Towards a Unified Setup to Simulate Mid-Latitude and Tropical Mesoscale Convective Systems at Kilometer-Scales

Andreas F. Prein[1], Ming Ge[1], Alexandra Ramos Valle[1], Dié Wang[2], Scott E. Giangrande[2]

[1]National Center for Atmospheric Research, Boulder, CO, U.S.A.
[2]Brookhaven National Laboratory, Upton, NY, U.S.A.

Key Points:

- A method is developed that accounts for spatiotemporal displacements in simulated U.S. and Brazilian mesoscale convective systems (MCSs).
- Central U.S. MCSs are found more sensitive to microphysics while planetary boundary layer (PBL) physics are more influential in Brazil.
- Optimal mid-latitude and tropical MCS model settings are identified.

Corresponding author: Andreas F. Prein, `prein@ucar.edu`

Abstract

Mesoscale convective systems (MCSs) are the main source of precipitation in the tropics and parts of the mid-latitudes and are responsible for high-impact weather worldwide. Studies showed that deficiencies in simulating mid-latitude MCSs in state-of-the-art climate models can be alleviated by kilometer-scale models. However, whether these models can also improve tropical MCSs and weather we can find model settings that perform well in both regions is understudied. We take advantage of high-quality MCS observations collected over the Atmospheric Radiation Measurement (ARM) facilities in the U.S. Southern Great Plains (SGP) and the Amazon basin near Manaus (MAO) to evaluate a perturbed physics ensemble of simulated MCSs with 4 km horizontal grid spacing. A new model evaluation method is developed that enables to distinguish biases stemming from spatiotemporal displacements of MCSs from biases in their reflectivity and cloud shield. Amazon MCSs are similarly well simulated across these evaluation metrics than SGP MCSs despite the challenges anticipated from weaker large-scale forcing in the tropics. Generally, SGP MCSs are more sensitive to the choice of model microphysics, while Amazon cases are more sensitive to the planetary boundary layer (PBL) scheme. Although our tested model physics combinations had strengths and weaknesses, combinations that performed well for SGP simulations result in worse results in the Amazon basin and vice versa. However, we identified model settings that perform well at both locations, which include the Thompson and Morrison microphysics coupled with the Yonsei University (YSU) PBL scheme and the Thompson scheme coupled with the Mellor–Yamada–Janjic (MYJ) PBL scheme.

1 Introduction

Mesoscale convective systems (MCSs) play an important role in Earth's energy balance (Houze Jr, 2018) and are essential for Earth's water cycle in the tropics (Nesbitt et al., 2006; Feng, Leung, et al., 2021) and mid-latitude regions (Fritsch et al., 1986; Feng et al., 2016; Feng, Leung, et al., 2021). These systems are prolific rain producers and are the main cause of warm-season flooding (R. S. Schumacher & Johnson, 2005; Rasmussen et al., 2014; Pokharel et al., 2018). Observations of MCSs over the continental U.S. indicate that extreme precipitation rates associated with MCSs have significantly increased during the past decades (Feng et al., 2016) and MCSs are predicted to further intensify in the future climate (A. F. Prein et al., 2017). Nevertheless, a major bottleneck for predicting possible climate change effects on climate extremes has been related to the poor representation of MCS intensity (e.g., precipitation rates, updraft strength) and spatiotemporal evolution in state-of-the-art models (Wang et al., 2020; Donner et al., 2016; Lin et al., 2022). Improving our MCS modeling capabilities on weather, seasonal, and climate time scales is essential to advance the credibility of model predictions.

The frontier of global and regional atmospheric modeling has reached convection-permitting scales (horizontal grid spacings $\Delta x \leq 4$ km; Satoh et al. (2019)). While global large eddy (LES) simulations on climate timescales are far out of reach, convection-permitting "gray-zone" decadal simulations are already feasible in regional models (e.g., Liu et al. (2017); Berthou et al. (2020)) and will soon be achievable with global models (e.g., Stevens et al. (2019)). Convection-permitting models (CPMs) can explicitly simulate deep convective clouds, which revolutionizes our ability to simulate and predict severe weather and climate extremes (A. F. Prein et al., 2015; Clark et al., 2016). These CPMs substantially improve the simulation of MCSs including their propagation direction and speed, evolution, size, and associated extreme precipitation (A. F. Prein et al., 2020). Although this progress is encouraging, CPMs have difficulties simulating MCS cold pool and draft dynamics that modulate the MCS lifecycle and development, especially in weakly-forced environments (Haberlie & Ashley, 2019a; Wang et al., 2020).

In this study, we investigate the ability of the Advanced Research Weather Research and Forecasting (AR-WRF or short WRF) model (Skamarock & Klemp, 2008; Powers et al., 2017) at convection-permitting resolution (4 km horizontal grid spacing) to simulate MCSs that overpassed the U.S. Department of Energy's (DOE) Atmospheric Radiation Measurement (ARM) (Mather & Voyles, 2013) sites in the U.S. Southern Great Plains (SGP, Lamont, Oklahoma) (Sisterson et al., 2016) and the Amazon basin (MAO, Manaus, Brazil) (Martin et al., 2016). MCSs in those regions initiate and develop under very different environmental conditions, which promotes distinct convection lifecycle characteristics and associated precipitation behaviors (Wang et al., 2019).

Moreover, Wang et al. (2019) recently showed that MCSs over the SGP and MAO sites feature similar rainfall rates and accumulations, though having larger stratiform rainfall contributions in the U.S. SGP events. Similarly, the convective cold pools have comparable strength in both regions. However SGP MCSs show more intense convective updrafts, deeper and stronger convective downdrafts, and larger mass flux than MAO MCSs (Wang et al., 2020). The SGP MCS investigated for this study predominantly occur in the springtime (see Table 1) and feature strong large-scale forcing (e.g., frontal passages), which is distinctly different from MAO MCSs that typically occur under much weaker synoptic-scale forcing. These differences in synoptic-scale and thermodynamic forcings motivate this study to explore the skill of CPMs and their sensitivities to model physics in simulating MCSs in both regions. There have been several studies that investigate the sensitivity of simulated MCSs to the model microphysics (Xue et al., 2017; Feng et al., 2018; Shpund et al., 2019; Fan et al., 2017; Han et al., 2019) and planetary boundary layer (PBL) schemes (Stephan & Alexander, 2014) using kilometer-scale models. Most of these studies focus on U.S. mid-latitude MCSs and do not assess sensitivities across climate zones.

The MCS overpasses over the SGP and MAO sites that are described in Wang et al. (2019) provide a unique opportunity for in-depth analyses of the performance of CPMs in simulating MCSs in these two regimes. However, using ARM observations for model evaluation is challenging due to the limited spatial coverage of the observations and spatiotemporal displacements of MCSs in the simulations. Traditional evaluation methods cannot be used in such situations due to the so-called "double-penalty problem" (Roberts & Lean, 2008; A. Prein et al., 2013). In these cases, simulation performance is penalized twice compared to null behaviors, once for not properly simulating the primary event and once for including secondary features that were not observed. We present a new method that identifies the spatiotemporal displacement of simulated storms, which allows us to disentangle displacement errors from other modeling errors. We use this method to test the sensitivity of simulated MCSs at the SGP and MAO sites to perturbed model PBL- and micro-physics schemes.

2 Data and Methods

2.1 MCS Case Selection

MCSs in the U.S. Great Plains are well studied using observations (Fritsch et al., 1986; Houze Jr, 2018; Song et al., 2019; Haberlie & Ashley, 2019b; Wang et al., 2019) and models (Trier et al., 2010; Feng et al., 2018; A. F. Prein et al., 2020). U.S. MCSs have a strong seasonality and are most frequent in the southern Great Plains during spring and in the northern Great Plains during summer (Li et al., 2021). Spring MCSs are frequently related to frontal passages, feature an enhanced Great Plains low-level jet, and are typically squall lines (Song et al., 2019). Spring MCSs occur under large convective available potential energy (CAPE) and convective inhibition (CIN) anomalies, while summertime MCSs occur under much weaker synoptic-scale forcing (Feng, Song, et al., 2021). Even kilometer-scale models can have difficulties in simulating MCSs under weakly forced summertime conditions (A. F. Prein et al., 2020).

MCSs are the dominant rain-producer in the Amazon basin, which features about 7,200 MCSs per year (Rehbein et al., 2018; Anselmo et al., 2021) contributing approximately 50 % of the annual rainfall (Feng, Leung, et al., 2021). MCSs in the Amazon can have various morphologies and sizes but typically develop in lines (Anselmo et al., 2021). The number of MCSs during the ARM Green Ocean Amazon (GoAmazon) campaign period (2014–2015), whose data we leverage in this study, was about 50 % lower than the 2000–2013 climatology (Rehbein et al., 2019) probably due to positive equatorial Pacific sea surface temperatures and reduced moisture transport into the Amazon basin. For additional context, the larger-scale regimes associated with GoAmazon mature MCS events are summarized by Wang et al. (2019) and more generically in Giangrande et al. (2020).

We investigate MCS events over the U.S. DOE ARM (Mather & Voyles, 2013) SGP (Sisterson et al., 2016) site and the 2014/15 GoAmazon field campaign site near Manaus Brazil (MAO; Martin et al. (2016, 2017); Giangrande et al. (2017)). In previous work, we identified 16 MCS overpasses over the SGP site and 44 over the MAO sites (Wang et al., 2019). From these cases, we simulate a sub-set of 13 cases at the SGP site and 41 cases at the MAO site based on observational data availability. After comparing to observed radar reflectivity (Z) and satellite brightness temperature (BT) we select 11 well-simulated MAO MCSs that occurred in different seasons (4 cases in the wet seasons, 4 cases in the transition seasons, and 3 cases in the dry seasons; see Table 1). These MCSs have various morphologies (e.g., propagating lines vs. convection organization over the site). The dates and characteristics of the selected SGP and MAO cases are shown in Table 1.

2.2 Observational Datasets

We use S-band Z (10-cm wavelength, 3 GHz) and satellite-derived cloud BT for model evaluation. For MAO MCS cases, volumetric radar observations collected during the GoAmazon2014/15 campaign by the SIPAM S-band radar are used. These data are quality controlled and interpolated to a three-dimensional grid (C. Schumacher & Funk, 2018). Data are available every 12 minutes on a 2 km horizontal grid, which covers a maximum area of 480 km×480 km. The vertical grid spacing is 500 m at constant altitude plan position indicator (CAPPI) with a vertical grid spacing of 500 m up to 20 km height. We test the sensitivity of our model evaluation using scans at 2 km, 4 km, and 6 km altitudes. Those levels are chosen due to the complete data coverage over the radar site (higher levels have missing data over the radar location). The 2 km CAPPIs (those that sample to the furthest distance) are limited to 180 km in range, which informed our decision about the default scanning distance that we will introduce in Section 2.4. For the SGP MCS events, we use the GridRad (Homeyer & Bowman, 2017) product that merges Z from the U.S. National Weather Service NEXRAD WSR-88D high-resolution S-band Doppler weather radars (Ansari et al., 2018). GridRad provides hourly, three-dimensional, Z on an 0.02° horizontal and 1 km vertical grid between 1995 to 2018 covering the Contiguous United States. We use the same scanning distance as for the Amazon MCS evaluation for consistency, but the larger radar coverage in the U.S. also allows testing the sensitivity of our analysis to the scanning distance setting. The NEXRAD and SIPAM are quality controlled and accurate within 1-2 dBz (Wang et al., 2019).

The BT observations are derived from the GOES-13 satellite Chanel 4 (Knapp & Wilkins, 2018). Channel 4 measurements (often referred to as the "Cirrus" Band; approx. central wavelength is 1.37 µm) were selected owing to their high contrast in identifying anvil clouds and their good agreement with simulated BT data (Feng, Leung, et al., 2021). The only exception where we substituted GOES13 data with GOES15 data is June 5, 2013, MCS case that over-passed the SGP site due to missing data in GOES13. GOES data are provided hourly with all data mapping to the nearest hour on an equal angle grid with a spacing of 0.04°.

BT needs to be estimated from WRF output since it is not a standard output variable, either by running a radiative transfer model (NCEP, 2020) or by applying empirical relationships between the top of the atmosphere long-wave outgoing radiation and BT (Yang & Slingo, 2001; Wu & Yan, 2011). We tested both approaches and found that they result in very similar estimates, especially over the convective regions, which are the focus of this study. Since empirical estimates are significantly cheaper to perform compared to running a radiative transfer model, we use the equation presented in Wu and Yan (2011) for the calculation of the BTs using WRF top of the atmosphere outgoing long-wave radiating. This approach is similar to what has been used in existing MCS studies (Feng, Leung, et al., 2021).

2.3 Model Setup

We use the WRF model version 4.1.5 (Skamarock & Klemp, 2008; Powers et al., 2017) for the simulations. The simulation domains are shown in Fig. 1. Each domain consists of 500×500 grid cells and 96 stretched vertical levels. The horizontal grid spacing is approximately 4 km, which is sufficient to capture bulk MCS properties reasonably well at computational affordable costs (A. F. Prein et al., 2020). Each simulation is started 24-hours before an MCS overpass at the corresponding ARM site (Wang et al., 2019) and has a total runtime of 36-hours. The initial and lateral boundary conditions are derived from hourly pressure level data from the fifth generation European Center for Medium-Range Weather Forecasting (ECMWF) reanalysis (ERA5; Hersbach et al. (2020)). We use the Noah-MP land surface model (Niu et al., 2011), the RRTMG shortwave and longwave radiation scheme (Iacono et al., 2008) and do not use a convection parameterization scheme. We test three options each for the microphysics and PBL parameterizations, to be described below. These schemes were selected as they represent several of the most widely tested/used options and feature different levels of complexity and underlying assumptions.

For our microphysical sensitivity tests, we chose the Thompson (Thompson et al., 2004), the Morrison (Morrison et al., 2009), and the P3 (Morrison & Milbrandt, 2015) schemes. All of these are bulk microphysics schemes that vary in their representation of hydrometeors. The Thompson scheme uses two moments for cloud water, rain, and graupel/hail hydrometeors and one moment for ice and snow; which allows it to predict graupel/hail, water, and rain density. However, Thompson representations for ice properties is otherwise limited compared to the other schemes tested. The Morrison microphysics scheme is more complex than the Thompson scheme since it also represents two moments of ice and snow. The P3 scheme follows the full 2-moment implementation of the Morrison scheme but includes a detailed prediction of ice particle properties (conceptually similar to Jensen et al. (2017)). This change avoids the artificial classification of frozen hydrometeors into ice, snow, and graupel/hail categories. This scheme has a conceptual advantage over the Morrison and Thompson schemes but is less widely used and tested.

For the PBL parameterization sensitivity testing, we considered the Yonsei University (YSU; Hong et al. (2006)), the Mellor–Yamada–Janjic (MYJ; Janjić (1994); Mesinger (1993)), and Mellor–Yamada Nakanishi Niino Level 2.5 schemes (MYNN2.5; Nakanishi and Niino (2006, 2009)). YSU is a non-local scheme that uses a first-order closure and has an improved simulation of deeper vertical mixing in buoyancy-driven PBLs and shallower mixing in strong-wind regimes compared to successor PBL schemes (e.g., the MRF scheme; Hong and Pan (1996)). However, YSU features systematic biases that may include issues such as PBLs that deepen too vigorously for springtime deep convective environments, resulting in an underestimation of near-surface buoyancy (Coniglio et al., 2013). In contrast, the MYJ parameterization is a local 1.5-order closure scheme, which improves PBL simulations compared to its preceding schemes (Mellor & Yamada, 1982), without increases in computational costs. However, MYJ tends to undermix PBLs for locations upstream of convection (Coniglio et al., 2013). Finally, we employ the MYNN2.5, which is a local scheme that uses a 1.5-order closure and improves the PBL depiction compared to non-local schemes (e.g., YSU) during springtime in environments that support deep convection (Coniglio et al., 2013). Similar to MYJ, the local formulations of MYNN2 do not fully account for deep vertical mixing.

