Research Article

The Sensitivity of Heavy Precipitation to Horizontal Resolution, Domain Size, and Rain Rate Assimilation: Case Studies with a Convection-Permitting Model

Xingbao Wang, Peter Steinle, Alan Seed, and Yi Xiao

Research and Development Branch, Bureau of Meteorology, Melbourne, VIC 3008, Australia

Correspondence should be addressed to Xingbao Wang; xingbao.wang@bom.gov.au

Received 5 October 2015; Revised 16 November 2015; Accepted 17 November 2015

Academic Editor: Hann-Ming H. Juang

Copyright © 2016 Xingbao Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Australian Community Climate and Earth-System Simulator (ACCESS) is used to test the sensitivity of heavy precipitation to various model configurations: horizontal resolution, domain size, rain rate assimilation, perturbed physics, and initial condition uncertainties, through a series of convection-permitting simulations of three heavy precipitation (greater than 200 mm day⁻¹) cases in different synoptic backgrounds. The larger disparity of intensity histograms and rainfall fluctuation caused by different model configurations from their mean and/or control run indicates that heavier precipitation forecasts have larger uncertainty. A cross-verification exercise is used to quantify the impacts of different model parameters on heavy precipitation. The dispersion of skill scores with control run used as “truth” shows that *the impacts of the model resolution and domain size on the quantitative precipitation forecast are not less than those of perturbed physics and initial field uncertainties in these not intentionally selected heavy precipitation cases*. The result indicates that model resolution and domain size should be considered as part of probabilistic precipitation forecasts and ensemble prediction system design besides the model initial field uncertainty.

1. Introduction

Many studies show the importance of higher resolution in improving the forecast skill. The increased resolution has the benefit of reducing numerical truncation, of explicitly resolving dynamical interactions for wider range of spatial scales, and permits the simulation of fine-scale details and hence of relying less strongly on parameterization of sub-grid scale processes. Weisman et al. [1], Done et al. [2], Lean et al. [3], Schwartz et al. [4], and many others have shown that convection-permitting models yield qualitatively more realistic precipitation fields and are quantitatively more skillful than lower-resolution simulations with parameterized convection. The results from short-range weather forecasting [5–7] also indicate that such convection-permitting models outperform those coarser resolution models that require a convective parameterization. However, due to the high computational cost of integration of high resolution model, a necessary price for higher resolution is using limited horizontal domain. The limited domains models need artificial lateral

boundaries, which introduce additional uncertainties and errors into the weather forecast. The emphasis on horizontal resolution in most mesoscale simulation is based on the implicit assumption that errors from artificial lateral boundaries and the effects of limited domain size may not play a significant role in the dynamics of the mesoscale system. Theoretically, one would like to suppose that grid spacing is fine enough to resolve the phenomena of interest and domain size is large enough that the placement of the boundaries would not unduly contaminate the solution, at least for the time scales of interest. However, studies have shown that both of the domain size and horizontal resolution of limited area model influence the spectrum of resolved scale and the nature of scale interaction in the model dynamics [8, 9]. Stevens et al. [10] used a cloud model studying the sensitivity of shallow cumulus convection to model domain and resolution and found that domain size and resolution have big impacts on the cloud structure and their statistical characteristics. Laprise et al. [11] discussed many limitations and sources of error associated with limited area models, such as the resolution

jump between regional models and its driving data, frequency of Lateral Boundary Conditions (LBCs), and imperfections in LBC data. Leduc and Laprise [12] studied the issue of domain size in the context of regional climate modelling using perfect model approach and suggested the regulations against the use of too small and too large domains.

During the past years, the mesoscale forecasting community has paid relatively less attention to the uncertainties associated with domain size than the uncertainty associated with initial field, model resolution, and model physics. A comparative, comprehensive, and quantitative estimation of the relative role of the uncertainties of domain size, grid spacing, and initial condition in the simulation of mesoscale events is valuable for operational center, like Australian Bureau of Meteorology, to make decision on decreasing their domain size in return for gaining higher resolution or vice versa. Goswami et al. [13] had investigated this issue using 10 km or coarser resolutions, MM5 simulations, with respect to three high impact weather (heavy rainfall) events in the tropics to establish the model's performance. Their results show that domain size plays a role that is as important as that of grid spacing and initial condition in the simulation of high impact weather events. In this study, we use convection-permitting ACCESS model with 1.5 km grid spacing as benchmark to quantitatively compare the impacts of model domain size and resolution on heavy precipitation forecast with the initial field uncertainty and perturbed physics through different configurations of ACCESS. The key motivation for this work is to compare the impacts of model horizontal resolution and domain size with the initial field errors in heavy precipitation situations. Besides the intercomparison between different model configurations with control run, we also compare the model results with radar observation. The accuracy of these forecasts is assessed by exploiting the verification metrics of quantitative precipitation forecasts (QPF). The study is performed for three intense precipitation events, which cause flood events in eastern Australia.

