



RESEARCH ARTICLE

10.1029/2019MS001741

Key Points:

- The nesting grid ratio can be as large as 30 without significant effect
- An intermediate nesting grid is undesirable
- A scale-aware cumulus parameterization in a coarse grid may degrade precipitation forecast skill in the nested convection-resolving grid

Correspondence to:

X.-Z. Liang and M. Zeng,
xliang@umd.edu;
swordzmj@qq.com

Citation:

Liang, X.-Z., Li, Q., Mei, H., & Zeng, M. (2019). Multi-grid nesting ability to represent convections across the gray zone. *Journal of Advances in Modeling Earth Systems*, 11, 4352–4376. <https://doi.org/10.1029/2019MS001741>

Received 15 MAY 2019

Accepted 18 OCT 2019

Accepted article online 13 NOV 2019

Published online 16 DEC 2019

Multi-Grid Nesting Ability to Represent Convections Across the Gray Zone

Xin-Zhong Liang^{1,2} , Qi Li^{2,3} , Haixia Mei^{2,4} , and Mingjian Zeng⁴

¹Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD, USA, ²Earth System Science Interdisciplinary Center, University of Maryland, Riverdale Park, MD, USA, ³College of Atmospheric Science, Nanjing University of Information Science and Technology, Nanjing, China, ⁴Jiangsu Institute of Meteorological Sciences, Jiangsu, China

Abstract This study investigated the multi-grid nesting ability of a limited area model to effectively represent convections across the gray zone, the resolution around 1–10 km where both cumulus parameterization and explicit convection are problematic. It evaluated the sensitivity of Meiyu rainfall forecasts in Jiangsu, China to model configurations of grid nesting and convection treatment. These configurations consisted of grid spacings from 30, 15, 9, 5, 3 to 1 km, single or double or triple nested grids, and the traditional Kain-Fritsch (KF) or scale-aware Grell-Freitas cumulus parameterization or the explicit convection in the outer domain [O]. In single nesting [O], coarse grids (>3–5 km) required parameterization to represent organized cumuli, while explicitly resolving convections in finer grids were necessary to improve forecasts. In double nesting [O] using cumulus parameterization at 30–9 km with the inner domain [I] using explicit convection at 1 km, the nesting ratio could be as large as 30 without significantly impacting [I] forecasts. This suggests a pragmatic approach to avoid the challenge in representing convections across the gray zone. Using Grell-Freitas may improve mean [O] rainfall distributions, but this was not true for [I] forecasts due to counter errors in space and time, which were larger than using KF and at coarser grids. Triple nesting with a middle 3- or 5-km grid was unnecessary and could even degrade [I] forecasts. Nesting [O] using KF to parameterize cumuli at 15 km with [I] explicitly resolving convections at 1 km achieved the best overall rainfall forecast in Jiangsu.

Plain Language Summary This study investigated the multi-grid nesting ability of a limited area model to effectively represent convections across the gray zone, the resolution around 1–10 km where both cumulus parameterization and explicit solution are problematic. It evaluated the sensitivity of Meiyu rainfall forecasts in Jiangsu, China to model configurations of grid nesting and convection treatment. These configurations consisted of grid spacings from 30, 15, 9, 5, 3 to 1 km, single or double or triple nested grids, and the traditional or scale-aware cumulus parameterization or the explicit convection. Parameterization in 30–9 km grids is required to represent organized cumuli, while explicitly resolving convections in cloud-permitting grids around 1 km is necessary to improve forecasts. This coupling can be achieved through double nesting, in which the grid ratio could be as large as 30 without significantly impacting forecasts. Triple nesting with a middle grid is unnecessary and can even degrade forecasts. The result suggests a pragmatic approach to avoid the challenge in representing convections across the gray zone. Using a scale-aware cumulus scheme may improve mean rainfall distributions in the outer coarse grid, but this may not increase forecast skill in the nesting fine grid due to counter errors in space and time.

1. Introduction

Mesoscale models are used for increasingly high-resolution simulations, with demonstrated benefits (Prein et al., 2015). In a wide range of applications such as numerical weather prediction (Mass et al., 2002), regional climate downscaling (Giorgi, 2006), and air quality modeling (Krol et al., 2005), the most common practice of increasing resolution is the use of a limited area model (LAM) driven by a general circulation models (GCM) with nested domains that telescopically zoom into the target region while reducing grid spacing. The LAM must be configured properly, or else its downscaling performance may be even worse than the driving GCM. Successful downscaling requires appropriate choices of physics representation, spatial resolution, and the nesting grid ratio, as well as careful treatment of lateral boundary conditions (LBCs) (Giorgi & Mearns,

©2019. The Authors.

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

1999; Liang et al., 2012; Warner et al., 1997; Xue et al., 2014). The latter especially has many well-documented sensitive features, including the location and width of the buffer zones (and so the outer domain size), the relaxation algorithm, and the update frequency (Denis et al., 2003; Giorgi et al., 1993; Liang et al., 2001; Seth & Giorgi, 1998). The choices of physics representation and spatial resolution are interlinked because many parameterizations are scale-dependent, especially for convection (Molinari & Dudek, 1992; Weisman et al., 1997). While increasing resolution is desirable, the transition from the driving GCM coarse grid to the LAM finest target grid remains a critical issue.

Telescoping with multiple nest grids offers a pragmatic approach to reach high-resolution modeling with reasonable computational cost. However, this requires determining the number of nest grids needed and the ratios at which they will remain cost effective. Mesoscale modeling has been traditionally using a nesting grid ratio of 3 (Biswas et al., 2014; Christensen et al., 1998; Gopalakrishnan et al., 2012; Harris & Durran, 2010; Kwon & Hong, 2017; Leung & Qian, 2003; Skamarock et al., 2008; Warner & Hsu, 2000; Yu & Lee, 2011; Zhang et al., 1986). Though many studies simply used a ratio between 2 and 9 (Barthlott & Hoose, 2015; Liang et al., 2019; Prein et al., 2015), few tested the result sensitivity. Denis et al. (2003) and Antic et al. (2006) showed that for a LAM at fixed 45-km spacing, the driving GCM grid could be increased to yield a nesting ratio as high as 12 without notable impacts. Similarly, Amengual et al. (2007) concluded that refining the GCM input data resolution with a grid ratio from 10 to 3 did not improve the result of a LAM at 30-km spacing. These three studies varied the GCM resolution but fixed the LAM grid at the medium range of the mesoscale (~40 km) to test the nesting ratio sensitivity. Further refinement from the mesoscale toward the convection-permitting scale (<4 km) will require another resolution jump by a factor of 10. Beck et al. (2004) showed that adding a 50-km intermediate nest did not much improve skill but that direct nesting from 120- to 12-km grids appeared adequate. Similarly, Matte et al. (2016) indicated that adding a 50-km nest between 200-km driving and 16-km inner grids produced little mean difference but reduced temporal correlation in the transient-eddy component by weakening the LBC control. Brisson et al. (2016) also found that a 7-km intermediate nest between 25- and 2.8-km grids was redundant, although the 25-km outer grid was essential for skillfully downscaling mesoscale signals from ~80-km driving data. All these studies tended to suggest that a nesting grid ratio around 10 could be cost-effective for downscaling from the driving GCM fields (synoptic) to the LAM outer (mesoscale) and to inner (convective) domains. They, however, did not account for the impact of using different physics representations in different grids, which may ultimately affect the choice of the nesting ratio.

It is extremely challenging to use a unified physics representation across the range of spatial scales involved in synoptic, mesoscale, and convective resolutions, which varies by a factor of 100. Representing convection in the gray zone, the grid scale around 1–10 km where parameterized and resolved convective clouds exist simultaneously, is especially difficult (Arakawa, 2004; Gerard, 2007; Molinari & Dudek, 1992; Randall et al., 2003). Each physics parameterization is developed for optimal skill at a limited spatial range and thus its performance is sensitive to horizontal resolution (Dudek et al., 1996; Field et al., 2017; Giorgi & Marinucci, 1996; Jung & Arakawa, 2004; Yu & Lee, 2011). This scale-sensitivity is also found in explicit convection modeling in the gray zone and even at grid spacing below 1 km (Bryan et al., 2003; Weisman et al., 1997). As such, the choice of cumulus treatment in the coarse outer grid can greatly affect the intensity and timing of the inner grid explicitly resolved convection rainfall through induced subsidence, accumulated heating effect on the environmental flow, and other mesoscale processes (Biswas et al., 2014; Warner & Hsu, 2000). Recent scale-aware schemes for cumulus parameterization in the gray zone have been demonstrated to significantly improve precipitation forecasts as compared to those using traditional parameterization or the explicit convection solution (Arakawa & Wu, 2013; Grell et al., 2014; Mahoney, 2016; Zheng et al., 2016; Gao et al., 2017; J. Han et al., 2017; Kwon & Hong, 2017; Freitas et al., 2018). While these scale-aware schemes generally alleviate the excessive precipitation problem, the impact of their use in multiple nesting grids on rainfall spatial structures and temporal characteristics is not fully evaluated or understood.

This study comprehensively evaluates the impact of grid nesting and convection treatment on operational numerical weather prediction in Jiangsu Province in China. It focuses on quantitative precipitation forecasts in a select month-long period of representative weather systems governing various convections over the region and examines grid spacings in the range of 30, 15, 9, 5, 3 to 1 km with single, double, or triple nested grids and varying physics configurations. It addresses the multi-grid nesting ability to represent convections across the gray zone and seeks to determine the most cost-effective nesting grid ratio in the presence of

traditional, scale-aware, or no cumulus parameterization in the outer domain. The evaluation is concentrated on rainfall spatial structures and temporal characteristics in Jiangsu, which is enclosed within the inner domain.

2. Model Description and Domain Configuration

The LAM used in this study is the Weather Research and Forecasting (WRF) model (Skamarock et al., 2008), version 3.9. Figure 1 illustrates the model's computational domains based on the Lambert conformal map projection centered at 35.18°N, 110°E and three possible nested grids of progressively smaller spacing. The outer domain (D1) was designed to realistically capture the synoptic and mesoscale circulations that are essential to regional climate variations (and so statistical weather characteristics) in China (Liang et al., 2019; Liu et al., 2008). It contains a total of 232×172 grid points at 30-km spacing, and for a finer resolution, the grid number increases by a squared factor of a spacing reduction while minimizing change to the domain size. The LBCs were specified within the buffer zones along the four edges of the domain using a dynamic relaxation technique with exponential nudging coefficients that decrease toward the inner boundaries (Giorgi et al., 1993). They were updated at every 6-hr interval by the Global Forecast System (GFS) analysis made available at the 0.5° (~55 km) grid from the U.S. National Centers for Environmental Prediction. This update frequency is adequate for dependable downscaling (Amengual et al., 2007; Antic et al., 2006; Denis et al., 2003). The buffer zone width was chosen to cover approximately two grid points of the driving GFS data, equivalent to 5, 9, and 15 points of D1 for the 30-km, 15-km, and 9- or 5-km spacings, respectively. These four sets of grid spacing were tested for D1 resolution effects.

The intermediate domain (D2) was designed with the intent to better resolve the mesoscale circulations that govern the weather systems affecting Jiangsu precipitation. Following Liang et al. (2001), when possible, its boundaries were located in prevailing centers in which Jiangsu precipitation is teleconnected to upper tropospheric circulations and encompass dominant near-surface mesoscale features such as topographic characteristics, low-level jets, and water vapor pathways. These teleconnection centers and mesoscale features were identified through correlation analyses with interannual anomalies and daily fluctuations during 1980–2015. The inner domain (D3) was nested to explicitly resolve the convective processes that determine the precipitation spatial structures and temporal characteristics in Jiangsu. Its western boundary was located away from Jiangsu at a distance approximately twice the province width, while the other three boundaries were set closer by about one of the width. Given the prevalence of westerly winds, including more distance from Jiangsu on the western lateral inflow boundary allows local processes permitted by the finer resolution to fully develop (Matte et al., 2017). While D3 used grid spacing of 1 km as the target resolution for Jiangsu forecasts, D2 was tested with 5 and 3 km or eliminated in order to determine the impact on the D3 skill.

As adopted in most operational forecasts, the one-way (parasitic) nesting strategy was used in this study to downscale from D1 to D2 to D3 or directly from D1 to D3, without considering possible feedbacks from the child to parent domains. In addition to orographic blending along the domain interface at the setup, the concurrent nesting during the integration of the child domain involved initialization at each fine grid point and specification of prognostic variables solely in the outer row and column of the fine grid by spatial (monotone) and temporal (linear) interpolation from the parent grid forecasts (Skamarock et al., 2008). In practice, the nesting is often done not concurrently but sequentially from the parent to child domains. This is done by the *ndown* procedure in WRF. We will examine the sensitivity of concurrent versus sequential nesting procedures along with varying physics configurations in a separate paper.

The existing Jiangsu operational forecast system is based on an earlier WRF version 3.5.1 and uses the same outer domain D1 at 15-km grid spacing with a nested child domain at 3-km spacing, which is a reduced version of D2 with inwardly shifted northern and eastern boundaries to barely include the Korean Peninsula in the corner. It uses the same GFS analysis for initial conditions, the real-time GFS forecasts for 6-hourly updates of the LBCs in D1, and the *ndown* procedure for one-way (sequential) nesting with the reduced D2. The model has 51 terrain-following (Eta) vertical levels, with the top at 50 hPa. This study adopted the same vertical resolution in all domains and kept the model physics and domain configurations close to the operational system so that the result can be directly applied for upgrade in the future. Therefore, the control physics configuration also followed that of the existing operational system, which includes RRTM longwave and Goddard shortwave radiation (Chou & Suarez, 1999; Mlawer et al., 1997), Thompson microphysics

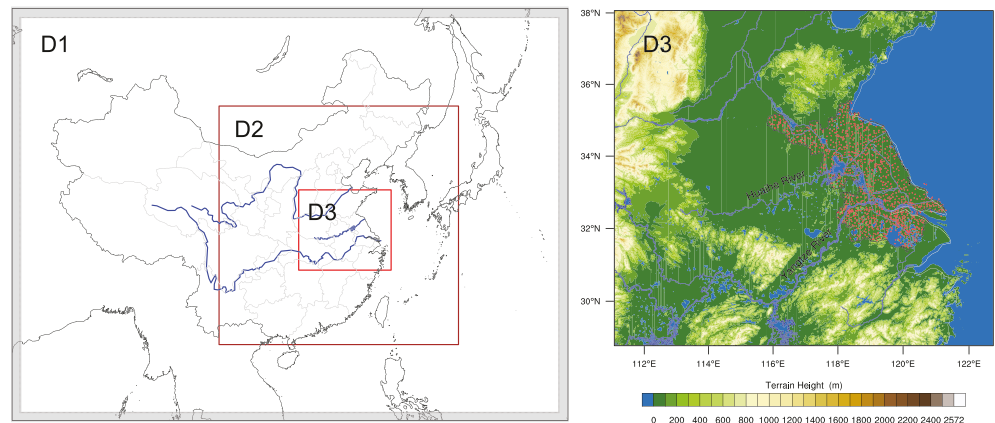


Figure 1. The WRF computational outer (D1, O), middle (D2, M), and inner (D3, I) domains for all forecasts made in this study. Enlarged is D3, overlaid with the terrain height (m) distribution at 1-km grid spacing and 1,238 automated monitoring stations distributed over Jiangsu (red points). The hatched edge areas around D1 are the buffer zones, where LBCs are specified from GFS forecasts.

(Thompson et al., 2008), Kain-Fritsch cumulus (Kain, 2004), Mellor-Yamada-Janjic (MYJ) planetary boundary layer (PBL; Janjic, 2001), and unified NOAH land surface (Tewari et al., 2004). As discussed below, the cumulus parameterization was eliminated to yield an explicit convection solution in D3 (always) and D2 (optional) and was replaced with alternative schemes in D1 sensitivity experiments.

