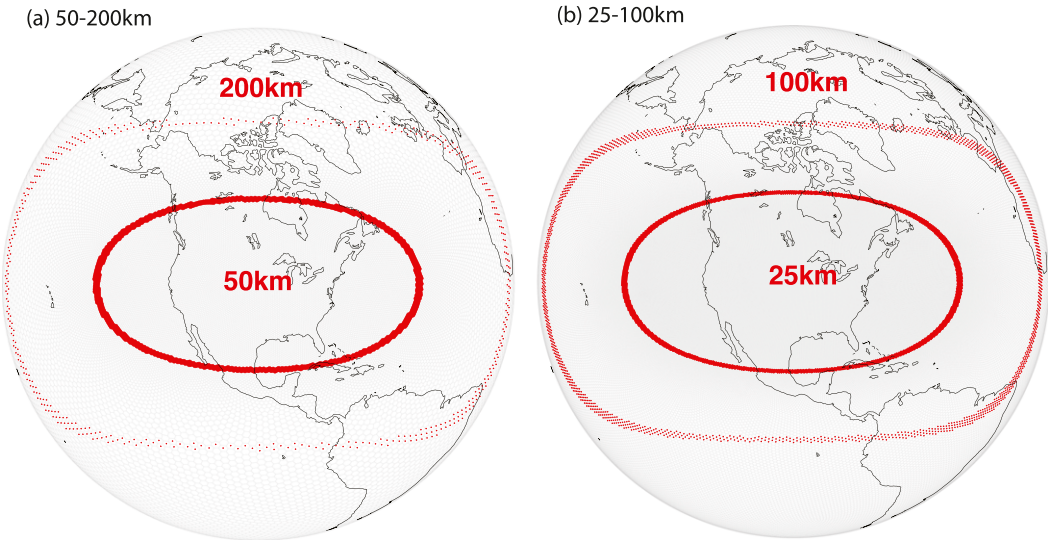

FIG. 1. Variable-resolution mesh for the two MPAS-CAM simulations used in this study. Regional refinement of horizontal grid spacing at (a) 50 km ("MP50km") and (b) 25 km ("MP25km") over CONUS are shown in the inside mesh. The grid spacing is smoothly coarsened to 200 km in (a) and 100 km in (b) outside of North America.

simulated favorable MCS large-scale environments and associated precipitation evolution; Summary of the key findings and discussion on guidance for future model development are provided in section 6.

2. Model setup and observational data

a. Model setup

This study uses the beta version of the Community Earth System Model (CESM) version 2, to which the nonhydrostatic MPAS–Atmosphere (Skamarock et al. 2012) version 4 has been added as an experimental option for dynamical cores in CAM. A similar model code was used by Gettelman et al. (2018) to test the VR capability of another dynamical core, the spectral element model. Two regionally refined global meshes were created with a high-resolution region centered over the CONUS (Fig. 1). As shown in Fig. 1, the grid spacing of the VR grids (25–100 and 50–200 km, hereafter referred to as MP25km and MP50km) varies smoothly from low- to high-resolution, taking advantages of the grid generation used by MPAS (Ju et al. 2011; Ringler et al. 2010).

We use the physics package of CAM5.4, which is an interim version toward CAM version 6 (Bogenschutz et al. 2018) and was available at the time MPAS coupling was done. Physics parameterizations in CAM5.4 are summarized in Table 1. A major difference between CAM5.4 and the previous version CAM5.0 (the atmosphere component of CESM1) is the prognostic precipitation mass and number concentrations (rain and snow; rimmed ice particles are not considered) in the new cloud microphysics scheme, MG2 (Gettelman and Morrison 2015). Using prognostic concentrations reduces model assumptions necessary for a diagnostic approach (e.g., advection of precipitating particles is assumed to be negligible), so it is more appropriate for high-resolution simulations.

Following Bacmeister et al. (2014) and the HighResMIP protocol, aerosols are prescribed in this study. The monthly mean aerosol mass concentrations for year 2000 conditions derived from a previous CAM4 simulation on a 1° (0.9° × 0.125°) grid with the three-moment modal aerosol scheme (Liu et al. 2012) are used as input. With the prescribed aerosols, aerosol number concentrations are calculated using a prescribed relationship between the mass and number concentrations, and both concentrations are passed to the cloud microphysics for droplet activation and ice nucleation (Gettelman et al. 2015). Therefore, both direct and indirect aerosol effects are considered in the gridscale cloud microphysics but likely with larger uncertainty than modeling using prognostic, interactive aerosols. The simpler microphysics in the shallow and deep convection schemes do not consider the impact of varying aerosol number concentrations. Such a simple representation of aerosols may limit the areal coverage of anvil clouds (Chakraborty et al. 2016) and possible convection invigoration by aerosols (Chen et al. 2020).

Parameters for the physics parameterizations are tuned based on a global CAM5.4 simulation with the finite-volume

TABLE 1. Physics parameterizations used in CAM5.4 and their references.

Physics	Reference
Deep convection	Neale et al. (2008) and Zhang and McFarlane (1995)
Shallow convection	Park and Bretherton (2009)
Boundary layer	Bretherton and Park (2009)
Cloud macrophysics	Park et al. (2014)
Cloud microphysics	Gettelman and Morrison (2015)
Radiation	Iacono et al. (2008)
Prescribed aerosol	Bacmeister et al. (2014) and Kiehl et al. (2000)

TABLE 2. Convective time scales used in the Zhang–McFarlane convection scheme for the two sets of simulations.

Grids	Physics Δt (s)	Dynamics Δt (s)	Convective time scale (s)
50–200 km	900	150	1800
25–100 km	600	90	1200

dynamical core on the 1° grid. No attempt has been made to tune the parameter values differently at each resolution, so the parameter values are not necessarily optimal for the two VR simulations conducted in this study. It is noted that resolution-dependent tuning seems to be necessary to take advantage of increased resolution for many aspects of the simulated hydrological cycles (Xie et al. 2018). But following the HighResMIP protocol for a clean evaluation of the impact of horizontal resolution, we do not tune model parameters differently at each resolution. Exceptions are time step lengths, numerical diffusion coefficients, and the convective time scale used in the Zhang–McFarlane convection scheme (Neale et al. 2008; Zhang and McFarlane 1995) (Table 2). Although the Zhang–McFarlane scheme does not explicitly address scale awareness, especially with regard to the convective time scale, adjusting the time scale in proportion to the physics time step length has been shown to mitigate some of its resolution sensitivity (Gross et al. 2018; Williamson 2013). Typical physics time step and convective time scale are 1800 and 3600 s, respectively, for global simulations at ~1° grid spacing. Note that the dynamical time steps are constrained by the smallest grid spacings of the variable-resolution mesh.

Our simulations follow the setup of the Atmosphere Model Intercomparison Project (''FAMIPC5'' compset of CESM), in which the atmosphere and land models are active while sea surface temperature (SST) and sea ice cover are prescribed based on observations. We use the SST and sea ice cover from the ERA-Interim reanalysis (Dee et al. 2011). The 6-hourly ERA-Interim data are averaged to daily for used as model input. The land model is the Community Land Model version 4 (Lawrence et al. 2011) with prescribed, static land cover types and vegetation properties (e.g., leaf area index) roughly representing the conditions for year 2000. The simulations were run from 1989 to 2013, but only the last 10 years (2004–13) are analyzed to take advantage of better availability of observational data (section 2b).

As suggested by previous studies, we do not expect many aspects of MCSs to be well simulated by this model configuration at resolutions requiring parameterized convection. We emphasize again that the focus of the study is to describe and demonstrate the utility of our new diagnostic approach targeted at simulated MCSs and associated large-scale environments. In particular, the approach is designed to reveal whether the intensity and diurnal cycle of precipitation are simulated for the correct reasons. Along with the model configuration similar to those in HighResMIP, the results presented below would be useful to many users and developers of global climate models.

b. Observation datasets

The observational dataset used in this study is a recently developed 13-yr (2004–16) high-resolution (~4 km, 1 hourly) MCS database (Feng et al. 2019). This dataset synthesizes geostationary infrared brightness temperature T_b, Next Generation Radar Network (NEXRAD) 3D radar reflectivity and Stage IV radar-based precipitation estimates with rain gauge bias correction. Long-lived and intense MCSs (i.e., robust MCSs) were defined as a large cold cloud system (CCS; $T_b < 241$ K) area exceeding 6×10^4 km^2, with a precipitating feature (PF) of at least 100 km, containing convective feature radar reflectivity > 45 dBZ at any vertical level, and persisting for at least 6 h. Feng et al. (2019) showed that these robust MCSs account for 50%–60% warm season (March–August) total precipitation over the majority of the central United States. These MCSs are the main focus in the current study.

To match the horizontal grid spacing of high-resolution GCMs, including the two sets of VR-MPAS simulations in this study, the 4-km MCS database is down-sampled to three sets of coarser resolution (12, 25, 50 km) datasets using the Earth System Modeling Framework (ESMF) regridding software (https://www.ncl.ucar.edu/Applications/ESMF.shtml) and the NetCDF Operators (Zender 2019). Brightness temperature T_b and precipitation (i.e., flux variables) are regridded using conservative mapping, and MCS location masks are regridded using nearest neighbor to preserve their tracked numbers. The regridding procedure essentially creates a reference MCS dataset that depicts MCS T_b and precipitation characteristics at the respective coarsened spatial resolutions.

3. Methodological development for model evaluation

a. Methodology to track MCS across resolutions

While the coarsened MCS database provides a reference observational dataset at the equivalent model grid spacing, it is difficult to directly compare with the model simulations because the MCS tracking uses radar reflectivity at 4 km resolution, which is not available from the model outputs. To facilitate evaluation of model-simulated MCS, we develop a new method to track MCSs using observed T_b and precipitation data. The procedure is described in detail in this section.

Since T_b and precipitation are two variables commonly observed by various remote sensing platforms (e.g., satellite and radar), we choose these two variables to describe deep cumulonimbus cloud systems with intense precipitation that achieve mesoscale dimension and longevity, hence satisfying the MCS definition. In addition, model simulations commonly output outgoing longwave radiation (OLR, which can be converted to T_b) and surface precipitation, allowing more direct comparisons with observations. An MCS tracking algorithm that can be consistently applied to these two variables in the observations and model simulations is appealing.

The main workflow for tracking MCSs is similar to the procedure of the Flexible Object Tracker (FLEXTRKR; Feng et al. 2018, 2019). Tracking is performed on CCS associated with deep convection. The CCS segmentation follows the detect and spread method detailed in Feng et al. (2018) to iteratively grow a cold cloud core with $T_b < 225$ K outwards to reach 241 K. CCSs between two successive time steps (1 h apart) are tracked if they overlap for more than 50% of their area. Tracking is terminated if no features between two time

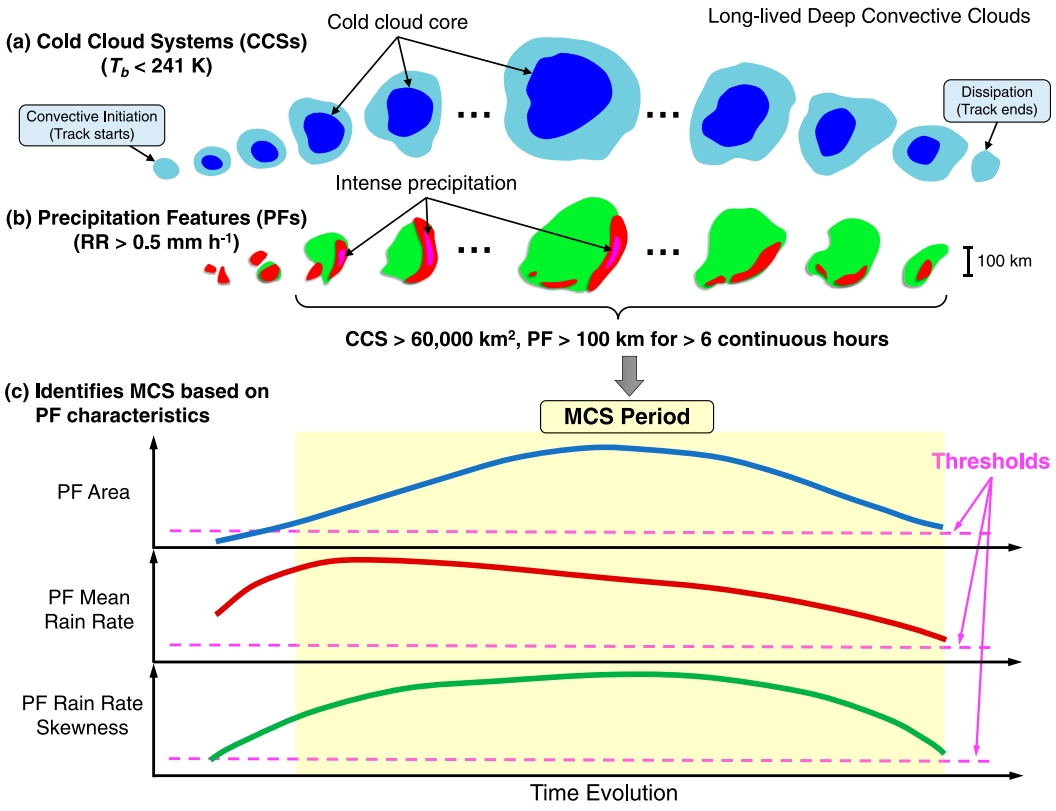

FIG. 2. Schematic of MCS identification using joint CCS and PF characteristics developed in this study. (a) Depiction of the evolution of long-lived deep convective clouds from cloud-top view perspective, (b) depiction of the evolution of collocated precipitation associated with the deep convective clouds, and (c) identification of MCS using both CCS and PF characteristics. The yellow shading in (c) denotes the MCS period as defined by CCS > 60 000 km^2, with a PF major axis length larger than 100 km for longer than 6 continuous hours. During this period, the PF area, mean rain rate, and rain-rate skewness must exceed the thresholds denoted by the magenta dashed lines to qualify as an MCS.

steps satisfy the overlap criteria. After deep convective clouds are tracked using CCSs, we use precipitation data within the CCS to better characterize the evolution of the convective system (Figs. 2a,b). This is where the identification of MCS differs from previous work. Feng et al. (2019) used persistence of radar defined PF size and echo intensity at 4-km resolution to identify MCSs. It is not clear how MCS-associated PF characteristics change at coarser resolutions such as those investigated in this study. Conceptually (Fig. 2b), MCS consists of both convective features that produce heavy precipitation over a relatively small area and a stratiform region that produces moderate-to-light precipitation over a large area during the genesis and mature stages of its life cycle (Feng et al. 2019). Therefore, it is expected that MCS typically contains a large PF with moderately high mean rain rate, especially during the upscale growth stage where most of the precipitation are convective, and a positive skewness for the rain rate distribution during the mature stage when mixed convective and stratiform rainfall coexist.