The paper is organized as follows. Section 2 provides a summary of the meteorological and hydrological aspects for the three heavy precipitation events. Section 3 presents a brief description of the convection-permitting ACCESS model, its data assimilation, and configurations for the sensitivity tests. Section 4 describes the characteristics of the simulated precipitation systems. Section 5 documents the verification of the simulated precipitation using bias score, bias-adjusted threat score, and Fractions Skill Score. Section 6 illustrates the sensitivity of precipitation to the model configurations with the verification metrics. Finally, concluding remarks and the implications of these results for QPF and ensemble forecast are presented in Section 7.

2. Case Overview

Figure 1(a) shows the height of topography and geographical location affected by these heavy precipitation events. The first is the Toowoomba/Brisbane Flood of 10–11 January 2011. According to the analysis of Australian National Climate Center [14], the heaviest rainfalls were in the areas at north and west of Brisbane (Figure 1(b)). The highest daily

precipitation totals observed in the Bureau's regular network were 298.0 mm at Peachester and 282.6 mm at Maleny on 10 January, while the highest three-day totals were 648.4 mm at Mount Glorious and 617.5 mm at Peachester. Intense short-period falls also occurred during the event, with one-hour falls in excess of 60 mm occurring on both 10 and 11 January at numerous stations in various locations at north and west of Brisbane.

The second case is the flooding over south-east Queensland on 27 to 28 January 2013 (Figure 1(c)). Extratropical cyclone Oswald and an associated monsoon trough passed over parts of Queensland and New South Wales for a number of days, with widespread impact, including severe storms, flooding, and tornadoes. The third case is a flood over Sydney which happened on 22 and 23 February 2013 caused by a deep low-pressure system. The heaviest rain, in excess of 300 millimetres, fell in north of Port Macquarie (Figure 1(d)).

3. Model Description and Experiments Design for Sensitivity Tests

3.1. Model Description. The forecast tests are performed using ACCESS [15] which has been operational since August 2010 over both global and regional domains. The dynamics and physics of ACCESS are based on the UK Met-Office Unified Model [16]. ACCESS uses a semi-implicit, semi-Lagrangian numerical scheme to solve the nonhydrostatic, deep atmosphere dynamics. The horizontal grid is rotated in latitude-longitude and uses Arakawa C-grid staggering, while the vertical grid is terrain-following hybrid-height coordinates with Charney-Phillips staggering, 70 levels, and top model level height of about 40 km. The model uses a comprehensive set of parameterizations, including a convection scheme based on Gregory and Rowntree [17], a nonlocal boundary layer scheme [18], a surface layer scheme [19], a radiation scheme [20], and a mixed-phase cloud microphysics scheme [21]. Figure 2(a) shows the nesting of ACCESS-G, ACCESS-R, and ACCESS-A with 80 km, 37.5 km, and 12 km grid spacing, respectively. All of these models use 6-hour (between -3 h and $+3$ h) 4D variational data assimilation [22]. In ACCESS, the analysis increments are calculated outside the model by the 4DVAR and the increments are added to the model at the start of the model run. The convection-permitting ACCESS model used in this study is a modified version of the UK Met-Office convection-permitting variable grid model (UKV), which is nested in a coarser resolution model with larger domain and sequentially initiated (allowing 3-hour delay for model spin-up) with the IC and LBC provided by the coarser resolution model. As is standard with ACCESS, the nesting is one-way; that is, there is no feedback from the higher-resolution nests to the coarser grid model.