3. Evaluation Period and Experiment Design

To facilitate a robust evaluation, it is important to select sufficient cases that represent a wide spectrum of regional precipitation systems. Since the skill of weather forecasts is typically evaluated using statistics of a characteristic period such as a month, it is desirable to select a continuous period that contains a wide variety of convection or precipitation cases. Precipitation systems associated with the Asian summer monsoon are

prevalent in Jiangsu Province. During the monsoon, southerly winds bring warm and moist oceanic air masses inland to frequently produce heavy precipitation in south and east China. After its onset in the South China Sea in May, the monsoon progresses northward to the Yangtze-Huaihe River basin during mid-June to mid-July, and there maintains a quasi-stationary subtropical front to produce persistent heavy rainfall, referred to as Meiyu (Ding, 1991; Samel et al., 1999; Tao, 1987; Zhao et al., 2018). Since it is located in the east of the basin, Jiangsu receives concentrated heavy rainfall and frequent rainstorms during Meiyu. Thus, the Meiyu period is preferred for the forecast evaluation.

Meiyu was abnormal in 2016, and its accumulated precipitation in Jiangsu approximately doubled the climatological mean. Under the influence of both the decay of the previous extreme winter El Niño event and the persistent abnormal spring-summer warming in the entire Indian Ocean basin, the western Pacific subtropical high was extremely strong during the Meiyu period. The abnormally strong moisture flux from the southwest flank of the subtropical high converged in the Yangtze and Huaihe River Valleys with the weak cold air from the north, leading to a longer Meiyu period and greatly increased total precipitation (Zhao et al., 2018). On average, Meiyu in Jiangsu continues from 18 June to 10 July, but in 2016 it persisted from 19 June to 20 July, breaking multiple historical records in rainstorm occurrence and coverage as well as rainfall intensity and total amount.

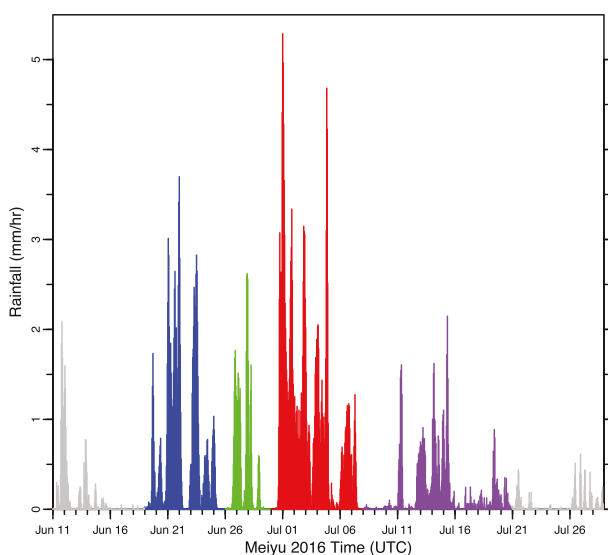


Figure 2. Evolution of hourly rainfall (mm hr^{-1}) averaged over Jiangsu Province for Meiyu in 2016. Colored are four subperiods with intense (first three) and dispersive (last) rainstorms, consisting of a total of 32 days from 19 June to 20 July for the forecasts.

Figure 2 shows the evolution of hourly rainfall averaged over Jiangsu Province for Meiyu in 2016. There were three concentrated periods of intense rainstorms. The first was from 19 to 23 June, when intense rains fell in the areas between the Yangtze and Huaihe Rivers and the southeastern part of south Jiangsu. The second was from 27 to 28 June, when the rainband moved southward to the areas along the Yangtze River and south Jiangsu. The third was from 30 June to 7 July, when the rainband remained steady in the southern part between the Yangtze and Huaihe Rivers and the areas along the Yangtze River and south Jiangsu. After 8 July, Jiangsu experienced mainly dispersive rainstorms, with particularly heavy rainfall from 11 to 13 July. Therefore, the Meiyu period from 19 June (onset) to 20 July (ending), a total of 32 days, was selected for forecast experiments in this study. Clark et al. (2018) used a similar period length (24 days) in the U.S. Spring Forecasting Experiment to evaluate the performance of a wide variety of convective-permitting models. The period, while not particularly long, is sufficient for a rigorous evaluation of the various grid nesting and convection treatment configurations in this study. Each forecast was initialized at 12:00 UTC and integrated for 36 hr. Forecast skill was evaluated based on hourly rainfall data from 1,238 stations in the automated monitoring network which had no missing records during the entire period. These stations were evenly distributed over Jiangsu (Figure 1), with an approximate spacing of 9 km.

This study explored Jiangsu precipitation forecast sensitivity to different combinations of nesting grids, grid spacings, and convection treatments. Table 1 provides a complete list of the experimental configurations, except for those designed to specifically address whether the scale-awareness of cumulus parameterization makes differences in section 8. In the outer domain, D1, the control KF cumulus scheme and the Grell and Freitas (2014) scale-aware cumulus parameterization were tested for grid spacings at 30, 15, 9, and 5 km. For grid spacings at 9 and 5 km in the gray zone, also tested in D1 was the EC solution, that is, using no cumulus parameterization. It is generally understood that EC cannot provide a correct solution for grid spacings above 10 km, where cumulus parameterization becomes necessary (Molinari & Dudek, 1992). In the gray zone, both parameterized and resolved convections coexist (Arakawa, 2004; Gerard, 2007), and scale-aware schemes are currently preferred. However, whether cumulus parameterization should be used for grid spacings below 4 km still remains controversial. Weisman et al. (1997) found that grid spacing of 4 km was sufficient for explicit mesoscale modeling, and Kain et al. (2008), Schwartz et al. (2009), and Fierro et al. (2009) showed that further decreasing grid spacing to 2 or 1 km yielded no significant improvement. On the other hand, Han and Hong (2018) showed that EC at 3-km grid seriously underestimated light rain events though improving forecast for heavier rainfall, and Petch et al. (2002) and Bryan et al. (2003) suggested that grid sizes below 1 km were necessary to explicitly predict convective systems. More recent studies, such as Barthlott and Hoose (2015), Mahoney (2016), Leutwyler et al. (2017), and their cited references, indicated that the results from further grid refinement were mixed, depending on specific applications. Given limited computing resources, this study focused on grid spacing of D3 at 1 km, and D2 at 3 and 5 km, all of which use EC without parameterization. The double nesting experiments also evaluated the impact of using KF, GF, and EC in D1 for grid spacings at 9 and 5 km on the D3 forecast skill.

Therefore, there were total 23 configurations, each consisting of a specific choice of convection treatment in the outer domain and a selection of single, double, or triple grid nesting with varying grid spacings. It is difficult in comparison to name a specific configuration in a meaningful way without getting too cumbersome. For convenience, the following notation is used to describe these configurations. Refer to Figure 1 for the telescoping domains of the three nesting grids. By single nesting, we mean that only the outer D1 grid was activated during the forecast and nested in the grid of (i.e., driven by) the GFS forcing. In contrast, double nesting means either the middle D2 grid or the inner D3 grid was also activated and nested in the outer D1 grid. On the other hand, triple nesting means all D1, D2, and D3 grids were activated and nested simultaneously during the forecast. To further simplify this, each grid is identified by a prefix (O, M, I) for the (outer D1, middle D2, inner D3) domain followed by grid spacing in km. Any nested grid is appended in sequential order using the same method. For example, O15M3I1 refers to a triple nesting between D1 at 15 km, D2 at 3 km, and D3 at 1 km. Similarly, O9I1 denotes a double nesting between D1 at 9 km and D3 at 1 km, while O30 represents a single nesting of D1 at 30 km (in the GFS forcing grid). Since both D2 and D3 always adopt EC, a specific convection treatment (KF, GF, EC) is only specified for D1 experiments. This allows replacing unambiguously the O prefix with the convection treatment (KF, GF, EC). Meanwhile, 1 after I can be safely omitted because D3 always uses grid spacing at 1 km. Thus, KF15M3I is same as the aforementioned triple nesting O15M3I1 using the KF cumulus scheme in D1. Since there is no feedback in

Table 1
Summary of WRF Nested Forecast Experiment Configurations

Driving data for LBCs	Outer domain [O]:CUP or EC				Middle domain [M]:EC		Inner domain [I]:EC
	Convection treatment	Grid spacing (km)	Domain size (grids)	Buffer zone width	Grid spacing (km)	Domain size (grids)	Grid spacing = 1 km domain size (grids)
GFS analysis 6-hourly 0.5° (~55 km)	KF	30	232 × 172	4			1,171 × 1,021
		15	463 × 343	14			1,156 × 1,021
		15	463 × 343	14	3	996 × 991	1,153 × 1,000
		9	772 × 571	14			1,162 × 1,000
		5	1,387 × 1,027	14			1,151 × 1,006
	GF	30	232 × 172	4			1,171 × 1,021
		15	463 × 343	14			1,156 × 1,021
		15	463 × 343	14	5	598 × 595	1,156 × 1,006
		15	463 × 343	14	3	996 × 991	1,153 × 1,000
		9	772 × 571	14			1,162 × 1,000
	EC	5	1,387 × 1,027	14			1,151 × 1,006
		9	772 × 571	14			1,162 × 1,000
		5	1,387 × 1,027	14			1,151 × 1,006

Note. The LBCs update frequency of the WRF outer domain keeps up with the available GFS forcing data (i.e., 6 hourly). Nested LBCs update frequency keeps up with its own parent domain's integration time step since all employ one-way concurrent nesting. To ensure stable integration, the time step is set to 4 times of the grid size (km) for the configurations with nesting ratio of 15:1 or 30:1, while it is set to 6 times for others.

one-way nesting, the result from a parent grid can stand alone without noting its child grid. For example, KF15M3 is the result of D2 at 3 km nested in D1 at 15 km using the KF cumulus scheme, although it may contain a child D3 at 1 km as in KF15M3I. Physics parameterizations other than the convection treatment are kept identical among all grids.

4. Spatial Distribution

Figure 3 compares the 32-day mean 24-hr accumulative rainfall forecasts from all D1 configurations. Also shown are the concurrent observed rainfall distributions at 1,238 stations and the GFS forecast at the 0.25° (~28 km) grid. The operational GFS model uses horizontal T1534 spectral truncation, providing gridded outputs at 0.5° and 0.25°, respectively, for the pressure-level and surface variables. This study used the former for the LBCs driving WRF (following the operational practice) and the latter for precipitation forecast evaluation. GFS uses a cumulus parameterization scheme based on a simplified Arakawa-Schubert quasi-equilibrium closure (Han & Pan, 2011), its forecast is denoted by GFS28 for brevity. For both GFS and WRF forecasts, the initial 12-hr spin-up was excluded, so the results represented the mean of the next day 0:00–24:00 UTC. All model forecasts, regardless of grid spacing, were linearly interpolated onto the monitoring stations to facilitate the comparison. One distinct feature during the forecast period was the major rainband observed in south Jiangsu, oriented from the southwest to northeast side of the Yangtze River. To evaluate how well each configuration captures this distinct feature, we compared the average over the rainband (arb), which encompasses all monitoring stations that observed rainfall larger than 15 mm day⁻¹. The arb error (arbe) is shown in Figure 3, along with the spatial pattern correlation coefficient (corr) and root-mean-square error (RMSE) between each forecast and observations at all 1,238 stations in Jiangsu.

The driving GFS forecast substantially underestimated total precipitation in the major rainband by 27%, while WRF downscaling results overestimated it by 3–52% depending on the configuration of convection treatment and grid spacing. When using KF, WRF overestimated arb by 32–39%, with spatial pattern correlation slowly increasing from 0.44 to 0.57 as grid spacing decreased from 30 to 5 km. A similar tendency occurred when using GF, which generally showed higher spatial pattern skill than KF (corr increased by 0.12, 0.18, 0.19, and 0.08 in the respective 30-, 15-, 9-, and 5-km grids). However, GF still produced substantial RMSE, even larger than KF in 30- and 15-km grids and further expanded the rainband. Consequently, GF more greatly overestimated the rainband than KF (larger arbe ~43–52%) in all grids except 5 km. Using EC reduced RMSE by 40–44% relative to using KF and GF at both 9- and 5-km grids, except that

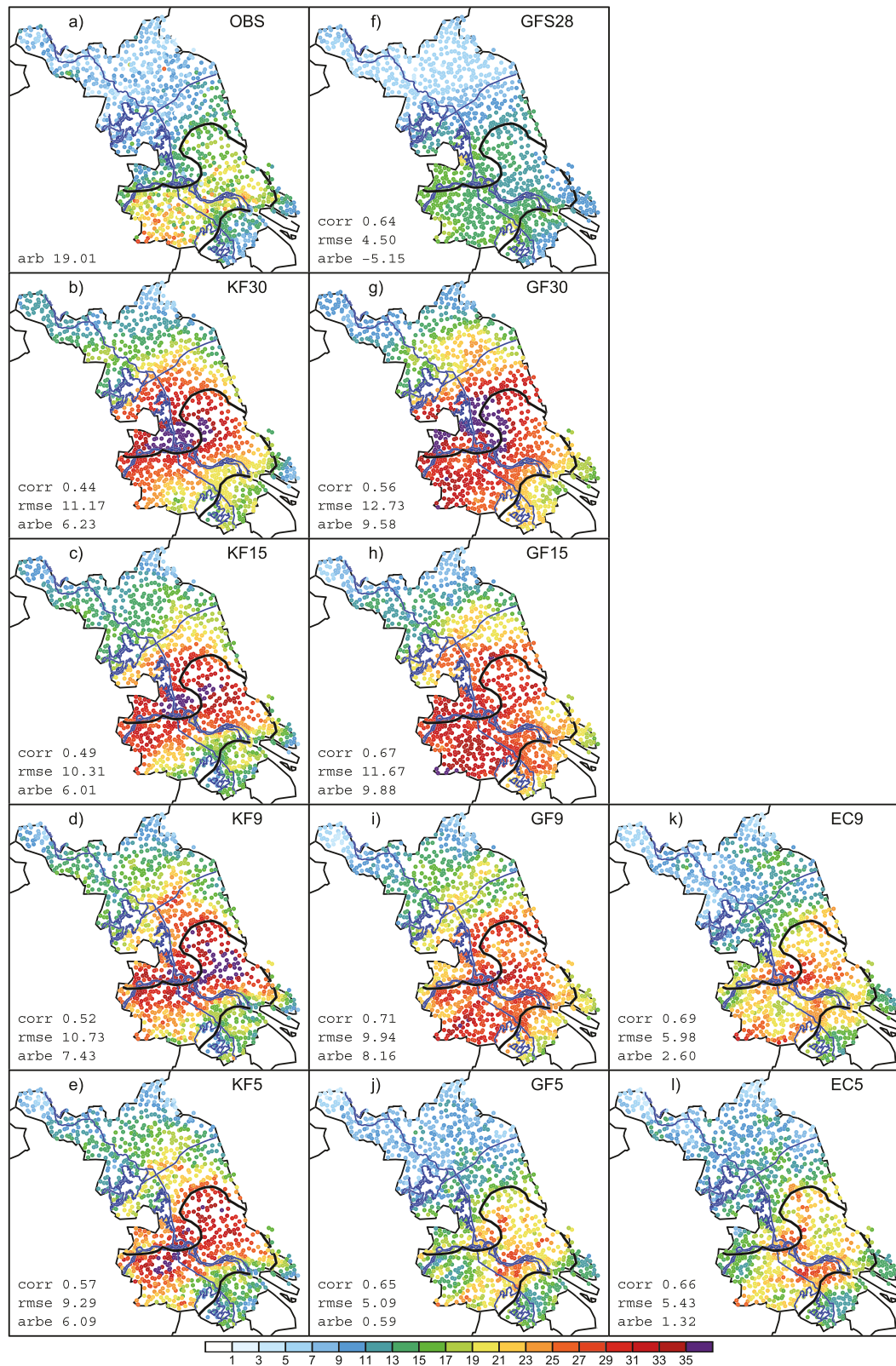


Figure 3. Geographic distributions over Jiangsu of the daily rainfall (mm day^{-1}) averaged for the entire period of 32 days observed (OBS) and forecasted by GFS and WRF using all D1 configurations of grid spacing and convection treatment. The thick black curve encloses the observed rainband consisting of all the stations with average rainfall larger than 15 mm day^{-1} . Listed are the average over the rainband (arb) and the forecast's arb error (arbe), spatial pattern correlation coefficient (corr), and root-mean-square error (RMSE) as compared with observations.