Our aim is to derive key PF quantities that can be used to identify MCSs at different resolutions. To quantify the PF characteristics, we make use of the 4-km MCS database that has been regridded to coarser resolutions described in section 2b. For each MCS tracked by the 4-km radar data, we calculate their PF area, major axis length, mean rain rate, and rain rate skewness within the MCS location masks using the respective coarsened precipitation dataset (Fig. S1 in the online supplemental material). Only MCSs that spent over half their lifetime in the central United States are included in this calculation. Figure 3 shows several bottom percentile values for these four PF parameters at 25 km resolution as a function of MCS lifetime. In Fig. 3, a majority of the radar tracked MCSs have specific PF parameters higher than those indicated by the dots. For example, the 5th percentile value of MCS PF area for MCS lifetime of 15 h is approximately 7500 km^2. That means 95% of the radar tracked MCSs lasting 15 h contain a PF larger than 7500 km^2. It is apparent that all PF parameters increase with MCS lifetime, suggesting that longer-lived MCSs generally have larger and more heavily precipitating PFs, with more positively skewed rain rate distributions (i.e., more convective-like precipitation).

Linear fit lines to each of the PF values were determined and their intercepts and slopes are provided in the legends of Fig. 3. Similar fits for 12- and 50-km resolution are shown in Figs. S2

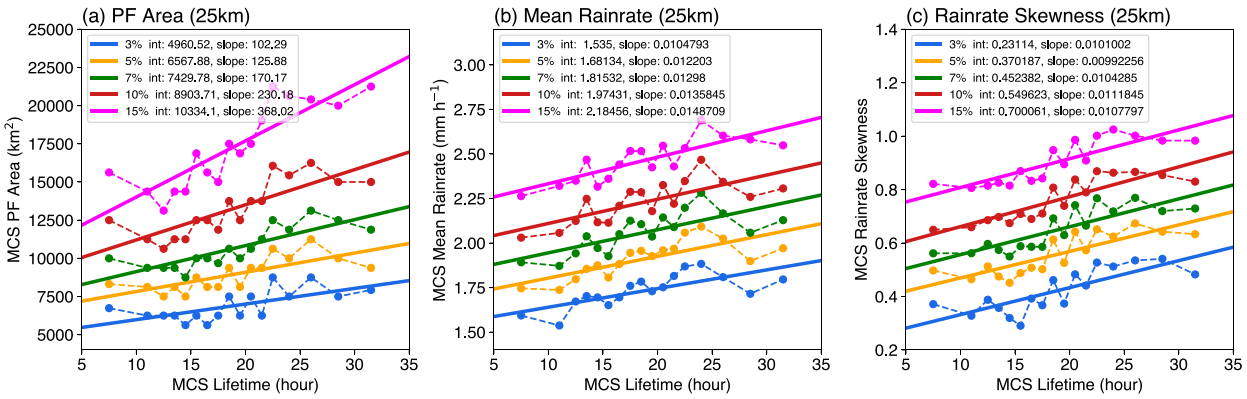

FIG. 3. MCS PF characteristics at 25-km resolution as a function of MCS lifetime: (a) PF area, (b) PF mean rain rate, and (c) PF rain-rate skewness. Dots are specific percentile values shown in the legend, and solid lines are linear fits to each set of percentile values. The fit intercepts and slopes are shown in the legends.

and S3. These fit lines are used as thresholds to identify MCSs (Fig. 2c). With these collocating CCS and PF characteristics, we define an MCS in this study as a convective system with:

1) CCS $> 6 \times 10^4$ km^2 containing a PF with major axis length > 100 km (same as the standard FLEXTRKR at 4 km);
2) PF area, mean rain rate, and rain rate skewness larger than the lifetime dependent thresholds provided by the linear fits in Fig. 3 and Fig. S3;
3) both conditions 1 and 2 are met continuously for at least 6 h.

A conceptual depiction of our MCS definition is shown in Fig. 2.

The idea of using PF characteristics to identify MCSs in the absence of 3D radar reflectivity data is similar to our previous work in Feng et al. (2016), where we used hourly precipitation data alone to track MCS in constructing a long-term historical MCS database over the CONUS. Compared to that previous method, our current methodology in tracking MCSs is more stringent because it requires both deep convective cloud systems (i.e., CCS) and heavy precipitating features (PFs) to achieve mesoscale dimension and persists for longer than the typical isolated convective storms, both of which are key distinctions of MCSs. The PF criteria are designed to exclude large cloud and precipitation bands associated with frontal systems and/or midlatitude cyclones that would satisfy the size and duration requirement, but the precipitation is not convective in nature and should not be included as MCSs.

To find the optimal PF parameter thresholds that will result in an MCS database more consistent with our reference 4-km MCS database, we tested various combinations of these PF parameters under each resolution separately. This is done manually by varying the PF parameter values for different sets of percentile values together (e.g., choosing percentile values of 3rd, 5th, and 10th for all three PF parameters) and comparing the resulting MCS characteristics to those of the reference 4-km dataset. Obtaining the smallest error in the spatial distribution of MCS precipitation amount and the fraction to total precipitation (Fig. 5) are more important than obtaining a similar number of MCS as shown in Fig. 4. Choosing lower PF thresholds results in more but weaker MCS identified and a positive bias in the MCS precipitation amount; on the other hand, choosing higher PF thresholds retains only the stronger MCS resulting in underestimating MCS precipitation amount. The PF parameters that produce the smallest error in MCS precipitation amount compared with the 4-km reference MCS database are shown in Table 3. The PF characteristics thresholds are adjusted for each set of resolutions. As the resolution coarsens, the PF area thresholds increase, mean rain rate intensity and skewness thresholds decrease, representing weaker PF with reduced spatial variability at coarser resolutions. We note that except for the MCS precipitation amount, the impact of the PF parameter values on the overall MCS characteristics (see section 3b) is relatively small, suggesting that our results are not particularly sensitive to the exact PF thresholds used.

b. *Evaluation of MCS tracking across resolutions*

The seasonal cycle of MCSs from observations at four different resolutions in the central United States is shown in Fig. 4. Overall, tracking of MCSs using PF at three coarser resolutions provides a consistent MCS seasonal cycle with MCS activity increasing from March, peaking in June, and decreasing in the fall. There are some observable differences between the PF tracking and the 4-km reference dataset. The total number of MCS systematically decreases with lower resolution, particularly from March to June. This is expected because the T_b and precipitation data are smeared at coarser resolution, making nearby MCSs that are separable at high resolution indistinguishable, particularly at 50-km resolution. Although the VR-MPAS simulations in this study do not include 12-km horizontal grid spacing, simulations at this refinement grid spacing are ongoing and would benefit from the MCS tracking algorithm tuned for this resolution.

The spatial distribution of the average number of MCS, MCS precipitation amount, and fraction of MCS-to-total precipitation from March–August are shown in Fig. 5. The spatial distributions in these important MCS quantities are quite consistent across the four resolutions, except the spatial

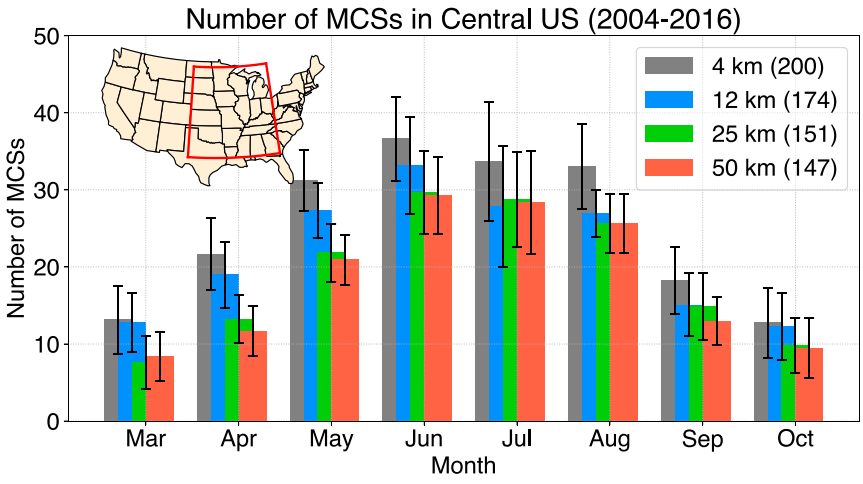

FIG. 4. Monthly average number of observed MCS occurrences in the central United States (red box in the inset map) per year between 2004 and 2016. The color bars show the averaged number tracked at each respective resolution, and the error bars show the standard deviation from the 13 years. The numbers in the figure legend are the average number of MCSs from March to October.

distribution of the number of MCSs seems to be slightly higher at coarser resolution. While this seems to contradict the lower number of MCSs shown in Fig. 4, the spatial distribution of the number of MCSs in Fig. 5 is also affected by the size of MCSs. As the resolution decreases, neighboring clouds and PFs that are separated at 4-km resolution could be smeared and appear as one larger MCS that covers a larger area during its lifetime. The smearing effect is supported by comparison of the MCS characteristics across four resolutions (Fig. S4). While the distribution of MCS lifetime is very consistent between the coarse resolution tracking applied to the reference dataset, the MCS PFs are successively larger and have weaker mean rain rate (except for 12 km, which remains comparable with the reference 4 km) and smaller rain rate skewness with decreasing resolution. Therefore, the spatial distribution shows a slightly higher number of MCSs at coarser resolution, although the impact on MCS precipitation amount and fraction of MCS-to-total precipitation is quite small (Fig. 5).

Besides statistical agreement, we also find that a majority of the major individual MCS events are consistently captured in our tracking across all four resolutions with high statistical skill scores (not shown). This is rather encouraging particularly for coarser resolutions such as 25 and 50 km. We demonstrated that given T_b (or OLR) and precipitation similar to those in observations, our algorithm can reliably track MCSs across resolutions ranging from 12 to 50 km that are consistent with the reference 4-km MCS database. The new MCS tracking algorithm tailored for resolutions at 25 and 50 km is applied to the VR-MPAS model simulations to demonstrate its utility in evaluating the performance of the MCS simulations in section 4.

c. *Methodology to evaluate large-scale environment favorable for MCS initiation*

Song et al. (2019) identified four favorable large-scale meteorological patterns (FLSMPs) associated with MCS initiation during boreal spring by applying self-organizing map (SOM) analysis to the North American Regional Reanalysis (NARR) dataset (Mesinger et al. 2006). In their application of SOM, neural network was used to discriminate the different environments for MCS initiation based on the Euclidean distance among the horizontal winds at three levels (200, 500, and 925 hPa) and specific humidity at two levels (500 and 925 hPa). Large-scale patterns that are similar to each other, as measured by the Euclidean distance among all input variables, are clustered through an iterative procedure to train the SOM. This procedure updates the ''winning'' pattern with the smallest Euclidean distance among the initiating nodes until the

TABLE 3. PF parameters used to identify MCSs at 12-, 25-, and 50-km resolution.

	12 km	25 km	50 km
Minimum rain rate to define a PF	1 mm h^{-1}	0.5 mm h^{-1}	0.5 mm h^{-1}
PF area (km^2)	3491.06 + 78.868 × lifetime	10 334.1 + 368.02 × lifetime	14 675.3 + 350.918 × lifetime
PF mean rain rate (mm h^{-1})	2.388 99 + 0.013 658 2 × lifetime	2.184 56 + 0.014 870 9 × lifetime	1.777 64 + 0.016 907 3 × lifetime
PF skewness (unitless)	0.568 637 + 0.007 207 69 × lifetime	0.700 061 + 0.010 779 7 × lifetime	0.203 418 + 0.011 769 5 × lifetime

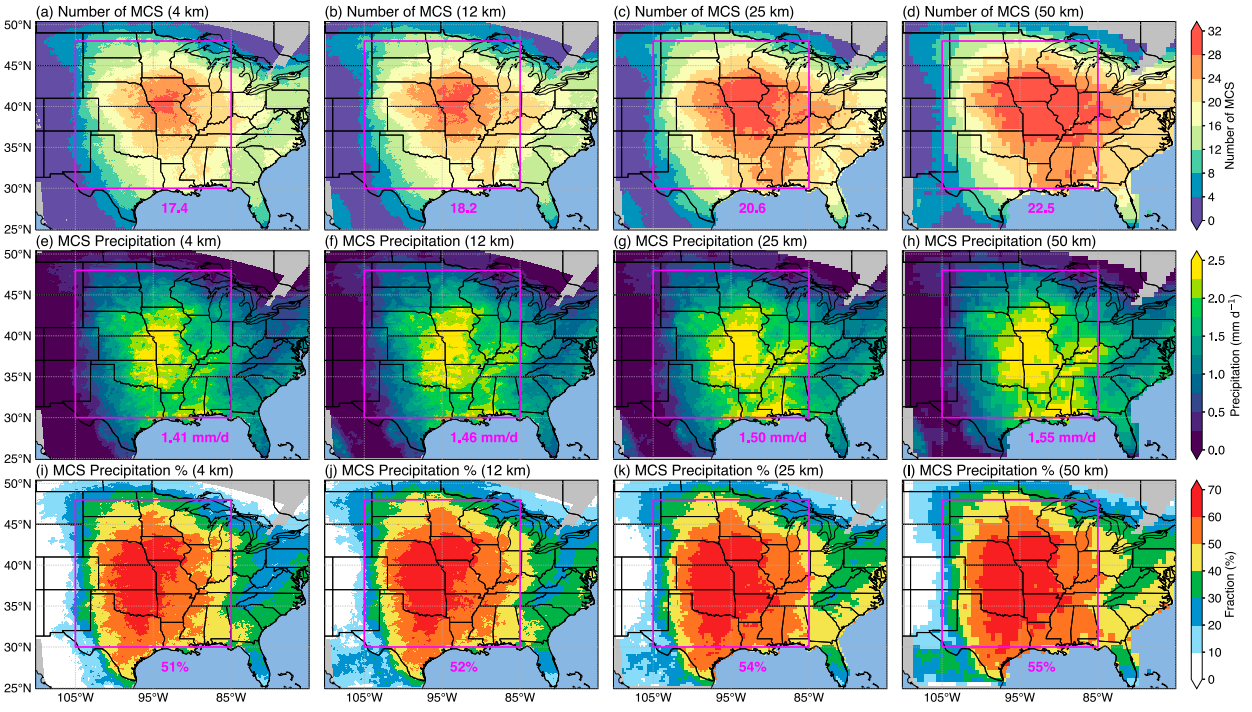

FIG. 5. Comparison of the spatial distribution of observed warm season (March–August) MCSs tracked in four different resolutions for 2004–16. Each column represents one resolution (from left to right) 4, 12, 25, and 50 km. (a)–(d) Number of MCSs, (e)–(h) MCS precipitation amount, and (i)–(l) MCS precipitation percentage to total precipitation. The number of MCSs in (a)–(d) is calculated by adding each swath of an MCS PF during its lifetime (counted as one sample over each grid point within the swath) over all 13 warm seasons divided by the total number of seasons (13). The magenta numbers are mean values in the central U.S. region shown by the magenta box.

patterns are stable and the nodes are self-organized. The resulting SOM types that are physically interpreted to provide dynamically and thermodynamically favorable large-scale environments for MCS development are called FLSMPs. They represent the mean 3D large-scale patterns from many individual cases with similar patterns. During spring, two FLSMPs are associated with frontal systems, which feature an anomalous cyclone to the west and an anomalous anticyclone to the east of the Great Plains in the upper levels and strong midlevel upward motion over the Great Plains. The other two FLSMPs are associated with enhanced LLJ over the Great Plains but differ in their meridional extent.