2.4 Model Evaluation

This study is motivated by the complex nature by which model biases arise due to spatiotemporal displacements that are, in parts, intrinsic to the simulation of deep convection from biases that are predominantly related to model deficiencies (e.g., model physics, grid spacing, numerics). We introduce a method that separates those bias components by identifying the time and location in the simulation that best aligns with the observed MCS overpass over the ARM sites. The model analysis is performed at the identified optimal, displaced location by using observations that are common to many regions where MCSs occur.

In Fig. 2, we provide a schematic of our approach. It starts by selecting a scan area (N), which is based on the spatial reach of the SIPAM S-band radar in Manaus ($\sim$180 km radius at 2 km height). A square box scan area with a side length of N=2.6° ($\sim$290 km) was selected over a circular region for computational efficiency. As previously described, Z CAPPIs at 2 km, 4 km, and 6km above ground level, are input, with the multiple heights included testing the sensitivity of the model evaluation to the height of the radar observation. Since BT and GridRAD observations are only available at full hours, we search for the time of maximum Z in the scan area for MAO MCSs and round the time to the closest full hour.

Next, we define a search time window (T) that corresponds to the maximal allowed temporal displacement in the model. Here we use $T\pm4$ hours around the time of the observed MCS overpass. We decided to constrain T to four hours since larger temporal displacements would likely result in MCSs that develop in significantly different environmental conditions than the observed MCSs. We used a similar rationale in defining a search area that is 2° degrees larger than the scan area in each direction (M =N+2°+2°; Fig. 2).

For each model output interval (t=10-minutes), we derive the simulated Z and BT within the search area and the search time window. For instance, we have 65×65 grid cells within the scan area and 165×165 grid cells within the search area, which results in 100×100 possibilities to shift the scan area within the search area. For each output interval, we calculate two skill scores for all possible shifts of the scan area within the search area.

The first skill score is the spatial pattern correlation coefficient (CC; Wilks (2011)) which evaluates the model skill in capturing the spatial pattern of simulated Z and BT without penalizing the model for systematic magnitude biases. The second skill score is the absolute cumulative distribution function difference (ACDFD; see Fig. 2 for an example). The ACDFD, in comparison, does not penalize the model for deficiencies in simulating spatial patterns, but solely focuses on the correct simulation of the Z and BT magnitudes.

This evaluation results in a displacement matrix containing $T\times(M-N)\times(M-N)$ skill scores for Z CC, BT CC, Z ACDFD, and BT ACDFD. Using the above example, these two skill scores are calculated 480,000 times for each MCS case, assuming a search time window of ±4-hours with 10-minute output intervals (48-time slices) and 100×100 possibilities to shift the scan area within the search area. To find the temporal (Δt) and spatial (Δx) displacement that corresponds to the location of the optimal model performance, we combine these four skill scores by normalizing their distributions to a mean of zero and a standard deviation of one. Next, we multiply the normalized ACDFD matrices by minus one, which means that larger values are better (similarly to the CC statistics). The average of the four derived matrices is calculated, resulting in a displacement matrix in which larger values correspond to an improved agreement between the model and observations. Finally, we search for the maximum in the displacement matrix to derive Δt and Δx and use the optimal location and time for model evaluation.

To better understand the impact of model physics on the MCS environments, we calculate CAPE, CIN, and vertical average hydrometeor mixing ratios in a $\pm$15 grid cell square around the optimal location. CAPE and CIN are calculated with the python wrf.cape_2d function that finds the level of maximum equivalent potential temperature height in the lowest 3,000 m above ground. Next, a parcel with 500 m depth centered on this height is defined and used for the CAPE and CIN calculation.

3 Results

3.1 Idealized Tests

Before we apply the model evaluation method to our simulated MCSs, we test its performance on four idealized cases. These cases are similar to cases used in previous studies for testing model evaluation methods (e.g., see Fig. 2 in Davis et al. (2009)) and exemplify how the derived skill scores are affected by specific biases in the simulation. To simplify the analysis, we remove the time dimension and only consider one variable (Z). A summary of these tests is provided in the list below.

- The first case represents a simulated MCS that is identical to the observed case, but eastward displaced by 4° (Fig. 3a). The algorithm can detect the displacement, which is 350 km (4° longitude at 36.6° North; the latitude of the SGP site). As expected, the CC=1 and the ACDFD=0 dBZ when accounting for the shift.
- The second case (Fig. 3b) is identical to the first case, but the simulated Z values are 10 dBZ higher than the observed ones. The method can identify the displacement without any problems and returns a perfect correlation coefficient (CC=1) and an ACDFD of 6 dBZ. The reason why the latter is not 10 dBZ is because we are associating cloud-free areas remain zero dBZ, in both observation and simulations.
- The third test case (Fig. 3c) features a displacement bias of 5° to the east and has a simulated MCSs width that is double the observed one. The algorithm detects an eastward displacement of 579.2 km (6.6°) and returns a moderate CC of 0.58 indicating that the MCS spatial patterns are erroneously simulated. The ACDFD score has a value of 6 dBZ resulting from the wider areas of simulated Z.
- The final test case (Fig. 3d) features a simulated storm that is identical to the observed one but shifted by 7° to the east and rotated by 90°. The rotational bias is reflected in a low correlation coefficient (CC=0.1), while the ACDFD score is zero due to the correct simulation of Z magnitudes in the scan area. The simulated storm is identified as shifted eastward and slightly southward. The southward shift stems from the asymmetry in the storm's Z values.

3.2 Evaluation of Simulated MCSs

After gaining confidence in our evaluation method based on idealized settings, we apply the evaluation algorithm to real-world MCS simulations over the SGP and MAO. Fig. 4 shows a representative example of the algorithm's input and output for the June 6, 2014 case at 9:00 UTC at the SGP site. The observed MCS shows a large anvil cloud shield (Fig. 4a) associated with a squall line (Fig. 4g). The simulated cloud shield (Fig. 4b) is smaller than the observed one although the model overestimates Z at 4 km altitude (Fig. 4h). The location where the modeled MCS is most similar to the observed one is 180 km displacement towards the northeast (distance between red and blue dots in Fig. 4a,b) and occurs 120 minutes later (Fig. 4m). Comparisons between BTs in the observed (Fig. 4c) and simulated (Fig. 4d) scan areas show that simulated cloud tops are warmer (Fig. 4f; ACDFD=7.2 K), and the spatial pattern correlation is CC=0.41 (Fig. 4e). The spatial structures of Z in the scan areas are better simulated (Fig. 4i,j) with a slight higher CC of 0.54 (Fig. 4k); however, the simulated Z values are higher than observed (Fig. 4l; ACDFD=3.9 dBZ).

Fig. 4m shows the spatial maximum values of the normalized ACDFD Z, ACDFD BT, Z CC, and BT CC scores for every 10-minutes model output within the search time window. Note that the normalized ACDFD values have been multiplied by minus one, which means that larger values are better for all skill scores. While BT CC is the closest to the observations at t=0 minutes, the maxima in the other skill scores are delayed. Equally weighting the four skill scores results in a temporal delay of 120 minutes between the observed and the simulated MCSs. Fig. 4n shows the displacement matrix at t=120-minutes with the maximum value being highlighted as a blue dot. The displacement matrix component from the four skill scores are shown in Fig. 4o–r, each showing a displacement of the simulated MCS towards the northeast.

3.3 Model Physics Sensitivities

Fig. 5 shows observational and simulated results for the Nov. 17, 2014, MCS event at the MAO site to exemplify the impacts of different microphysics and PBL schemes on the simulated cloud and Z fields. The simulated fields are shown at the time of the optimal comparison to the observations. This event features a line of clouds produced by a weakly forced line of convection (Fig. 5a). Somewhat unexpectedly for a tropical MCS event, most of the tested simulations can capture the basic characteristics of this case. Clear outliers are the simulations that use the MYNN2.5 PBL scheme, which produces wide-spread, disorganized clouds. Additionally, MYNN2.5 seems to produce clouds that are strongly influenced by the Amazon River (especially visible in the simulations using the Thompson and Morrison microphysics schemes), which is not evident from the observations.

Fig. 6 shows a 'heat map' that provides an overview of the four skill scores including the spatial and temporal displacements for all tested physics combinations (MCS cases for Z CAPPIs at 2 km above the ground). The large case-to-case variability at both locations is prominent, which appears larger than the sensitivity to the selected physics (a more detailed analysis on this topic is presented below). Additionally, there is little correlation between skill scores. This implies that well simulated BT patterns (e.g., the SGP case on June 18, 2016) do not infer well simulated Z patterns or ACDFD values. This lack of consistency is surprising and should be further investigated in follow-up studies. Another surprising result is that skill scores are similar for the MAO and SGP MCSs despite their different environmental conditions. Our initial expectation was that the SGP MCSs might be better simulated due to the involvement of mid-latitude disturbances that help to initiate and organize the systems. However, such a difference is not obvious besides there being a slightly smaller spatiotemporal displacements for SGP MCSs. The largest difference between MCSs at these two locations are the lower (better) ACDFD Z scores over MAO, which is in part due to the stronger and larger frontally driven MCSs in the SGP (i.e., relative Z differences might be more similar). Another noteworthy difference between the two regions is that SGP MCSs tend to be simulated too early (on average), while MAO MCSs are simulated slightly too late for most physics settings. There is little systematic effect from evaluating Z at different altitudes (not shown). We note that slightly higher average CC values are observed for the SGP events that used the 4 km CAPPIs. However, ACDFD scores for Z are the worst (highest) at this altitude. We attribute these discrepancies to radar 'bright band' signatures (observed Z enhancement owing to aggregation and melting) in observed Z factor expected near the melting layer, something that model microphysics are struggling to simulate and standard forward model-radar operators do not capture.

Due to the large case-to-case variability, we average the skill scores of all cases and the 2 km, 4 km, and 6 km Z heights to better isolate the impact of the model physics on the simulation performance (Fig. 7). Limited consistency exists between schemes that perform well for the SGP and MAO events. For instance, simulations using the MYNN2.5 scheme have the highest BT CCs for SGP cases, yet feature the lowest CCs for MAO MCSs (Fig. 7b). This result is in agreement with previous examples as in Fig. 5. Interestingly, a lower skill score in simulating cloud top structures (Fig. 7c) does not directly translate to a lower skill score in simulating Z patterns (Fig. 7a).

A more rigorous assessment of the sources of variability in the SGP and MAO cases based on variance decomposition (Déqué et al., 2007; A. F. Prein et al., 2011) is shown in Fig. 8. Case to case variability is the largest source of uncertainty contributing between one-third to two-thirds of the total variability (Fig. 8a–l). The variability stemming from the choice of microphysics or PBL scheme is comparatively small. The choice of the PBL scheme is most important (contributing $\sim$20 % to the total variability) for the ACDFD score of BT in the Amazon (Fig. 8g), while the microphysics scheme selection is most influential in simulating the Z ACDFD score in the SGP (Fig. 8f). The sizes of the rings in Fig. 8 indicate the total variability. The ACDFD scores of Z (Fig. 8e,f) are smaller for cases in MAO, while the opposite is true for temporal and spatial displacements (Fig. 8i–l).

We repeated the variance decomposition by averaging over all cases within a region to highlight the other sources of variability besides the case-to-case variability (Fig. 8m–x). This reveals major differences between modeling sensitivities of MAO and SGP MCS cases. For instance, the PBL scheme substantially impacts Amazon brightness temperature CC (45 %; Fig. 8o) and ACDFD scores (80 %; Fig. 8s). On the other hand, SGP Z ACDFD scores are very sensitive to the microphysics parameterization (60 %; Fig. 8r) and Z CCs change substantially with height at which Z is measured (60 %; Fig. 8n). Concerning the total variability at the two locations, BT ACDFD score and the temporal and spatial displacement variabilities at the MAO site are substantially larger than those at the SGP site.

To better understand the impact of the tested physics settings on the MCS simulations at the MAO and SGP sites, we calculate the evolution of CAPE, CIN, and mean cloud condensates at the corresponding optimal locations averaged over all cases (Fig. 9). One of the most noticeable differences is that MAO pre-MCS environments have lower CIN values when using the MYNN2.5 PBL scheme (Fig. 9c). This indicates that using MYNN2.5 results in enhanced mixing at the top of the PBL and supports the development of wide-spread convection, such as seen in Fig. 5. Similarly, CAPE values are typically smaller in pre-MCS environments when using MYNN2.5, although the differences are not as clear as for CIN (Fig. 9a). The impact of model physics on CAPE and CIN is less systematic for SGP MCSs (Fig. 9b,d) except for the post-MCS environmental CAPE, which is the lowest when using Thompson microphysics and the highest when using the Morrison scheme.

Using the MYNN2.5 scheme at the MAO site results in a local maximum of cloud condensates at $\sim$3 km height in the pre-MCS environments and during the MCS overpass (Fig. 9e,f), while using the YSU and MYJ schemes leads to a less pronounced peak that is at a lower altitude. Consistent with the above analysis, this indicates that the MYNN2.5 scheme produces a deeper and strongly mixed PBL. Such differences are not obvious for SGP MCS cases (Fig. 9h–j).

3.4 Best Performing Model Settings

In this section, we calculate the average skill scores ranks for each physics combination. This allows combining the six individual scores to a single average rank skill score that ranges from zero (best performing option for all tested physics) to one (worst-performing option), with 0.5 indicating the average performance. For more details, see Section 2.4.

Overall, the results plotted in Fig. 10a shows that the Thompson microphysics performs best for SGP MCSs on average, while cases that use the Morrison and P3 schemes perform worse than average. The average sensitivity to the PBL scheme is smaller compared to microphysics sensitivities, although individual scores such as the temporal displacement show a clear sensitivity to the PBL scheme. Our simulations indicate the Thompson scheme performs best in capturing the Z distribution (ACDFD score), while it performs below average concerning spatial displacements (particularly in combination with the YSU scheme).

The MAO MCS cases are more sensitive to the selection of the PBL scheme than SGP cases, although there is a clear effect of a microphysics-PBL scheme interaction as well (Fig. 10b). As shown before, using the MYNN2.5 scheme results in a sub-optimal performance independent of the microphysics parameterization. The main contributors to the poor performance are its deteriorated simulation of Z statistics and BT correlation coefficients. Overall, the best performance is achieved with the Thompson and MYJ scheme, followed by Morrison-YSU, and Thompson-YSU. Interestingly, the worst-performing physics combination in one region can perform best in the other region and vice-versa. This can be seen for the spatial and temporal displacement scores when using Thompson-YSU.

From our input sensitivity testing, we find that three physics combinations perform above average independent of the Z height that is used for the analysis. Those are all PBL combinations with the Thompson scheme and Morrison-YSU. If the performance of an individual score is more important than the average performance, other physics combinations might be more appropriate such as the P3-YSU combination, which results in the smallest MCS displacement.

In addition to the tests of CAPPI altitudes, we test the sensitivity of model performance to radar coverage/capture domain. These tests are only possible at the SGP site using NEXRAD GridRad product given the available radar networks (Fig.11). Overall, sensitivities are generally smaller than the physics sensitivity. This implies that the selection of well-performing physics options is not affected by this setting.

4 Summary, Discussion, and Conclusion

We present a new model evaluation method that allows us to differentiate spatiotemporal displacement biases from biases in the simulated structure and intensity of the MCSs. We evaluate the skill of kilometer-scale models in simulating MCSs using SGP and MAO radar and GOES satellite observations. We are particularly interested in the impact of the model microphysics and PBL scheme on the simulations of mid-latitude and tropical MCSs. The main results of this study are as follows.