3.2. Latent Heating Nudging (LHN) for Rain Rate Data Assimilation and IC Uncertainties. One of the significant sources of error in convection-permitting forecasts is the lack of mesoscale information in IC derived from a coarse resolution model. The operational radiosonde system is not dense enough to provide sufficient mesoscale observation for the O (1 km) grid spacing model. Radar data assimilation

FIGURE 1: (a) The height of topography (in meter) and geographical location of three precipitation events; (b) daily rainfalls (in mm) for 09-10 January 2011, (c) 26-27 January 2013, and (d) 22-23 February 2013.

is important for incorporating some mesoscale information into the convection-permitting model. Radar data provide consistent spatial and temporal information that cannot be readily gleaned from gauge information alone. The estimation of the rainfall in the Australian Bureau of Meteorology is the blended radar and gauge rainfall accumulations [23]. In this study, the gauge-adjusted radar-based rainfall rates are assimilated through the LHN. The LHN is a method of forcing a model with measured precipitation rates from radars and/or gauges to improve analysis and short-range forecast of precipitation. More details on the method can be found in Jones and Macpherson [24]. The general idea of the

scheme is to rescale the vertical profiles of latent heating using the ratio of observed and modeled precipitation:

$$\left(\frac{\partial T}{\partial t}\right)_{\text{LHN}} = \left(\frac{\partial T}{\partial t}\right)_{\text{mo}} \frac{RR_{\text{analysis}}}{RR_{\text{mo}}}. \quad (1)$$

The corrections are calculated at analysis grid points and then interpolated to the model grid before being added to the model fields. The LHN makes the analysis correction to model temperature fields at every time step from the start of the forecast for 2 hours to incorporate the mesoscale information containing in the radar rain rates. Horizontal

FIGURE 2: (a) Domains configuration of ACCESS-G, ACCESS-R, and ACCESS-A; (b) four model domains for the convection-permitting forecasts, D1: 738 km $\times$ 720 km, D2: 1230 km $\times$ 1200 km, D3: 2050 km $\times$ 2000 km, and D4: 3280 km $\times$ 3200 km.

filtering of increments on the model grid may be used in the LHN processes.

3.3. Experiments Setup. To quantify the impacts of model resolution and domain size on QPF under real synoptic forcing, a series of numerical experiments are set up for the three heavy precipitation cases. A set of model runs with horizontal grid spacing of 4 km, 2.5 km, 1.5 km, 0.5 km, and the same domain size are undertaken to assess the sensitivity of QPF to horizontal resolutions. Meanwhile, the impacts of domain size on QPF are analysed using a series of additional experiments with 738 km $\times$ 720 km, 1230 km $\times$ 1200 km, 2050 km $\times$ 2000 km, and 3280 km $\times$ 3200 km domains and at fixed 1.5 km grid spacing. One set of domains are shown in Figure 2(b). Specifically, the domain encompasses regions which were significantly affected by heavy precipitation. All of the convection-permitting simulations use the same physical parameterizations. The ICs and LBCs are from the forecasts of ACCESS-A unless otherwise stated. The experiments differing in their spatial resolutions and domain sizes and with or without the LHN are initialized at 06 UTC and cover forecast lead time up to 36 h. Table 1 contains a summary of these experiments. Each experiment is labelled by the nomenclature Eijk. Integer i represents which domain is used in the simulation and can take on the values 1, 2, 3, or 4 corresponding to D1, D2, D3, or D4 in Figure 2(b). Integer j denotes what horizontal resolution is used. Numbers 0, 1, 2, and 4 mean grid sizes 0.5 km, 1.5 km, 2.5 km, and 4 km, respectively. Integer k is the key for using LHN or not ($k = 1$ indicates that LHN is invoked and $k = 0$ otherwise). E210 is taken as the control run since this is the default configuration for operational use.

For the sake of comparison between the impacts of different model parameters and the IC error, we carry out two

other experiments E210ICR1 and E210ICR2 by perturbing model ICs with ACCESS-A 4DVAR increment fields, in which the background error and observational information have been incorporated, by 10% and 20% of the increment value. In addition, there are many physical processes, for example, convections, boundary layer exchanges, and cloud physics that occur on scales too small to be directly resolved by the model and need to be parameterized. These parameterizations involve a number of empirical-adjustable parameters and thresholds which are given somewhat arbitrary values. The perturbed physics using random parameters (RP) to account for the uncertainty associated with these empirical parameters and to simulate the nondeterministic processes not explicitly accounted for by the different parameterizations [25]. For instance, the critical relative humidity parameter that controls the functional dependence of cloud cover fraction on relative humidity is allowed to vary in time (as a first-order autoregressive process) between judiciously chosen upper and lower limits. E210RP1 and E210RP2 are experiments using two perturbed physics schemes with different random parameters sets called RP1 and RP2. RP1 perturbs the parameters associated with convection entrainment rate, convective available potential energy closure time scale, large scale precipitation, gravity wave drag, and boundary layer physics. RP2 perturbs the ice falling speed in large scale cloud scheme and Charnock parameter in boundary layer in addition to RP1.