To evaluate the model's ability in reproducing the four types of FLSMP observed during boreal spring, we compare the large-scale environment in the simulations and observations associated with each FLSMP and its frequency of occurrence. First, we regrid the simulated zonal and meridional winds at three levels (925, 500, and 200 hPa) and the specific humidity at two levels (925 and 500 hPa) to the same 1° × 1° resolution region (20°–55°N, 70°–110°W) that defines the FLSMPs derived from NARR (Song et al. 2019). A cosine-latitude weighting is adopted when the spatial dimensions of the variables are collapsed into a single dimension for calculating the Euclidean distance. Then, at each model output instance (i.e., every 6 h), we calculate the Euclidean distance between the model 3D environmental fields (same as those used in NARR) and each of the four observed FLSMP types. If the shortest Euclidean distance to a particular type of FLSMP among the four FLSMP types is smaller than a threshold, the large-scale environment is considered to be similar to this type of FLSMP. The threshold is used to ensure a reasonable level of similarity between the large-scale environment and the FLSMP and taken as the average Euclidean distance between each sample among one SOM type and the FLSMP for this SOM type, roughly equal to 6500. To give equal weighting to each of the environmental fields, each variable is normalized as follows:

$$V_n = (V_i - \overline{V_i})/\sigma_{V_i}, \tag{1}$$

where V_i and V_n are the variable at grid point i before and after normalization, respectively; and $\overline{V_i}$ and σ_{V_i} are the mean and standard deviation of the variable at the same grid point within a season (e.g., spring), respectively.

In Eq. (1), $\overline{V_i}$ and σ_{V_i} from observations are used to normalize both the model and observed (NARR) variable. To evaluate the effect of model biases, we also use $\overline{V_i}$ and σ_{V_i} from simulations to normalize the simulated variable to quantify the impact of model biases on the Euclidean distance. Normalization can also be performed using a combination of model and observed $\overline{V_i}$ and σ_{V_i} for different variables to isolate the effect of model biases in the specific variables (e.g., winds versus moisture) or the specific statistical moments (e.g., mean versus standard deviation) on the Euclidean distance. Different normalization procedures are used in this study and the results are discussed in section 5b.

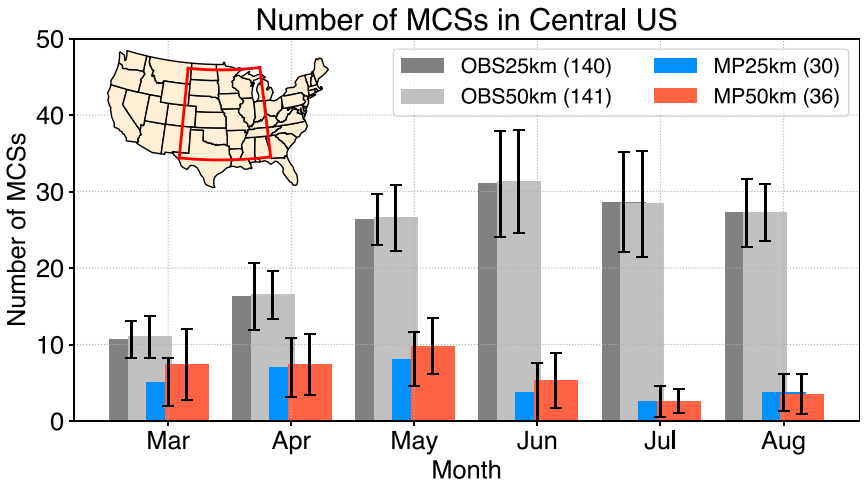

FIG. 6. As in Fig. 4, but for observations and MPAS-CAM simulations (MP25km and MP50km) during 2004–13. Different from Fig. 4, the numbers are for MCS initiated within the red box region in the inset map.

4. Evaluation of model-simulated MCSs

In this section, we evaluate the VR-MPAS simulated MCS directly against observations at equivalent resolutions. We track MCSs in both observations and model simulations using the same algorithm detailed in section 3. The consistency in resolution and tracking algorithm facilitates the comparison and interpretation between simulations and observations. The simulated OLR and precipitation are used as inputs for FLEXTRKR to track convective clusters/objects. OLR is converted to T_b following the empirical relationship provided by Yang and Slingo (2001). Since we are tracking cloud-top and precipitation signatures as proxies of organized convection in the simulations, we consider the tracked precipitation objects "MCS-like" features. These proxies mean that the model-simulated precipitation from deep convective clouds evolve similarly to those produced by MCSs in reality, but the proxies do not guarantee that the model is able to simulate all aspects of the observed MCSs such as their precipitation intensity and vertical heating profiles that are most impactful to surface hydrology and upscale effects on atmospheric circulations, respectively (Feng et al. 2018). Details of the simulated "MCS-like" features, which will simply be referred to as simulated MCSs, will be evaluated against observations. As multiple criteria are used in this study to define MCS, including size, duration, and intensity of both cloud and PF, we also examine the model evaluation sensitivity to MCS criteria in section 4c.

a. MCS frequency and precipitation

The seasonal cycle of the number of simulated MCSs in the central United States is shown in Fig. 6. It is clear that the average number of MCSs are significantly underestimated in VR-MPAS, particularly from May to August. The model biases are not evidently reduced with higher resolution, as the observed PF characteristics at 25 km is larger and stronger than those at 50 km (see Table 3), resulting in fewer simulated PF in MP25km that satisfy the MCS criteria.

Figure 7 shows the spatial distribution of the model biases in total precipitation, MCS precipitation and MCS number in spring (March–May). Both simulations generally underestimate the total precipitation over the central United States, but overestimate it over the Rocky Mountains and eastern United States. The dry bias for MCS precipitation is much larger, particularly over the Southern Great Plains (SGP) region, where MCS precipitation is underestimated by ~50%. The bias in the number of MCSs over the central United States is similar between MP25km and MP50km (~52%). Similar bias in MCS precipitation in that region between MP25km and MP50km is mainly caused by underestimations of the frequency (~40%) rather than intensity (~16%) of MCS precipitation (Fig. S5). Compared to spring, the biases are much larger in the summer (Fig. S6). Not only do the two simulations miss the majority of MCSs and their associated precipitation (~86%) in the central United States, the total precipitation is also underestimated by 17% in that region.

Song et al. (2019) compared different types of observed FLSMP and found that synoptic patterns associated with the passage of strong baroclinic waves during spring are much more skillful in estimating the occurrence of MCSs than those during the summer. Their findings are consistent with those reported by Jankov and Gallus (2004) and Squitieri and Gallus (2016) that forecasting of MCS rainfall is more skillful under strong large-scale forcing conditions typically associated with cold fronts and LLJ, while model performance is poor in midsummer under warm or stationary front. Song et al. (2019) hypothesized a more important role of smaller-scale disturbances such as surface fluxes and shortwave troughs (e.g., midtropospheric perturbations, Wang et al. 2011a,b) that may limit the predictability of summer MCSs. Considering the finest model resolution used in this study (25 km) is likely not sufficient to resolve the smaller-scale perturbations, and parameterized convection poorly represents convective processes and precipitation over land, we

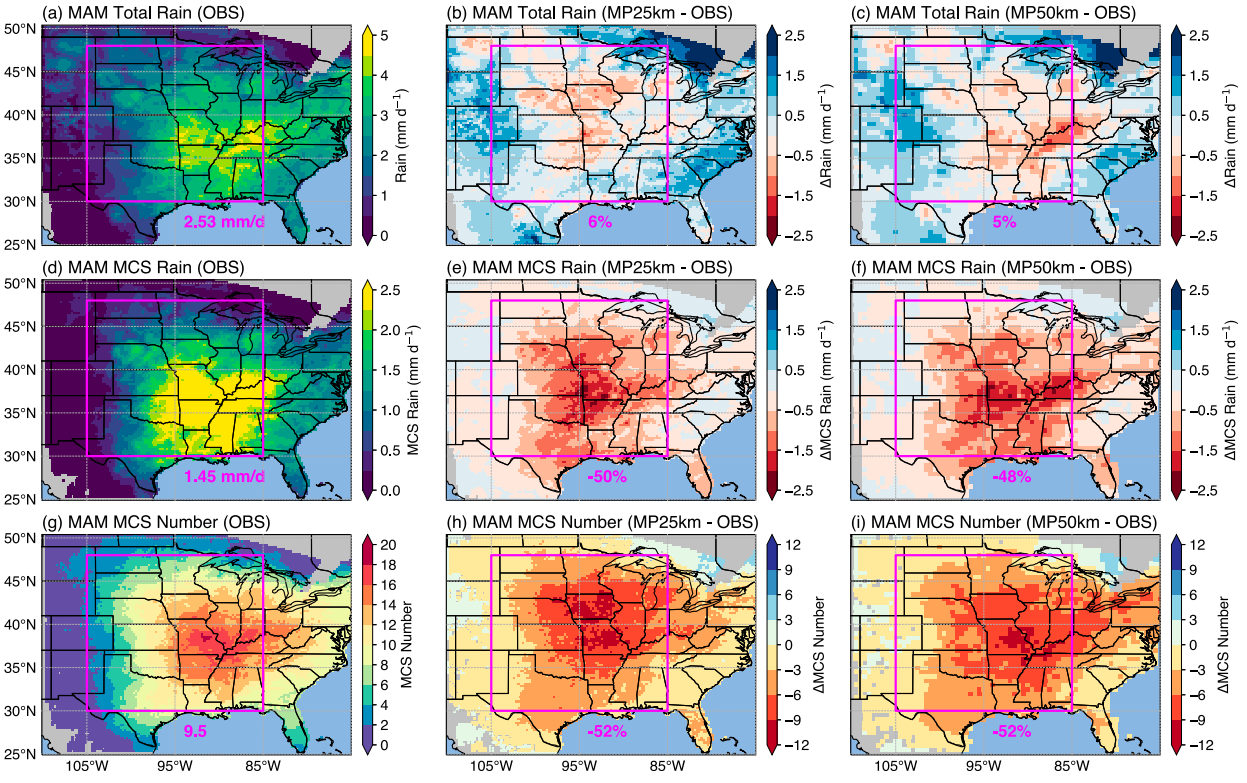

FIG. 7. Comparison of the spatial distribution of spring (MAM) MCSs between observations (OBS) and MPAS-CAM simulations. (left) OBS25km, (center) MP25km − OBS25km, and (right) MP50km − OBS50km. (a)–(c) Total precipitation, (d)–(f) MCS precipitation, and (g)–(i) number of MCSs. The magenta numbers are mean values from observations in (a), (d), and (g) and mean relative biases from the simulations in the central U.S. region shown by the magenta box.

focus our subsequent analysis on the spring season with stronger large-scale forcings for MCSs.

b. *Diurnal cycle and properties of simulated MCSs*

The composite diurnal cycle of total and MCS precipitation as functions of longitude over the central United States is shown in Fig. 8. In observations, both total and MCS precipitation shows eastward propagation over the Great Plains, with total precipitation peaking between 1800 and 2400 LT. MCS precipitation peaks several hours later around local midnight, accounting for 60%–70% of the total nocturnal precipitation. The diurnal amplitude of the observed precipitation maximizes in the Great Plains and significantly decreases outside, particularly for MCS precipitation. In contrast, both models simulated the total precipitation peaking in the early afternoon with similar amplitudes, followed by a much-dampened nocturnal peak of MCS precipitation that is shifted to the east of the Great Plains. The simulated MCS precipitation only accounts for 20%–40% total nocturnal precipitation.

Figure 9 shows the diurnal cycle of MCS initiation and precipitation amount averaged within the central United States. Precipitation amount is separated into MCS and non-MCS precipitation, where non-MCS precipitation includes precipitation from both isolated convection and nonconvective stratiform precipitation. Observed MCSs most frequently initiate between midafternoon to early evening (1400–2000 LT), and the results are comparable for the two resolutions. Model-simulated MCSs tend to initiate slightly more often during nocturnal and morning hours, although the diurnal amplitude (the first harmonic) is 2–8 times weaker than observed. After initiation, observed MCSs go through upscale growth for several hours and reach maximum precipitation around midnight, and the precipitation gradually decreases in the morning to reach a minimum around local noon. Observed MCS precipitation diurnal amplitude is 2 times that of non-MCS precipitation, which peaks around 1900–2000 LT. MCS and non-MCS precipitation each account for ~60% and ~40% of the observed total precipitation. In contrast, simulated non-MCS precipitation diurnal amplitude is more than twice stronger than observed, accounting for 75% of simulated total precipitation, and peaks 4 h earlier (~1500 LT), while simulated MCS precipitation has very little diurnal variation and only account for 25% total precipitation.