- Kilometer-scale models equally well simulate continental tropical and mid-latitude MCSs in terms of the spatial structure and intensity of the convection and the cloud top field. However, spatial and temporal displacements tend to be smaller in mid-latitudes, likely due to the ability of the model to capture the large-scale forcing-driven convection.
- Model physics that work well in mid-latitudes do not necessarily work well in the tropics and vice versa. For instance, simulations using the MYNN2.5 PBL scheme best simulate the pattern correlations of cloud BT in SGP but perform worst in the Amazon basin. Nevertheless, we can identify model settings that perform above average in both environments, such as the Thompson and Morrison microphysics with the YSU PBL scheme or the Thompson scheme with the MYJ PBL scheme.

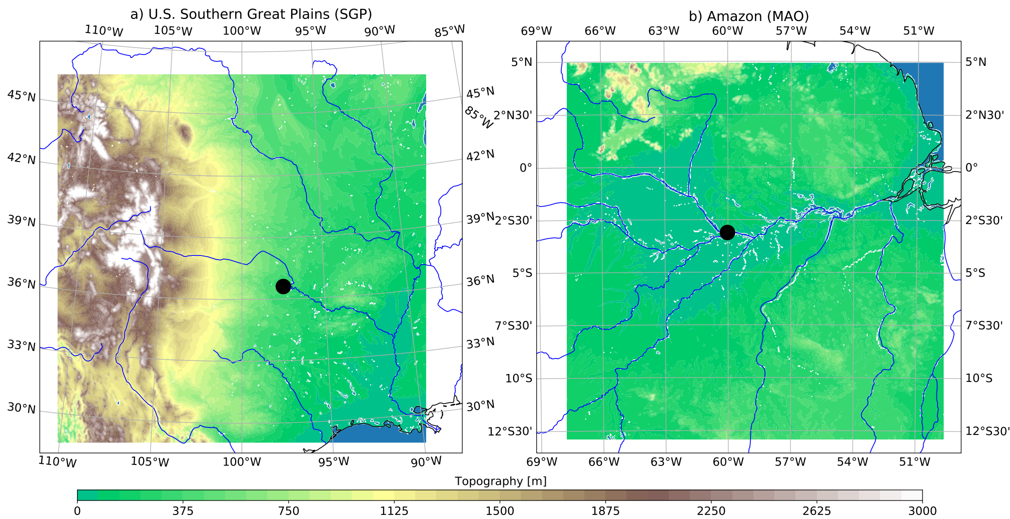


Figure 1. Computational 4 km WRF domains for simulating MCS cases in the U.S. Southern Great Plains (SGP; a) and Amazon basin (MAO; b). The colored contours show the model topography. Each domain consists of 500×500×96 grid cells. The black circle shows the location of the ARM SGP and MAO site.

Finding model setups that work well in different environments is critical for global kilometer-scale modeling efforts.

- SGP MCS simulations are most sensitive to changes in the microphysics, in agreement with previous studies (Stephan & Alexander, 2014). Amazon MCSs are more sensitive to the PBL scheme formulation. However, MCS case-to-case variability is the largest source of variability in our model performance evaluation, highlighting the necessity of an ensemble-based approach for model evaluation.
- There is little correlation between the model's performance in simulating Z and cloud BTs. This indicates that model physics are potentially tuned to capture one or the other, but have difficulties capturing both fields simultaneously in a physically sound way.

Future studies should focus on the simulation of the 3D structure of Z in kilometer-scale simulations to better understand potential biases in the vertical structure of MCSs due to deficiencies in simulating convective properties (e.g., up-and down-drafts as shown in Wang et al. (2020)) at 4 km grid spacing in combination with biases in the model physics. A high-priority research area is to better understand the ability of kilometer-scale models to simulate oceanic MCSs, particularly over the tropics due to their importance for the global water and energy cycle.

The results from this study will inform the model setup of additional MCS simulations in the U.S. Southern Great Plains and the Amazon basin at different horizontal grid spacings. These simulations will assess the bulk and structural convergence of MCS characteristics in these two environments and will help to improve the representation of MCSs in weather and climate models.

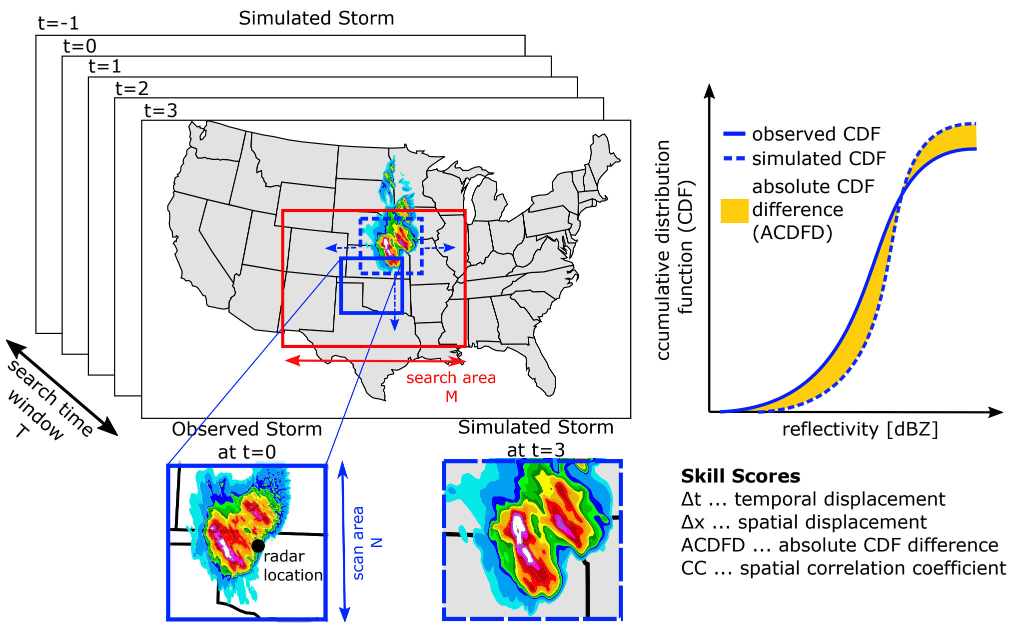


Figure 2. Schematic showing the evaluation framework. An observed storm that passes over a target location (here the SGP ARM site) is defined at time $t = 0$ or t_0 (lower left corner) within a scan area (N; e.g. areal extend of a radar station). A search time window (T) and search area (M; red rectangle in the large map) are defined. Simulated data are derived within the search time window ($t_0 - T/2 \ldots t_0 + T/2$) and within the search area M. Two scores are calculated for all possible shifts of the scan area within the search area and every time step (($M - N) \times (M - N) \times T$ combinations for each score). The scores are the spatial correlation coefficient (CC) and the absolute cumulative distribution function difference (ACDFD; orange area in right figure).

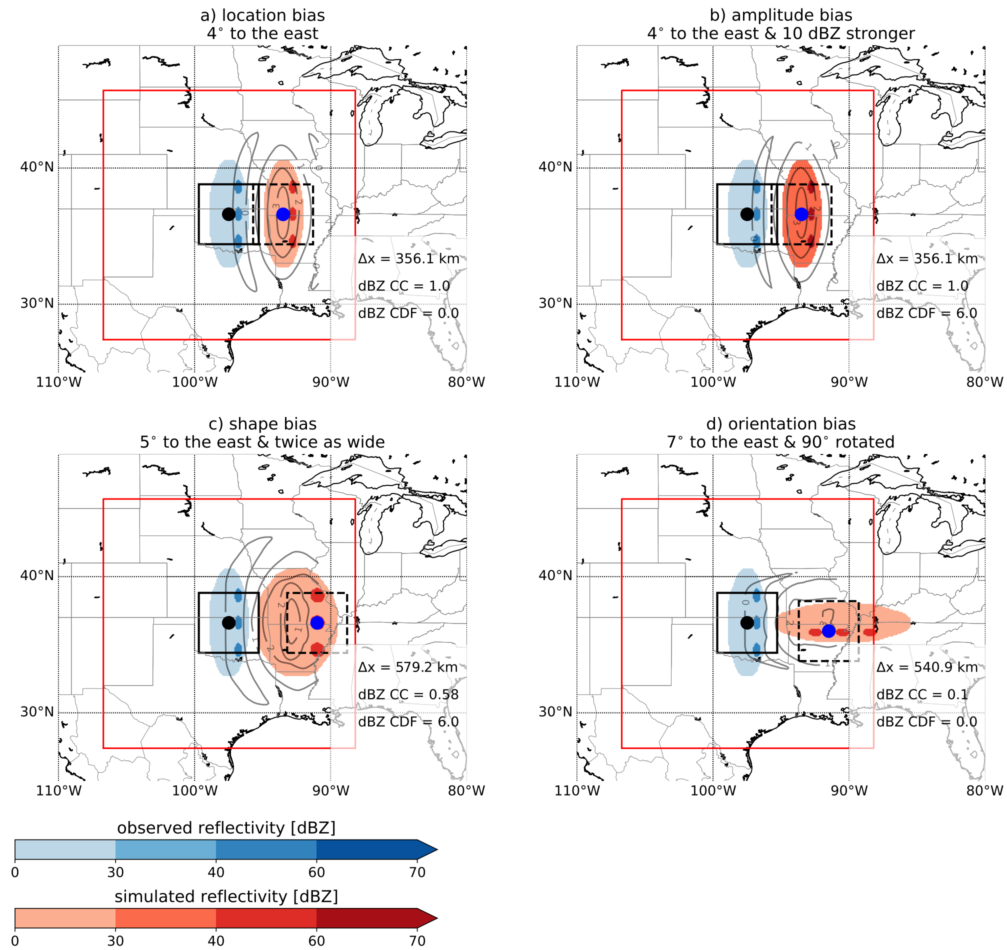


Figure 3. Ideal test case experiment showing the impact of displacements (a), intensity (b), shape (c), and rotational biases (d) on the spatial displacement (Δx), correlation coefficient (CC), and absolute cumulative distribution differences (CDF) skill score between an observed (blue) and simulated (red) storm system. The red rectangle shows the algorithm's search area (M=14 degrees), the hypothetical location of a radar site (here the ARM SGP site), the approximate reach of an S-Band radar (black) rectangle (4.4 degrees), displacement matrix (gray contour lines), and the optimal displacement location (blue circle and black dashed rectangle).

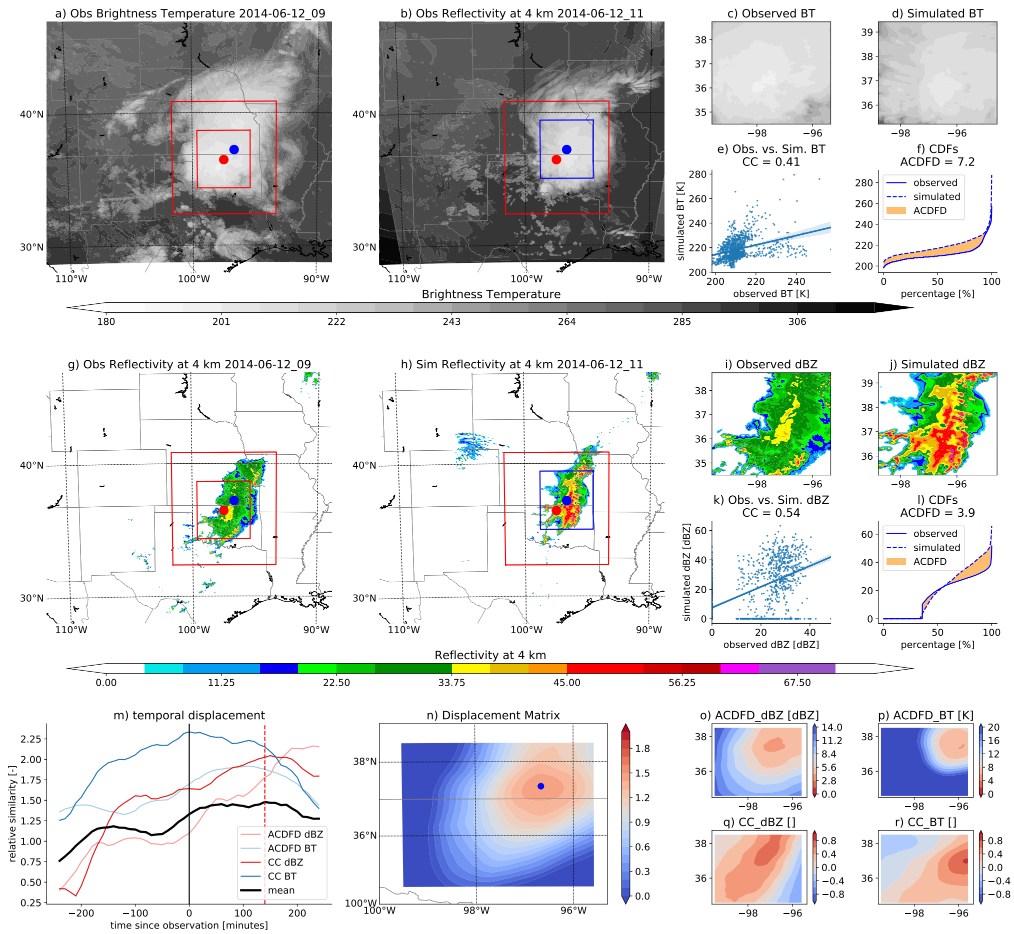


Figure 4. Observed brightness temperature (BT in K; a) and reflectively (Z in dBZ; g) field during an MCS overpass over the ARM SGP site (red dot in a,b,g,h) on June 12, 2014, 9 UTC. Simulated BT (b) and Z (h) field that is most similar to the observed fields using the Thompson microphysics and the YSU PBL scheme. a,b,g,h) The large red rectangle shows the search area and the small red rectangle the scan area. The blue rectangle in b,h shows the most similar simulated area compared to the observed scan area with the blue dot indicating the best estimate for the displacement error. Additionally shown is a zoomed-in version of the observed scan area and the most similar simulated BT (c,d) and Z (i,j). The scatter plot and Spearman correlation coefficient (CC) are shown for BT (c) and Z (k). Only every tenth point is shown in the scatter plot to improve visibility. The absolute commutative distribution differences (ACDFD; orange area) are shown for BT (f) and Z (l). m) Maxima of the normalized spatial fields of Z ACDFD (light red), BT ACDFD (light blue), Z CC (dark red), and BT CC (dark blue). The maxima of the averaged normalized spatial field of these four components is shown as a thick black line and the time displacement (peak value) of the simulated optimal field (120-minutes too late) is indicated with a red dashed line. n) Displacement matrix with the optimal simulated location shown as a blue circle and the four components of the displacement matrix including o) Z ACDFD, p) BT ACDFD, q) Z CC, and r) BT CC during the optimal displacement time.