GF performed slightly better than EC at 5 km. Thus, WRF's overestimation might be due to the use of cumulus parameterization, which led to the production of heavier rainfall over a broader band. The opposite occurred in GFS28, which greatly underestimated rainfall. KF simulated excessive ratios of parameterized to total rainfall, especially in the 30- and 15-km grids. On the other hand, GF generated notably smaller ratios, and GFS28 produced ratios similar to GF15 (supporting information, Figure S1). Thus, though the scale-aware GF allows more explicit rain by a factor of 2–3 (Figure S2), this did not significantly improve its skill in total precipitation over the traditional KF, especially in grids coarser than 5 km.

It is important to note that the two commonly used metrics, spatial corr and RMSE, are inadequate to identify differences in geographic structures among the forecasts. For example, the spatial pattern in Jiangsu had mixed skill, since EC performed better than KF at 9- and 5-km (corr increased by 0.17 and 0.09, respectively) but slightly worse than GF at 9-km (corr reduced by 0.02), although both RMSE and arbe suggested EC performed clearly the best. Both visually and according to arbe, the rainband produced by WRF with EC5 was significantly more realistic than that by GFS28, though corr and RMSE indicated mixed skill. Evaluation of more comprehensive metrics below, especially those that consider temporal characteristics, is needed.

Figure 4 compares 32-day mean 24-hr accumulative rainfall forecasts from all D2 and D3 configurations. The comparison led to three major points. First, KF(30, 15, 9, 5)I results showed that the resolution of the parent domain D1 using KF cumulus parameterization had a minor effect on the forecast in the child domain D3 at 1-km grid, producing corr of 0.60–0.64, RMSE of 4.83–5.10 mm day^{−1}, and arbe of −0.40 to 0.46 mm day^{−1} or −2.1 to 2.4%. GF(30, 15, 9, 5)I displayed slightly more resolution sensitivity, producing corr of 0.59–0.70, RMSE of 4.60–6.25 mm day^{−1}, and arbe of 0.21–2.13 mm day^{−1} or 1.1–11.2%. Thus, altering the nesting ratio from 2 to 11 for the driving to parent grids and from 5 to 30 for the parent to child grids had no significant effect on the forecast skill of average rainfall distribution in Jiangsu.

Second, KF5I and EC5I produced similar results, though EC5I overestimated more the rainband (arbe: 1.03 vs. −0.40 mm day^{−1}) but better captured the spatial pattern (corr: 0.70 vs. 0.62). GF performed slightly better (worse) than KF at 5 (9) km but had notably larger RMSE and arbe at both 30- and 15-km. Thus, using the scale-aware cumulus scheme GF in the parent domain had no benefit to using the traditional KF but resulted in greater overestimation in the child grid.

Third, KF15(I, M3I) results showed that adding the middle domain D2 at 3-km grid increased the overall skill in Jiangsu, leading to higher corr (0.60, 0.68) and smaller RMSE (5.10, 4.33), but also to greater underestimation of the major rainband, causing larger negative arbe (−0.29, −0.66). GF15(I, M3I) results indicate a greater improvement, with the addition of M3 increasing corr (0.59, 0.69) and decreasing both RMSE (6.25, 4.62) and arbe (1.80, 0.31). Thus, in terms of the inner domain I1 forecast, the middle grid M3 was unnecessary for KF15 but beneficial for GF15. Furthermore, the triple nesting I1 had little impact on the skill of either (KF, GF)15 M3, but it improved the skill of GF15M5, reducing both RMSE (4.55, 5.35) and arbe (0.13, 1.83). These mixed results between GF15M(3,5)I may indicate the sensitivity of grid nesting around 3–5 km. The fact that the incorporation of the middle grids M(3, 5) caused different skill changes suggests that grid nesting in the range of 3–5 km may be particularly sensitive.

5. Temporal Evolution

Figure 5 compares the averaged hourly rainfall evolution (including the spin-up period) forecasts from all D1 configurations. The average was done first over 1,238 stations and then for 32 days. Observations exhibited a clear rainfall diurnal cycle with double peaks at local times 08:00 (primary) and 15:00 (secondary) and reach the day's minimum at 20:00. The WRF forecasts had mixed results. The explicit convection solution at both 5- and 9-km grid spacings predicted the diurnal cycle reasonably well. EC5 captured both the morning and afternoon peak times correctly but significantly postponed the minimum by 5 hr. EC5 systematically overestimated the rainfall amount from early morning (05:00) to evening (23:00) but underestimated it after 04:00 of the next day. EC5 overestimated the afternoon peak more strongly than the morning peak, thus switching the primary to secondary in the forecast. EC9 produced a similar result, except that it more greatly overestimated the morning peak and the period after the afternoon peak. Both KF and GF cumulus parameterizations yielded greater but different errors than did the explicit solution, depending on forecast time and grid spacing. KF5 was close to EC5 before 09:00 and after 19:00 but substantially overestimated the time in between, creating a predominant daytime diurnal cycle with a peak at 14:00. KF at grid spacings of 9, 15, and

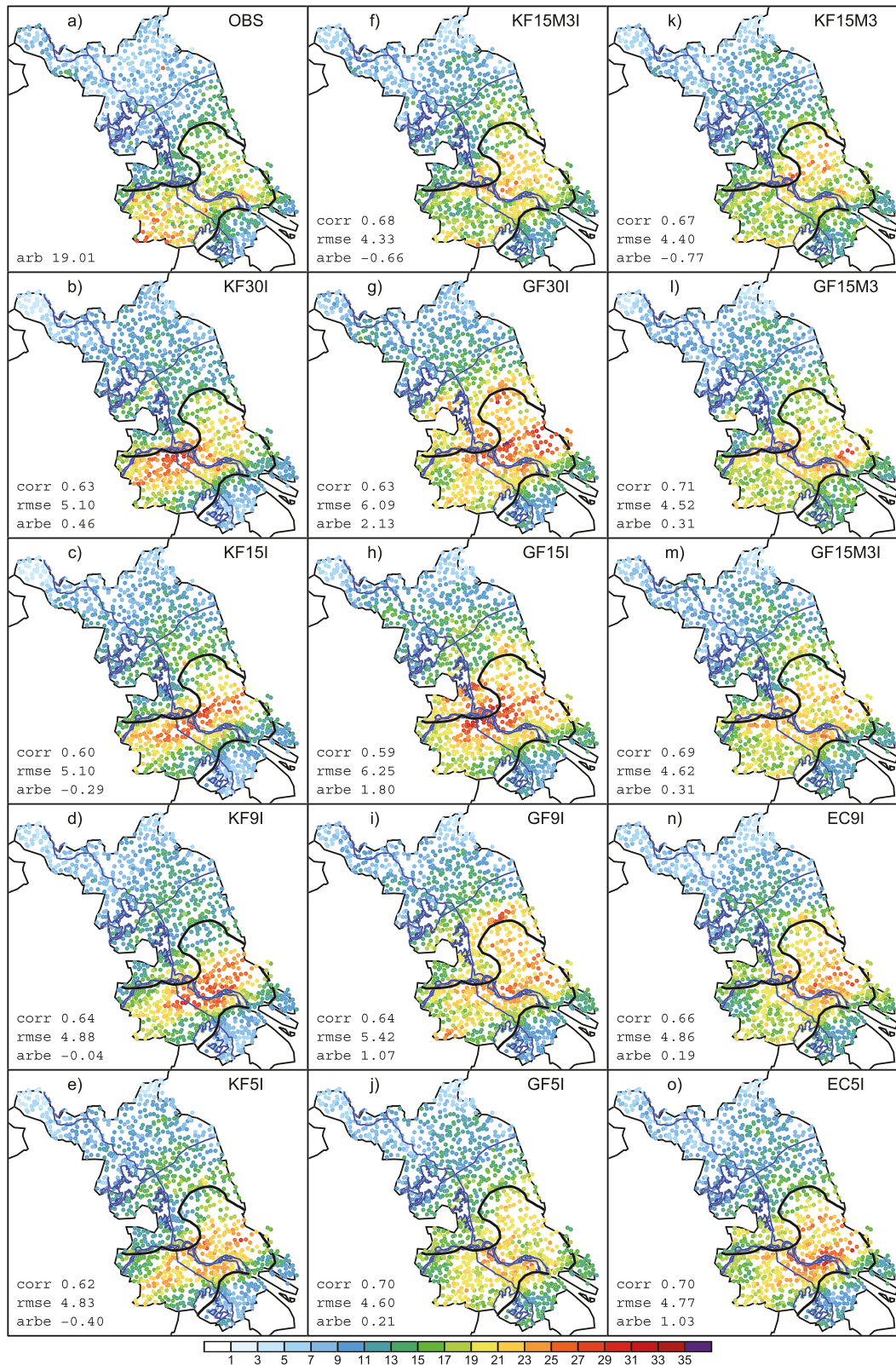


Figure 4. Same as Figure 3 except for WRF forecasts using all D2 and D3 configurations. Note that KF15M3 and KF15I denote for double nesting of respectively D2 at 3 km and D3 at 1 km using EC, while KF15M3I for triple nesting of both, all with D1 at 15 km using KF. Others are named similarly, referring to the convention in the last paragraph of section 4.

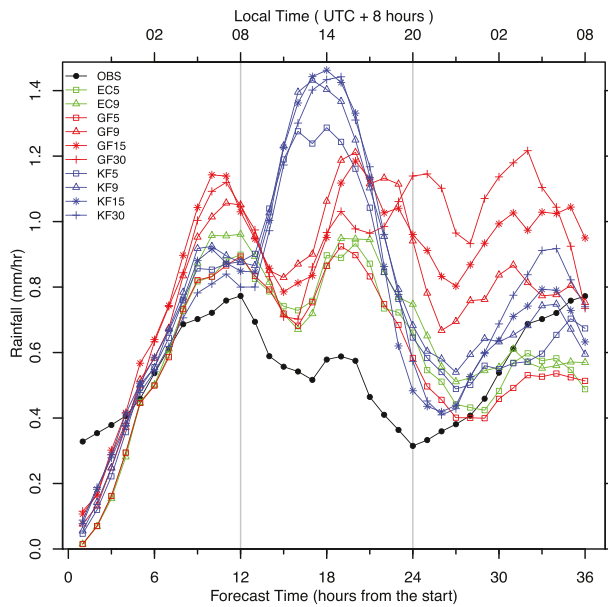


Figure 5. Variations with the forecast time (including the spin-up) of the hourly rainfall (mm hr^{-1}) averaged over all stations and for the entire period of 32 days observed (OBS) and forecasted by WRF using all D1 configurations. The corresponding local time is scaled at the top.

smaller ratios than KF; the ratio reduction was especially large in the morning before the afternoon peak but was less obvious in the nighttime, delaying the diurnal peak for several hours.

Figure 6 compares the averaged hourly rainfall evolution forecasts from all double nesting D1–D3 configurations. KF(30, 15, 9, 5)I exhibited a minor sensitivity to grid spacing of the parent domain D1. Their performance was even better than EC(9, 5)I that used explicit convection in D1. They systematically reduced EC's rainfall overestimation in most of the day (local time 02:00–20:00). This is in a sharp contrast to the D1 forecast, where KF overestimated daytime rainfall much more substantially than EC (Figure 5). On

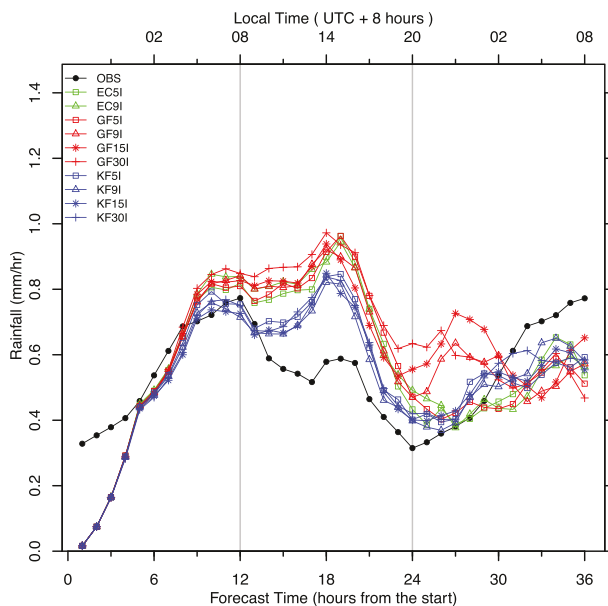


Figure 6. Same as Figure 5 except for WRF forecasts using all double nesting D1–D3 configurations.

30 km produced similar results, except with even higher peaks. Interestingly, KF captured the rising phase of the diurnal cycle after midnight well (albeit with larger rain rates), especially in the 15- and 30-km grids. Given the scale-aware assumption, GF5 resembled EC5 closely (albeit with smaller rain rates after 19:00) due to negligible effects from cumulus parameterization at 5 km. In contrast, GF9 was similar to EC9 in phase, but systematically overestimated rain rates, especially around the afternoon peak and the next forecast day's morning peak. As grid spacing increased from 9 to 15 to 30 km, GF progressively reduced daytime overestimations but further enhanced them in the nighttime. Consequently, GF proved much more sensitive to resolution than KF, which was opposite of what it was designed for.

Thus, the scale-aware GF does successfully reduce the traditional KF's overestimation of daytime rainfall by yielding more parameterized cumulus to explicit convection as resolution increases (Fowler et al., 2016; Freitas et al., 2018; Grell et al., 2014) but at the cost of greatly increasing nighttime precipitation. Such counter errors were stronger in coarser grids. These sensitivities were reflected in the parameterized to total precipitation ratio (Figure S3). KF generated notably higher ratios in the daytime than in the nighttime, following the diurnal cycle of total precipitation, but these ratios steadily decreased as grid spacing was reduced from 30, 15, 9 to 5 km. GF5 ratios were essentially zero for the entire forecast. In the other grids, GF had a similar resolution sensitivity but generally

smaller ratios than KF; the ratio reduction was especially large in the morning before the afternoon peak but was less obvious in the nighttime, delaying the diurnal peak for several hours.

the other hand, GF(15, 30, 9, 5)I showed a larger sensitivity to grid spacing, performing worse than KF(30, 15, 9, 5)I and even EC(9, 5)I, with larger overestimations during most of the day, especially in the evening. This is again in a sharp contrast to the D1 forecast, where GF largely reduced KF's daytime overestimation, albeit at the cost of causing excessive nighttime precipitation, especially at coarser grids. Therefore, the scale-aware cumulus parameterization GF reduced the traditional KF's rainfall overestimation in the parent domain but worsened the forecast in the child domain using explicit convection, producing heavier rainfall throughout the day. Note that EC(9, 5)I performed persistently worse than all KF(30, 15, 9, 5)I during the first 24-hr forecast. Thus, it is necessary to represent organized convection by parameterization while explicitly resolving smaller scale convection. This can be effectively done using a double nesting configuration with respectively parameterized (15–30 km) and explicit convection (1 km) grids.