Comparisons of the simulated MCS characteristics with observations are shown in Fig. 10. The simulated MCS lifetimes are longer than observations at comparable resolutions (Fig. 10a). A larger fraction of simulated MCSs have lifetime between 25 and 50 h compared to observations, and the averaged simulated lifetime is longer by 3–4 h (12%–18% overestimated). Composite evolution of MCS PF area, mean

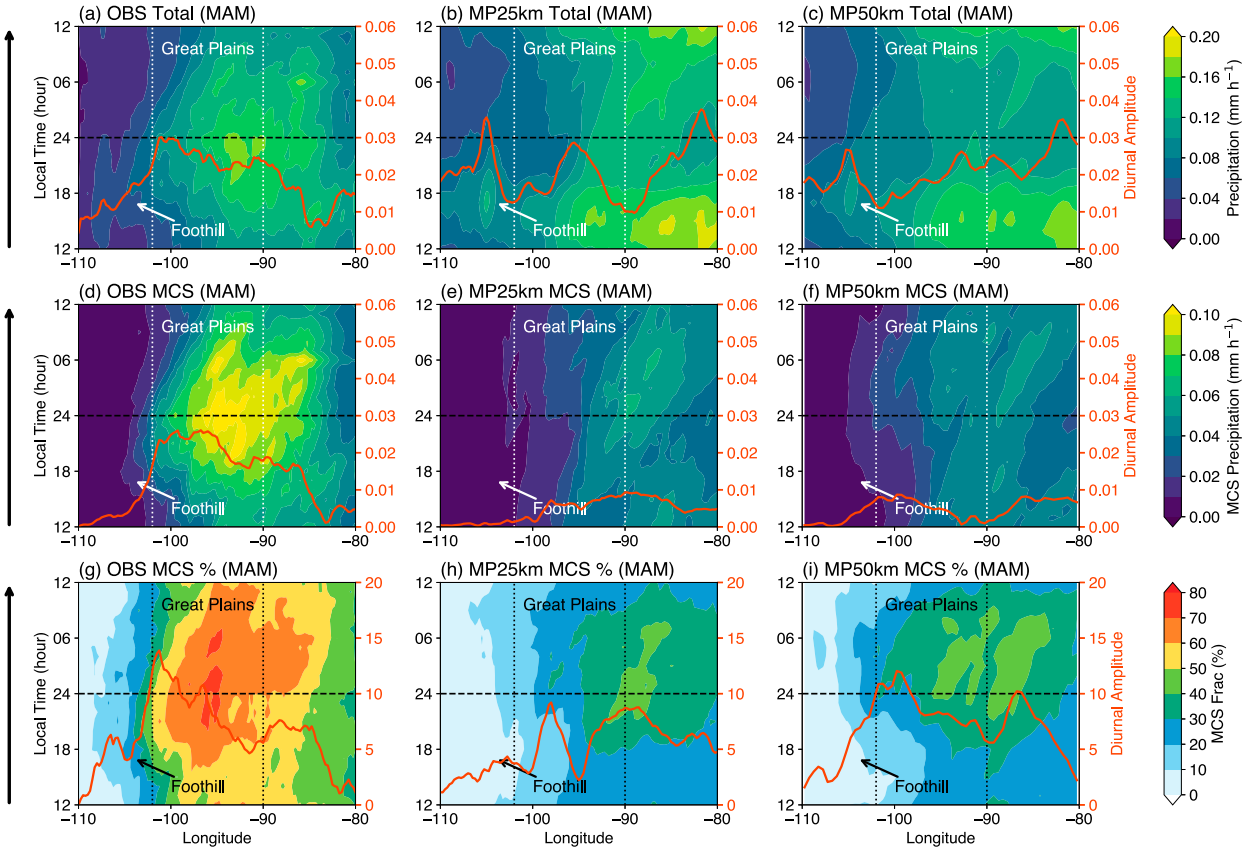

FIG. 8. Diurnal cycle of precipitation as a function of longitude over the central United States (magenta box in Fig. 5) during spring. (left) Observations (OBS), (center) MP25km, and (right) MP50km. (a)–(c) Total precipitation, (d)–(f) MCS precipitation, and (g)–(i) percentage of MCS precipitation to total precipitation. Orange lines show the amplitude of the first harmonic of the mean diurnal cycle (for each variable) at each longitude. Approximate locations of the Rocky Mountain foothills and the Great Plains are marked.

precipitation rate and volumetric precipitation are shown in Figs. 10b–d as a function of normalized MCS time. The simulated PF areas are consistently overestimated throughout the life cycle of MCS by ~40%–80%, while simulated PF mean precipitation rates are underestimated by ~15%, consistent with weaker mean MCS precipitation intensity shown in Fig. S5h and S5i. The resulting averaged volumetric precipitation per MCS is overestimated by 50%–70% due to the larger error in PF area than mean intensity. The biases in PF area and volumetric rainfall are larger in MP25km than MP50km, suggesting the higher resolution alone (without tuning) does not produce more realistic MCS PFs. Since most of the observed MCSs in spring occur under FLSMP (Song et al. 2019), these results suggest that the model tends to produce continuous precipitation features that are too large and too weak compared to observations under strongly forced baroclinic environments. We will further examine the model-simulated precipitation under different types of FLSMP in more detail in section 5c.

c. *Sensitivity to MCS criteria*

As multiple criteria are used in this study to define MCS, including size, duration, and intensity of both cloud and PF, it is important to evaluate the sensitivity of the MCS model evaluation results to the different criteria used in the MCS definition described in section 3b. Such analysis is also informative of the aspects of MCS that the model tends to fail to reproduce, and hence potential solutions to address the model biases.

By relaxing each of the MCS criteria to approximately 50% of the thresholds used in the observations, we find that the model-simulated PF characteristics (i.e., PF area, PF mean rain rate, PF rain rate skewness) have the largest impact on the biases, followed by the PF duration. Among the three PF characteristics used to define MCS, the model PF most frequently fails the PF mean rain rate criteria (>80% failure), followed by the rain rate skewness (30%–50% failure), while only ~20% of the PF fails the PF area criteria (Table S1).

The sensitivity of the number of identified MCS in the simulations to various thresholds can be seen by comparing the increase in the MCS number relative to that using the same criteria as observations (Fig. S7). Reducing the CCS area and CCS duration has very little impact on the number of MCS identified. On the other hand, reducing the PF duration to 3 h increases the number of MCS by 60%–70%, while further reducing the PF characteristics criterion to the 7th percentile thresholds (vs 15th percentile thresholds in observations)

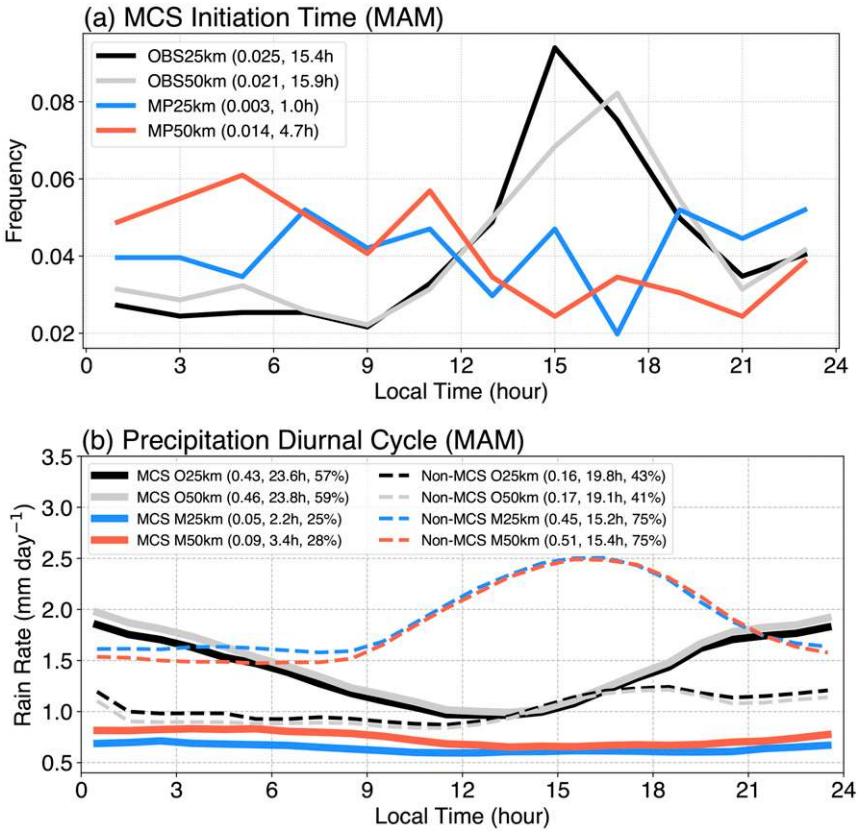

FIG. 9. Diurnal cycle of (a) MCS initiation time and (b) MCS and non-MCS precipitation averaged in the central United States during spring. Numbers in the legends in (a) are the amplitude and phase (peak timing) of the first harmonic of the diurnal cycle, and in (b) the amplitude, phase of the diurnal cycle, and the fraction of the total rainfall, respectively.

increases the MCS number by 1.6–1.8 times, resulting in a comparable number of MCSs with observations from March–May (biases from June–August are still larger than 50%). Physically, these results mean that the model can sometimes simulate large and intense PFs that resemble MCS, but their duration are shorter, and the precipitation intensity and variability are weaker than observations. By relaxing the PF duration and PF characteristics criteria, more simulated PFs can be identified as MCS.

Specifically, relaxing all MCS thresholds by approximately 50% for both the MPAS-CAM simulations and observations results in ~50% bias reduction in MCS precipitation frequency and amount (Figs. S8b,c,e,f), as the number of weaker PFs tracked as MCS in the simulations is larger than those in observations. Further, by relaxing the MCS thresholds for MPAS-CAM simulations by ~50% but keeping the same MCS thresholds in section 3a for observations to account for model deficiencies in simulating weaker precipitation intensity, the simulated and observed MCS numbers become more comparable and the precipitation amount biases are further reduced by ~15% (Figs. S9b,c,e,f). While relaxing the MCS criteria reduces the bias in the MCS number and MCS precipitation amount, other aspects of the biases remain, including the underestimated MCS precipitation intensity (Figs. S8 and S9h,i) and erroneous MCS and non-MCS precipitation diurnal cycle (Figs. S10 and S11). The sensitivity tests described here suggest that while changing the MCS criteria affects the simulated MCS evaluation results quantitatively, the major biases in model-simulated MCS characteristics presented in Sections 4a and 4b hold regardless of the thresholds used. Therefore, it is important to examine multiple aspects of the simulated MCS characteristics to better diagnose model deficiencies.

5. Simulated MCS favorable large-scale environments

a. Moisture and wind associated with the Great Plains LLJ

In this section, we examine the LLJ to illustrate the physical linkages between the model biases in large-scale environments and the total and MCS precipitation discussed in section 4. The link will be elaborated and generalized by using FLSMP in the next subsection. Previous studies have shown that the LLJ plays a critical role in controlling warm season heavy precipitation over the central United States (Harding and Snyder 2015; Higgins et al. 1997; Wang and Chen 2009). For example, more accurate simulation of LLJ under strongly forced synoptic environments result in better MCS precipitation forecast

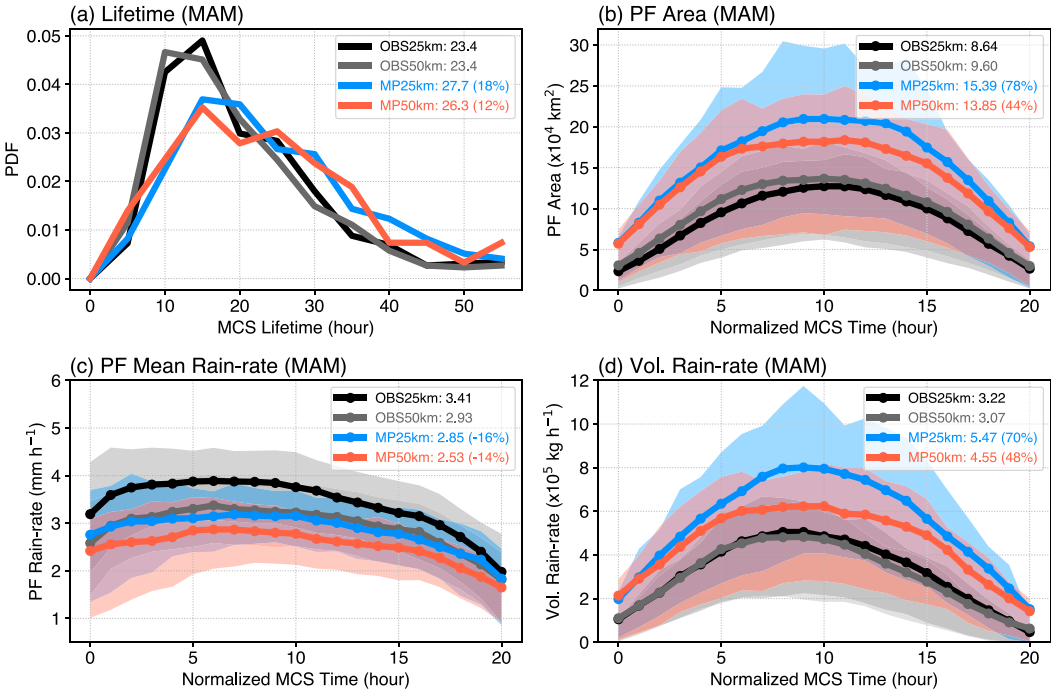

FIG. 10. (a) Frequency distribution of MCS lifetime, composite evolution of MCSs over the central United States during spring (MAM) for (b) PF area, (c) PF mean rain rate, and (d) volumetric rain rate. In (b)–(d), the x axis shows normalized MCS time, where hour 0 denotes convective initiation and hour 20 denotes dissipation. Lines show the average values, and shadings denote the 25th and 75th percentile values at each time. Numbers in the legends are the averaged values across the MCS lifetime, and the numbers in parentheses are average bias percentages.

skill (Squitieri and Gallus 2016). We use the NARR dataset as a reference. NARR has a 32-km horizontal resolution, which is comparable to the regional refinement grid spacing of MP25km. Berg et al. (2015) compared six commonly used reanalysis products against long-term radiosonde observations over the SGP region and found that NARR is among the best reanalysis products in representing the frequency and variability of the LLJ, although the moisture transport by the stronger LLJs are still underestimated by a factor of 2. The model outputs are regridded to the NARR grid to facilitate the comparison in this section.

Compared to NARR, we find that both the 50- and 25-km simulations underestimate the climatological meridional wind speed in the central United States, resulting in an overall low-level (below 700 hPa) dry bias over the Great Plains extending to the Appalachian Mountains (not shown). To quantitatively evaluate the VR-MPAS simulated LLJ and associated moisture transport, we follow the LLJ identification methodology by Bonner (1968) and Whiteman et al. (1997). A total of four LLJ categories are defined, based on the maximum wind speed below 700 hPa (i.e., the LLJ nose) and the wind speed reduction above the level of the maximum wind speed but below 700 hPa (Table 4). The methodology was applied to the instantaneous 3D wind fields from NARR and VR-MPAS outputs.