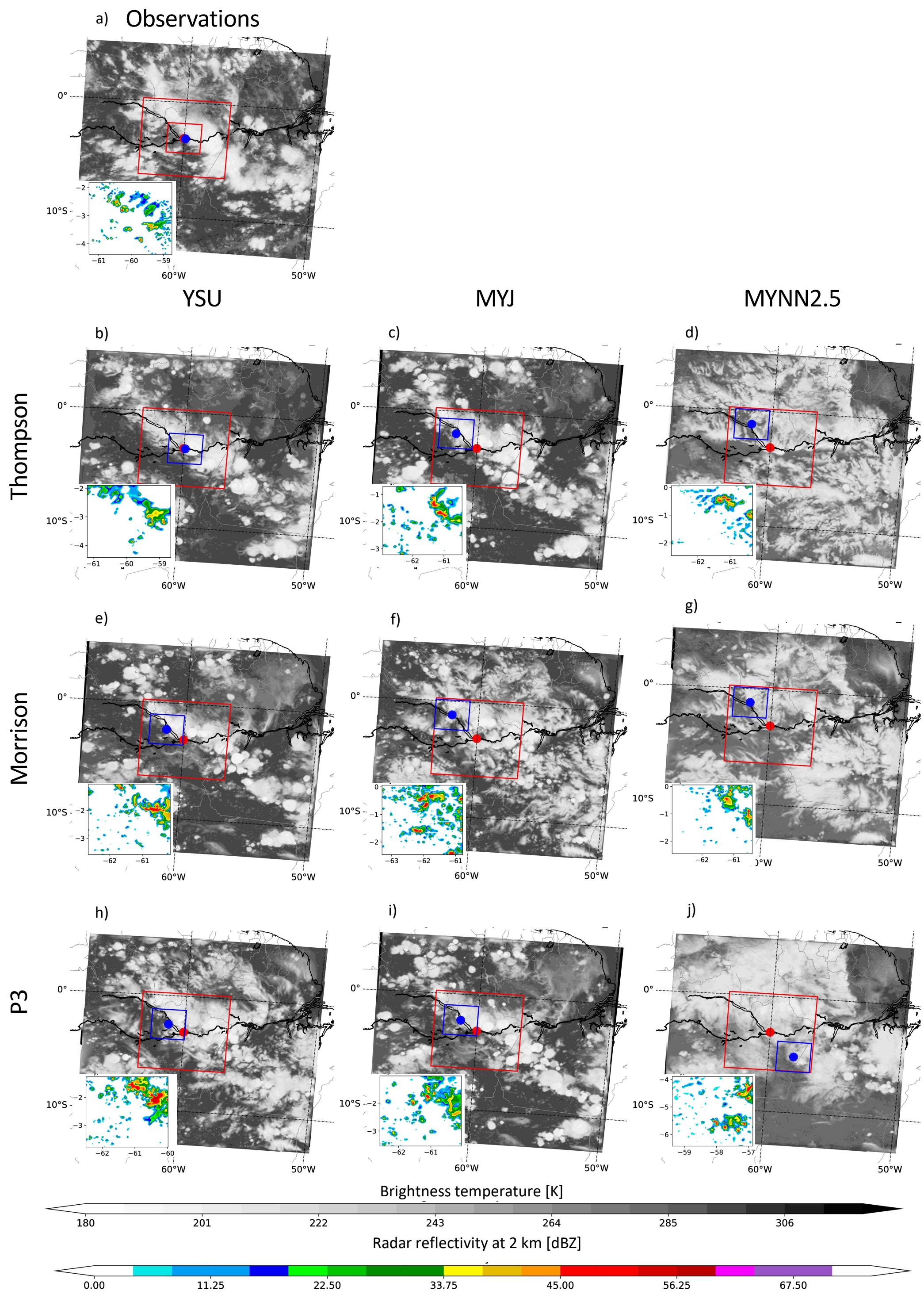


Figure 5. Example of the observed (a) and simulated (b–j) BT (gray contorts) and radar reflectively (inlet in lower left) at 2 km for the Nov. 17, 2014, MCS case in the Amazon. Results using the Thompson, Morrison, and P3 microphysics scheme are shown top down and YSU, MYJ, and MYNN2.5 planetary boundary layer scheme from left to right.

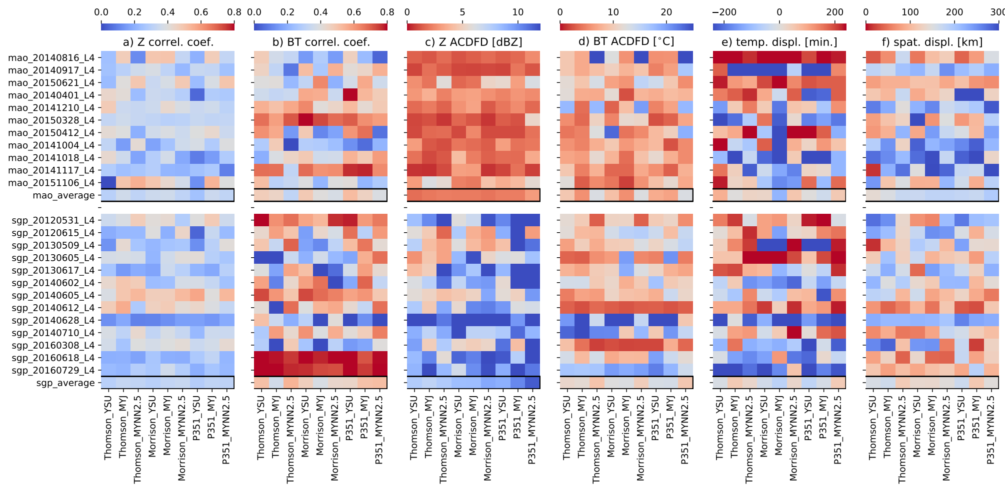


Figure 6. Heatmap of the reflectivity (Z) correlation coefficient (a), BT correlation coefficient (b), Z ACDFD (c), BT ACDFD (d), temporal displacement (e), and spatial displacement (f). Each panel shows results from the 9 different physics perturbations (columsn) and case experiments (rows) in the Amazon (mao; top block) and U.S. central Great Plains (sgp; bottom block). The average performance (mean over all cases) is highlighted in a black rectangle. Shown are results for a 2 km radar scanning heights.

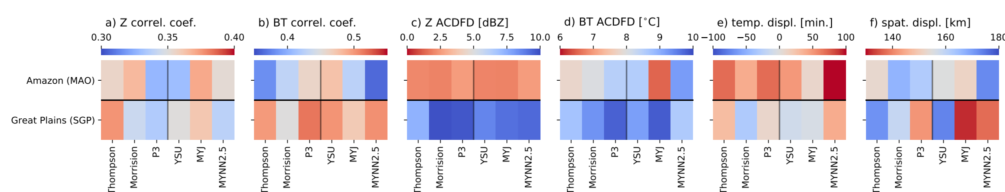


Figure 7. Average skill scores over all cases and reflectivity heights in Fig. 6. Note that the color bar ranges are narrower compared to Fig. 6.

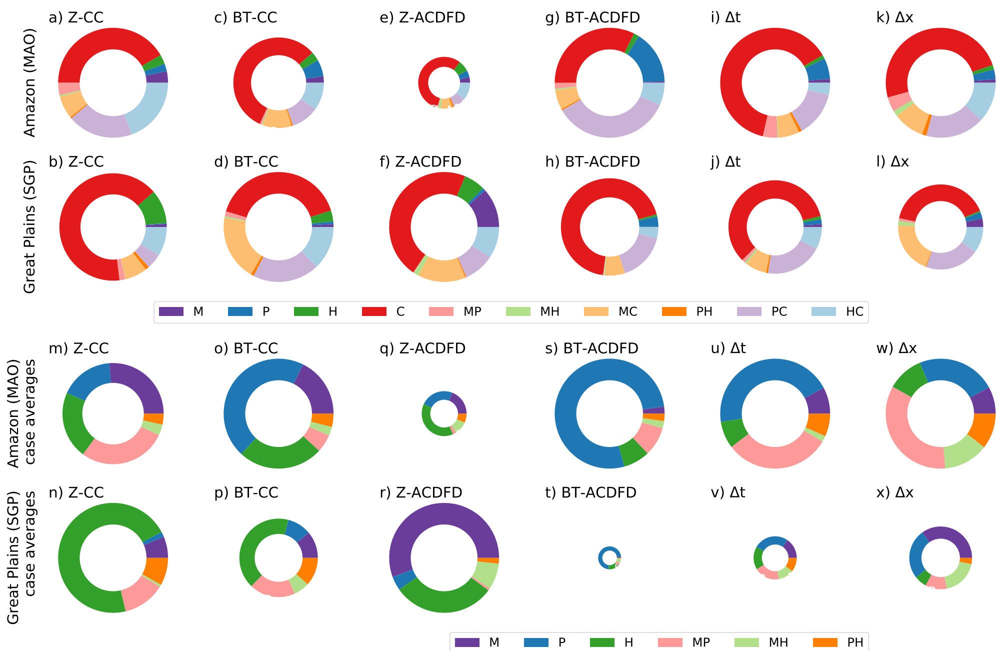


Figure 8. Variance decomposition showing the contribution of the PBL scheme (P), microphysics scheme (M), radar scanning height (H), case-to-case variability (C), and mixed terms (e.g., PM denotes the variability from the PBL and microphysics combination) to the total variance in the Z and BT correlation coefficient (CC), Z and BT absolute cumulative distribution function difference (ACDFD), temporal (Δt), and spatial displacement (Δx); from left to right. Results are shown including case-to-case variability (top two rows) and for the variability averaged over cases (bottom two rows). For each of these two options results from the Great Plains (SGP, top) and Amazon (MAO, bottom) simulations are shown. Circle sizes indicate the total amount of variability relative to the region with the larger variability (i.e., smaller circles show smaller total variability).

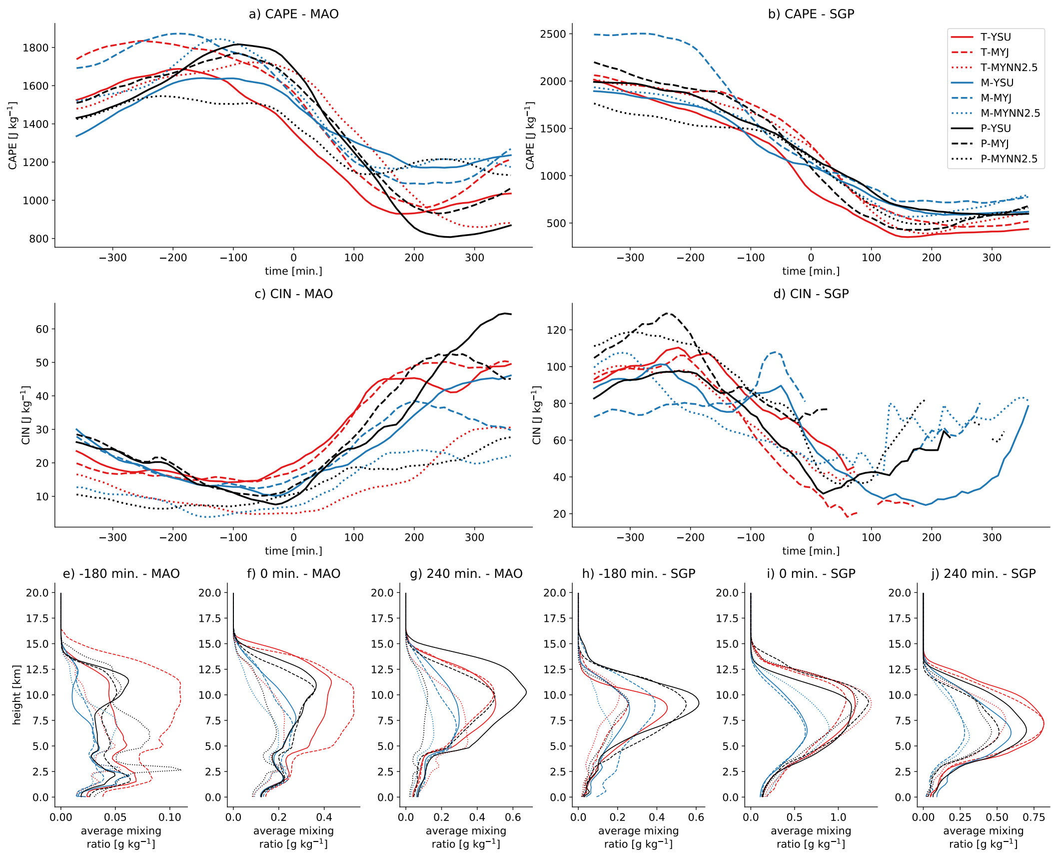


Figure 9. Temporal development of CAPE (a, b), CIN (c, d), and average cloud hydromether mixing ratios (e–j) for MAO (left column) and SGP MCSs (right column). Shown are the average values over all cases in each region at the optimal location (±15 grid cells) and relative to the optimal time in the simulations. Panels e–j show the average cloud (frozen and liquid) hydrometeor mixing rations within the ±15 grid cells region arround the optimal location 180 minutes before, at, and 240 minutes after the optimal time.

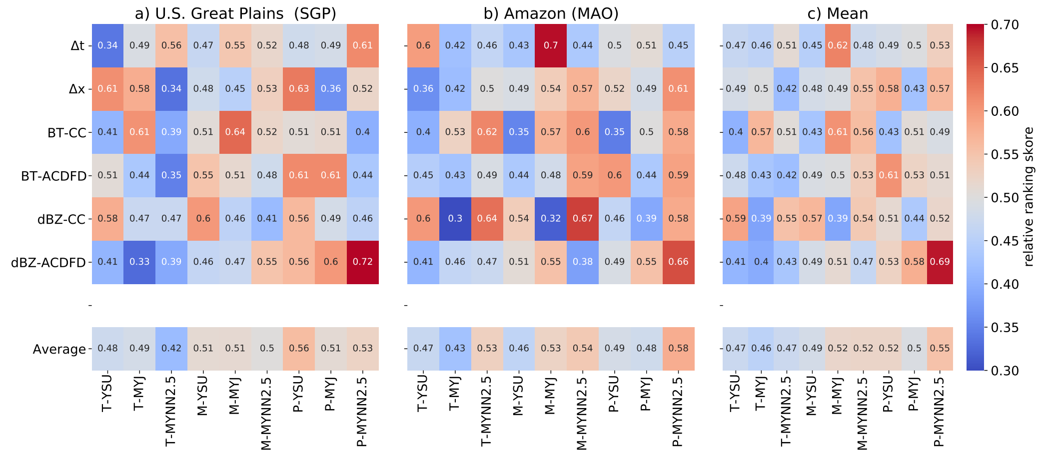


Figure 10. Average rank skill scores comparing the performance of the tested physics options (columns in each panel) over the SGP (a) and MAO (b) site including the mean scores of both locations (c) concerning temporal displacement (Δt), spatial displacement (Δx) BT correlation coefficients (BT-CC), BT absolute cumulative distribution function differences (BT-ACDFD), reflectivity (Z) correlation coefficient (Z-CC), and Z absolute cumulative distribution function differences (Z-ACDFD). The bottom row in each panel shows the mean score averaged over all metrics.

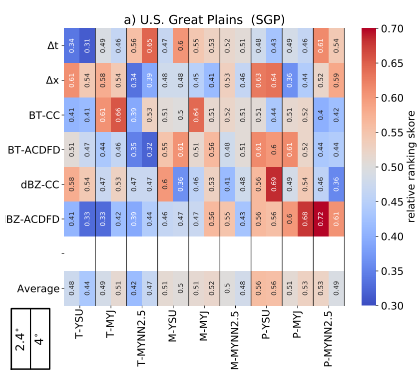


Figure 11. Similar to Fig. 10 but showing the impact of increasing the scan area from 2.4° (left sub-columns) to 4° (right sub-columns) for U.S. Great Plains cases. The radar scan height is 2 km.

Table 1. Selected MCS cases in the U.S. central Great Plains (SGP) and Amazon basin (MAO). The date and time indicates the overpass of the MCS over the corresponding ARM site as defined in Wang et al. (2019).