Note that both D1 and D1–D3 forecasted realistic rainfall after a spin-up of less than 6 hr. This occurred for all grids using parameterized or explicit convection, indicating that the characteristic time for the mesoscale circulation and local precipitation system to fully establish in response to the synoptic scale forcing through LBCs is 6 hr or shorter. Since forecast errors generally grow with time, our evaluation focused on the characteristics of the last 24-hr forecast (excluding the initial 12, rather than 6, hr) to maximize model performance differences.

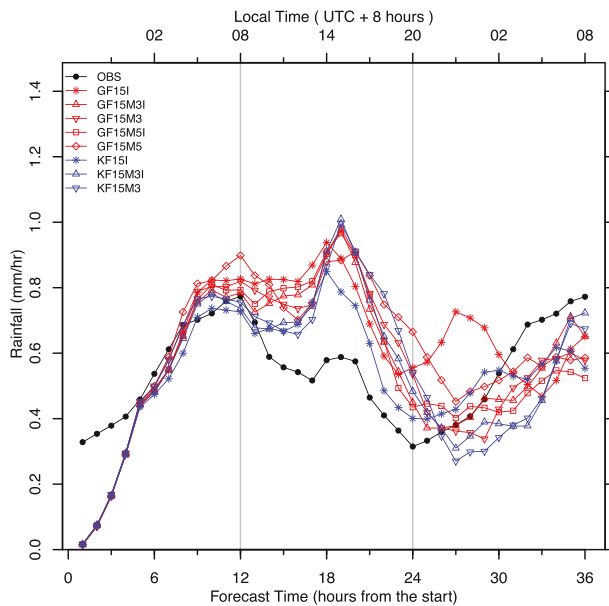


Figure 7. Same as Figure 5 except for WRF forecasts using all O15 and its nested configurations.

Figure 7 compares the averaged hourly rainfall evolution forecasts from all O15 nested configurations. KF15I was more realistic than KF15M3I after 18 hr. Thus, nesting a middle grid M3 brought no benefit but worsened the inner domain I1 forecast. The close similarity between KF15(M3I, M3) results indicates that the I1 forecast was essentially locked to that of its parent M3. As discussed earlier, GF15I performed notably worse than KF15I throughout the forecast period. Adding M3 or M5 improved GF15I morning and nighttime forecasts but degraded its afternoon skill. The triple nesting I1 had some mixed effects on the GF15M3 skill but generally improved the GF15M5 forecast. Thus, using the scale-aware parameterization GF in D1 yielded a mixed forecast skill in the inner domain. Nonetheless, KF15I is much better than KF15M3, indicating that the finer 1-km grid resolution is beneficial to precipitation forecast.

Figure 8 compares the averaged hourly rainfall evolution forecasts from O5 and O5I configurations. Using explicit convection in the parent domain, the double nesting EC5I improved the majority of the EC5 forecast, excluding a few noontime hours. As discussed earlier, given a small partition of parameterized to explicit convection, GF5I and GF5 were similar to EC5I and EC5, respectively. On the other hand, the improvement from KF5 to KF5I was much greater (especially in the daytime) and occurred systematically throughout the day when using KF cumulus parameterization. Therefore, the 1-km grid is a significant refinement for precipitation forecast to the 5-km grid whether using explicit or parameterized convection. In addition, KF5I is clearly more realistic than EC5I, indicating that in the 5-km grid organized convections still play a significant role and cannot be fully resolved but require parameterization. Considering the insensitivity or mixed difference between M3I and M3 as nested with KF15 or GF15 discussed previously, the 3-km grid appears to be the cutoff at which cumulus parameterization is required rather than an explicit convection solution.

6. Intensity Frequency

Figure 9 compares all D1 configurations' forecasts of the 24-hr accumulative rainfall quantiles to observations, excluding the initial 12-hr spin-up. The driving GFS forecast is also shown. The quantiles were based on all 1,238 stations and 32 days, for a total of 39,616 samples. These samples (regardless of their original locations and times) were rearranged in increasing order, and the actual rainfall values at the cumulative 1, 2, ... to 100% of the total samples were collected to form the sequential intensity quantiles. For clarity, the low quantiles are shown in an inset. The diagonal line represents a perfect forecast, whereas a curve above (below) this line depicts model overestimation (underestimation). In addition, a clear day success rate is added to depict the percentage of correctly modeled no-rain events out of the total number observed. Clearly, GFS substantially underestimated rain events with daily intensities greater than 5 mm while also systematically overestimating lighter rainfall and clear events. Of all the configurations, EC5 (and hence GF5) most realistically captured the overall distribution of the quantiles, except that it slightly underestimated rain events between 30 and 100 mm, compensating for slight overestimations at the lighter and heavier tails. EC9 shared a similar feature but tended to increase overestimation and so produce systematically larger precipitation. Interestingly, KF15 closely resembled EC5 for rainfall greater than 40 mm but more notably overestimated the lighter and clear events. KF30 performed slightly worse than KF15. KF5 performed realistically between 50 and 70 mm but overestimated both tails more than EC5 and GF5. This tendency was enhanced in KF9, which performed worse than EC9. In contrast, GF9 performed worse than KF9 (and hence the other KFs and ECs) due to systematic overestimation for all quantiles below 110 mm, and GF(15, 30) performed even worse. All WRF forecasts overestimated light rain events, and such overestimation increased as convection treatment changed from using EC to KF to GF and also grid spacing enlarged from 5, 9, 15 to 30 km. In other words, more light rain events were associated with coarser grids, more so when using GF than KF.

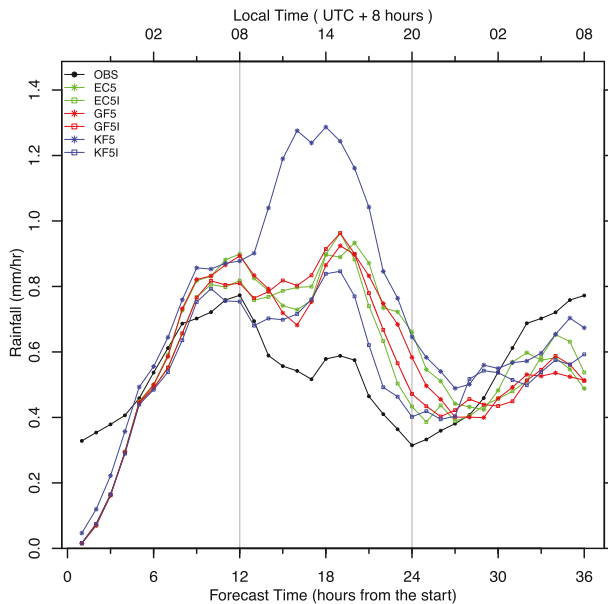


Figure 8. Same as Figure 5 except for WRF forecasts using all O5 and O5I configurations.

Shorter duration rain events presented a different situation. Figure 10 compares all D1 configurations' forecasts of the hourly rainfall quantiles to observations. The quantiles were based on all 1,238 stations, 32 days, and last 24 hr of forecast, for a total of 950,784 samples. Here, EC5 realistically captured light rainfall (<3 mm) and clear events, but overestimated the heavier range. EC9 performed similarly but with more overestimation. GF5 performed slightly better than EC5, while GF9 performed even better for rainfall events above 5 mm but slightly underestimating lighter rain events. GF15 closely resembled GF9, except that it overestimated light rainfall (<5 mm). GF30 resembled GF15 for rainfall below 10 mm but underestimated heavier events. On the other hand, all KF forecasts underestimated rainfall heavier than 5 mm, and such underestimation increased as grid spacing enlarged from 5, 9, 15 to 30 km; they also all overestimated lighter events. In many cases, these tendencies between KF and GF performance on the hourly quantiles were opposite to those of the 24-hr accumulative quantiles (Figure 9) discussed earlier. This result indicates that the forecast skill evaluation based on quantiles was sensitive not only to spatial resolution but also to the duration of the rain event. KF15 captured daily accumulation more realistically, while GF15 better captured hourly rainfall. These results could potentially differ if they were validated against data from a denser network of stations. The current data resolution was no more than 9 km and is hence inadequate to evaluate forecasts at grid spacing of 5 km or finer.

Figure 11 compares all double nesting D1–D3 configurations' forecasts of the daily rainfall quantiles against observations. EC5I (and thus GF5I) best captured the observed distribution, while KF5I underestimated intensity below 70 mm and overestimated heavier rainfall. EC9I resembles EC5I closely. Other KFs performed similarly, exhibiting little resolution sensitivity, except that they produced slightly greater overestimation of the heavier tail as the parent grid spacing increased. Other GFs generally overestimated rainfall, especially for intensities below 50 mm. All these differences were much smaller than those among the D1 forecasts (Figure 9). The sensitivity to parent grid spacing and convection treatment was further reduced for the hourly rainfall quantiles, where all forecasts were very close to observations (Figure S4). Adding M3 or M5 had trivial effects on the daily and hourly rainfall quantiles (not shown), indicating that the triple nesting was not necessary. Therefore, the double nesting with large grid ratios (15–30) performed similarly to, if not better than, the traditional use of more nesting grids with smaller increments (typically 3). This also effectively avoids the challenge of representing convection across the gray zone.

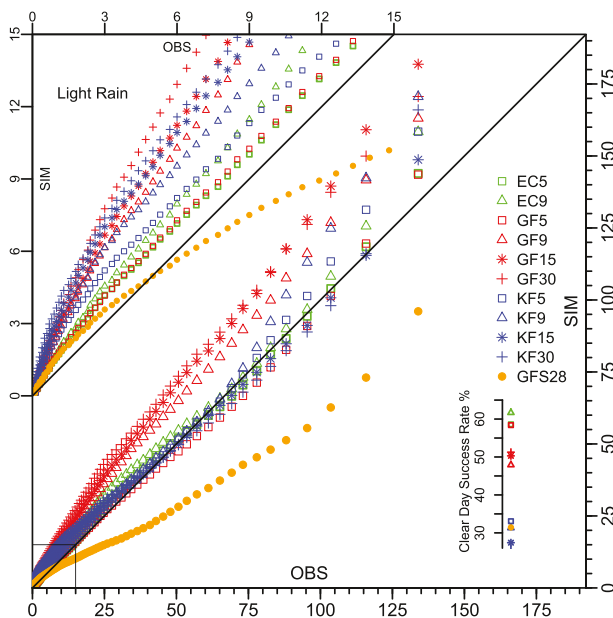


Figure 9. Scatter diagram of forecasted (SIM) against observed (OBS) daily rainfall (mm day⁻¹) quantiles. Compared are the forecasts by GFS and WRF using all D1 configurations. The diagonal line represents a perfect forecast. The upper-left inset enlarges the light rain portion, while the lower-right inset shows the clear day success rate (the percentage of correctly modeled over total observed no-rain events).

Therefore, the double nesting with large grid ratios (15–30) performed similarly to, if not better than, the traditional use of more nesting grids with smaller increments (typically 3). This also effectively avoids the challenge of representing convection across the gray zone.

7. Operational Skill

Our evaluation so far has been focused on time-averaged spatial patterns (section 4), space-averaged temporal variations (section 5), or spatiotemporal frequency distributions (section 6), none of which consider direct forecast-observation correspondence in time and space. This section focuses on such correspondence, which is typically used in operational forecast evaluation.

Figure 12 compares the spatiotemporal RMSE of hourly rainfall forecasts from all D1 configurations. The RMSE was calculated at each forecast hour and based on all 1,238 stations and 32 days, for a total of 39,616 samples. Also shown is the spatial standard deviation of observed rainfall, which was similarly calculated. The RMSE exhibited a strong diurnal

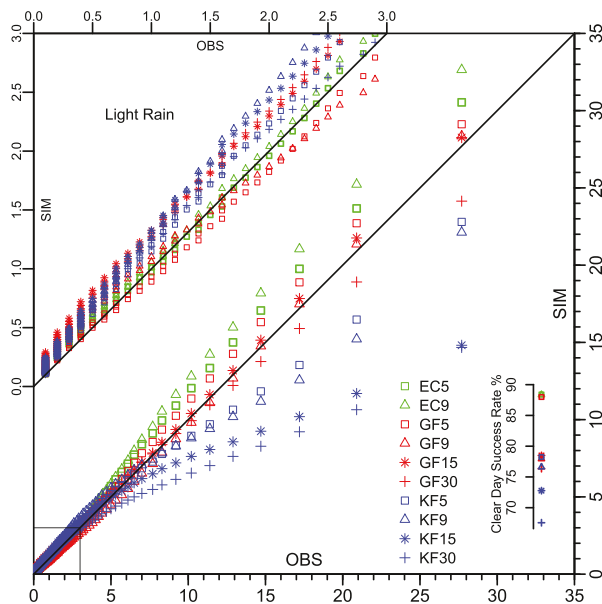


Figure 10. Same as Figure 9 except for hourly rainfall (mm hr^{-1}) quantiles.

variation with double peaks. It was greatly enhanced near the afternoon (secondary) peak as compared to the early morning (primary) peak of the mean rainfall diurnal cycle (Figure 5). This greater RMSE peak in the afternoon did not appear to have been caused by the greater penalty to the overestimation than underestimation of rainfall rate. Quite the opposite, KF produced notably larger mean rainfall rates but smaller RMSE than GF in the afternoon (Figure 5). The RMSE diurnal variation generally followed that of the observed spatial deviation. Larger spatial deviations in the afternoon than in the early morning corresponded to more local or scattered convection. The latter is generally more difficult to forecast and also more sensitive to the choice of explicit or parameterized treatment. Another factor leading to the RMSE enhancement is that forecast errors usually grow with time, which can be seen here through comparison of the same morning hours (02:00–08:00) of the first and second days of the forecast.

There were large differences in RMSE between forecasts with different configurations of grid spacing and convection treatment in the parent domain. EC(9, 5) yielded the worst forecasts and the largest RMSE over most of the period, except for the last 12 hr. GF(9, 5) performed almost as poorly, while KF(9, 5) generally had smaller RMSE. This may suggest that cumulus parameterization is necessary to represent organized convection in the gray zone. KF(30, 15) performed the best, with the lowest RMSE in the first 30 hr of forecast, especially between 22 and 27 hr. GF(30, 15) were systematically worse than KF(30, 15), especially between 21 and 30 hr. During this window, the difference in RMSE resembled that of the spatial average rain rates, in that KF was much closer to observations than GF (Figure 5). On the other hand, KF generated smaller RMSE than EC and GF between 14 and 22 hr, contrasting its much larger overestimation of the average rain rates (Figure 5). This contrast implies that EC and GF may predict rainfall or clear events in incorrect locations more often than KF, albeit producing smaller spatial mean biases.