Figure 11 shows the frequency of the four categories of LLJ from NARR and the biases (model − NARR) from the two sets of model simulations. Stronger LLJ (category 2–3) occurrence is confined more within the SGP compared to those from weaker LLJ (category 0–1) in NARR (Figs. 11a–d). Both simulations underestimate the frequency of the LLJ across all four categories over the Great Plains, but stronger LLJ frequency biases are significantly larger, particularly over the SGP.

Comparisons of the mean vertically integrated vapor transport (IVT) associated with the four categories of LLJ (from surface to 700 hPa) are shown in Fig. 12. Over the SGP region, IVT from LLJ category 1–3 accounts for the majority of the total IVT in NARR, each contributing similarly to the IVT amount except that category 1 has a broader spatial extent. These results are qualitatively consistent with those reported by Berg et al. 2015. Compared to NARR, the VR-MPAS simulations underestimated the LLJ IVT by 25%–50% over the SGP, particularly for the stronger LLJ categories. The above evaluation and the known influence of the LLJ on warm-season

TABLE 4. Criteria to define LLJ categories (adapted from Berg et al. 2015.

LLJ category	U_{max} (m s^{-1})	ΔU_z (m s^{-1})
0	≥10	≥5
1	≥12	≥6
2	≥16	≥8
3	≥20	≥10

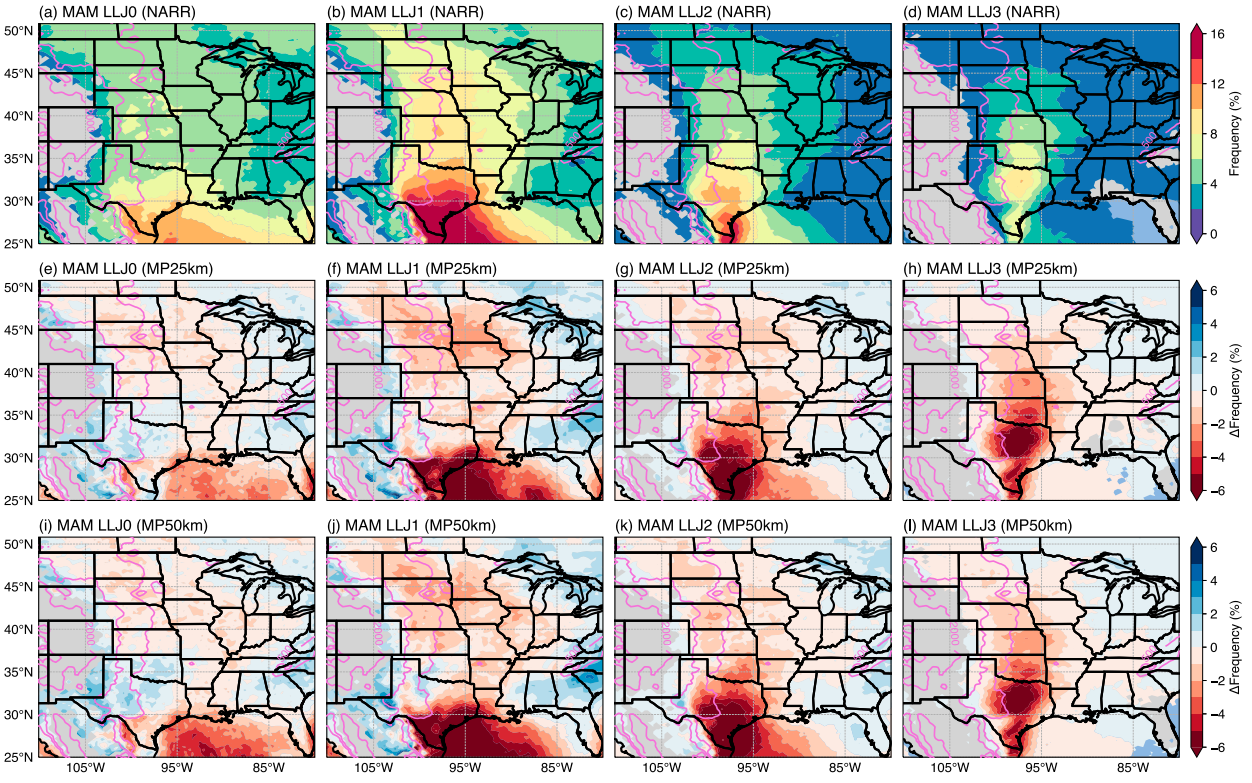

FIG. 11. Frequency of the four categories of LLJ during MAM identified from (a)–(d) NARR, (e)–(h) MP25km biases (MP25km − NARR), and (i)–(l) MP50km biases (MP50km − NARR). Terrain heights are in magenta contours (500, 1000, and 2000 m).

precipitation strongly suggest that the weaker LLJ and associated moisture transport contribute to the dry bias found in the total and MCS precipitation reported in section 4. This logic is supported by the FLSMP diagnosis, as shown in the next section, which further suggests that the LLJ is not the only large-scale environment that models need to capture to realistically simulate MCS.

b. *Simulated MCS favorable large-scale meteorological patterns*

Analysis of the LLJ suggests that biases in the large-scale environment may limit the ability of the model to simulate MCSs. Besides LLJ, other aspects of the large-scale environment have also been identified to favor MCS development. Here we examine how well VR-MPAS simulates different types of FLSMP for MCS initiation. The methodology described in section 3 is used to identify and compare the simulated and observed large-scale environments associated with each of the four types of observed FLSMP found by Song et al. (2019).

Figure 13 compares the composite of the NARR and MP25km large-scale environment associated with each FLSMP during MAM. Both observations and simulations are normalized using the NARR mean and standard deviation. The composite patterns for MP50km are very similar to MP25km and thus not shown. The results show that the model is able to produce the large-scale environment associated with different types of FLSMP that are comparable to that of NARR. Specifically, type 1 and 3 are associated with frontal systems that provide a large-scale lifting mechanism, and type 2 and 4 are related to enhanced Great Plains LLJ that provides anomalous low-level moisture (Song et al. 2019). This agreement is partly related to the use of the same threshold of 6500 to ensure comparable similarity between the observed and simulated large-scale environment and the observed FLSMP (see details in section 3c). Therefore, a more important question is how often VR-MPAS simulates large-scale environment comparable to the four types of FLSMP and how this frequency compares with that of observations.

The frequency of occurrence of observed MCSs initiation and observed and simulated large-scale environment associated with FLSMP during MAM are shown in Table 5. The observed number of MCS initiation is significantly lower than the observed number of FLSMP occurrence. This is because each MCS (on average lasting 20+ hours) is only counted once at its initiation, while the environment associated with FLSMP could last longer and is counted multiple times at 6-hourly interval during the entire MCS lifetime. Hence the significantly lower number of FLSMP than the number of MCS initiation does not imply a weak relationship between the FLSMP and MCS initiation. Song et al. (2019) already demonstrated significant correlation between the FLSMP and MCS initiation in spring. Therefore, comparing the frequency of large-scale environment associated with FLSMP between NARR and the VR-MPAS simulations provides an indication of the

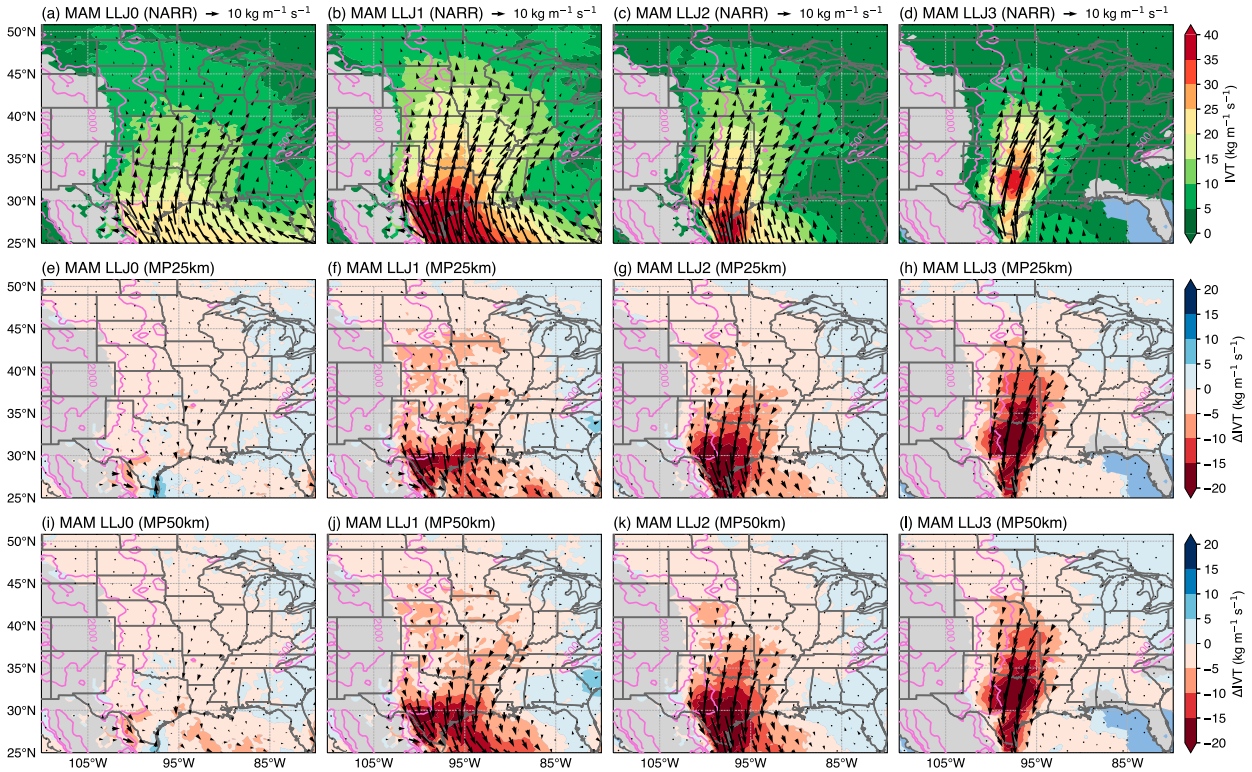

FIG. 12. Mean vertically integrated vapor transport (IVT) associated with four types of LLJ (from surface to 700 hPa) during MAM identified from (a)–(d) NARR, (e)–(h) MP25km biases (MP25km − NARR), and (i)–(l) MP50km biases (MP50km − NARR). Shadings are IVT magnitudes, and vectors show both the direction and magnitude of IVT.

contribution of biases in the large-scale environment to the biases in MCS initiation in the model.

Overall, the VR-MPAS simulated total occurrences associated with FLSMP are underestimated by 50%–60% compared to those from NARR. Among the four types of FLSMP, VR-MPAS shows larger biases (61%–79%) in simulating the occurrence of frontal systems (type 1 and 3) compared to the biases (22%–59%) of LLJ types (type 2 and 4). We examined the diurnal cycle of the large-scale environment and found a much weaker diurnal amplitude for both NARR and model simulations (Fig. S9) than that of MCS initiation (Fig. 9). This suggests that the favorable large-scale environments are less efficient at initiating MCSs during nocturnal and morning hours in the absence of surface heating from solar insolation that helps trigger convection. The VR-MPAS simulated FLSMP occurrences are underestimated across the diurnal cycle, especially for the FLSMPs featuring frontal systems.

To better understand the relative contributions of model biases in dynamics (i.e., biases in wind) and thermodynamics (i.e., biases in moisture) to modeling the environment favorable for MCSs, we conducted several sensitivity tests by varying how the Euclidean distance of the wind and moisture field between model simulation and NARR is calculated. In normalizing the simulations, we replaced the mean $\overline{V_i}$ and standard deviation σ_{V_i} [Eq. (1) in in section 3c] from NARR with those from VR-MPAS so the calculation of the Euclidean distance is based on model anomalies and the effect of model biases is removed. The frequency of occurrence of the simulated environment associated with each FLSMP is recalculated by replacing the mean and standard deviation of NARR with that of the simulation one at a time. This analysis only provides a first-order estimate of the relative contribution of dynamic (winds) versus thermodynamic (moisture) biases to the total bias, since low-level moisture and meridional winds are not independent over the Great Plains.

The results from the sensitivity tests are shown in Tables 6 and 7. Generally speaking, model biases in the low-level moisture dominate the occurrence biases of the environment associated with type 1 and 2 FLSMP, while the meridional wind biases have larger impact on the environment of type 1 and 3 FLSMP, as indicated by 80% or larger changes in the occurrences of these FLSMPs. The LLJ penetrates further inland and transports larger amount of moisture reaching the Northern Great Plains in type 1 and 2 than type 3 and 4 (Figs. 13a,b,e,f). Therefore, a dry bias would have larger effect on type 1 and 2 by limiting the moisture flux reaching the higher latitudes. For type 1 and 3, frontal convergence associated with the low-level wind likely plays a more important role (Figs. 13a,c,e,g). For type 4 (LLJ), biases in meridional wind and moisture equally contribute to biases in simulating the FLSMP (Figs. 13d,h). In addition, contribution from the biases in the mean state versus variability is also examined for each variable by using the simulated

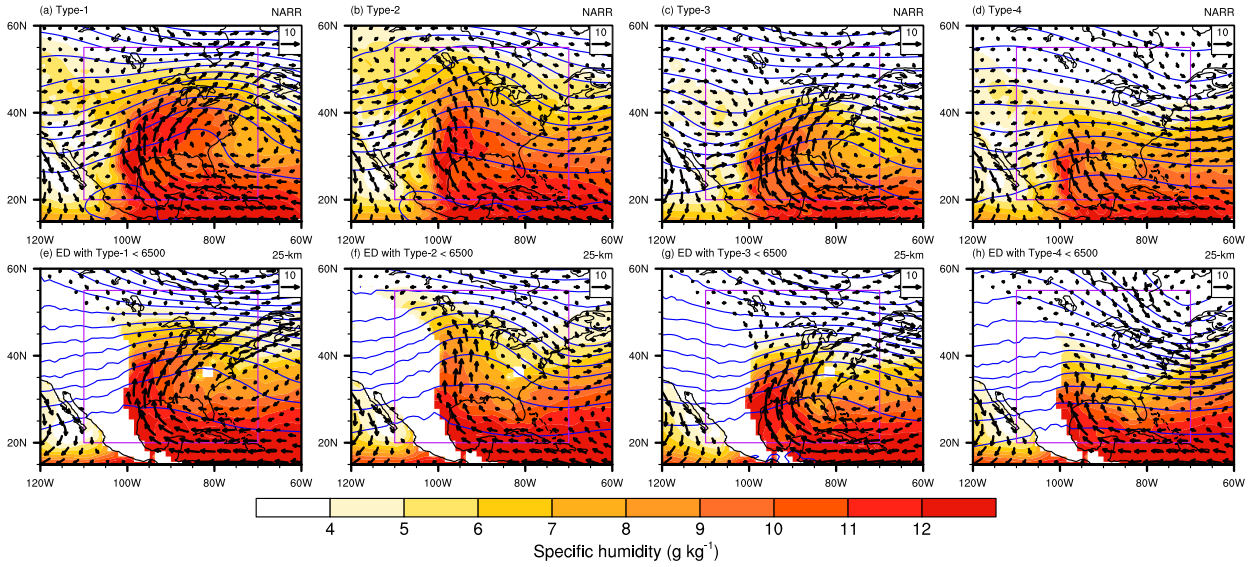

FIG. 13. Comparisons of NARR and MP25km large-scale environment associated with MCS favorable large-scale meteorological patterns (FLSMPs) (Song et al. 2019). (a)–(d) Composites of the NARR large-scale environment associated with four types of FLSMP; (e)–(h) as in (a)–(d), but for composites of the MP25km large-scale environment. Shadings are 925-hPa specific humidity (g kg^{-1}), vectors are 925-hPa winds (m s^{-1}), and contours are 500-hPa geopotential height (gpm). The magenta box shows the region used in calculating the Euclidean distance between the NARR and MP25km large-scale environment and the FLSMPs to determine their similarity. See text for more details about the methodology used in identifying the large-scale environment associated with each FLSMP.