Region	Date & Time [UTC]	Season	Morphology
SGP	2012.05.31 04:00	spring	squall line
SGP	2012.06.15 07:00	spring	squall line
SGP	2013.05.09 07:00	spring	squall line
SGP	2013.06.05 09:00	spring	bow echo
SGP	2013.06.17 07:00	spring	squall line
SGP	2014.06.02 04:00	spring	squall line
SGP	2014.06.05 12:00	spring	bow echo
SGP	2014.06.12 06:00	spring	squall line
SGP	2014.06.28 16:00	spring	weakly organized
SGP	2014.07.10 10:00	summer	training line
SGP	2016.03.08 15:00	spring	weak squall line
SGP	2016.06.18 10:00	spring	mesoscale convective complex
SGP	2016.07.29 09:00	summer	squall line
MAO	2014.08.16 14:00	dry	local, small system
MAO	2014.09.17 17:00	dry	squall line
MAO	2015.06.21 14:00	dry	squall line
MAO	2014.04.01 15:00	wet	training line
MAO	2014.12.10 14:00	wet	local, weakly organized
MAO	2015.03.28 15:00	wet	local, weakly organized
MAO	2015.04.12 12:00	wet	squall line
MAO	2014.10.04 13:00	transition	squall line
MAO	2014.10.18 14:00	transition	local, weakly organized
MAO	2014.11.17 18:00	transition	squall line
MAO	2015.11.06 12:00	transition	squall line

Acknowledgments

This research was supported by the US Department of Energy Atmospheric System Research, an Office of Science Biological and Environmental Research Program (grant no. DE-SC0020050).S. Giangrande and D. Wang of Brookhaven Science Associates, LLC, are supported under Contract DE-SC0012704 with the U.S. DOE. Data were obtained from the Atmospheric Radiation Measurement (ARM) user facility, a U.S. Department of Energy (DOE) Office of Science user facility managed by the Biological and Environmental Research Program. We would like to acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation under Cooperative Agreement No. 1852977. This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. ERA-5 reanalysis data can be accessed from the Copernicus Climate Data Store (`https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview`). GOES13 and GOES15 satellite brightness temperature observations can be accessed from `https://www.ncei.noaa.gov/data/geostationary-ir-channel-brightness-temperature-gridsat-b1`. GRIDRAD data can be downloaded from `https://rda.ucar.edu/datasets/ds841.0/`. Gridded GoAmazon2014/15 campaign SIPAM S-band radar data can be accessed from ARM's data archive `https://www.arm.gov/data/`. The WRF model is open source and can be downloade from `https://github.com/wrf-model/WRF`. The model simulations can be accessed upon request from the corresponding author. The code that was used to analyze and visualize the presented data can be accessed at `https://github.com/AndreasPrein/Unified_Setup_to_Simulate_Mid-Latitude_and_Tropical_MCSs`.

References

Ansari, S., Del Greco, S., Kearns, E., Brown, O., Wilkins, S., Ramamurthy, M., ... others (2018). Unlocking the potential of NEXRAD data through NOAA's Big Data Partnership. *Bulletin of the American Meteorological Society*, *99*(1), 189–204.

Anselmo, E. M., Machado, L. A., Schumacher, C., & Kiladis, G. N. (2021). Amazonian mesoscale convective systems: Life cycle and propagation characteristics. *International Journal of Climatology*.

Berthou, S., Kendon, E. J., Chan, S. C., Ban, N., Leutwyler, D., Schär, C., & Fosser, G. (2020). Pan-European climate at convection-permitting scale: a model intercomparison study. *Climate Dynamics*, *55*(1), 35–59.

Clark, P., Roberts, N., Lean, H., Ballard, S. P., & Charlton-Perez, C. (2016). Convection-permitting models: a step-change in rainfall forecasting. *Meteorological Applications*, *23*(2), 165–181.

Coniglio, M. C., Correia Jr, J., Marsh, P. T., & Kong, F. (2013). Verification of convection-allowing WRF model forecasts of the planetary boundary layer using sounding observations. *Weather and Forecasting*, *28*(3), 842–862.

Davis, C. A., Brown, B. G., Bullock, R., & Halley-Gotway, J. (2009). The method for object-based diagnostic evaluation (MODE) applied to numerical forecasts from the 2005 NSSL/SPC Spring Program. *Weather and Forecasting*, *24*(5), 1252–1267.

Déqué, M., Rowell, D., Lüthi, D., Giorgi, F., Christensen, J., Rockel, B., ... van den Hurk, B. (2007). An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. *Climatic Change*, *81*(1), 53–70.

Donner, L. J., O'Brien, T. A., Rieger, D., Vogel, B., & Cooke, W. F. (2016). Are atmospheric updrafts a key to unlocking climate forcing and sensitivity? *Atmospheric Chemistry and Physics*, *16*(20), 12983–12992.

Fan, J., Han, B., Varble, A., Morrison, H., North, K., Kollias, P., ... others (2017).

Cloud-resolving model intercomparison of an MC3E squall line case: Part I—Convective updrafts. *Journal of Geophysical Research: Atmospheres*, *122*(17), 9351–9378.

Feng, Z., Leung, L. R., Hagos, S., Houze, R. A., Burleyson, C. D., & Balaguru, K. (2016). More frequent intense and long-lived storms dominate the springtime trend in central US rainfall. *Nature communications*, *7*(1), 1–8.

Feng, Z., Leung, L. R., Houze Jr, R. A., Hagos, S., Hardin, J., Yang, Q., … Fan, J. (2018). Structure and evolution of mesoscale convective systems: Sensitivity to cloud microphysics in convection-permitting simulations over the United States. *Journal of Advances in Modeling Earth Systems*, *10*(7), 1470–1494.

Feng, Z., Leung, L. R., Liu, N., Wang, J., Houze Jr, R. A., Li, J., … Guo, J. (2021). A Global High-Resolution Mesoscale Convective System Database Using Satellite-Derived Cloud Tops, Surface Precipitation, and Tracking. *Journal of Geophysical Research: Atmospheres*, *126*(8), e2020JD034202.

Feng, Z., Song, F., Sakaguchi, K., & Leung, L. R. (2021). Evaluation of mesoscale convective systems in climate simulations: Methodological development and results from MPAS-CAM over the United States. *Journal of Climate*, *34*(7), 2611–2633.

Fritsch, J., Kane, R., & Chelius, C. (1986). The contribution of mesoscale convective weather systems to the warm-season precipitation in the United States. *Journal of Applied Meteorology and Climatology*, *25*(10), 1333–1345.

Giangrande, S. E., Feng, Z., Jensen, M. P., Comstock, J. M., Johnson, K. L., Toto, T., … others (2017). Cloud characteristics, thermodynamic controls and radiative impacts during the Observations and Modeling of the Green Ocean Amazon (GoAmazon2014/5) experiment. *Atmospheric Chemistry and Physics*, *17*(23), 14519–14541.

Giangrande, S. E., Wang, D., & Mechem, D. B. (2020). Cloud regimes over the Amazon Basin: perspectives from the GoAmazon2014/5 campaign. *Atmospheric Chemistry and Physics*, *20*(12), 7489–7507.

Haberlie, A. M., & Ashley, W. S. (2019a). Climatological representation of mesoscale convective systems in a dynamically downscaled climate simulation. *International Journal of Climatology*, *39*(2), 1144–1153.

Haberlie, A. M., & Ashley, W. S. (2019b). A radar-based climatology of mesoscale convective systems in the United States. *Journal of Climate*, *32*(5), 1591–1606.

Han, B., Fan, J., Varble, A., Morrison, H., Williams, C. R., Chen, B., … others (2019). Cloud-resolving model intercomparison of an MC3E squall line case: Part II. Stratiform precipitation properties. *Journal of Geophysical Research: Atmospheres*, *124*(2), 1090–1117.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … others (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049.

Homeyer, C. R., & Bowman, K. P. (2017). *Algorithm description document for version 3.1 of the three-dimensional gridded NEXRAD WSR-88D Radar (GridRad) dataset* (Tech. Rep.). Retrieved from `http://gridrad.org/pdf/GridRad-v3.1-Algorithm-Description.pdf`

Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly weather review*, *134*(9), 2318–2341.

Hong, S.-Y., & Pan, H.-L. (1996). Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Monthly weather review*, *124*(10), 2322–2339.

Houze Jr, R. A. (2018). 100 years of research on mesoscale convective systems. *Meteorological Monographs*, *59*, 17–1.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Cal-

culations with the AER radiative transfer models. *Journal of Geophysical Research: Atmospheres*, *113*(D13).
Janjić, Z. I. (1994). The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Monthly weather review*, *122*(5), 927–945.
Jensen, A. A., Harrington, J. Y., Morrison, H., & Milbrandt, J. A. (2017). Predicting ice shape evolution in a bulk microphysics model. *Journal of the Atmospheric Sciences*, *74*(6), 2081–2104.
Knapp, K. R., & Wilkins, S. L. (2018). Gridded satellite (GridSat) GOES and CONUS data. *Earth System Science Data*, *10*(3), 1417–1425.
Li, J., Feng, Z., Qian, Y., & Leung, L. R. (2021). A high-resolution unified observational data product of mesoscale convective systems and isolated deep convection in the United States for 2004–2017. *Earth System Science Data*, *13*(2), 827–856.
Lin, G., Jones, C. R., Leung, L. R., Feng, Z., & Ovchinnikov, M. (2022). Mesoscale convective systems in a superparameterized E3SM simulation at high resolution. *Journal of Advances in Modeling Earth Systems*, *14*(1), e2021MS002660.
Liu, C., Ikeda, K., Rasmussen, R., Barlage, M., Newman, A. J., Prein, A. F., ... others (2017). Continental-scale convection-permitting modeling of the current and future climate of North America. *Climate Dynamics*, *49*(1), 71–95.
Martin, S. T., Artaxo, P., Machado, L., Manzi, A. O., Souza, R. d., Schumacher, C., ... others (2017). The Green Ocean Amazon experiment (GoAmazon2014/5) observes pollution affecting gases, aerosols, clouds, and rainfall over the rain forest. *Bulletin of the American Meteorological Society*, *98*(5), 981–997.
Martin, S. T., Artaxo, P., Machado, L. A. T., Manzi, A. O., Souza, R. A. F. d., Schumacher, C., ... others (2016). Introduction: observations and modeling of the Green Ocean Amazon (GoAmazon2014/5). *Atmospheric Chemistry and Physics*, *16*(8), 4785–4797.
Mather, J. H., & Voyles, J. W. (2013). The ARM Climate Research Facility: A review of structure and capabilities. *Bulletin of the American Meteorological Society*, *94*(3), 377–392.
Mellor, G. L., & Yamada, T. (1982). Development of a turbulence closure model for geophysical fluid problems. *Reviews of Geophysics*, *20*(4), 851–875.
Mesinger, F. (1993). Forecasting upper tropospheric turbulence within the framework of the Mellor-Yamada 2.5 closure. *Res. Activ. Atmos. Oceanic Mod.*.
Morrison, H., & Milbrandt, J. A. (2015). Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part I: Scheme description and idealized tests. *Journal of Atmospheric Sciences*, *72*(1), 287–311.
Morrison, H., Thompson, G., & Tatarskii, V. (2009). Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one-and two-moment schemes. *Monthly weather review*, *137*(3), 991–1007.
Nakanishi, M., & Niino, H. (2006). An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Boundary-Layer Meteorology*, *119*(2), 397–407.
Nakanishi, M., & Niino, H. (2009). Development of an improved turbulence closure model for the atmospheric boundary layer. *Journal of the Meteorological Society of Japan. Ser. II*, *87*(5), 895–912.
NCEP. (2020). *UPP Users Guide V* (Tech. Rep.). National Centers for Environmental Prediction (NCEP). Retrieved from `https://dtcenter.org/community-code/unified-post-processor-upp/documentation`
Nesbitt, S. W., Cifelli, R., & Rutledge, S. A. (2006). Storm morphology and rainfall characteristics of TRMM precipitation features. *Monthly Weather Review*, *134*(10), 2702–2721.
Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., ... oth-

ers (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres*, *116*(D12).

Pokharel, B., Wang, S.-Y. S., Lin, Y.-H., Zhao, L., & Gillies, R. (2018). Diagnosing the atypical extreme precipitation events under weakly forced synoptic setting: the West Virginia flood (June 2016) and beyond. *Climate prediction S&T digest*, 8.

Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., ... others (2017). The weather research and forecasting model: Overview, system efforts, and future directions. *Bulletin of the American Meteorological Society*, *98*(8), 1717–1737.

Prein, A., Gobiet, A., Suklitsch, M., Truhetz, H., Awan, N., Keuler, K., & Georgievski, G. (2013). Added value of convection permitting seasonal simulations. *Climate Dynamics*, *41*(9-10), 2655–2677.

Prein, A. F., Gobiet, A., & Truhetz, H. (2011). Analysis of uncertainty in large scale climate change projections over Europe. *Meteorologische Zeitschrift*, *20*(4), 383.

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., ... others (2015). A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of geophysics*, *53*(2), 323–361.

Prein, A. F., Liu, C., Ikeda, K., Bullock, R., Rasmussen, R. M., Holland, G. J., & Clark, M. (2020). Simulating North American mesoscale convective systems with a convection-permitting climate model. *Climate Dynamics*, *55*(1), 95–110.

Prein, A. F., Liu, C., Ikeda, K., Trier, S. B., Rasmussen, R. M., Holland, G. J., & Clark, M. P. (2017). Increased rainfall volume from future convective storms in the US. *Nature Climate Change*, *7*(12), 880–884.

Rasmussen, K. L., Zuluaga, M. D., & Houze Jr, R. A. (2014). Severe convection and lightning in subtropical South America. *Geophysical Research Letters*, *41*(20), 7359–7366.

Rehbein, A., Ambrizzi, T., & Mechoso, C. R. (2018). Mesoscale convective systems over the Amazon basin. Part I: climatological aspects. *International Journal of Climatology*, *38*(1), 215–229.

Rehbein, A., Ambrizzi, T., Mechoso, C. R., Espinosa, S. A., & Myers, T. A. (2019). Mesoscale convective systems over the Amazon basin: The GoAmazon2014/5 program. *International Journal of Climatology*, *39*(15), 5599–5618.

Roberts, N. M., & Lean, H. W. (2008). Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, *136*(1), 78–97.

Satoh, M., Stevens, B., Judt, F., Khairoutdinov, M., Lin, S.-J., Putman, W. M., & Düben, P. (2019). Global cloud-resolving models. *Current Climate Change Reports*, *5*(3), 172–184.

Schumacher, C., & Funk, A. (2018, 4). *GoAmazon2014/5 Three-dimensional Gridded S-band Reflectivity and Radial Velocity from the SIPAM Manaus S-band Radar* (Tech. Rep.). Oak Ridge National Lab. (ORNL).

Schumacher, R. S., & Johnson, R. H. (2005). Organization and environmental properties of extreme-rain-producing mesoscale convective systems. *Monthly weather review*, *133*(4), 961–976.

Shpund, J., Khain, A., Lynn, B., Fan, J., Han, B., Ryzhkov, A., ... Gill, D. (2019). Simulating a mesoscale convective system using WRF with a new spectral bin microphysics: 1: Hail vs graupel. *Journal of Geophysical Research: Atmospheres*, *124*(24), 14072–14101.

Sisterson, D., Peppler, R., Cress, T., Lamb, P., & Turner, D. (2016). The ARM southern great plains (SGP) site. *Meteorological Monographs*, *57*, 6–1.

Skamarock, W. C., & Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *Journal of computational physics*, *227*(7), 3465–3485.

Song, F., Feng, Z., Leung, L. R., Houze Jr, R. A., Wang, J., Hardin, J., & Homeyer, C. R. (2019). Contrasting spring and summer large-scale environments associated with mesoscale convective systems over the US Great Plains. *Journal of Climate*, *32*(20), 6749–6767.

Stephan, C., & Alexander, M. J. (2014). Summer season squall-line simulations: Sensitivity of gravity waves to physics parameterization and implications for their parameterization in global climate models. *Journal of the Atmospheric Sciences*, *71*(9), 3376–3391.

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., ... others (2019). DYAMOND: the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. *Progress in Earth and Planetary Science*, *6*(1), 1–17.

Thompson, G., Rasmussen, R. M., & Manning, K. (2004). Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis. *Monthly Weather Review*, *132*(2), 519–542.

Trier, S., Davis, C., & Ahijevych, D. (2010). Environmental controls on the simulated diurnal cycle of warm-season precipitation in the continental United States. *Journal of the atmospheric sciences*, *67*(4), 1066–1090.