4. The General Characteristics of Simulated Precipitation System

Figure 3(a) shows the surface equivalent potential temperature, $\theta_e \approx (T + (L_v/C_p)q)(p_0/p)^{R_d/C_p}$ [26], and horizontal wind at 1 km height. Figure 3(b) shows the east-west vertical

FIGURE 3: (a) Surface θ_e (shading, in Kelvin) and horizontal wind at 1 km height, (b) East-west vertical cross section of 29S–25S averaged θ_e and $(u, w * 100)$ vector at 12 UTC 9 January 2011.

cross section of equivalent potential temperature and horizontal and vertical wind. Figure 3 indicates that lower-level easterly flow brought warm and moist air to Queensland and northeast New South Wales. Potential instability was evident over the northeast area of Queensland where there was prominent warm and moist air in the lower atmosphere, as indicated by the vertical distribution of θ_e in Figure 3(b). The potential instability and moderate vertical shear of wind across this region resulted in series of convective storms developing throughout 10 to 12 January 2011. The evolution of radar observed hourly rain rates (not illustrated here) indicates that there were a series of convective precipitation cells generated near the coast and moved from northeast to southwest. The heavy precipitation from these cells was the main cause of the flooding in the Toowoomba/Brisbane area. To deliver general understanding of the model results, at first the directly visual comparisons between the model and the radar observations are made. In latter sections, various metrics are used for their quantitative comparison. Figure 4 shows hourly precipitation from the radar observations and the corresponding forecasts in 8 and 12 UTC 9 January from E210, E211, E320, E220, and E221, respectively. Only the region with radar observation is illustrated. The 06 to 08 UTC is the LHN period for E211 and E221. We use hourly rainfall in 08 UTC to demonstrate the effect of LHN. The comparison between E211 (E221) and E210 (E220) with radar observations shows that the precipitation pattern and location are better in experiments with LHN than without it. Comparing E210, E220, and E320 in Figure 4, the impacts of domain sizes and horizontal resolutions are not significant before 6 h lead time. Comparatively, the impact of resolution (Figure 4(h) versus Figures 4(j) and 4(k)) is a little larger than domain size (Figure 4(j) versus Figure 4(k)) since the impact of LBC associated with domain size has not taken

effect yet. Figure 5 is the same as Figure 4 but shows hourly rainfall in 18 UTC 9 January and 00 UTC 10 January 2011. Figure 5 indicates that the impacts of resolution and domain size start to be significant after 12 h lead time. For example, the hourly rainfall differences among E210, E220, and E320 at 18 h lead time (Figures 5(b), 5(d), and 5(e)) are larger than the differences between E210 and E211 (Figure 5(b) versus Figure 5(c)) or between E220 and E221 (Figure 5(e) versus Figure 5(f)). The results in 00UTC 10 January 2011 are similar (Figure 5, rows 3 and 4) to 18 UTC 09 January 2011. Figures 6 and 7 show hourly precipitation of two other cases at 2 h and 12 h lead time, respectively. The results are quite similar to first case; that is, *the effect of LHN on hourly precipitation is more significant during the first few forecast hours and the resolution and domain size have larger impacts on the precipitation after 12-hour lead time. The impact of resolution is more significant than the domain size before 12 h lead time, while the domain size makes bigger impact after 12 h lead time.*

Figure 8 shows the 16 h forecast precipitation accumulation from 08 UTC January 9 to 00 UTC January 10 and radar observed (OBS) versus the corresponding model simulations in E210, E211, E310, E110, E220, E221, E320, E321, E100, E210RP1, E210RP2, E340, E440, and their arithmetic mean. We simply call it as “ensemble mean” (ENSM) in this study when we compare an average of several experiments with observation or individual experiments. In order to avoid the 2 h (from 06UTC to 08UTC) LHN period, we use 16 h (08UTC to 00UTC) rainfall instead of 18 h (08UTC to 02UTC). The overall comparison between the radar observed and the simulated 16 h rainfall indicates that the model in general overpredicts the accumulated precipitation. The overprediction is even larger as LHN data assimilation is used, though the LHN improves the position of precipitation.

FIGURE 4: Hourly rainfall in 2 h (row 1 and row 2) and 6 h (row 3 and row 4) forecast from 6 UTC 9 January 2011 against radar observation in (a) and (g).