$\overline{V_i}$ and σ_{V_i} separately in the normalization procedure. The results show that biases in the variability of the model-simulated large-scale environment dominate the total bias. Overall, analysis of the large-scale environment suggests that different model biases contribute to the biases of certain FLSMP type differently. Furthermore, comparing Table 6 with Table 7, removing the effect of model biases in the mean and standard deviation induces a larger difference in the occurrence of favorable environment at 25- than 50-km resolution. This difference indicates generally larger model biases in these variables for MP25km than MP50km.

c. *Evolution of precipitation under FLSMP*

To better understand the response of precipitation under different FLSMPs, we composited the temporal evolution of precipitation with respect to the occurrence of large-scale environment associated with each type of FLSMP, based on normalization using the NARR mean and standard deviation. Hence this represents a fair comparison of the precipitation under environments in both observation and model simulations comparable to the observed FLSMPs. From Fig. 14, the two types of FLSMP associated with frontal systems (type 1 and 3) typically produce more intense precipitation compared to the other two types associated with the Great Plains LLJ (type 2 and 4) in observations. This is particularly true for type 3 with a stronger frontal boundary as indicated by the drier northwesterly winds at 925 hPa (Fig. 13) and stronger 500-hPa vertical motion (Fig. 5 in Song et al. 2019). Observed precipitation in the central United States tends to increase sharply about 3 h after the initiation time determined by the FLSMP and peak 6–12 h later (Figs. 14a–d). The initial sharp increase of precipitation denotes the upscale growth of convection into MCS genesis, while the peak precipitation 6–12 h later is consistent with the mature stage of MCSs that produces the majority of rainfall.

In contrast, except under type 3 FLSMP, the simulated precipitation shows very little temporal change following the occurrence of large-scale environment associated with

TABLE 5. The average number of occurrences per season for observed MCS initiation (first row), observed FLSMP (second row), and MPAS simulated FLSMP for 25 km (third row) and 50 km (fourth row) during MAM. The ± values are the standard deviations denoting year-to-year variability. Parentheses in the last two rows show the percentage of model mean biases in FLSMP occurrence. The FLSMP occurrences are calculated using 6-hourly model outputs.

	Type 1 (frontal)	Type 2 (LLJ)	Type 3 (frontal)	Type 4 (LLJ)	All
OBS No. of MCS initiation	11.7 ± 7.6	12.3 ± 5.5	14.1 ± 5.9	11.2 ± 5.3	47.9 ± 5.5
NARR FLSMP	55.2 ± 22.6	31.9 ± 17.6	49.0 ± 17.9	57.5 ± 18.5	193.6 ± 21.7
MP25km FLSMP	11.7 ± 6.4 (−79%)	13.2 ± 11.8 (−59%)	17.0 ± 8.1 (−65%)	35.6 ± 17.9 (−38%)	77.5 ± 18.4 (−60%)
MP50km FLSMP	21.5 ± 18.4 (−61%)	24.8 ± 12.2 (−22%)	13.8 ± 9.1 (−72%)	33.0 ± 13.0 (−43%)	93.1 ± 27.3 (−52%)

TABLE 6. The average number of occurrences of large-scale environment associated with each type of FLSM per season for MP25km when normalized with different model variables. Parentheses show the percentage of mean occurrence changes compared to occurrences using normalization with observed values; q, u, and v denote specific humidity, zonal wind and meridional wind, respectively.

Normalization tests	Type 1 (frontal)	Type 2 (LLJ)	Type 3 (frontal)	Type 4 (LLJ)
All variables normalized with OBS	11.7	13.2	17.0	35.6
q_i: replace with model $\overline{q_i}$ and σ_{q_i}	21.9 (87%)	26.8 (103%)	23.5 (38%)	47.9 (35%)
q_i: replace with model $\overline{q_i}$	15.9 (36%)	16.2 (23%)	18.2 (7%)	34.1 (−4%)
q_i: replace with model σ_{q_i}	17.5 (50%)	20.5 (55%)	22.2 (31%)	49.1 (38%)
u_i: replace with model $\overline{u_i}$ and σ_{u_i}	21.2 (81%)	16.6 (26%)	25 (47%)	37.1 (4%)
u_i: replace with model $\overline{u_i}$	16.0 (37%)	11.1 (−16%)	18.4 (8%)	28.6 (−20%)
u_i: replace with model σ_{u_i}	15.8 (35%)	18.3 (39%)	22.7 (34%)	48.6 (37%)
v_i: replace with model $\overline{v_i}$ and σ_{v_i}	21.4 (83%)	19.7 (49%)	30.8 (81%)	50.2 (41%)
v_i: replace with model $\overline{v_i}$	15.6 (33%)	12.9 (−2%)	22.1 (30%)	33.1 (−7%)
v_i: replace with model σ_{v_i}	18.3 (56%)	20.8 (58%)	24.7 (45%)	53.8 (51%)
All σ_i replaced with model σ_{V_i}	29.1 (149%)	37.2 (182%)	37.0 (118%)	86.8 (144%)
All $\overline{V_i}$ replaced with model $\overline{V_i}$	27.0 (131%)	21.1 (60%)	25.1 (48%)	29.0 (−19%)

FLSMP, as rainfall remains weak 6–12 h later (Figs. 14e–h). This weak precipitation bias is consistent with the persistently weaker mean PF rain rate simulated during upscale growth and mature stage of MCSs (Fig. 10). Further, the model precipitation seems to have a weak maximum approximately 18–24 h after the occurrence of FLSMP that is absent in the observation. To examine this peculiar peak, we compare the composite diurnal cycle of precipitation under different FLSMPs in Fig. 15.

In observations, precipitation is consistently maximized between early evening (1800 LT) and early morning hours (0600 LT) among all four FLSMPs, with type 3 frontal systems producing the most rainfall amount further east of the Great Plains (90°–95°W), consistent with the frontal convergence zone being further eastward (Fig. 13c) compared to other types. In contrast, the majority of the simulated precipitation peaks during the daytime hours (0900–1800 LT), regardless of the FLSMP. Nonnegligible nocturnal precipitation is simulated only in frontal system (type 3) and when the LLJ is constrained over the SGP (type 4). Dividing the total precipitation into convective and large-scale precipitation in the model (not shown), convective precipitation contributes more to the total precipitation for the first two types, but large-scale precipitation dominates the total precipitation in type 3 (frontal system) and type 4 (LLJ). These results agree with previous studies (e.g., Zhang 2002, 2009) that precipitation simulated by the Zhang–McFarlane convection parameterization is strongly tied to the diurnal timing of the surface-based convective available potential energy (CAPE), rather than other forcings such as the elevated instability associated with the LLJ, which could be decoupled from the surface during nocturnal hours (Geerts et al. 2017).

6. Summary and discussion

In this study, we develop a process-oriented approach to systematically evaluate the performance of climate models in simulating warm-season MCS-like precipitation features and their spring favorable large-scale meteorological patterns (FLSMPs) over the United States. The new method is applied to two sets of 10-yr simulations with refined horizontal grid spacings at 50 and 25 km over North America using VR-MPAS coupled with the CAM5.4 physics.

To directly evaluate model-simulated MCS-like precipitation with observations, a new observation-driven methodology to track MCSs across three different resolutions targeting next-generation GCMs (50, 25, 12 km) has been developed. The method jointly uses infrared brightness temperature (or outgoing longwave radiation) and surface precipitation, both commonly available in observations and model output, to track

TABLE 7. As in Table 6, but for MP50km.

Normalization tests	Type 1 (frontal)	Type 2 (LLJ)	Type 3 (frontal)	Type 4 (LLJ)
All variables normalized with OBS	21.5	24.8	13.8	33
q_i: replace with model $\overline{q_i}$ and σ_{q_i}	32.7 (52%)	36.4 (47%)	16.3 (18%)	42.7 (29%)
q_i: replace with model $\overline{q_i}$	27.7 (29%)	27.5 (11%)	11.3 (−18%)	33.8 (2%)
q_i: replace with model σ_{q_i}	25.6 (19%)	30.4 (23%)	17.4 (26%)	42.3 (28%)
u_i: replace with model $\overline{u_i}$ and σ_{u_i}	33.4 (55%)	28.2 (14%)	17.8 (29%)	37.8 (15%)
u_i: replace with model $\overline{u_i}$	27.3 (27%)	21.7 (−13%)	11.8 (−14%)	28.6 (−13%)
u_i: replace with model σ_{u_i}	25.7 (20%)	32.0 (29%)	18.6 (35%)	45.6 (38%)
v_i: replace with model $\overline{v_i}$ and σ_{v_i}	33.3 (55%)	29.5 (19%)	24.2 (75%)	47.4 (44%)
v_i: replace with model $\overline{v_i}$	25.1 (17%)	22.5 (−9%)	16.2 (17%)	32.1 (−3%)
v_i: replace with model σ_{v_i}	27.4 (27%)	31.8 (28%)	20.2 (46%)	49.3 (49%)
All σ_i replaced with model σ_{V_i}	33.2 (54%)	47.4 (91%)	30.1 (118%)	82.6 (150%)
All $\overline{V_i}$ replaced with model $\overline{V_i}$	39.7 (85%)	25.9 (4%)	18.8 (36%)	32.0 (−3%)

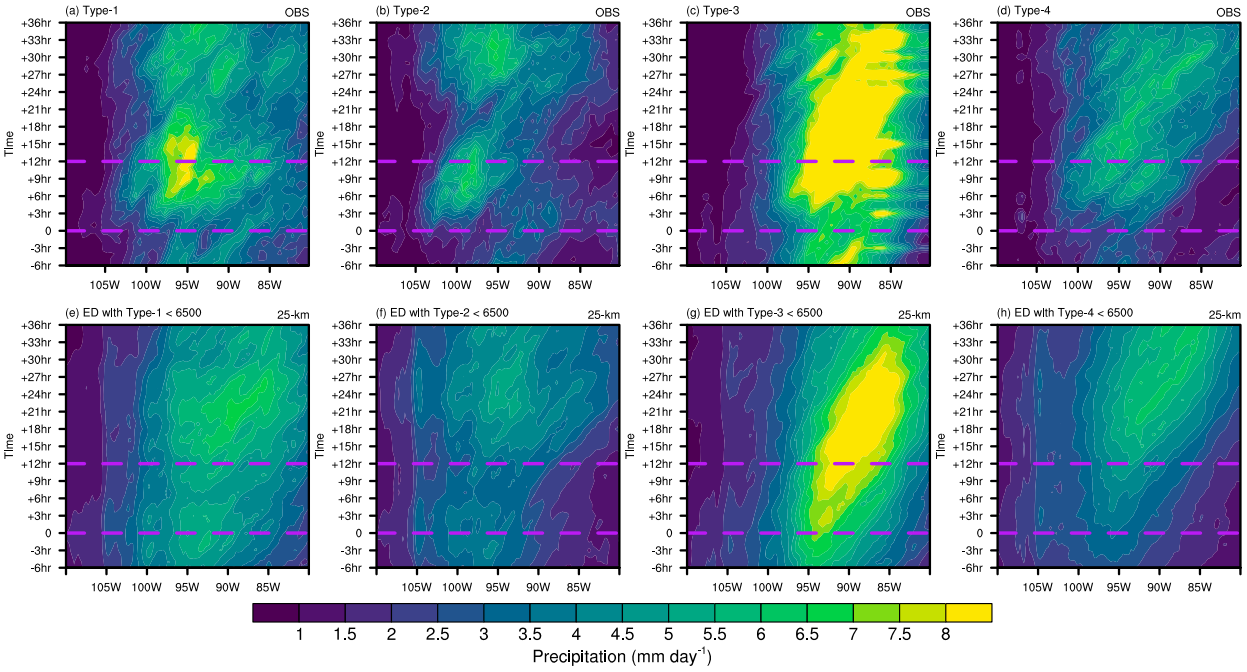

FIG. 14. Composite precipitation temporal evolution as a function of longitude associated with four types of FLSMP during spring. (a)–(d) Observations and (e)–(h) MP25km simulation. The two purple dashed lines show the relative time of FLSMP occurrence at 0 and 12 h after, respectively.

and identify MCSs. We show that this new tracking algorithm is able to track MCSs consistently across the three resolutions to reproduce the MCS frequency, seasonal, and diurnal cycle, impact on precipitation (Figs. 4 and 5), as well as a majority of individual MCS events from the reference 4-km U.S. radar MCS database (Feng et al. 2019). This result is particularly encouraging because the algorithm can be consistently applied to observations and global/regional model simulations at moderate-to-high resolution to facilitate direct comparisons.

Applying the new MCS tracking algorithm to the VR-MPAS simulations and observations at equivalent resolution, we are able to quantify model biases from multiple aspects of MCS. First, the number of simulated MCSs in the central United States is significantly underestimated in both simulations, particularly from May to August (Fig. 6). Without resolution-specific tuning, the model biases do not seem to be reduced with the higher 25-km grid spacing. As a result, the simulated precipitation amount, particularly MCS precipitation in the central United States, is underestimated by ~50% during spring (Fig. 7), and the model misses most of the MCS precipitation (~86%) during summer, resulting in low total precipitation amount bias by 17%.