Figure 13 compares the spatiotemporal RMSE of hourly rainfall forecasts from all double nesting D1–D3 configurations. The sensitivity to the parent grid spacing and convection treatment was notably reduced. However, errors in the coarser parent grid were not improved, perhaps signifying that local rainfall events

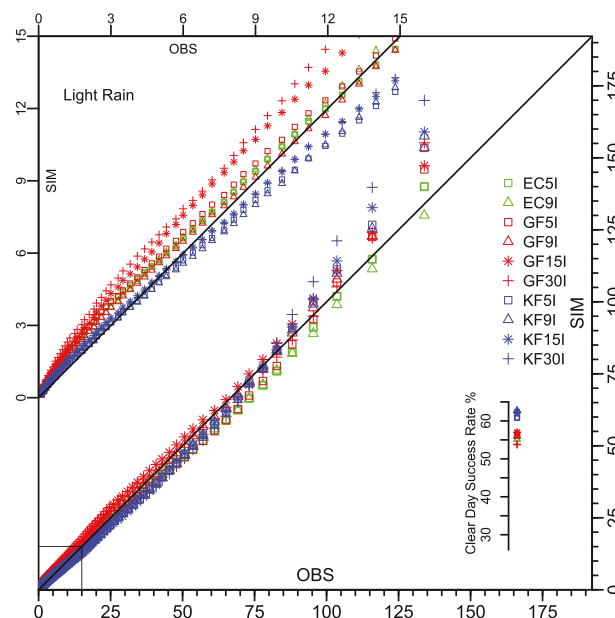


Figure 11. Same as Figure 9 except for WRF forecasts using all double nesting D1–D3 configurations.

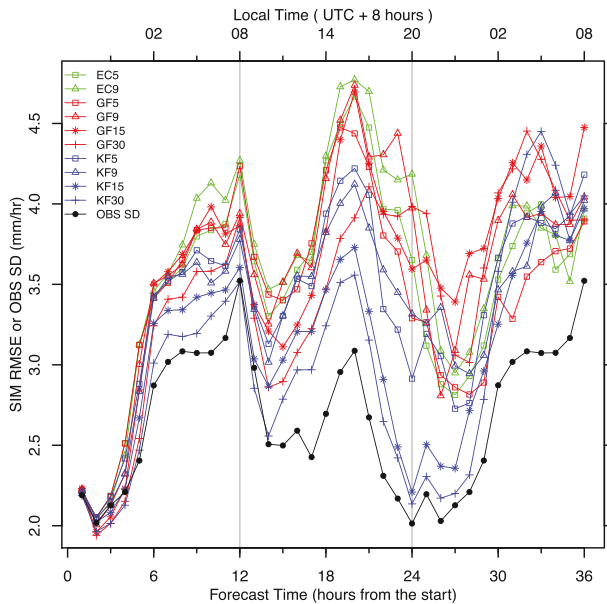


Figure 12. Variations with the forecast time (including the spin-up) of hourly rainfall (mm hr^{-1}) RMSE based on all stations and 32 days of the forecasts by WRF using all D1 configurations. Also shown is the observed spatial standard deviation (OBS SD). The corresponding local time is scaled at the top.

are difficult to predict and observational stations are not dense enough. KF15I performed better than KF5I, except between 18 and 27 forecast hours. GF15I performed more poorly than KF15I in most hours. EC5I performed worse than KF(15, 5)I for the initial 24 hr but better afterward, indicating that organized or parameterized convection may be more important in the daytime than the nighttime. Most of the differences between child grids KF(15, 5)I, GF15I, and EC5I can be traced back to the differences in the RMSE of their parent grids (Figure 12), indicating that the influence of nesting is nontrivial. However, GF30I performed the worst during the initial 24 hr but performed better than GF15I between 25 and 28 hr, despite the fact that their parent grids GF(30, 15) performed oppositely (Figure 12). The sensitivity to parent grid spacing was small using KF but remained large after 18 hr using GF. Thus, the influence of the parent grid was more complicated when using GF than KF. In agreement with our earlier finding, triple nesting using an intermediate grid did not improve RMSE (Figure S5). The addition of M3 increased KF15I errors between 6–12 and 18–24 forecast hours, corresponding to the primary and secondary rainfall diurnal peak phases, respectively. The addition of M3 and M5 also increased GF15I errors between 18 and 23 hr. The only apparent gain with these additional grids was the reduction of forecast errors between 24 and 33 hr.

Figure 14 compares the frequency distributions of station-wise correlations of observed hourly rainfall variations with forecasts from all D1 configurations. The correlations were calculated using the time series of 32 days (a total of 768 hourly samples) at all 1,238 individual stations. These distributions are depicted by density functions, below which the integrated area equals to 1. The inset depicts their added values respective to EC5, following Kanamitsu and DeHaan (2011). The added value of a specific configuration is defined as the total area enclosed by its respective function to the right wing of the EC5 function. Positive values indicate more stations with high correlations, and thus greater skill than EC5, while negative values denote the opposite. KF(30, 15, 9) showed large improvements over EC5, with added values larger than 0.1 (or 10% of stations). GF9 improved by a similar amount, while GF30 gained even more skill (~18%). However, GF15 made no

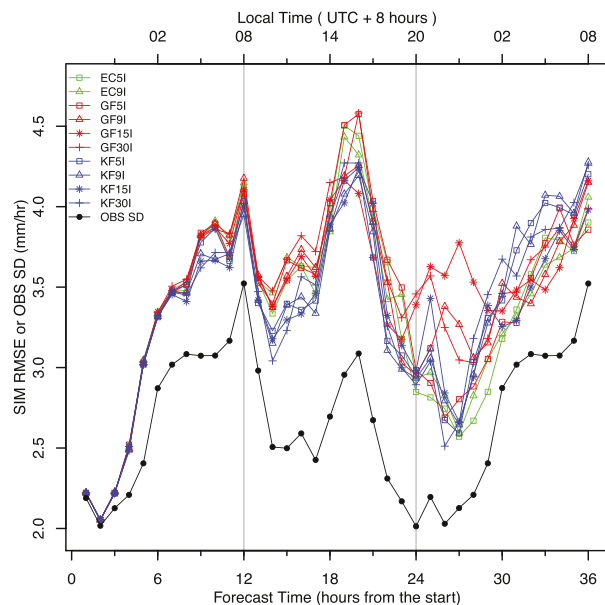


Figure 13. Same as Figure 12 except for WRF forecasts using all double nesting D1–D3 configurations.

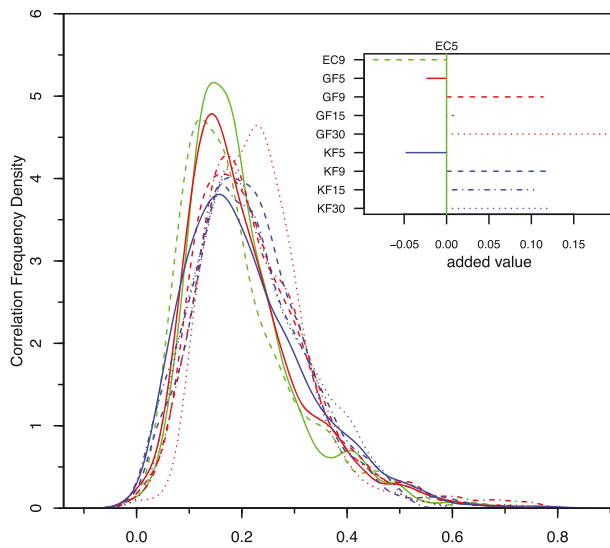


Figure 14. Frequency density functions of observed hourly rainfall temporal (24 hr $\times$ 32 days) correlations at all 1,238 stations with WRF forecasts using all D1 configurations. The inset depicts their added values, defined as the total area enclosed by their respective functions to the right wing of the EC5 function. The color dash legend for the model configurations is shown on the left and top axes of the inset.

improvement over EC5, KF5, and GF5 degraded somewhat, and EC9 lost significant skill ($\sim 10\%$). This strongly indicates that cumulus parameterization is needed for grid spacing above 5 km but is not beneficial below that threshold. Furthermore, double nesting with a 1-km grid (Figure 15) added more value (9–11%) to EC5 for KF(30,15)I but reduced the value (7–11%) for GF(30,15)I. Double nesting GF5I also degraded EC5I forecast skill (6%), but KF5I added a little (2%). Thus, the explicit solution was favorable for convection treatment at grid spacing below 5 km, and parent cumulus parameterization had a nontrivial influence on the child grid forecast. On the other hand, triple nesting with a middle grid M3 reduced both (KF, GF)15I values by $\sim 5\%$, while adding M5 had little effect on GF15I (Figure S6).

Operational verification for deterministic precipitation prediction most often uses skill scores based on the 2×2 contingency frequencies of dichotomous forecasts, as illustrated in Table 2. In particular, the threat score (TS) and bias score (BS) are most popular (Wilks, 2011). While TS and BS measure, respectively, the accuracy and discrepancy between frequencies of forecasts and observations, we here also adopt the tetrachoric correlation coefficient (TCC) to depict their association (Juras & Pasarić, 2006). To alleviate their sensitivity to climatological frequency variations and under- or over-biased forecasts (Brill & Mesinger, 2009; Fekri & Yau, 2016; Hamill & Juras, 2006), we calculated these scores for each of six mutually exclusive precipitation categories (mm day^{-1}): clear [0, 0.1), light [0.1, 10), moderate [10, 25), large [25, 50), heavy [50, 100), and extreme [100 beyond). Daily rainfall forecast was paired with the respective observation at each individual station, and a binary value was assigned to depict whether or not both belonged to a specific category. Following the operational practice, TS and BS were first calculated from these daily binary frequencies (a, b, c, d) over all 1,238 stations and then averaged over all 32 days to yield the final scores. Since TCC is inferred iteratively by integrating its bivariate normal probability density function to give the joint forecast-observation frequency (a), all 1,238 stations and 32 days (for a total 39,616 samples) of the binary values were used to calculate (a, b, c, d) once before inferring the final score. This reduces uncertainty due to inadequate samples. As a companion to TCC, the polychoric correlation coefficient (PCC) offers the overall skill score for the association between forecast and observation frequencies across all precipitation categories.

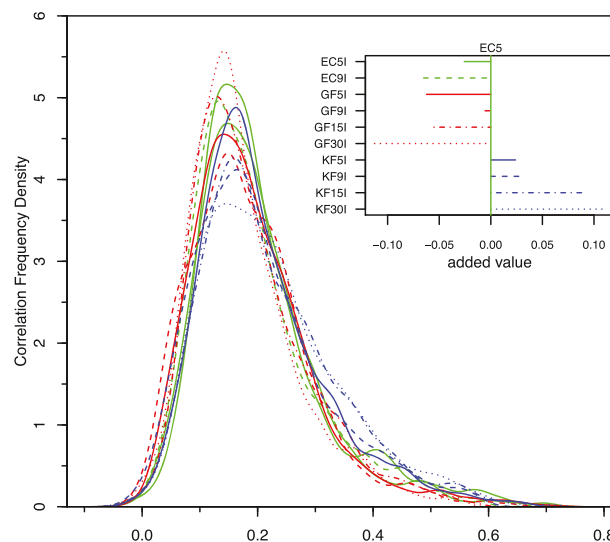


Figure 15. Same as Figure 14 except for WRF forecasts using all double nesting D1–D3 configurations.

Table 2
Definition of the Skill Scores Based on Dichotomous Forecasts

Rainfall Event $r_L \leq r < r_U$	Observation			Variable
	Yes	No	Sum	
	Yes	a	b	F
	No	c	d	1-F
Forecast	Sum	O	1-O	1
• a = hit • b = false alarm • c = miss • d = correct negative • r = rainfall in the bound $[r_L, r_U)$ for a given category				
Skill score definition				Reference
Threat score (TS) = $\frac{a}{a+b+c}$				Wilks (2011)
Bias score (BS) = $\frac{a+b}{a+c}$				0 ≤ TS ≤ 1 <1: under-forecast >1: over-forecast
Tetrachoric correlation coefficient (TCC) = the correlation coefficient that defines the bivariate normal probability density function to yield the integration over the four quadrants with the given relative frequencies equal a.				-1 ≤ TCC ≤ 1 -1 ≤ PCC ≤ 1
Polychoric correlation coefficient (PCC) = the correlation coefficient (like TCC) between standard normal deviates across all exclusive categories.				Juras and Pasaric (2006)
True skill statistics (TSS) = $\frac{a}{a+c} - \frac{b}{b+d}$				-1 ≤ TSS ≤ 1 Wilks (2011)
Information-theoretical score (ITS) = $\frac{C^+ + C^-}{S(O)}$				0 ≤ ITS ≤ 1 Fekri and Yau (2016)
$C^+ = a \log \frac{a}{(a+b)(a+c)} + d \log \frac{d}{(d+b)(d+c)}$				
$C^- = b \log \frac{b}{(b+a)(b+d)} + c \log \frac{c}{(c+d)(c+a)}$				
$S(O) = -(a+c) \log(a+c) - (b+d) \log(b+d)$				
Symmetric extremal dependence index (SEDI) = $\frac{\log F - \log H - \log(1-F) + \log(1-H)}{\log F + \log H + \log(1-F) + \log(1-H)}$				-1 ≤ SEDI ≤ 1 Ferro and Stephenson (2011)
$H = \frac{a}{a+c}, F = \frac{b}{b+d}$				

Figure 16 compares TS and BS scores of daily rainfall forecasts from all double nesting D1–D3 configurations. The observed frequencies for clear sky and light rain were 38.73% and 33.58%, respectively. For the both categories, KF(30, 15, 9, 5)I had systematically higher TS scores than GF(30, 15, 5)I and even EC5I. The corresponding BS scores were close to 1 (realistic) for KF, a little smaller (underestimated) for GF forecasts of clear conditions, and slightly larger (overestimated) for both GF and KF forecasts of light rain events. TS scores were substantially reduced as precipitation intensity increased, from ~0.4 (clear) and ~0.3 (light) to ~0.09 (heavy) and ~0.05 (extreme), although BS scores remained close to 1. The observed frequency of the heavy or extreme category was relatively small (6.67%) or rare (2.06%), for which KF still forecasted better than GF. The moderate and large rainfall categories with the observed frequencies of 11.27% and 7.69% were more problematic. TS scores remained low (~0.08–0.1) but BS scores varied from 3 to 6, indicating notable over-forecasts of the frequencies. That is, the forecasts had many more false alarms than misses. In both cases, the forecasts using KF were more realistic than those using GF or EC. Overall, KF(15,9)I performed best. The addition of M3 degraded the KF15I performance, which was especially evident for clear, light, moderate, and extreme events (Figure S7). Adding M(3, 5) also degraded the GF15I performance for moderate and heavy events, albeit with slight improvements for clear and light events and few changes for others.