Wang, D., Giangrande, S. E., Feng, Z., Hardin, J. C., & Prein, A. F. (2020). Updraft and downdraft core size and intensity as revealed by radar wind profilers: MCS observations and idealized model comparisons. *Journal of Geophysical Research: Atmospheres*, *125*(11), e2019JD031774.

Wang, D., Giangrande, S. E., Schiro, K. A., Jensen, M. P., & Houze Jr, R. A. (2019). The characteristics of tropical and midlatitude mesoscale convective systems as revealed by radar wind profilers. *Journal of Geophysical Research: Atmospheres*, *124*(8), 4601–4619.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.

Wu, X., & Yan, J. (2011). Estimating the outgoing longwave radiation from the FY-3B satellite visible infrared radiometer Channel 5 radiance observations. *Chinese Science Bulletin*, *56*(32), 3480–3485.

Xue, L., Fan, J., Lebo, Z. J., Wu, W., Morrison, H., Grabowski, W. W., ... others (2017). Idealized simulations of a squall line from the MC3E field campaign applying three bin microphysics schemes: Dynamic and thermodynamic structure. *Monthly Weather Review*, *145*(12), 4789–4812.

Yang, G.-Y., & Slingo, J. (2001). The diurnal cycle in the tropics. *Monthly Weather Review*, *129*(4), 784–801.

Figure 1.

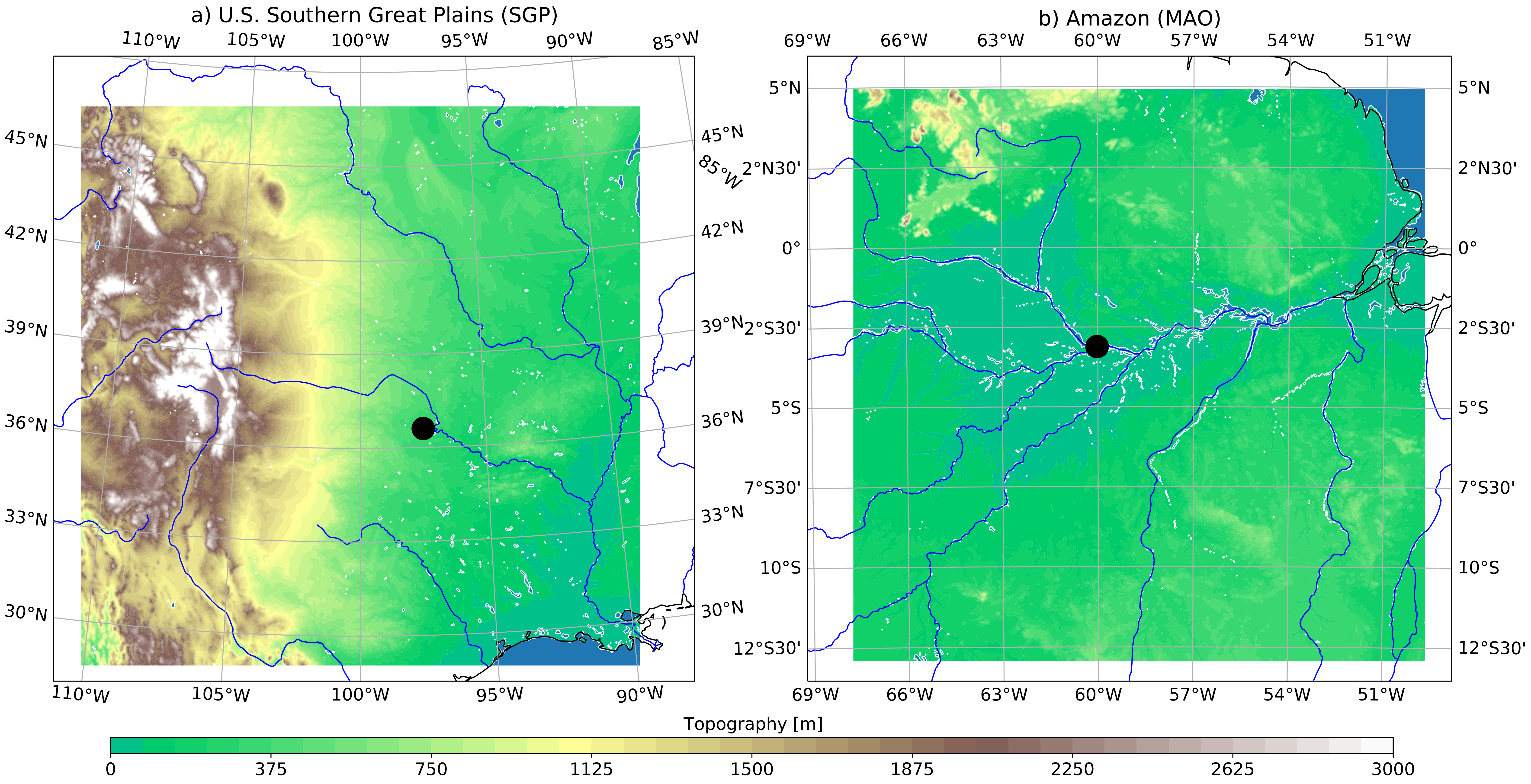
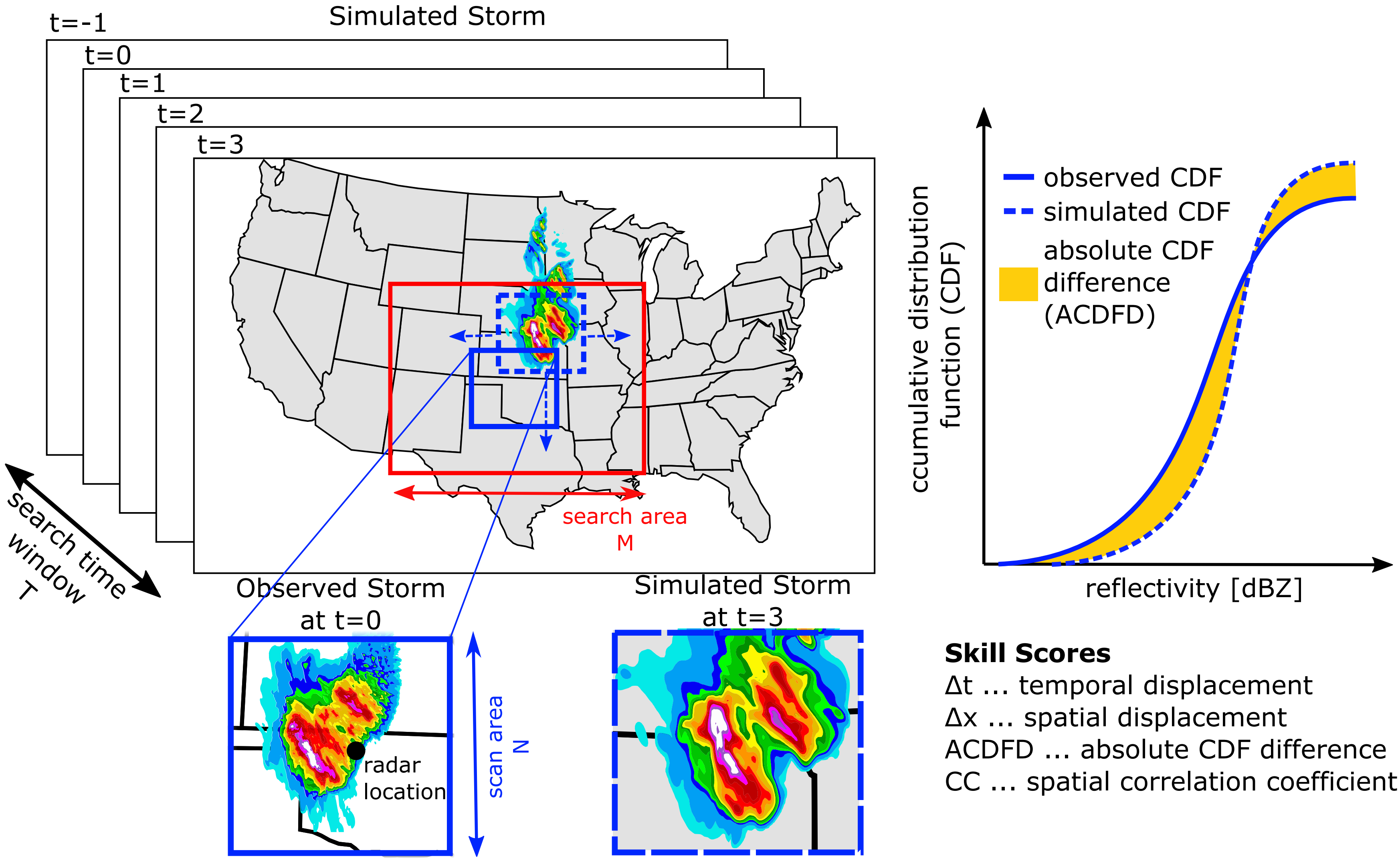
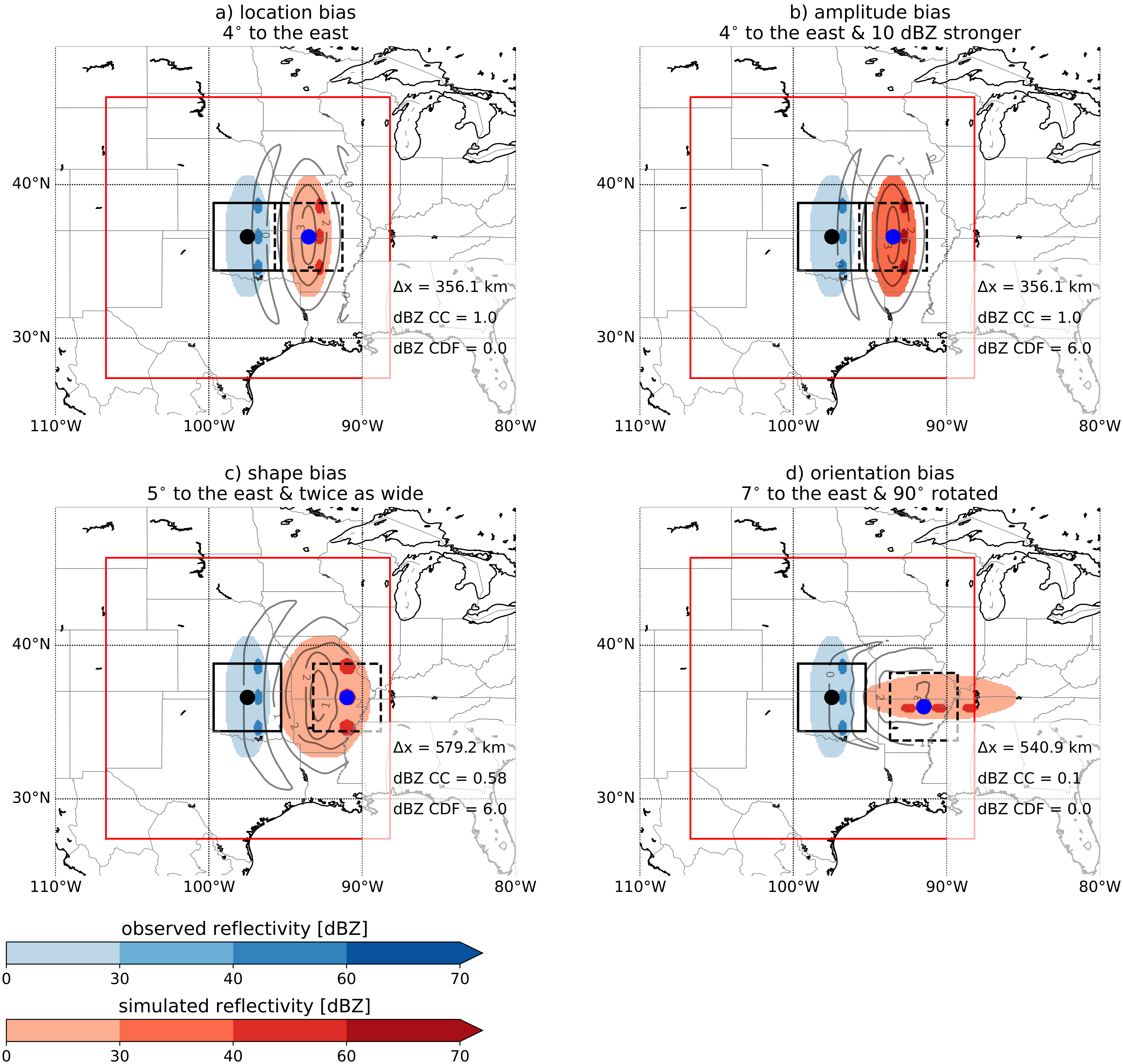

a) U.S. Southern Great Plains (SGP)
b) Amazon (MAO)
110°W
105°W
100°W
95°W
90°W
85°W
45°N
42°N
39°N
36°N
33°N
30°N
69°W
66°W
63°W
60°W
57°W
54°W
51°W
5°N
2°N30'
0°
2°S30'
5°S
7°S30'
10°S
12°S30'
Topography [m]
0
375
750
1125
1500
1875
2250
2625
3000

Figure 2.



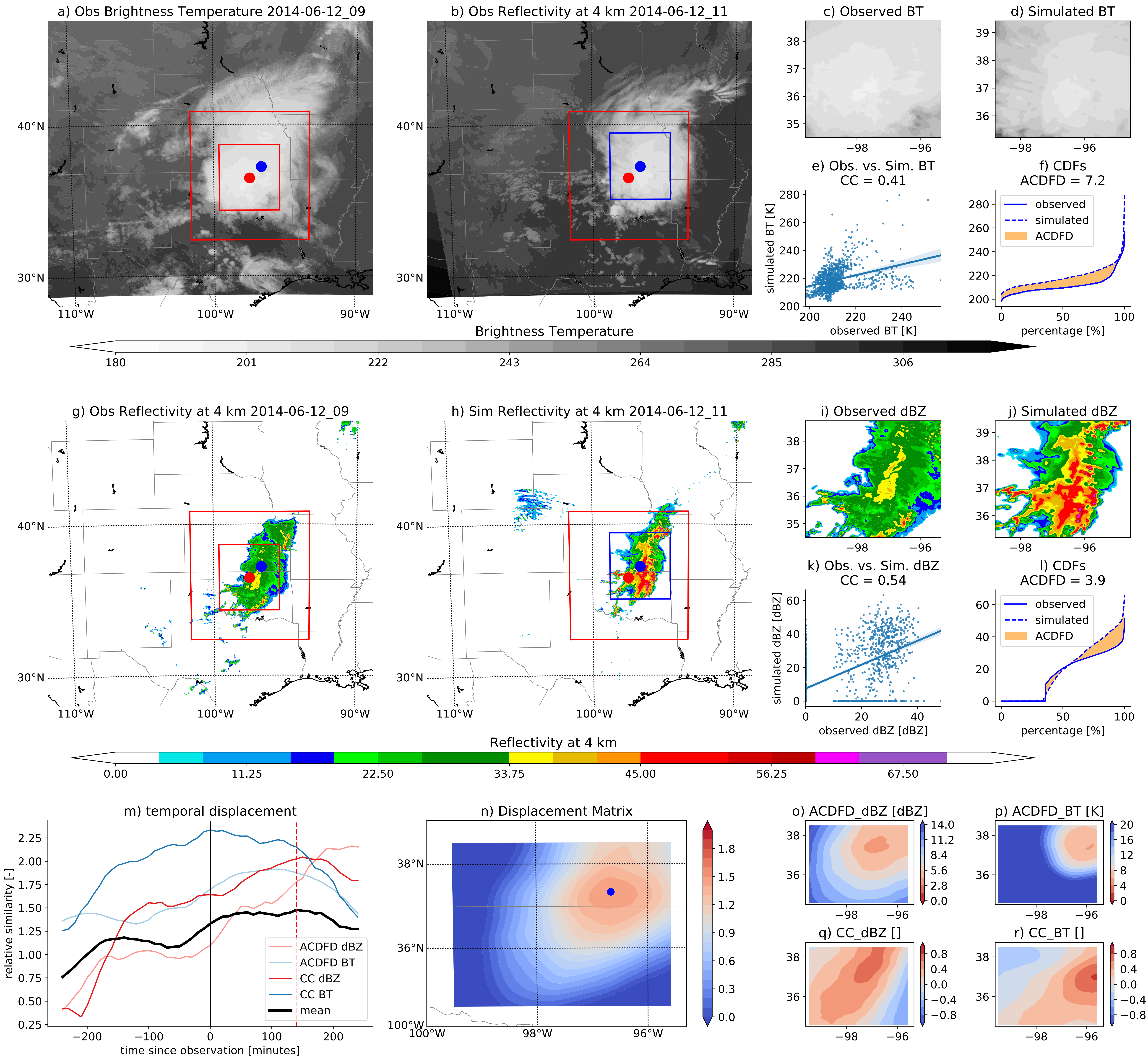
Simulated Storm
t=-1
t=0
t=1
t=2
t=3
search area
M
search time
window
T
Observed Storm
at t=0
radar
location
scan area
N
Simulated Storm
at t=3
cumulative distribution
function (CDF)
observed CDF
simulated CDF
absolute CDF
difference
(ACDFD)
reflectivity [dBZ]
Skill Scores
Δt ... temporal displacement
Δx ... spatial displacement
ACDFD ... absolute CDF difference
CC ... spatial correlation coefficient

Figure 3.