The high temporal resolution (hourly) of our MCS database allows us to quantify model errors at subdaily time scales as well. Spring MCSs in VR-MPAS tend to initiate slightly more often during nocturnal and morning hours, in contrast to late afternoon initiation in observations (Fig. 9). The simulated MCS precipitation has very little diurnal variation, while non-MCS precipitation diurnal amplitude is more than twice stronger than observed and accounts for 75% of the simulated total precipitation as opposed to 40% in observations. The model produces average MCS lifetimes 3–4 h longer than observations, while the simulated MCS PF areas are larger by 40%–80% and PF mean rain rates are lower by ~15%, resulting in a 50%–70% overestimation of volume rainfall per MCS (Fig. 10). Simulation at 25 km has larger biases in MCS PF characteristics than those at 50 km, again suggesting higher model resolution does not produce more realistic MCS PFs without resolution-specific tuning or scale-aware physics parameterizations.

To better understand the potential cause of the biases in simulating MCSs and associated precipitation, it is also necessary to examine several aspects of the modeled large-scale environments. This is particularly important for spring MCSs, which are more predictable by the large-scale environments (Song et al. 2019). For the VR-MPAS model used in this study, both the 50- and 25-km simulations underestimate the LLJ frequency and associated moisture transport, particularly for the stronger LLJs (Figs. 11 and 12). The LLJ bias likely play a role in the weaker climatological low-level (below 700 hPa) meridional wind speed and drier low-level environment.

Looking beyond the LLJ, we develop and demonstrate the use of a diagnostic framework to evaluate the favorable large-scale meteorological patterns (FLSMP) for MCS development (Song et al. 2019). In this framework, we compare the simulated and observed large-scale environment associated with four types of FLSMP that are found to be skillful in explaining the observed MCS variability in spring (Song et al. 2019). The result reveals that VR-MPAS has larger biases (61%–79%) in simulating the frequency of frontal systems compared to the

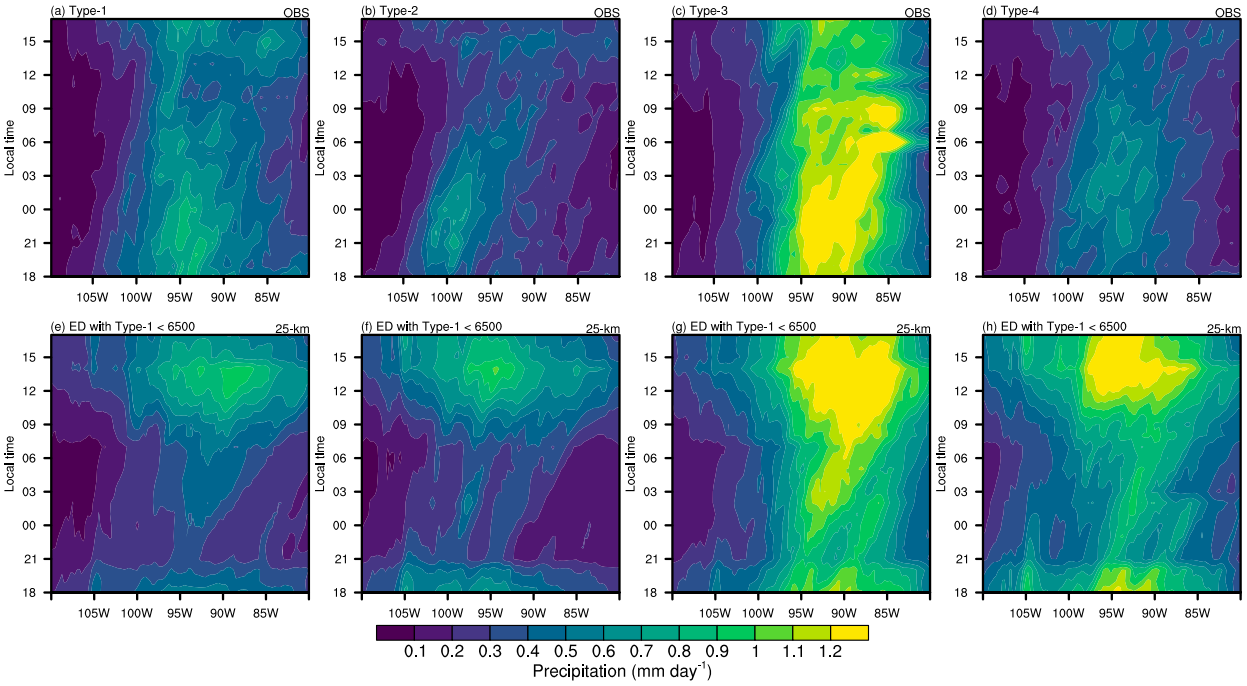

FIG. 15. Composite precipitation diurnal cycle as a function of longitude associated with four types of FLSMP during spring. (a)–(d) Observations and (e)–(h) MP25km simulation.

LLJ type FLSMPs (22%–59%). Our framework also allows us to tease out the relative contributions of model biases in dynamics and thermodynamics. And we find that the model low-level moisture biases have larger contribution to the biases of two types of FLSMP, where the moisture biases are limiting moisture transport to reach further inland in the central United States. In contrast, meridional wind biases have more significant impact on the other two types of FLSMP, where the strength of the low-level convergence is more important. In general, bias in the model-simulated variability of the large-scale environment dominates the total bias compared to the mean state bias. In particular, we find that VR-MPAS at both 50 and 25 km have weaker low-level meridional wind variability (not shown), which is consistent with the underestimation of the frequency and/or intensity of frontal systems.

The FLSMP analysis further finds that the simulated precipitation evolves differently under different FLSMPs than observations. While observed precipitation tends to increase sharply about 3 h after detection of the FLSMP and peak 6–12 h later, the VR-MPAS precipitation shows very little temporal variability, as it remains weak throughout the MCS life cycle and persists for too long, except for precipitation associated with frontal systems (Fig. 14). A majority of the simulated precipitation under different FLSMPs peaks during daytime, in contrast to the observed peak between early evening and early morning peak (1800–0600 LT).

The results in this study suggest that VR-MPAS at 50 and 25 km can simulate the different types of observed FLSMPs associated with spring MCSs, but the frequency is underestimated primarily due to the weaker Great Plains LLJ and drier low-level environment. Combined with the biases in the MCS characteristics, the understanding gained from the newly developed methodologies has already guided us to investigate several aspects of the particular configuration of MPAS coupled to CAM physics used in this study for improving simulation of MCS in the future. For example, the low-level wind bias could be related to the turbulent mountain stress parameterization introduced in the CAM5 physics, which is used in our VR-MPAS simulations (Neale et al. 2010; Richter et al. 2010). Previous studies showed that this parameterization helps reduce model biases in the large-scale circulation but makes the near-surface wind too weak over land, particularly over regions with complex topography or areas covered by grass/crop that has small surface roughness lengths (Liang et al. 2017; Lindvall et al. 2013). Sensitivity simulations conducted with the turbulent mountain stress turned off result in a visible difference in the seasonal-mean precipitation over the Great Plains (not shown). The diurnal cycle biases revealed by the MCS tracking and FLSMP analysis strengthen the general notion about the parameterized convection, particularly the drizzling tendency and the overly strong response to diurnal solar forcing. Recent developments in the ZM convection scheme allowing air parcels to launch above the boundary layer enable the model to capture nocturnal elevated convection (Wang et al. 2015; Xie et al. 2019) and may alleviate the precipitation diurnal timing biases. For the spring MCSs, more directly linking the convective trigger or closure to large-scale forcings may be helpful (Bechtold et al. 2014; Zhang 2002, 2009). We are currently investigating the performance of the Grell–Freitas convection parameterization implemented in the CAM-MPAS framework.

The Grell–Freitas convection scheme includes the aforementioned closures as well as scale awareness over the convection gray-zone (Freitas et al. 2018; Grell and Freitas 2014).

While the new developments briefly discussed above seem promising, we reiterate our notion that simply increasing the intensity and/or nocturnal precipitation in model simulations does not necessarily mean that the model captures the processes underlying MCSs. In the newly developed method reported here, we track the time, location, and other physical characteristics of MCS-like features along with diagnosis of the large-scale environment in which MCS is embedded, or FLSMP and compare them against observations at equivalent mesoscale resolutions. By exposing model biases from multiple perspectives, we are able to look deeper at the model fidelity in the underlying processes. The potential improvements in simulating MCS precipitation under various FLSMPs and at higher mesoscale resolution will be reported in future studies.

As global and climate-oriented model simulations start to resolve the mesoscale processes, it is expected that the processes critical to regional hydrological cycles and extreme events, such as MCSs, are represented more realistically than in previous generation models (Roberts et al. 2018). What is also needed is the methodology to analyze such mesoscale processes simulated in the next-generation climate models. The process-oriented evaluation and diagnostic approach developed in this study will be useful for identifying the cause of model biases in simulating MCS precipitation and understanding the performance of model physics for improving modeling of MCSs in the central United States.

Acknowledgments. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research program under Award DE-SC0016605 "A framework for improving analysis and modeling of Earth system and intersectoral dynamics at regional scales." Some data analysis and writing of the manuscript are also supported by the DOE Office of Science BER as part of the Regional and Global Modeling and Analysis program through the Water Cycle and Climate Extremes Modeling (WACCEM) scientific focus area. The simulations described in this study were performed using computational resources provided by the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract DE-AC02-05CH11231.

Data availability statement. The 4-km MCS database is obtained from the U.S. Department of Energy Atmospheric Radiation Measurement program (https://doi.org/10.5439/1571643). NCO software (Zender 2019) is used for regridding various observational datasets and model simulations output. Most of the model outputs and observational datasets used in this study are available from the data archive for the Framework for Assessing Climate's Energy Water-Land nexus by Targeted Simulations (FACETS) project (https://portal.nersc.gov/archive/home/projects/m2645/www/public_data_pnnl/).

REFERENCES

Bacmeister, J. T., and Coauthors, 2014: Exploratory high-resolution climate simulations using the Community Atmosphere Model (CAM). *J. Climate*, **27**, 3073–3099, https://doi.org/10.1175/JCLI-D-13-00387.1.

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

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

Bogenschutz, P. A., A. Gettelman, C. Hannay, V. E. Larson, R. B. Neale, C. Craig, and C. C. Chen, 2018: The path to CAM6: Coupled simulations with CAM5.4 and CAM5.5. *Geosci. Model Dev.*, **11**, 235–255, https://doi.org/10.5194/gmd-11-235-2018.

Bonner, W. D., 1968: Climatology of the low level jet. *Mon. Wea. Rev.*, **96**, 833–850, https://doi.org/10.1175/1520-0493(1968)096<0833:COTLLJ>2.0.CO;2.

Bretherton, C. S., and S. Park, 2009: A new moist turbulence parameterization in the Community Atmosphere Model. *J. Climate*, **22**, 3422–3448, https://doi.org/10.1175/2008JCLI2556.1.

Caldwell, P. M., and Coauthors, 2019: The DOE E3SM coupled model version 1: Description and results at high resolution. *J. Adv. Model. Earth Syst.*, **11**, 4095–4146, https://doi.org/10.1029/2019MS001870.

Chakraborty, S., R. Fu, S. T. Massie, and G. Stephens, 2016: Relative influence of meteorological conditions and aerosols on the lifetime of mesoscale convective systems. *Proc. Natl. Acad. Sci. USA*, **113**, 7426–7431, https://doi.org/10.1073/pnas.1601935113.

Chen, D., and A. Dai, 2019: Precipitation characteristics in the Community Atmosphere Model and their dependence on model physics and resolution. *J. Adv. Model. Earth Syst.*, **11**, 2352–2374, https://doi.org/10.1029/2018MS001536.

Chen, Q., J. Fan, Y. Yin, and B. Han, 2020: Aerosol impacts on mesoscale convective systems forming under different vertical wind shear conditions. *J. Geophys. Res. Atmos.*, **125**, e2018JD030027, https://doi.org/10.1029/2018JD030027.

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.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

Demory, M.-E., P. L. Vidale, M. J. Roberts, P. Berrisford, J. Strachan, R. Schiemann, and M. S. Mizielinski, 2014: The role of horizontal resolution in simulating drivers of the global hydrological cycle. *Climate Dyn.*, **42**, 2201–2225, https://doi.org/10.1007/s00382-013-1924-4.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Feng, Z., L. R. Leung, S. Hagos, R. A. Houze, C. D. Burleyson, and K. Balaguru, 2016: More frequent intense and long-lived storms dominate the springtime trend in central US rainfall. *Nat. Commun.*, **7**, 13429, https://doi.org/10.1038/ncomms13429.

——, and Coauthors, 2018: Structure and evolution of mesoscale convective systems: Sensitivity to cloud microphysics in

convection-permitting simulations over the United States. *J. Adv. Model. Earth Syst.*, **10**, 1470–1494, https://doi.org/10.1029/2018MS001305.

——, and Coauthors, 2019: Spatiotemporal characteristics and large-scale environments of mesoscale convective systems east of the Rocky Mountains. *J. Climate*, **32**, 7303–7328, https://doi.org/10.1175/JCLI-D-19-0137.1.

Freitas, S. R., G. A. Grell, A. Molod, M. A. Thompson, W. M. Putman, C. M. Santos e Silva, and E. P. Souza, 2018: Assessing the Grell–Freitas convection parameterization in the NASA GEOS modeling system. *J. Adv. Model. Earth Syst.*, **10**, 1266–1289, https://doi.org/10.1029/2017MS001251.

Geerts, B., and Coauthors, 2017: The 2015 Plains Elevated Convection At Night field project. *Bull. Amer. Meteor. Soc.*, **98**, 767–786, https://doi.org/10.1175/BAMS-D-15-00257.1.

Gettelman, A., and H. Morrison, 2015: Advanced two-moment bulk microphysics for global models. Part I: Off-line tests and comparison with other schemes. *J. Climate*, **28**, 1268–1287, https://doi.org/10.1175/JCLI-D-14-00102.1.

——, ——, S. Santos, P. Bogenschutz, and P. M. Caldwell, 2015: Advanced two-moment bulk microphysics for global models. Part II: Global model solutions and aerosol–cloud interactions. *J. Climate*, **28**, 1288–1307, https://doi.org/10.1175/JCLI-D-14-00103.1.