In the D1 configurations (Figure 17), KF showed strong and consistent dependence on resolution. As grid spacing increased from 5, 9, 15 to 30 km, TS continually decreased for clear-light events while BS changed little; BS continually increased for moderate-large events while TS remained unchanged; TS and BS both decreased steadily for extreme events; and TS increased while BS decreased for heavy events, with the exception of KF5. Thus, KF performed better at finer resolutions, producing generally more accurate forecasts with smaller false alarm rates, except for heavy (extreme) events, where KF30 was more realistic (producing less false alarms). On the other hand, GF exhibited no clear trend of resolution dependence, varying

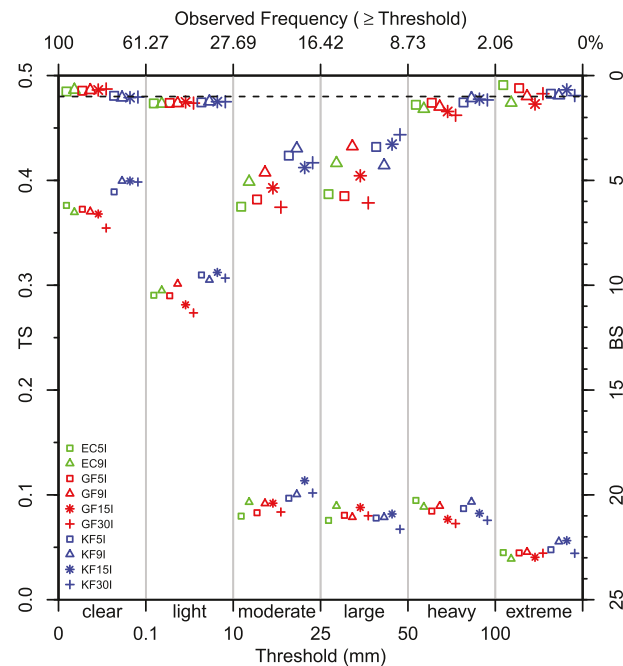


Figure 16. TS and BS scores (separated into six operationally used exclusive categories) of daily rainfall forecasts from WRF using all double nesting D1–D3 configurations. TS scores are illustrated in the lower portion with the scale at the left axis, while BS scores are shown in the upper portion with the scale at the right axis. The respective threshold (mm) is given at the bottom axis, while the corresponding observed frequency (%) for all events with rainfall exceeding the threshold is shown at the top axis. For clarity, different configurations are displayed by a small horizontal offset, and BS scores use a larger size of central markers than those for TS scores.

substantially among rainfall categories. EC(5, 9) and GF5 (equivalent to EC5 by design) produced the highest TS scores for clear-light events, and the BS scores closest to 1 for large-heavy-extreme events. These results indicate that using explicit convection at finer resolutions may increase forecast accuracy

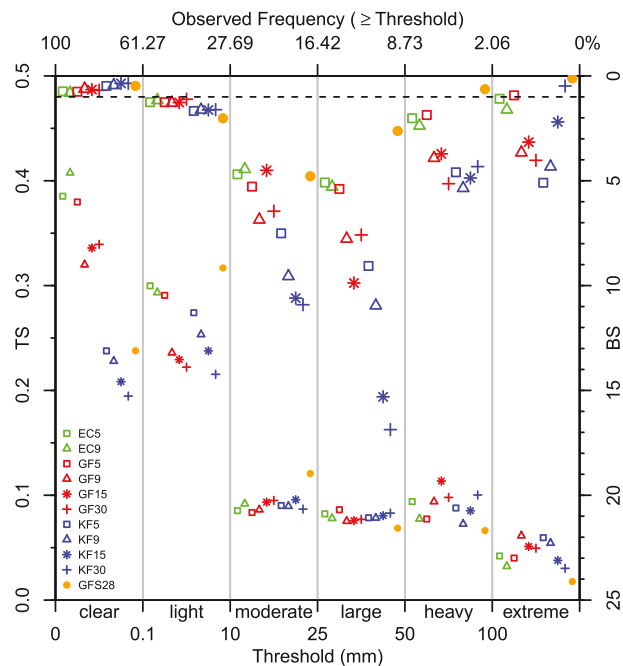


Figure 17. Same as Figure 16 except for WRF forecasts using all D1 configurations.

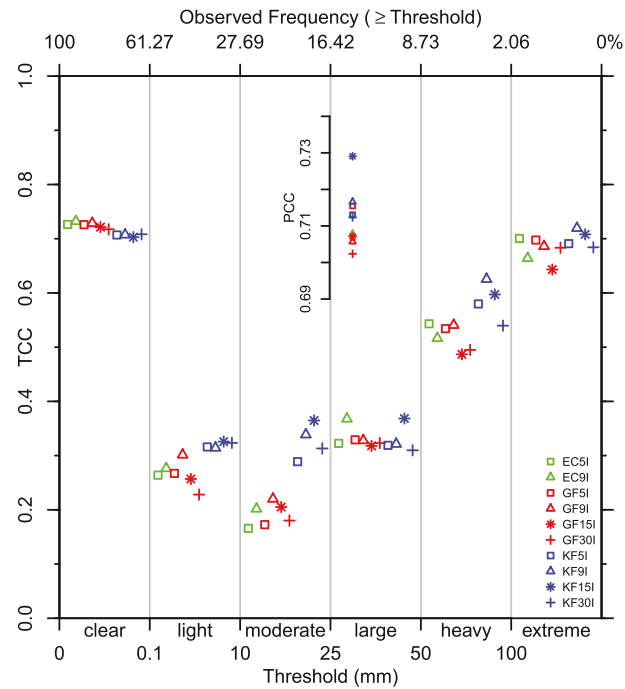


Figure 18. Same as Figure 16 except for only TCC scores. The upper-middle inset depicts the overall PCC scores across the entire rainfall range.

and reduce false alarms. Surprisingly, the driving GFS produced higher TS scores than all downscaling WRF forecasts at 5–30 km grids for light-moderate events but lower scores for all other categories. This implies that GFS achieves more accurate forecasts of light-moderate rainfall at the expense of under-forecasting clear and heavy-extreme events.

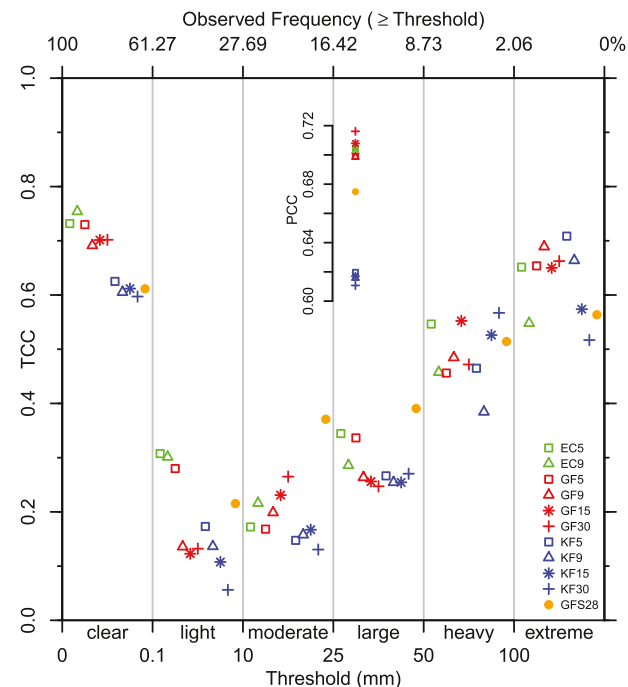


Figure 19. Same as Figure 16 except for only TCC scores of WRF forecasts using all D1 configurations. The upper-middle inset depicts the overall PCC scores across the entire rainfall range.

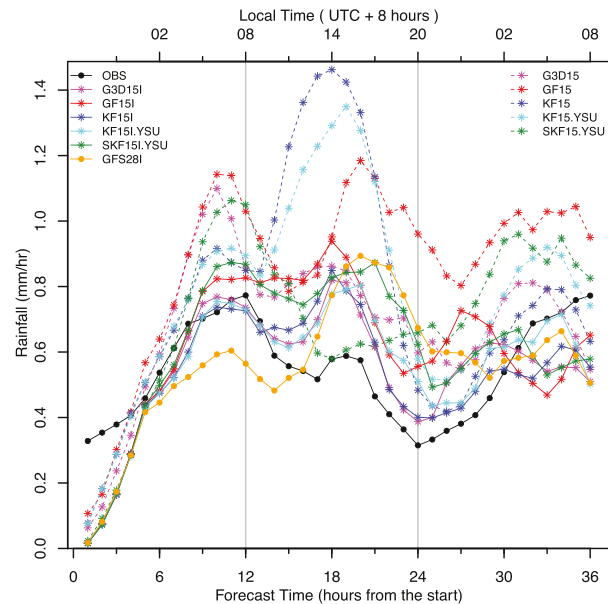


Figure 20. Same as Figure 5 except for WRF forecasts using KF, SKF, G3D, and GF [O15] and [O15I] configurations. Also included is single nesting GFS28 directly to [I].

Figure 18 compares TCC and PCC scores of daily rainfall forecasts from all double nesting D1–D3 configurations. For clear events, TCC scores were very high for both KF (~0.71) and GF (~0.72) in all grids. Extreme events also received high TCC scores, ranging from 0.68 to 0.73 when using KF and 0.64 to 0.70 using GF. The scores were lower for heavy events, ranging from 0.52 to 0.63 using KF and 0.49 to 0.52 using GF. For both heavy and extreme categories, KF9I scored the highest, followed closely by KF15I, (EC, GF)5I performed more poorly, and GF15I showed the least skill. Thus, KF(9, 15)I showed high forecast skills for clear, extreme, and heavy rainfall events and captured these events with spatiotemporal occurrences that strongly correlated with observations. Such correlations dropped significantly in the other rainfall categories. TCC for light-moderate rains ranged from 0.30 to 0.36 when using KF and was further reduced by GF to only 0.17 to 0.25. For the range of moderate-large rains, the reduced TCC scores (indicating weaker associations) corresponded to increased BS values (indicating over-forecasts). In this range, KF15I produced higher TCC scores, clearly outperforming all others. KF15I also received the highest PCC score across the entire rainfall range, while GF15I ranked second lowest. The addition of M3 reduced KF15I's TCC scores (especially greatly for light-moderate and heavy-extreme events) and so the overall PCC skill (Figure S8). In contrast, the addition of M(3, 5) generally increased GF15I's scores, except for moderate events.

In the D1 configurations (Figure 19), TCC revealed strong resolution dependence. When increasing grid spacing from 5, 9, 15 to 30 km, KF's scores continuously decreased for clear-light and extreme events, whereas GF's scores steadily increased for moderate events but decreased for large events. For clear-light and large events, (EC, GF)5 generally received the highest TCC scores, indicating that explicitly resolving convection at finer grids improved the forecast skill for these events. Other situations exhibited mixed sensitivity to resolution. For example, KF produced its highest TCC scores at 30 km for heavy rains but at 5 km for extreme rains. The driving GFS produced higher TCC scores than all downscaling WRF forecasts at 5–30 km grids for moderate-large rain events, but the opposite occurred for other categories.

While the combination of TS, BS, and TCC scores can fully recover all information contained in the four contingency frequencies (a, b, c, d) of dichotomous precipitation forecasts for all six categories, we also compared other skill measures discussed in recent research (Table 2). These include the symmetric extremal dependence index (Ferro & Stephenson, 2011) and the true skill statistics and information-theoretical score (Fekri & Yau, 2016). Our comparison based on these skill measures reached essentially the same conclusion about the relative performances and sensitivities among all configurations as discussed above. Our conclusions differed only marginally in the degree of how one configuration departs from other(s), but the tendencies (better or worse) remained consistent. The robustness regardless of skill indices may result from our use of large data samples and exclusive precipitation categories.

8. Conclusion and Discussion

The primary objective of this study was to investigate the LAM's multi-grid nesting ability to effectively represent convections across the gray zone. Hence, we systematically evaluated the sensitivity of regional precipitation forecasts to various LAM configurations of grid nesting and convection treatment. This was demonstrated by comparing WRF forecasts similar to the operational numerical weather prediction system in Jiangsu Province of China. We selected the strong 2016 Meiyu period from 19 June to 20 July in order to evaluate forecast skill across a wide variety of convection and precipitation systems. We examined forecast skills using grid spacing from 30, 15, 9, 5, 3 to 1 km with single, double, or triple nested grids, in which the outer domain [O] applied either the traditional KF or scale-aware GF cumulus parameterizations or explicit convection solution EC. We also tested the effect of reducing the nesting grid ratio from [O] to the inner domain [I] at 1-km grid by including a middle domain [M] at 5- or 3-km grid. The evaluation concentrated on the ability to forecast rainfall spatial structures and temporal characteristics in Jiangsu, including time-averaged spatial patterns (section 4), space-averaged temporal variations (section 5), spatio-temporal intensity frequency distributions (section 6), and operationally used direct forecast-observation correspondences in time and space (section 7). We used a comprehensive set of statistical measures, including spatial and temporal correlations, root-mean-square errors, quantiles, frequency distributions and added values, and diurnal cycles, as well as operational skill scores and other advanced research indices, to quantify performance or sensitivity differences among the forecasts. Based on these analyses, we reached the following main conclusions.

First, quantitative precipitation forecasts strongly depend on model configurations of [O] grid spacing and convection treatment. EC5 most realistically captured the overall distribution of daily accumulative rainfall quantiles, while other configurations overestimated light rain events, with overestimation increasing with grid spacing and from EC to KF to GF. For daily rainfall, EC5 produced the highest TS (accuracy) scores for clear-light events and the best BS (bias) scores for large-heavy-extreme events. It also received the highest TCC (association) scores for clear-light and large rain events. Hence, explicitly resolving convection with finer grids increased forecast accuracy, reduced false alarms, and improved spatiotemporal correlations, especially for clear-light events. However, during most forecast hours EC9 yielded the largest RMSE, closely followed by GF9, while KF9 produced smaller errors. Overall, compared to the baseline of EC5, KF(9, 15, 30) and GF (9, 30) improved hourly rainfall correlation to observations at over 10% of stations, whereas EC9 reduced skill at a similar fraction of stations, and both (KF, GF)5 degraded about a half of the fraction. Therefore, cumulus parameterization was needed for coarse grids and in the gray zone but was not beneficial below 5 km.

Second, double nesting [O-I] using a large grid ratio significantly improved precipitation forecast skill and largely reduced sensitivity to [O] grid spacing. Varying the nesting ratio from 2 to 11 for the driving to [O] grids and 5 to 30 for [O] to [I] grids produced little impact on 1-km rainfall forecast skill. KF(30, 15, 9, 5)I performed even better than EC5I, systematically reducing rainfall overestimation in most of the day and frequency biases in clear-light events. This is in sharp contrast to [O] forecasts using KF, which overestimated daytime rainfall much more than EC5. Thus, it appears necessary to represent organized cumulus systems by parameterization in coarse [O] grids while nesting with fine [I] grids that can explicitly resolve smaller scale convections. KF15I also showed high forecast skills for clear and heavy-extreme rain events, whose occurrences in space and time strongly correlated with observations. Such correlations dropped notably for other rainfall categories, in which KF15I still received higher TCC scores and clearly outperformed other configurations. Across the entire rainfall range, KF15I showed the highest PCC score of all configurations and improved over EC5I hourly forecast-observation correlations at more than 10% of stations. Therefore, nesting [O] using KF to parameterize the organized cumulus system at grid spacing of 15 km with [I] explicitly resolving convection at 1 km achieved the best overall rainfall forecast in Jiangsu.

Third, using scale-aware cumulus parameterization may improve mean rainfall distribution in the parent domain without benefitting quantitative precipitation forecasts in the child domain. The scale-aware cumulus scheme GF by design yields more parameterized cumulus to explicit convection and so simulates notably smaller parameterized to total precipitation ratios than the traditional KF scheme. As a result, GF predicted a more realistic mean distribution than KF, increasing the pattern correlation with observations across all [O] grids. However, its RMSE remained substantial and was even larger than KF at the 30- and 15-km

grids, in which GF forecasted an expanded band of heavier precipitation. GF did reduce KF's overestimation of daytime rainfall but greatly enhanced nighttime overestimation. Such counter errors were stronger in coarser grids. For double nesting [O-I], GF over-forecasted light rain events more than KF. The forecast RMSE sensitivity to [O] grid spacing, while small using KF, remained large after 18 hr using GF. In most hours, GF15I had larger RMSE than KF15I. By TS, BS, and TCC scores, GF15I performed systematically worse than KF15I, and across the entire range of rainfall it also received the second lowest PCC score, while KF15I ranked the best of all [O-I] configurations. KF15I improved over EC5I hourly forecast-observation correlation at more than 7% stations, while GF15I reduced the skill. Therefore, use of traditional rather than scale-aware cumulus parameterization produced a more credible [O] mesoscale circulation that drove [I] explicit convection for a more skillful rainfall forecast in Jiangsu.