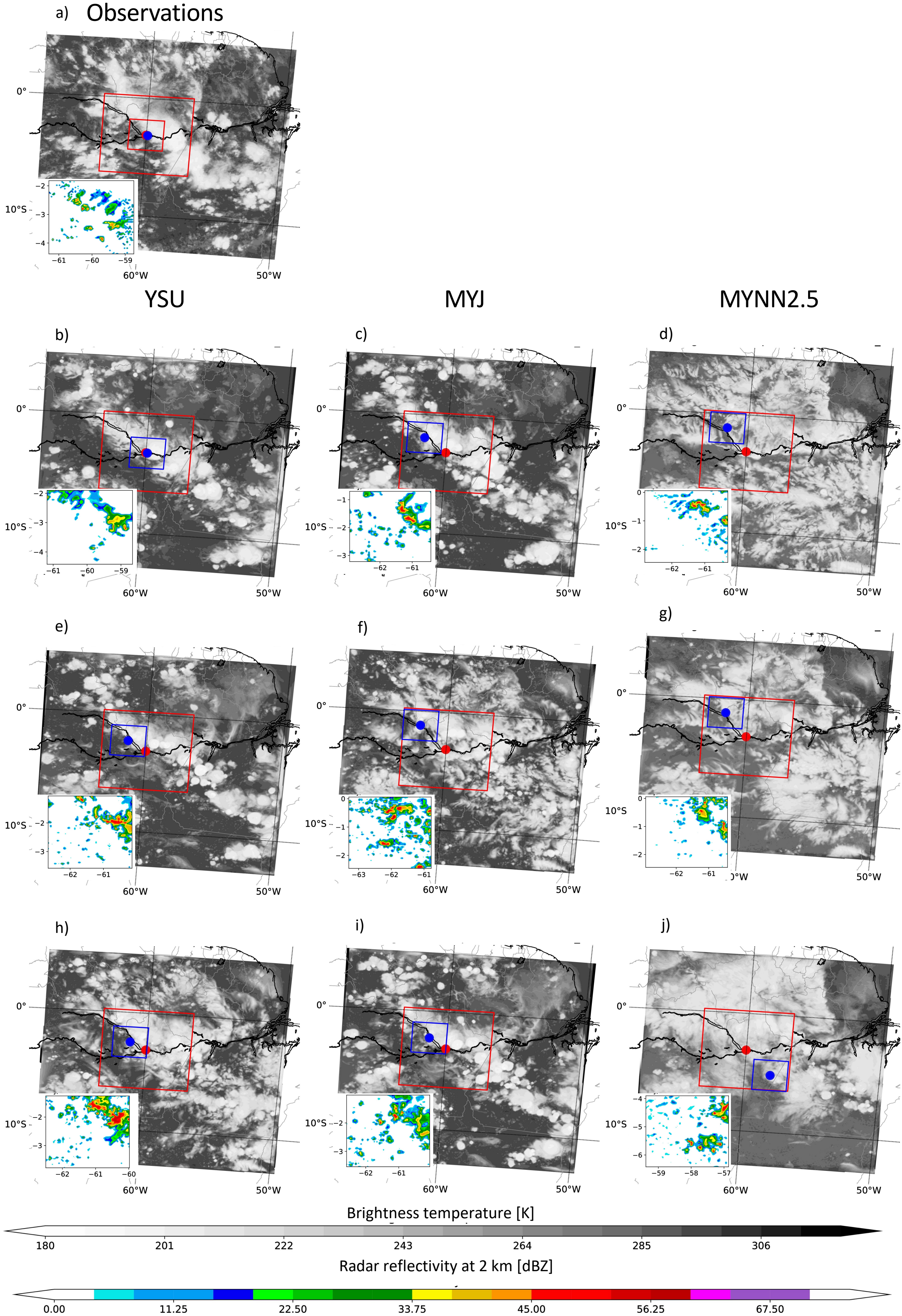
a) location bias
4° to the east
Δx = 356.1 km
dBZ CC = 1.0
dBZ CDF = 0.0
b) amplitude bias
4° to the east & 10 dBZ stronger
Δx = 356.1 km
dBZ CC = 1.0
dBZ CDF = 6.0
c) shape bias
5° to the east & twice as wide
Δx = 579.2 km
dBZ CC = 0.58
dBZ CDF = 6.0
d) orientation bias
7° to the east & 90° rotated
Δx = 540.9 km
dBZ CC = 0.1
dBZ CDF = 0.0
40°N
30°N
110°W
100°W
90°W
80°W
observed reflectivity [dBZ]
0 30 40 60 70
simulated reflectivity [dBZ]
0 30 40 60 70

Figure 4.



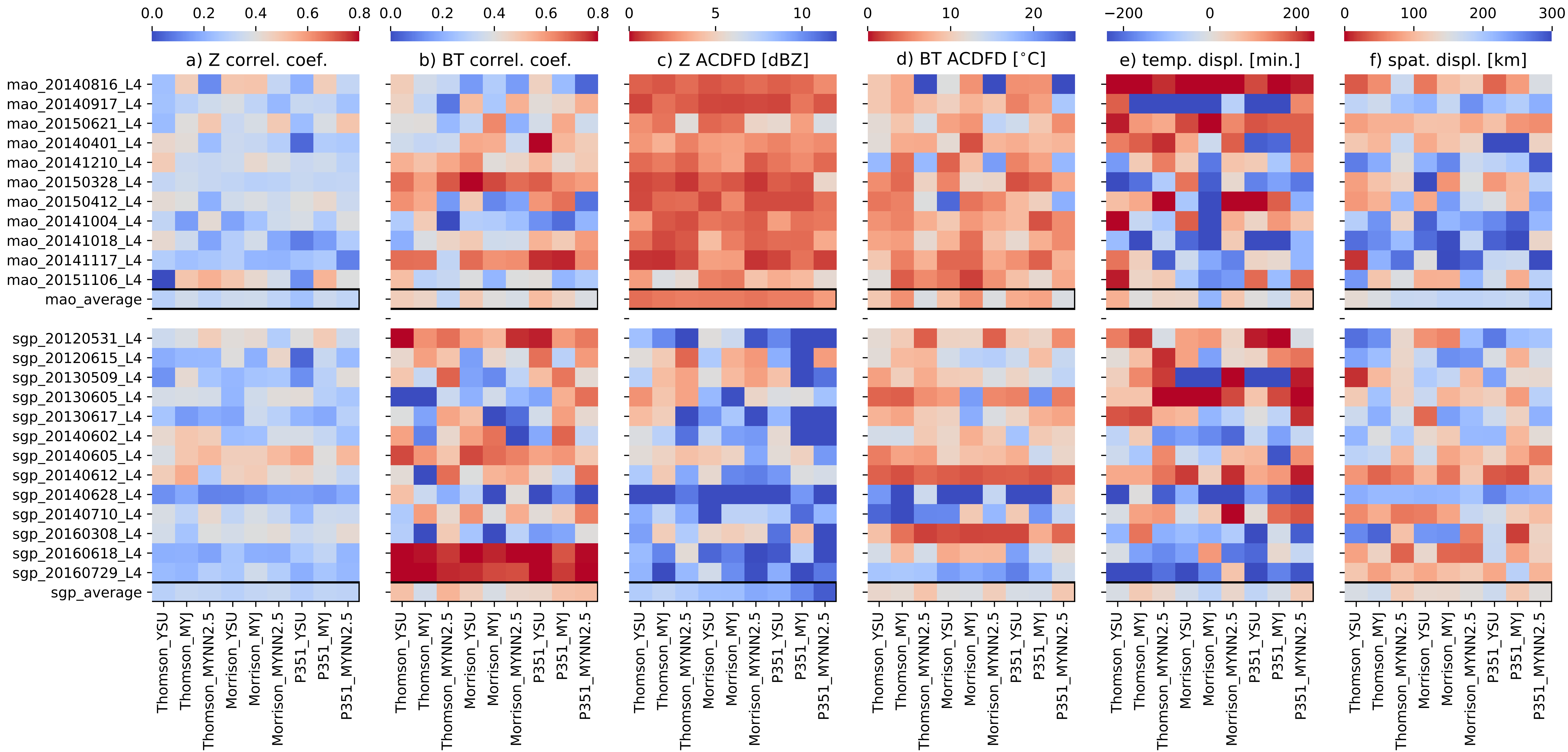
a) Obs Brightness Temperature 2014-06-12_09
b) Obs Reflectivity at 4 km 2014-06-12_11
c) Observed BT
d) Simulated BT
e) Obs. vs. Sim. BT
CC = 0.41
f) CDFs
ACDFD = 7.2
observed
simulated
ACDFD
simulated BT [K]
observed BT [K]
percentage [%]
Brightness Temperature
g) Obs Reflectivity at 4 km 2014-06-12_09
h) Sim Reflectivity at 4 km 2014-06-12_11
i) Observed dBZ
j) Simulated dBZ
k) Obs. vs. Sim. dBZ
CC = 0.54
l) CDFs
ACDFD = 3.9
simulated dBZ [dBZ]
observed dBZ [dBZ]
Reflectivity at 4 km
m) temporal displacement
n) Displacement Matrix
o) ACDFD_dBZ [dBZ]
p) ACDFD_BT [K]
q) CC_dBZ []
r) CC_BT []
relative similarity [-]
time since observation [minutes]
ACDFD dBZ
ACDFD BT
CC dBZ
CC BT
mean

Figure 5.



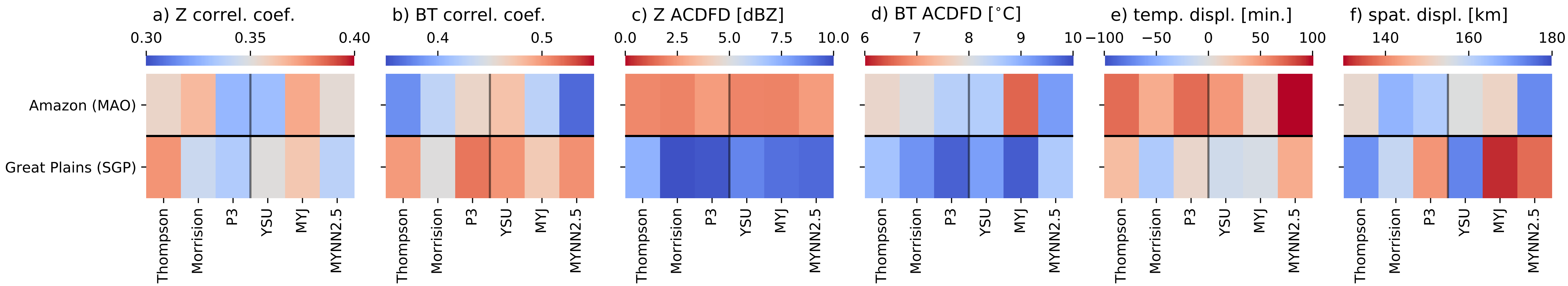
a) Observations
YSU
MYJ
MYNN2.5
Thompson
Morrison
P3
b)
c)
d)
e)
f)
g)
h)
i)
j)
0°
10°S
60°W
50°W
Brightness temperature [K]
180
201
222
243
264
285
306
Radar reflectivity at 2 km [dBZ]
0.00
11.25
22.50
33.75
45.00
56.25
67.50

Figure 6.



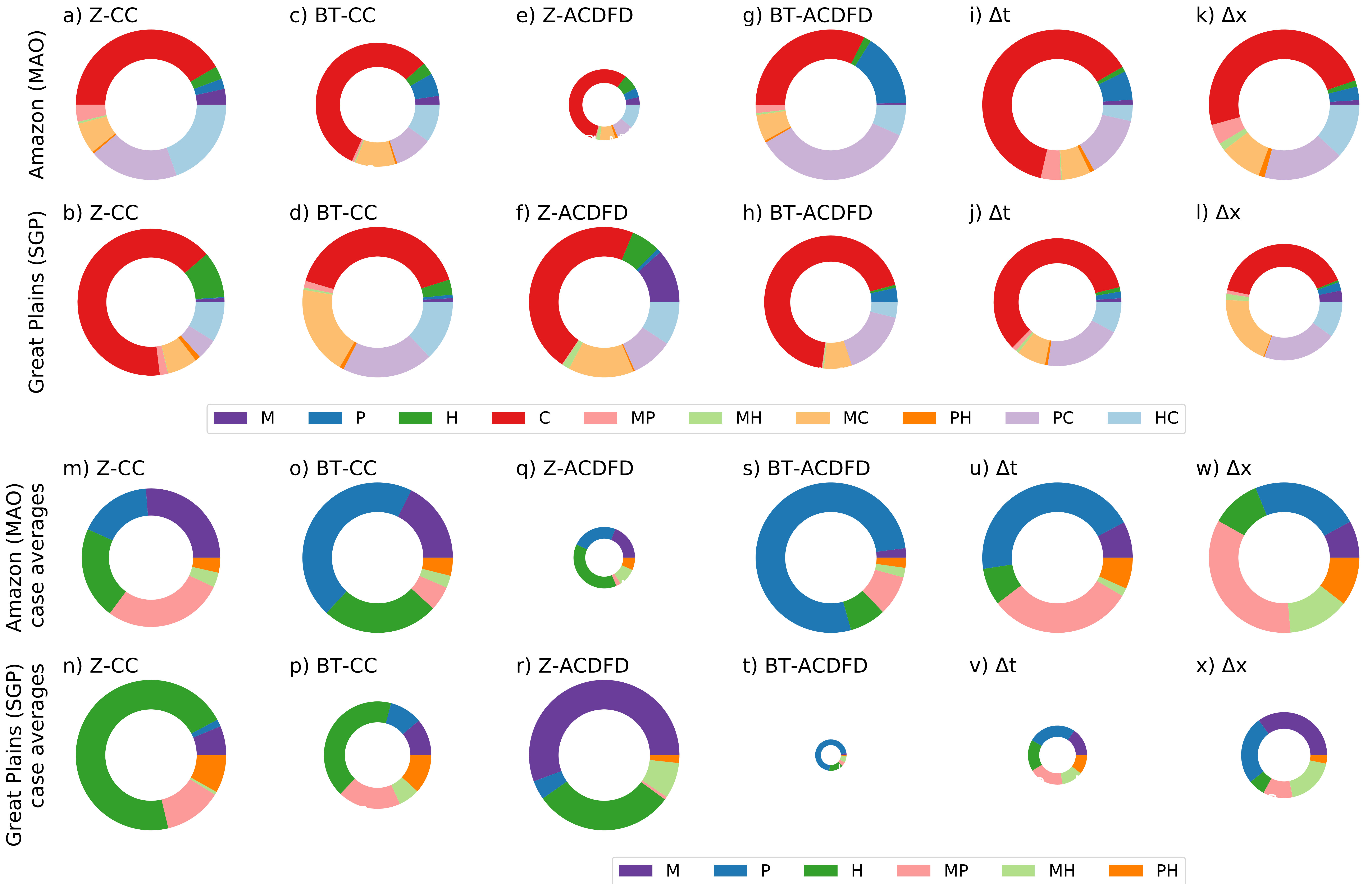
a) Z correl. coef.
b) BT correl. coef.
c) Z ACDFD [dBZ]
d) BT ACDFD [°C]
e) temp. displ. [min.]
f) spat. displ. [km]
mao_20140816_L4
mao_20140917_L4
mao_20150621_L4
mao_20140401_L4
mao_20141210_L4
mao_20150328_L4
mao_20150412_L4
mao_20141004_L4
mao_20141018_L4
mao_20141117_L4
mao_20151106_L4
mao_average
sgp_20120531_L4
sgp_20120615_L4
sgp_20130509_L4
sgp_20130605_L4
sgp_20130617_L4
sgp_20140602_L4
sgp_20140605_L4
sgp_20140612_L4
sgp_20140628_L4
sgp_20140710_L4
sgp_20160308_L4
sgp_20160618_L4
sgp_20160729_L4
sgp_average
Thomson_YSU
Thomson_MYJ
Thomson_MYNN2.5
Morrison_YSU
Morrison_MYJ
Morrison_MYNN2.5
P351_YSU
P351_MYJ
P351_MYNN2.5

Figure 7.