——, and Coauthors, 2018: Regional climate simulations with the Community Earth System Model. *J. Adv. Model. Earth Syst.*, **10**, 1245–1265, https://doi.org/10.1002/2017MS001227.

Grell, G. A., and S. R. Freitas, 2014: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmos. Chem. Phys.*, **14**, 5233–5250, https://doi.org/10.5194/acp-14-5233-2014.

Gross, M., and Coauthors, 2018: Physics–dynamics coupling in weather, climate, and earth system models: Challenges and recent progress. *Mon. Wea. Rev.*, **146**, 3505–3544, https://doi.org/10.1175/MWR-D-17-0345.1.

Gutjahr, O., and Coauthors, 2019: Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP). *Geosci. Model Dev.*, **12**, 3241–3281, https://doi.org/10.5194/gmd-12-3241-2019.

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.

Haberlie, A. M., and W. S. Ashley, 2019: A radar-based climatology of mesoscale convective systems in the United States. *J. Climate*, **32**, 1591–1606, https://doi.org/10.1175/JCLI-D-18-0559.1.

Harding, K. J., and P. K. Snyder, 2015: The relationship between the Pacific–North American teleconnection pattern, the Great Plains low-level jet, and north-central U.S. heavy rainfall events. *J. Climate*, **28**, 6729–6742, https://doi.org/10.1175/JCLI-D-14-00657.1.

Held, I. M., and B. J. Soden, 2006: Robust responses of the hydrological cycle to global warming. *J. Climate*, **19**, 5686–5699, https://doi.org/10.1175/JCLI3990.1.

Higgins, R. W., Y. Yao, E. S. Yarosh, J. E. Janowiak, and K. C. Mo, 1997: Influence of the Great Plains low-level jet on summertime precipitation and moisture transport over the central United States. *J. Climate*, **10**, 481–507, https://doi.org/10.1175/1520-0442(1997)010<0481:IOTGPL>2.0.CO;2.

Houze, R. A., 2004: Mesoscale convective systems. *Rev. Geophys.*, **42**, RG4003, https://doi.org/10.1029/2004RG000150.

——, 2018: 100 years of research on mesoscale convective systems. *A Century of Progress in Atmospheric and Related Sciences: Celebrating the American Meteorological Society Centennial*, *Meteor. Monogr.*, No. 59, Amer. Meteor. Soc., https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0001.1.

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.

Jankov, I., and W. A. Gallus Jr., 2004: MCS rainfall forecast accuracy as a function of large-scale forcing. *Wea. Forecasting*, **19**, 428–439, https://doi.org/10.1175/1520-0434(2004)019<0428:MRFAAA>2.0.CO;2.

Ju, L., T. Ringler, and M. Gunzburger, 2011: Voronoi tessellations and their application to climate and global modeling. *Numerical Techniques for Global Atmospheric Models*, P. Lauritzen et al., Eds., Springer, 313–342.

Kiehl, J. T., T. L. Schneider, P. J. Rasch, M. C. Barth, and J. Wong, 2000: Radiative forcing due to sulfate aerosols from simulations with the National Center for Atmospheric Research Community Climate Model, version 3. *J. Geophys. Res.*, **105**, 1441–1457, https://doi.org/10.1029/1999JD900495.

Kooperman, G. J., M. S. Pritchard, and R. C. J. Somerville, 2013: Robustness and sensitivities of central U.S. summer convection in the super-parameterized CAM: Multi-model intercomparison with a new regional EOF index. *Geophys. Res. Lett.*, **40**, 3287–3291, https://doi.org/10.1002/grl.50597.

Lawrence, D. M., and Coauthors, 2011: Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. *J. Adv. Model. Earth Syst.*, **3**, M03001, https://doi.org/10.1029/2011MS00045.

Liang, Y., L. Wang, G. J. Zhang, and Q. Wu, 2017: Sensitivity test of parameterizations of subgrid-scale orographic form drag in the NCAR CESM1. *Climate Dyn.*, **48**, 3365–3379, https://doi.org/10.1007/s00382-016-3272-7.

Lin, G., J. Fan, Z. Feng, W. I. Gustafson, P. L. Ma, and K. Zhang, 2019: Can the Multiscale Modeling Framework (MMF) simulate the MCS-associated precipitation over the Central United States? *J. Adv. Model. Earth Syst.*, **11**, 4669–4686, https://doi.org/10.1029/2019MS001849.

Lin, Y., W. Dong, M. Zhang, Y. Xie, W. Xue, J. Huang, and Y. Luo, 2017: Causes of model dry and warm bias over central U.S. and impact on climate projections. *Nat. Commun.*, **8**, 881, https://doi.org/10.1038/s41467-017-01040-2.

Lindvall, J., G. Svensson, and C. Hannay, 2013: Evaluation of near-surface parameters in the two versions of the atmospheric model in CESM1 using flux station observations. *J. Climate*, **26**, 26–44, https://doi.org/10.1175/JCLI-D-12-00020.1.

Liu, X., and Coauthors, 2012: Toward a minimal representation of aerosols in climate models: Description and evaluation in the community atmosphere model CAM5. *Geosci. Model Dev.*, **5**, 709–739, https://doi.org/10.5194/gmd-5-709-2012.

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. *J. Geophys. Res. Atmos.*, **123**, 2888–2909, https://doi.org/10.1002/2017JD027194.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, https://doi.org/10.1175/BAMS-87-3-343.

Moncrieff, M. W., 2010: The multiscale organization of moist convection and the intersection of weather and climate. *Climate*

Dynamics: Why Does Climate Vary? D.-Z. Sun and F. Bryan, Eds., Amer. Geophys. Union, 3–26.

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. *J. Geophys. Res. Atmos.*, **123**, 2655–2683, https://doi.org/10.1002/2017JD027199.

Neale, R. B., J. H. Richter, and M. Jochum, 2008: The impact of convection on ENSO: From a delayed oscillator to a series of events. *J. Climate*, **21**, 5904–5924, https://doi.org/10.1175/2008JCLI2244.1.

——, and Coauthors, 2010: Description of the NCAR Community Atmosphere Model (CAM 5.0). NCAR Tech. Note NCAR/TN-486+STR, 283 pp., www.cesm.ucar.edu/models/cesm1.1/cam/docs/description/cam5_desc.pdf.

Park, S., and C. S. Bretherton, 2009: The University of Washington shallow convection and moist turbulence schemes and their impact on climate simulations with the Community Atmosphere Model. *J. Climate*, **22**, 3449–3469, https://doi.org/10.1175/2008JCLI2557.1.

——, ——, and P. J. Rasch, 2014: Integrating cloud processes in the Community Atmosphere Model, version 5. *J. Climate*, **27**, 6821–6856, https://doi.org/10.1175/JCLI-D-14-00087.1.

Prein, A. F., C. Liu, K. Ikeda, S. B. Trier, R. M. Rasmussen, G. J. Holland, and M. P. Clark, 2017: Increased rainfall volume from future convective storms in the U.S. *Nat. Climate Change*, **7**, 880–884, https://doi.org/10.1038/s41558-017-0007-7.

——, ——, ——, R. Bullock, R. M. Rasmussen, G. J. Holland, and M. Clark, 2020: Simulating North American mesoscale convective systems with a convection-permitting climate model. *Climate Dyn.*, **55**, 95–110, https://doi.org/10.1007/s00382-017-3993-2.

Pritchard, M. S., and R. C. J. Somerville, 2009a: Assessing the diurnal cycle of precipitation in a multi-scale climate model. *J. Adv. Model. Earth Syst.*, **1**, 12, https://doi.org/10.3894/JAMES.2009.1.12.

——, and ——, 2009b: Empirical orthogonal function analysis of the diurnal cycle of precipitation in a multi-scale climate model. *Geophys. Res. Lett.*, **36**, L05812, https://doi.org/10.1029/2008GL036964.

Richter, J. H., F. Sassi, and R. R. Garcia, 2010: Toward a physically based gravity wave source parameterization in a general circulation model. *J. Atmos. Sci.*, **67**, 136–156, https://doi.org/10.1175/2009JAS3112.1.

Ringler, T. D., J. Thuburn, J. B. Klemp, and W. C. Skamarock, 2010: A unified approach to energy conservation and potential vorticity dynamics for arbitrarily-structured C-grids. *J. Comput. Phys.*, **229**, 3065–3090, https://doi.org/10.1016/j.jcp.2009.12.007.

Roberts, M. J., and Coauthors, 2018: The benefits of global high resolution for climate simulation: Process understanding and the enabling of stakeholder decisions at the regional scale. *Bull. Amer. Meteor. Soc.*, **99**, 2341–2359, https://doi.org/10.1175/BAMS-D-15-00320.1.

——, and Coauthors, 2019: Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments. *Geosci. Model Dev.*, **12**, 4999–5028, https://doi.org/10.5194/gmd-12-4999-2019.

Sakaguchi, K., and Coauthors, 2015: Exploring a multiresolution approach using AMIP simulations. *J. Climate*, **28**, 5549–5574, https://doi.org/10.1175/JCLI-D-14-00729.1.

Skamarock, W. C., J. B. Klemp, M. G. Duda, L. D. Fowler, S.-H. Park, and T. D. Ringler, 2012: A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tesselations and C-grid staggering. *Mon. Wea. Rev.*, **140**, 3090–3105, https://doi.org/10.1175/MWR-D-11-00215.1.

Song, F., Z. Feng, L. R. Leung, R. A. Houze Jr., J. Wang, J. Hardin, and C. R. Homeyer, 2019: Contrasting spring and summer large-scale environments associated with mesoscale convective systems over the U.S. Great Plains. *J. Climate*, **32**, 6749–6767, https://doi.org/10.1175/JCLI-D-18-0839.1.

Squitieri, B. J., and W. A. Gallus Jr., 2016: WRF forecasts of Great Plains nocturnal low-level jet-driven MCSs. Part I: Correlation between low-level jet forecast accuracy and MCS precipitation forecast skill. *Wea. Forecasting*, **31**, 1301–1323, https://doi.org/10.1175/WAF-D-15-0151.1.

Stevenson, S. N., and R. S. Schumacher, 2014: A 10-year survey of extreme rainfall events in the central and eastern United States using gridded multisensor precipitation analyses. *Mon. Wea. Rev.*, **142**, 3147–3162, https://doi.org/10.1175/MWR-D-13-00345.1.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Trenberth, K. E., 2011: Changes in precipitation with climate change. *Climate Res.*, **47**, 123–138, https://doi.org/10.3354/cr00953.

——, A. Dai, R. M. Rasmussen, and D. B. Parsons, 2003: The changing character of precipitation. *Bull. Amer. Meteor. Soc.*, **84**, 1205–1218, https://doi.org/10.1175/BAMS-84-9-1205.

Van Weverberg, K., and Coauthors, 2018: CAUSES: Attribution of surface radiation biases in NWP and climate models near the U.S. Southern Great Plains. *J. Geophys. Res. Atmos.*, **123**, 3612–3644, https://doi.org/10.1002/2017JD027188.

Wang, S.-Y., and T.-C. Chen, 2009: The late-spring maximum of rainfall over the U.S. central plains and the role of the low-level jet. *J. Climate*, **22**, 4696–4709, https://doi.org/10.1175/2009JCLI2719.1.

——, ——, and E. S. Takle, 2011a: Climatology of summer mid-tropospheric perturbations in the U.S. northern plains. Part II: Large-scale effects of the Rocky Mountains on genesis. *Climate Dyn.*, **36**, 1221–1237, https://doi.org/10.1007/s00382-010-0765-7.

——, ——, and J. Correia, 2011b: Climatology of summer mid-tropospheric perturbations in the U.S. northern plains. Part I: Influence on northwest flow severe weather outbreaks. *Climate Dyn.*, **36**, 793–810, https://doi.org/10.1007/s00382-009-0696-3.

Wang, Y.-C., H.-L. Pan, and H.-H. Hsu, 2015: Impacts of the triggering function of cumulus parameterization on warm-season diurnal rainfall cycles at the atmospheric radiation measurement Southern Great Plains site. *J. Geophys. Res. Atmos.*, **120**, 10 681–10 702, https://doi.org/10.1002/2015JD023337.

Wehner, M. F., and Coauthors, 2014: The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1. *J. Adv. Model. Earth Syst.*, **6**, 980–997, https://doi.org/10.1002/2013MS000276.

Whiteman, C. D., X. Bian, and S. Zhong, 1997: Low-level jet climatology from enhanced rawinsonde observations at a site in the Southern Great Plains. *J. Appl. Meteor.*, **36**, 1363–1376, https://doi.org/10.1175/1520-0450(1997)036<1363:LLJCFE>2.0.CO;2.

Williamson, D. L., 2013: The effect of time steps and time-scales on parametrization suites. *Quart. J. Roy. Meteor. Soc.*, **139**, 548–560, https://doi.org/10.1002/qj.1992.

Xie, S., and Coauthors, 2018: Understanding cloud and convective characteristics in version 1 of the E3SM atmosphere model. *J. Adv. Model. Earth Syst.*, **10**, 2618–2644, https://doi.org/10.1029/2018MS001350.

——, and Coauthors, 2019: Improved diurnal cycle of precipitation in E3SM with a revised convective triggering function. *J. Adv. Model. Earth Syst.*, **11**, 2290–2310, https://doi.org/10.1029/2019MS001702.

Yang, G. Y., and J. Slingo, 2001: The diurnal cycle in the tropics. *Mon. Wea. Rev.*, **129**, 784–801, https://doi.org/10.1175/1520-0493(2001)129<0784:TDCITT>2.0.CO;2.

Yang, Q., R. A. Houze, L. R. Leung, and Z. Feng, 2017: Environments of long-lived mesoscale convective systems over the Central United States in convection permitting climate simulations. *J. Geophy. Res. Atmos.*, **122**, 13 288–13 307, https://doi.org/10.1002/2017JD027033.

Zender, C. S., 2019: netCDF Operators (NCO) version 4.8.0. Zenodo, https://doi.org/10.5281/zenodo.61341.

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.

——, 2009: Effects of entrainment on convective available potential energy and closure assumptions in convection parameterization. *J. Geophys. Res.*, **114**, D07109, https://doi.org/10.1029/2008JD010976.

——, and N. A. McFarlane, 1995: Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centre general circulation model. *Atmos.–Ocean*, **33**, 407–446, https://doi.org/10.1080/07055900.1995.9649539.

Zhao, C., and Coauthors, 2016: Exploring the impacts of physics and resolution on aqua-planet simulations from a nonhydrostatic global variable-resolution modeling framework. *J. Adv. Model. Earth Syst.*, **8**, 1751–1768, https://doi.org/10.1002/2016MS000727.