Fourth, triple nesting by adding a middle grid across the gray zone appears unnecessary and can even degrade forecasts in the inner domain due to detrimental influence of uncertain convection treatment. Nesting M3 suppressed the ability of the explicit convection solution at 1-km to improve [I] rainfall forecasts, bringing trivial benefit to both daily and hourly intensity quantiles but worsening the diurnal cycle by locking it in phase with M3. In particular, the addition of M3 increased KF15I errors during the primary and secondary diurnal peak phases, respectively, in the 6–12 and 18–24 forecast hours. This reduced KF15I skill, which was especially evident for clear, light, moderate, and extreme events. The addition of M3 also enlarged GF15I errors in the secondary phase and reduced its skills for moderate and heavy events. On the other hand, KF5I's improvement of the rainfall diurnal cycle over KF5 was systematic and much greater than that of EC5I over EC5. Thus, organized convections still played a significant role at the 5-km grid and could not be explicitly resolved but required parameterization. Therefore, the 3-km grid seems to be the threshold above which cumulus parameterization becomes more beneficial than the explicit convection solution. The double nesting with large grid ratios (15–30) can achieve performance similar to, if not better than, the traditional use of more nesting grids of smaller increments (~3). This configuration also offers a pragmatic approach to avoid the challenge of representing convection across the gray zone.

An important question raised by a reviewer is whether the scale-awareness of cumulus parameterization causes the GF and KF performance differences. This is difficult to address since the two cumulus schemes were based on substantially different formulations. Disabling the scale-awareness in GF is not straightforward as it involves process coupling and scheme tuning. More complicated is that all available schemes implement only specific but limited algorithms that cannot represent the general outcome of the scale-awareness. We chose to compare a scale-aware version of KF (hereafter SKF) developed by Zheng et al. (2016) and the non-scale-aware G3D, an improved version of Grell and Dévényi (2002), upon which GF was built (Grell & Freitas, 2014). A same set of the forecast experiments (section 3) was conducted using SKF and G3D in [O] 15-km grid nested with [I] 1-km grid. However, SKF was currently coupled with a PBL scheme of YSU (Hong et al., 2006), which differs from MYJ used in this study. Thus, an additional experiment was done using KF in [O] 15-km grid while YSU replacing MYJ in both [O] and [I] grids. Following a similar convention, these runs are named as G3D15, SKF15.YSU, KF15.YSU in single nesting, and G3D15I, SKF15I.YSU, KF15I.YSU in double nesting.

Figure 20 compares the averaged hourly rainfall evolution forecasts from all [O15] configurations. Replacing MYJ with YSU caused KF15 to overestimate less daytime but more nighttime rainfall rates, shifting the secondary peak one hour late and the minimum one hour early. Coupling with the same YSU scheme, SKF15 dramatically reduced KF15's overestimation in the daytime but significantly enhanced that in the nighttime, causing a stronger primary peak and a prolonged valley from 14:00 to 22:00. On average, SKF15 reduced KF15's (RMSE, arbe) from (10.41, 8.42) to (7.28, 3.37) mm day⁻¹ while maintaining a same corr (0.61). Substantial counter errors occurred between the daytime and nighttime, most pronouncedly from KF to SKF, less so from MYJ to YSU. In contrast, G3D resembled GF in the diurnal phase but remarkably reduced GF's overestimation in the early morning and more systematically so in the afternoon through evening. On average, GF15 increased G3D15's (RMSE, arbe) from (7.02, 1.72) to (11.67, 9.88) mm day⁻¹ while producing a higher corr (0.52, 0.67). Daily average scores (TS, BS, TCC, PCC) showed that SKF15 had overall higher skills than KF15, especially for light-large rainfall events, and so did GF15 than G3D, especially for clear-moderate and heavy-extreme events. In double nesting [I], forecast errors and sensitivities to cumulus schemes were greatly reduced. Switching MYJ to YSU had little effect in the first 13 forecast hours, notably reduced KF15's overestimation for the next 5 hr, but then more systematically enhanced that till the 30th hour.

Changing KF to SKF, however, led to persistent, larger overestimation between 6 and 30 hr and then larger underestimation. A surprisingly similar result was obtained by changing G3D to GF. Thus, SKF15I performed overall more poorly than KF15I and so did GF15I than G3D15I, both of which are opposite to their [O] skill contrast. That is, these scale-aware algorithms may improve average performance in [O15] but with large counter errors in space and time and more likely degrade forecast skills in double nesting [I]. We speculate that convective instability or available potential energy once generated must be consumed in certain pace and, if held unrealistically, would be transported to release in wrong places and times and thus produce an incorrect spatiotemporal distribution of rainfall. While cumulus parameterization is intended to realistically consume the instability and provide its heating/moistening feedback to the environment, the parameterized precipitation contains large uncertainty due to many factors such as closure assumptions and microphysical interactions. As such, KF or G3D may over-predict precipitation in [O], but as long as it can correctly represent convective effects on the mesoscale circulation that drives the nesting [I] grid, the explicit convection therein can produce more skillful rainfall forecast.

Another question concerns the aggressive nesting ratios (15–30), which are suspected to cause significant artifacts near the lateral boundaries of the nesting domain and thus contaminate the interior solution especially when using one-way nesting (Harris & Durran, 2010; Harris & Lin, 2013, 2014; Park et al., 2014). Figure S9 compares the 32-day mean 24-hr accumulative rainfall distributions over the whole D3 domain from all double and triple nesting configurations. As a reference, the observed distribution is from an objective analysis at 0.1° grid by merging satellite retrievals with measurements at available automated monitoring stations (Pan et al., 2012). This analysis underestimated Jiangsu rainfall from only but denser stations, with arbe of $-0.72 \text{ mm day}^{-1}$ or -3.8% . There were no apparent distortions near the boundaries. This could be confirmed from the close similarity near the edges across all [O] grids using a same convection treatment, including EC(5, 9)I, KF(5, 9, 15, 30)I, and GF(5, 9, 15, 30)I. Differences near the edges did occur among using different convection schemes, where EC, GF, and SKF overestimated while KF and G3D slightly underestimated. Note that adding the intermediate grids M(3, 5) introduced larger differences from (KF, GF)15I. Nonetheless, all these edge differences should not affect forecasts in Jiangsu, which is sufficiently away from the boundaries (Matte et al., 2017). Therefore, these large nesting ratios, although aggressive, should not be a problem and may cause some distortions near the boundaries of only a few grids. Even using the GFS data at $\sim 28 \text{ km}$ to directly drive WRF in D3 at 1-km grid, the edge effect was still minor. This single nesting improved Jiangsu forecast, with (corr, RMSE, arbe) from GFS28 (0.64, 4.50, -5.15) to GFS28I (0.74, 4.70, 1.51). Daily average scores (TS, BS, TCC, PCC) showed that GFS28I had higher skills across all rainfall categories than (KF, GF) 30I. However, it was less skillful than KF(30, 15)I for the diurnal cycle forecast, largely underestimated the early morning primary peak and substantially overestimated the afternoon secondary peak through the evening valley (Figures 6 and 20). This implies that using KF in WRF [O] was better than the cumulus scheme in GFS.

The findings summarized above are subject to several limitations. In particular, all forecast skills were evaluated against hourly rainfall observations from 1,238 automated monitoring stations that had full data records for the entire 32-day period. The validation data resolution is no more than 9 km and is inadequate for evaluating forecasts at finer grids. To facilitate the model comparison, all forecasts, regardless of grid spacing, were interpolated linearly onto the stations. This may have led to incomparability between model (grid average) and station (point wise) representations and warrants a more objective analysis to reduce the effect of such data uncertainty. Some results may depend on the spatiotemporal density of available data. For example, forecast skill based on intensity quantiles is sensitive not only to spatial resolution but also to rain duration. While we found that KF15 more realistically captured daily accumulation, and GF15 better captured hourly rainfall, this result may change when validating using data of higher frequency from denser stations. Moreover, all forecasts were made using the same vertical resolution in all nesting domains and the same control physics configuration as in the existing operational system. They differed only in the model configurations of horizontal grid spacing and convection treatment. Our additional experiments show that further refining vertical resolution provided little benefit to Jiangsu's rainfall forecasts. However, there is certainly room for skill enhancement by improving model physics representations. For example, a significant forecast sensitivity to the PBL parameterization was identified when replacing MYJ with YSU (discussed earlier) or the Pleim (2007) scheme. The latter substantially reduced afternoon-earlier evening forecast errors but further increased nighttime-morning overestimations. It generated much smaller convective available potential energy, with bigger reductions during the daytime, causing the explicit convection to play a

larger role than cumulus parameterization. These sensitivities coupled with system feedback processes to lead to varied rainfall responses, which we will address in upcoming papers. Nonetheless, they do not affect our four main conclusions above.

Acknowledgments

The simulations and analyses were conducted on supercomputers, including the Maryland Advanced Research Computing Center's Bluecrab, the University of Maryland Deepthought HPC clusters, and the Computational and Information Systems Lab of the National Center for Atmospheric Research. The research was supported by the Jiangsu Institute of Meteorological Sciences subcontract to Nanjing University of Information Science and Technology (NJCAR2016ZD03), Jiangsu 333 Project (BRA2018100), and the China Scholar Council fellowship to Qi Li and the National Natural Science Foundation of China Grant (41805077) to Haixia Mei for their visits at the University of Maryland. Liang was partially supported by U.S. National Science Foundation Innovations at the Nexus of Food, Energy, and Water Systems under Grants EAR1639327 and EAR1903249. We sincerely thank Jennifer Kennedy for thorough editing. We all thank Lucas Harris and other anonymous reviewers for their constructive comments and suggestions. The observational data and model simulations used in this study is accessible at <ftp://earthserver.umd.edu/publish>.

References

- Amengual, A., Romero, R., Homar, V., Ramis, C., & Alonso, S. (2007). Impact of the lateral boundary conditions resolution on dynamical downscaling of precipitation in mediterranean Spain. *Climate Dynamics*, 29(5), 487–499. <https://doi.org/10.1007/s00382-007-0242-0>
- Antic, S., Laprise, R., Denis, B., & de Elia, R. (2006). Testing the downscaling ability of a one-way nested regional climate model in regions of complex topography. *Climate Dynamics*, 26(2–3), 305–325. <https://doi.org/10.1007/s00382-005-0046-z>
- Arakawa, A. (2004). The cumulus parameterization problem: Past, present, and future. *Journal of Climate*, 17(13), 2493–2525. [https://doi.org/10.1175/1520-0442\(2004\)017<2493:RATCPP>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2493:RATCPP>2.0.CO;2)
- Arakawa, A., & Wu, C.-M. (2013). A unified representation of deep moist convection in numerical modeling of the atmosphere. Part I. *Journal of the Atmospheric Sciences*, 70(7), 1977–1992. <https://doi.org/10.1175/JAS-D-12-0330.1>
- Barthlott, C., & Hoose, C. (2015). Spatial and temporal variability of clouds and precipitation over Germany: Multiscale simulations across the “gray zone”. *Atmospheric Chemistry and Physics*, 15(21), 12,361–12,384. <https://doi.org/10.5194/acp-15-12361-2015>
- Beck, A., Ahrens, B., & Stadlbacher, K. (2004). Impact of nesting strategies in dynamical downscaling of reanalysis data. *Geophysical Research Letters*, 31, L19101. <https://doi.org/10.1029/2004GL020115>
- Biswas, M. K., Bernardet, L., & Dudhia, J. (2014). Sensitivity of hurricane forecasts to cumulus parameterizations in the HWRF model. *Geophysical Research Letters*, 41, 9113–9119. <https://doi.org/10.1002/2014GL02071>
- Brill, K. F., & Mesinger, F. (2009). Applying a general analytic method for assessing bias sensitivity to bias-adjusted threat and equitable threat scores. *Weather and Forecasting*, 24(6), 1748–1754. <https://doi.org/10.1175/2009WAF2222272.1>
- Brisson, E., Demuzere, M., & Van Lipzig, N. P. M. (2016). Modelling strategies for performing convection-permitting climate simulations. *Meteorologische Zeitschrift*, 25(2), 149–163. <https://doi.org/10.1127/metz/2015/0598>
- Bryan, G. H., Wyngaard, J. C., & Fritsch, J. M. (2003). Resolution requirements for the simulation of deep moist convection. *Monthly Weather Review*, 131(10), 2394–2416. [https://doi.org/10.1175/1520-0493\(2003\)131<2394:RRFTSO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<2394:RRFTSO>2.0.CO;2)
- Chou, M. D., & Suarez, M. J. (1999). A solar radiation parameterization for atmospheric studies. Volume 15.
- Christensen, O. B., Christensen, J. H., Machenhauer, B., & Botzet, M. (1998). Very high-resolution regional climate simulations over Scandinavia—Present climate. *Journal of Climate*, 11(12), 3204–3229. [https://doi.org/10.1175/1520-0442\(1998\)011<3204:VHRRCS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<3204:VHRRCS>2.0.CO;2)
- Clark, A. J., Jirak, I. L., Dembek, S. R., Creagager, G. J., Kong, F., Thomas, K. W., et al. (2018). The community leveraged unified ensemble (CLUE) in the 2016 NOAA/hazardous weather testbed spring forecasting experiment. *Bulletin of the American Meteorological Society*, 99(7), 1433–1448. <https://doi.org/10.1175/BAMS-D-16-0309.1>
- Denis, B., Laprise, R., & Caya, D. (2003). Sensitivity of a regional climate model to the resolution of the lateral boundary conditions. *Climate Dynamics*, 20(2), 107–126. <https://doi.org/10.1007/s00382-002-0264-6>
- Ding, Y. (1991). Monsoons over China. *Atmospheric and Oceanographic Sciences Library*, 419(2), 252.
- Dudek, M. P., Liang, X.-Z., & Wang, W.-C. (1996). A regional climate model study of the scale dependence of cloud-radiation interactions. *Journal of Climate*, 9(6), 1221–1234. [https://doi.org/10.1175/1520-0442\(1996\)009<1221:ARCMO>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<1221:ARCMO>2.0.CO;2)
- Fekri, M., & Yau, M. K. (2016). An information-theoretical score of dichotomous precipitation forecast. *Monthly Weather Review*, 144(4), 1633–1647. <https://doi.org/10.1175/MWR-D-15-0225.1>
- Ferro, C. A. T., & Stephenson, D. B. (2011). Extremal dependence indices: Improved verification measures for deterministic forecasts of rare binary events. *Weather and Forecasting*, 26(5), 699–713. <https://doi.org/10.1175/waf-d-10-05030.1>
- Field, P. R., Brožková, R., Chen, M., Dudhia, J., Lac, C., Hara, T., et al. (2017). Exploring the convective grey zone with regional simulations of a cold air outbreak. *Quarterly Journal of the Royal Meteorological Society*, 143(707), 2537–2555. <https://doi.org/10.1002/qj.3105>
- Fierro, A. O., Rogers, R. F., Marks, F. D., & Nolan, D. S. (2009). The impact of horizontal grid spacing on the microphysical and kinematic structures of strong tropical cyclones simulated with the WRF-ARW model. *Monthly Weather Review*, 137(11), 3717–3743. <https://doi.org/10.1175/2009MWR2946.1>
- Fowler, L. D., Skamarock, W. C., Grell, G. A., Freitas, S. R., & Duda, M. G. (2016). Analyzing the Grell–Freitas convection scheme from hydrostatic to nonhydrostatic scales within a global model. *Monthly Weather Review*, 144(6), 2285–2306. <https://doi.org/10.1175/MWR-D-15-0311.1>
- Freitas, S. R., Grell, G. A., Molod, A., Thompson, M. A., Putman, W. M., Santos e Silva, C. M., & Souza, E. P. (2018). Assessing the Grell–Freitas convection parameterization in the NASA GEOS modeling system. *Journal of Advances in Modeling Earth Systems*, 10(6), 1266–1289. <https://doi.org/10.1029/2017MS001251>
- Gao, Y., Leung, L. R., Zhao, C., & Hagos, S. (2017). Sensitivity of U.S. summer precipitation to model resolution and convective parameterizations across gray zone resolutions. *Journal of Geophysical Research: Atmospheres*, 122(5), 2714–2733. <https://doi.org/10.1002/2016JD025896>
- Gerard, L. (2007). An integrated package for subgrid convection, clouds and precipitation compatible with meso-gamma scales. *Quarterly Journal of the Royal Meteorological Society*, 133(624), 711–730. <https://doi.org/10.1002/qj.58>
- Giorgi, F. (2006). Regional climate modeling: Status and perspectives. *Journal de Physique IV (Proceedings)*, 139(1), 101–118. <https://doi.org/10.1051/jp4:2006139008>
- Giorgi, F., & Marinucci, M. R. (1996). A investigation of the sensitivity of simulated precipitation to model resolution and its implications for climate studies. *Monthly Weather Review*, 124(1), 148–166. [https://doi.org/10.1175/1520-0493\(1996\)124<0148:AIOTSO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<0148:AIOTSO>2.0.CO;2)
- Giorgi, F., Marinucci, M. R., Bates, G. T., & De Canio, G. (1993). Development of a second-generation regional climate model (RegCM2). Part II: Convective processes and assimilation of lateral boundary conditions. *Monthly Weather Review*, 121(10), 2814–2832. [https://doi.org/10.1175/1520-0493\(1993\)121<2814:DOASGR>2.0.CO;2](https://doi.org/10.1175/1520-0493(1993)121<2814:DOASGR>2.0.CO;2)
- Giorgi, F., & Mearns, L. O. (1999). Introduction to special section: Regional climate modeling revisited. *Journal of Geophysical Research*, 104(D6), 6335–6352. <https://doi.org/10.1029/98JD02072>
- Gopalakrishnan, S. G., Goldenberg, S., Quirino, T., Zhang, X., Marks, F. Jr., Yeh, K. S., et al. (2012). Toward improving high-resolution numerical hurricane forecasting: Influence of model horizontal grid resolution, initialization, and physics. *Weather and Forecasting*, 27(3), 647–666. <https://doi.org/10.1175/WAF-D-11-00055.1>