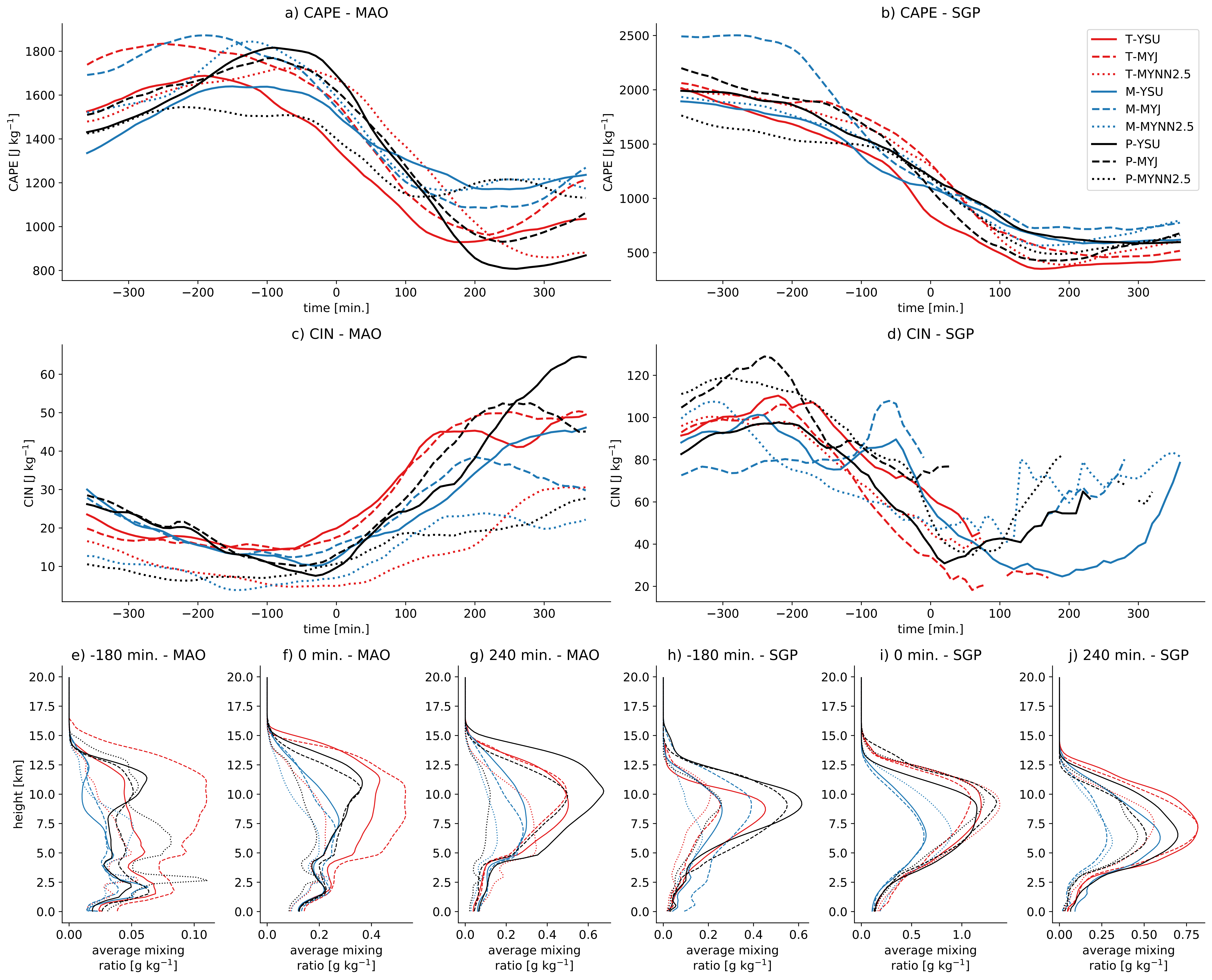
a) Z correl. coef.
0.30
0.35
0.40
b) BT correl. coef.
0.4
0.5
c) Z ACDFD [dBZ]
0.0
2.5
5.0
7.5
10.0
d) BT ACDFD [°C]
6
7
8
9
10
e) temp. displ. [min.]
−100
−50
0
50
100
f) spat. displ. [km]
140
160
180
Amazon (MAO)
Great Plains (SGP)
Thompson
Morrison
P3
YSU
MYJ
MYNN2.5

Figure 8.



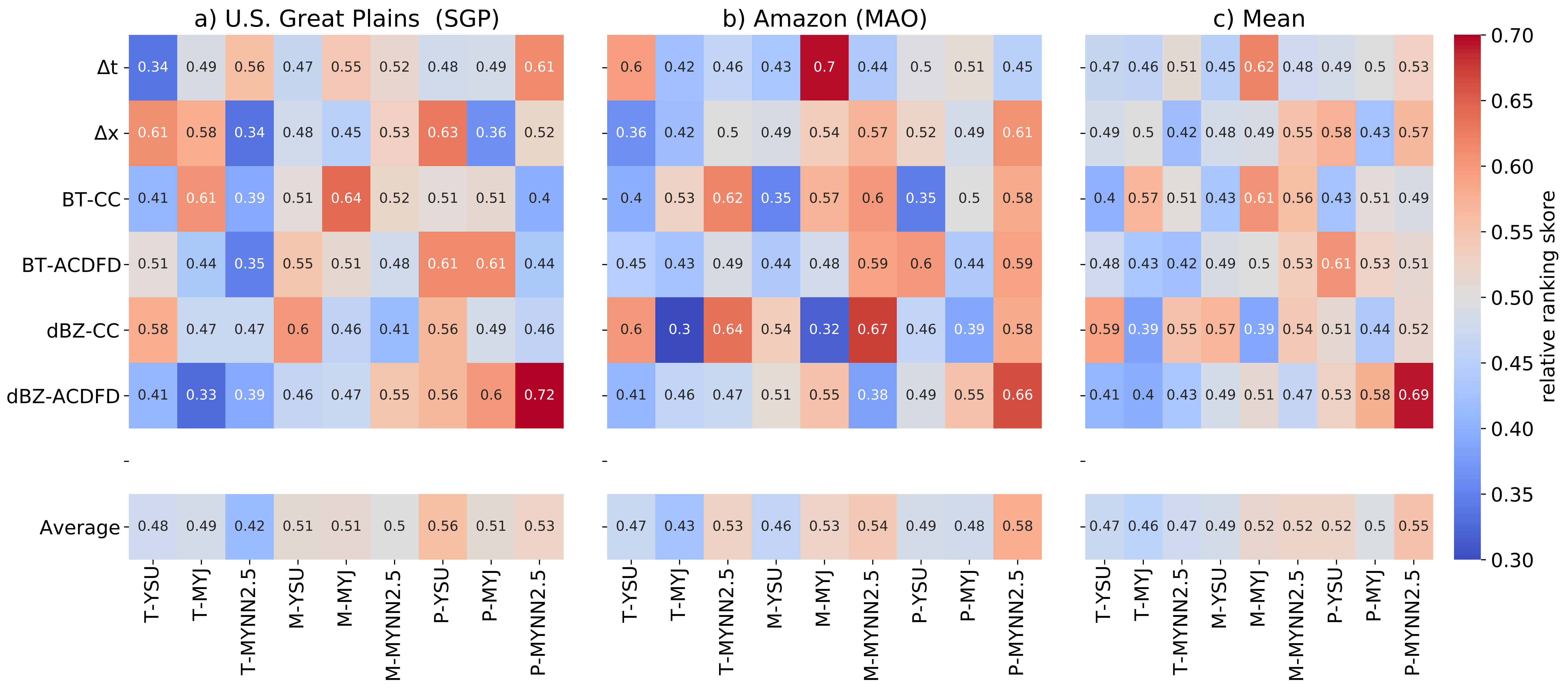
a) Z-CC
c) BT-CC
e) Z-ACDFD
g) BT-ACDFD
i) Δt
k) Δx
Amazon (MAO)
b) Z-CC
d) BT-CC
f) Z-ACDFD
h) BT-ACDFD
j) Δt
l) Δx
Great Plains (SGP)
M
P
H
C
MP
MH
MC
PH
PC
HC
m) Z-CC
o) BT-CC
q) Z-ACDFD
s) BT-ACDFD
u) Δt
w) Δx
Amazon (MAO) case averages
n) Z-CC
p) BT-CC
r) Z-ACDFD
t) BT-ACDFD
v) Δt
x) Δx
Great Plains (SGP) case averages
M
P
H
MP
MH
PH

Figure 9.



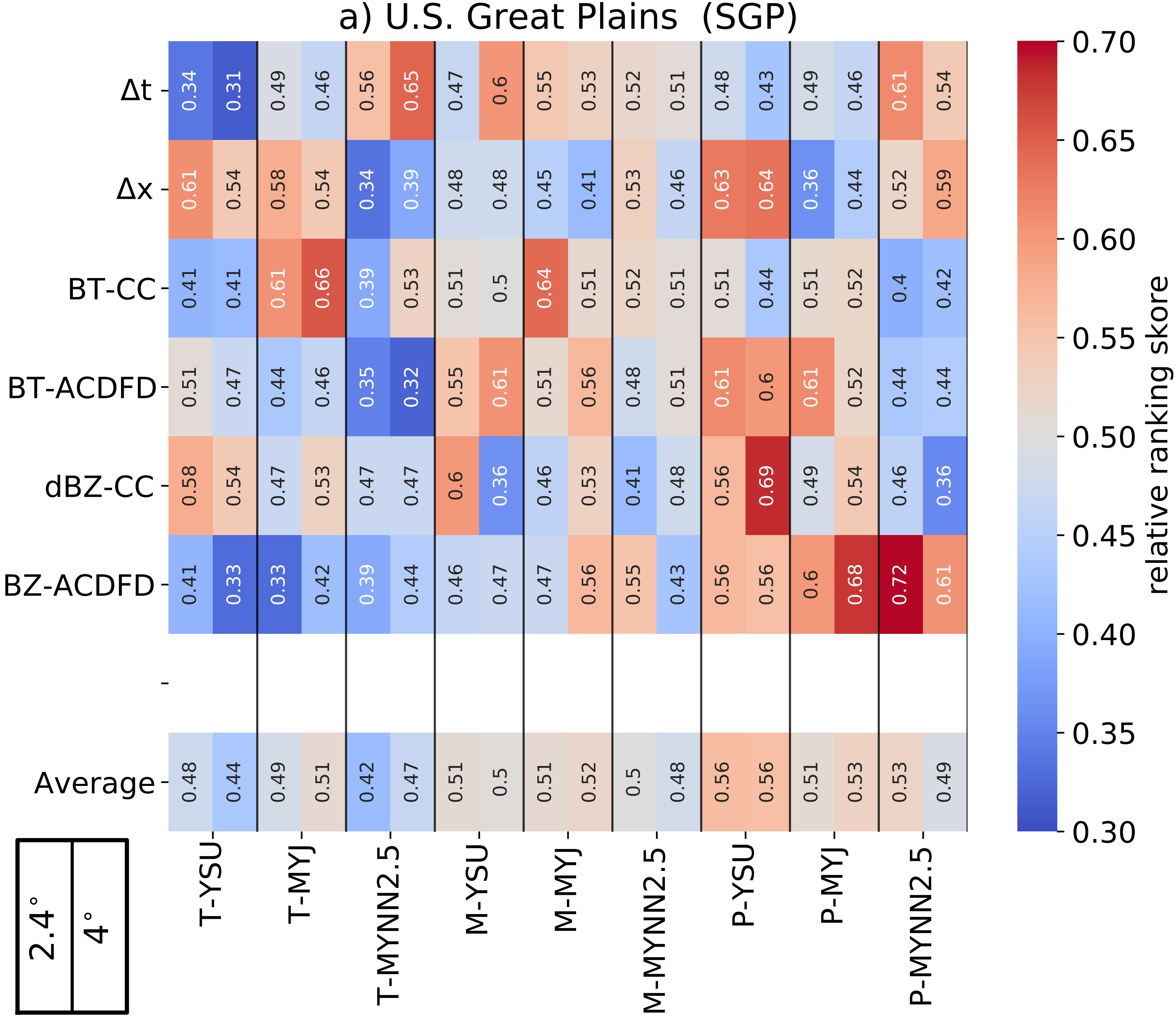
a) CAPE - MAO
b) CAPE - SGP
c) CIN - MAO
d) CIN - SGP
e) -180 min. - MAO
f) 0 min. - MAO
g) 240 min. - MAO
h) -180 min. - SGP
i) 0 min. - SGP
j) 240 min. - SGP
CAPE [J kg^{-1}]
CIN [J kg^{-1}]
time [min.]
height [km]
average mixing ratio [g kg^{-1}]
T-YSU
T-MYJ
T-MYNN2.5
M-YSU
M-MYJ
M-MYNN2.5
P-YSU
P-MYJ
P-MYNN2.5

Figure 10.



a) U.S. Great Plains (SGP)
b) Amazon (MAO)
c) Mean
Δt
Δx
BT-CC
BT-ACDFD
dBZ-CC
dBZ-ACDFD
Average
T-YSU
T-MYJ
T-MYNN2.5
M-YSU
M-MYJ
M-MYNN2.5
P-YSU
P-MYJ
P-MYNN2.5
relative ranking skore
0.30
0.35
0.40
0.45
0.50
0.55
0.60
0.65
0.70

Figure 11.


a) U.S. Great Plains (SGP)
Δt
Δx
BT-CC
BT-ACDFD
dBZ-CC
BZ-ACDFD
Average
T-YSU
T-MYJ
T-MYNN2.5
M-YSU
M-MYJ
M-MYNN2.5
P-YSU
P-MYJ
P-MYNN2.5
2.4°
4°
relative ranking skore
0.30
0.35
0.40
0.45
0.50
0.55
0.60
0.65
0.70