- Grell, G. A., & Dévényi, D. (2002). A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters*, 29(14), 1693. <https://doi.org/10.1029/2002GL015311>
- Grell, G. A., & Freitas, S. R. (2014). A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmospheric Chemistry and Physics*, 14, 5233–5250. <https://doi.org/10.5194/acp-14-5233-2014>
- Grell, G. A., Freitas, S. R., Studies, C., Paulista, C., & Paulo, S. (2014). A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmospheric Chemistry and Physics*, 14(10), 5233–5250. <https://doi.org/10.5194/acp-14-5233-2014>
- Hamill, T. M., & Juras, J. (2006). Measuring forecast skill: Is it real skill or is it the varying climatology? *Quarterly Journal of the Royal Meteorological Society*, 132(621C), 2905–2923. <https://doi.org/10.1256/qj.06.25>
- Han, J., & Pan, H.-L. (2011). Revision of convection and vertical diffusion schemes in the NCEP global forecast system. *Weather and Forecasting*, 26(4), 520–533. <https://doi.org/10.1175/WAF-D-10-05038.1>
- Han, J., Wang, W., Kwon, Y. C., Hong, S.-Y., Tallapragada, V., & Yang, F. (2017). Updates in the NCEP GFS cumulus convection schemes with scale and aerosol awareness. *Weather and Forecasting*, 32(5), 2005–2017. <https://doi.org/10.1175/WAF-D-17-0046.1>
- Han, J.-Y., & Hong, S.-Y. (2018). Precipitation forecast experiments using the Weather Research and Forecasting (WRF) model at gray-zone resolutions. *Weather and Forecasting*, 33(6), 1605–1616. <https://doi.org/10.1175/WAF-D-18-0026.1>
- Harris, L. M., & Durran, D. R. (2010). An idealized comparison of one-way and two-way grid nesting. *Monthly Weather Review*, 138(6), 2174–2187. <https://doi.org/10.1175/2010MWR3080.1>
- Harris, L. M., & Lin, S.-J. (2013). A two-way nested global-regional dynamical core on the cubed-sphere grid. *Monthly Weather Review*, 141, 283–306. <https://doi.org/10.1175/MWR-D-11-00201.1>
- Harris, L. M., & Lin, S.-J. (2014). Global-regional nested grid climate simulations in the GFDL high resolution atmospheric model. *Journal of Climate*, 27, 4890–4910. <https://doi.org/10.1175/JCLI-D-13-00596.1>
- Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, 134, 2318–2341. <https://doi.org/10.1175/MWR3199.1>
- Janjic, Z. I. (2001). Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP Meso model, NCEP Office Note, No. 437, 61 pp.
- Jung, J.-H., & Arakawa, A. (2004). The resolution dependence of model physics: Illustrations from nonhydrostatic model experiments. *Journal of the Atmospheric Sciences*, 61(1), 88–102. [https://doi.org/10.1175/1520-0469\(2004\)061<0088:TRDOMP>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<0088:TRDOMP>2.0.CO;2)
- Juras, J., & Pasarić, Z. (2006). Application of tetrahoric and polychoric correlation coefficients to forecast verification. *Geofizika*, 23(1), 59–82. Retrieved from <http://meetings.copernicus.org/www.cosis.net/abstracts/EGU06/01407/EGU06-J-01407.pdf>
- Kain, J. S. (2004). The Kain–Fritsch convective parameterization: An update. *Journal of Applied Meteorology*, 43(1), 170–181. [https://doi.org/10.1175/1520-0450\(2004\)043<0170:TKCPAU>2.0.CO;2](https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2)
- Kain, J. S., Weiss, S. J., Bright, D. R., Baldwin, M. E., Levit, J. J., Carbin, G. W., et al. (2008). Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Weather and Forecasting*, 23(5), 931–952. <https://doi.org/10.1175/WAF2007106.1>
- Kanamitsu, M., & DeHaan, L. (2011). The added value index: A new metric to quantify the added value of regional models. *Journal of Geophysical Research*, 116, D11106. <https://doi.org/10.1029/2011JD015597>
- Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., et al. (2005). The two-way nested global chemistry-transport zoom model TM5: Algorithm and applications. *Atmospheric Chemistry and Physics*, 5(2), 417–432. <https://doi.org/10.5194/acp-5-417-2005>
- Kwon, Y. C., & Hong, S.-Y. (2017). A mass-flux cumulus parameterization scheme across gray-zone resolutions. *Monthly Weather Review*, 145(2), 583–598. <https://doi.org/10.1175/MWR-D-16-0034.1>
- Leung, L. R., & Qian, Y. (2003). The sensitivity of precipitation and snowpack simulations to model resolution via nesting in regions of complex terrain. *Journal of Hydrometeorology*, 4(6), 1025–1043. [https://doi.org/10.1175/1525-7541\(2003\)004<1025:TSOPAS>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<1025:TSOPAS>2.0.CO;2)
- Leutwyler, D., Lüthi, D., Ban, N., Fuhrer, O., & Schär, C. (2017). Evaluation of the convection-resolving climate modeling approach on continental scales. *Journal of Geophysical Research: Atmospheres*, 122, 5237–5258. <https://doi.org/10.1002/2016JD026013>
- Liang, X.-Z., Kunkel, K. E., & Samel, A. N. (2001). Development of a regional climate model for U.S. midwest applications. Part I: Sensitivity to buffer zone treatment. *Journal of Climate*, 14(23), 4363–4378. [https://doi.org/10.1175/1520-0442\(2001\)014<4363:DOARCM>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<4363:DOARCM>2.0.CO;2)
- Liang, X.-Z., Sun, C., Zheng, X., Dai, Y., Xu, M., Choi, H. I., et al. (2019). CWRf performance at downscaling China climate characteristics. *Climate Dynamics*, 52(3–4), 2159–2184. <https://doi.org/10.1007/s00382-018-4257-5>
- Liang, X.-Z., Xu, M., Yuan, X., Ling, T., Choi, H. I., Zhang, F., et al. (2012). Regional Climate–Weather Research and Forecasting model. *Bulletin of the American Meteorological Society*, 93(9), 1363–1387. <https://doi.org/10.1175/BAMS-D-11-00180.1>
- Liu, S., Liang, X.-Z., Gao, W., & Zhang, H. (2008). Application of Climate–Weather Research and Forecasting model (CWRf) in China: Domain optimization. *Chinese Journal of Atmospheric Sciences (in Chinese)*, 32(3), 457. <https://doi.org/10.3878/j.issn.1006-9895.2008.03.04>
- Mahoney, K. M. (2016). The representation of cumulus convection in high-resolution simulations of the 2013 Colorado Front Range flood. *Monthly Weather Review*, 144(11), 4265–4278. <https://doi.org/10.1175/MWR-D-16-0211.1>
- Mass, C. F., Ovens, D., Westrick, K., & Colle, B. A. (2002). Does increasing horizontal resolution produce more skillful forecasts? *Bulletin of the American Meteorological Society*, 83(3), 407–430. [https://doi.org/10.1175/1520-0477\(2002\)083<0407:DIHRPM>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2)
- Matte, D., Laprise, R., & Thériault, J. M. (2016). Comparison between high-resolution climate simulations using single- and double-nesting approaches within the Big-Brother experimental protocol. *Climate Dynamics*, 47(12), 3613–3626. <https://doi.org/10.1007/s00382-016-3031-9>
- Matte, D., Laprise, R., Thériault, J. M., & Lucas-Picher, P. (2017). Spatial spin-up of fine scales in a regional climate model simulation driven by low-resolution boundary conditions. *Climate Dynamics*, 49(1–2), 563–574. <https://doi.org/10.1007/s00382-016-3358-2>
- Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, 102(D14), 16,663–16,682. <https://doi.org/10.1029/97JD00237>
- Molinari, J., & Dudek, M. (1992). Parameterization of convective precipitation in mesoscale numerical models: A critical review. *Monthly Weather Review*, 120(2), 326–344. [https://doi.org/10.1175/1520-0493\(1992\)120<0326:POCPIM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<0326:POCPIM>2.0.CO;2)
- Pan, Y., Shen, Y., Yu, J., & Zhao, P. (2012). Analysis of the combined gauge-satellite hourly precipitation over China based on the OI technique. *Acta Meteorologica Sinica*, 70(6), 1381–1389.
- Park, S.-H., Klemp, J. B., & Skamarock, W. C. (2014). A comparison of mesh refinement in the global MPAS-A and WRF models using an idealized normal-mode baroclinic wave simulation. *Monthly Weather Review*, 142, 3614–3634.

- Petch, J. C., Brown, A. R., & Gray, M. E. B. (2002). The impact of horizontal resolution on the simulations of convective development over land. *Quarterly Journal of the Royal Meteorological Society*, 128(584), 2031–2044. <https://doi.org/10.1256/003590002320603511>
- Pleim, J. E. (2007). A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. *Journal of Applied Meteorology and Climatology*, 46(9), 1383–1395. <https://doi.org/10.1175/JAM2539.1>
- Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., et al. (2015). A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of Geophysics*, 53, 323–361. <https://doi.org/10.1002/2014RG000475>
- Randall, D., Khairoutdinov, M., Arakawa, A., & Grabowski, W. (2003). Breaking the cloud parameterization deadlock. *Bulletin of the American Meteorological Society*, 84(11), 1547–1564. <https://doi.org/10.1175/BAMS-84-11-1547>
- Samel, A. N., Wang, W.-C., & Liang, X.-Z. (1999). The monsoon rainband over China and relationships with the Eurasian circulation. *Journal of Climate*, 12(1), 115–131. <https://doi.org/10.1175/1520-0442-12.1.115>
- Schwartz, C. S., Kain, J. S., Weiss, S. J., Xue, M., Bright, D. R., Kong, F., et al. (2009). Next-day convection-allowing WRF model guidance: A second look at 2-km versus 4-km grid spacing. *Monthly Weather Review*, 137(10), 3351–3372. <https://doi.org/10.1175/2009MWR2924.1>
- Seth, A., & Giorgi, F. (1998). The effects of domain choice on summer precipitation simulation and sensitivity in a regional climate model. *Journal of Climate*, 11(10), 2698–2712. [https://doi.org/10.1175/1520-0442\(1998\)011<2698:TEODCO>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<2698:TEODCO>2.0.CO;2)
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., et al. (2008). A description of the advanced research WRF version 3 (No. NCAR/TN-475+STR). Retrieved from http://www2.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf
- Tao, S.-Y. (1987). A review of recent research on the East Asian summer monsoon in China. *Monsoon Meteorology*, 60–92.
- Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M. A., Mitchell, K., et al. (2004). Implementation and verification of the unified NOAA land surface model in the WRF model. In 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction (Vol. 1115).
- Thompson, G., Field, P. R., Rasmussen, R. M., & Hall, W. D. (2008). Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Monthly Weather Review*, 136(12), 5095–5115. <https://doi.org/10.1175/2008MWR2387.1>
- Warner, T., Peterson, R., & Treadon, R. E. (1997). A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bulletin of the American Meteorological Society*, 78(11), 2599–2617. Retrieved from [https://doi.org/10.1175/1520-0477\(1997\)078<2599:ATOLBC>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<2599:ATOLBC>2.0.CO;2)
- Warner, T. T., & Hsu, H.-M. (2000). Nested-model simulation of moist convection: The impact of coarse-grid parameterized convection on fine-grid resolved convection. *Monthly Weather Review*, 128(7), 2211–2231. [https://doi.org/10.1175/1520-0493\(2000\)128<2211:NMSOMC>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<2211:NMSOMC>2.0.CO;2)
- Weisman, M. L., Skamarock, W. C., & Klemp, J. B. (1997). The resolution dependence of explicitly modeled convective systems. *Monthly Weather Review*, 125(4), 527–548. [https://doi.org/10.1175/1520-0493\(1997\)125<0527:TRDOEM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<0527:TRDOEM>2.0.CO;2)
- Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*. Amsterdam: Elsevier/Academic Press.
- Xue, Y., Janjic, Z., Dudhia, J., Vasic, R., & De Sales, F. (2014). A review on regional dynamical downscaling in intraseasonal to seasonal simulation/prediction and major factors that affect downscaling ability. *Atmospheric Research*, 147–148, 68–85. <https://doi.org/10.1016/j.atmosres.2014.05.001>
- Yu, X., & Lee, T.-Y. (2011). Role of convective parameterization in simulations of heavy precipitation systems at grey-zone resolutions—Case studies. *Asia-Pacific Journal of Atmospheric Sciences*, 47(2), 99–112. <https://doi.org/10.1007/s13143-011-0001-3>
- Zhang, D.-L., Chang, H.-R., Seaman, N. L., Warner, T. T., & Fritsch, J. M. (1986). A two-way interactive nesting procedure with variable terrain resolution. *Monthly Weather Review*, 114(7), 1330–1339. [https://doi.org/10.1175/1520-0493\(1986\)114<1330:ATWINP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1986)114<1330:ATWINP>2.0.CO;2)
- Zhao, J., Chen, L., & Wang, D. (2018). Characteristics and causes analysis of abnormal Meiyu in China in 2016. *Chinese Journal of Atmospheric Sciences (in Chinese)*, 42(5), 1055–1066. <https://doi.org/10.3878/j.issn.1006-9895.1708.17170>
- Zheng, Y., Alapaty, K., Herwehe, J. A., Del Genio, A. D., & Niyogi, D. (2016). Improving high-resolution weather forecasts using the Weather Research and Forecasting (WRF) model with an updated Kain–Fritsch scheme. *Monthly Weather Review*, 144(3), 833–860. <https://doi.org/10.1175/MWR-D-15-0005.1>