FIGURE 5: Hourly rainfall in 12 h (row 1 and row 2) and 18 h forecast (row 3 and row 4) from 6 UTC 9 January 2011 against radar observation in (a) and (g).

FIGURE 6: Hourly rainfall in 2 h (rows 1 and 2) and 12 h (rows 3 and 4) forecast from 6 UTC 26 January 2013 against radar observation in (a) and (g).

FIGURE 7: Hourly rainfall in 2 h (rows 1 and 2) and 12 h (rows 3 and 4) forecast from 6 UTC 22 February 2013 against radar observation in (a) and (g).

FIGURE 8: 16 h (08 UTC 9 January 2011 to 00 UTC 10 January 2011) accumulated rainfall against radar observation.

TABLE 1: Explanation of list experiments.

Domain	Grids	LBC	Resolution	LHN	ICR	Nomenclature
D1	1476×1440	ACCESS-A	0.5 km	No	No	E100
D1	492×480	ACCESS-A	1.5 km	No	No	E110
D2	820×800	ACCESS-A	1.5 km	No	No	E210 (CTL)
D2	820×800	ACCESS-A	1.5 km	No	No	E210RP1
D2	820×800	ACCESS-A	1.5 km	No	No	E210RP2
D2	492×480	ACCESS-A	2.5 km	No	No	E220
D3	1368×1334	ACCESS-A	1.5 km	No	No	E310
D3	820×800	ACCESS-A	2.5 km	No	No	E320
D3	505×500	ACCESS-A	4.0 km	No	No	E340
D4	820×800	ACCESS-A	4.0 km	No	No	E440
D2	820×800	E440	1.5 km	No	No	E210LBC440
D1	492×480	ACCESS-A	1.5 km	Yes	No	E111
D2	820×800	ACCESS-A	1.5 km	Yes	No	E211
D2	492×480	ACCESS-A	2.5 km	Yes	No	E221
D3	820×800	ACCESS-A	2.5 km	Yes	No	E321
D3	505×500	ACCESS-A	4.0 km	Yes	No	E341
D2	820×800	ACCESS-A	1.5 km	No	ICR1	E210ICR1
D2	820×800	ACCESS-A	1.5 km	No	ICR2	E210ICR2

FIGURE 9: TSA of (a) 4 h (8–12 UTC 9 January 2011) and (b) 16 h (8 UTC 9 January 2011 to 00 UTC 10 January 2011) accumulated precipitation against radar observation with different model configurations.

5. Evaluation with Radar Observations

The verification of model forecasts is a key element of operational NWP and is essential for its continuous improvement. Skill scores usually serve as metrics to measures and identify systematic model errors and forecast uncertainties, though there is no single skill score that completely characterizes the accuracy of the meteorological data set produced by the model.

5.1. Categorical Skill Scores. A standard method of validating model precipitation is to categorically compare model forecasted precipitation with observations. Generally, the measures are computed on the basis of contingency tables (Table 2) constructed by comparing point values. Based on the counts, a variety of skill scores such as bias score and threat score can be computed [27]. In this study, the model forecast of rainfall accumulation is interpolated to the radar rainfall analysis grid (1.5 km spacing Cartesian grid with radar at the origin) with bilinear interpolation approach and the model forecasted rainfall is set to missing for those grid points where the observed precipitation is not available [23]. Each grid is then classified according to the criteria listed in Table 2. The bias score, $BS = (a + b)/(a + c)$, and the bias-adjusted threat score [28], $TSA = ((a + c)^{1/BS} - c^{1/BS})/((a + c)^{1/BS} + c^{1/BS})$, are computed with the radar observations as “truth.” $BS = 1$ and $TSA = 1$ mean the forecast is perfect. TSA satisfies $0 \leq TSA \leq 1$ and its value is nearly the same as the critical success index, $CSI = a/(a + b + c)$, in most situations but relatively insensitive to bias score in some extreme situations. We have also tried the bicubic interpolation method and found that the skill scores are not sensitive to the interpolation approach.

TABLE 2: Contingency table for precipitation event.

	Observation YES	Observation NO
Forecast YES	a	b
Forecast NO	c	d

Figure 9(a) shows TSA with different thresholds for 4 h forecasting of rainfall accumulation from 08 UTC 9 to 12 UTC 9 January 2011. The reason why we use 4 h is to avoid the 2 h LHN period from 06 UTC to 08 UTC. The figure indicates that the experiments with LHN (E211 and E221) tend to have a better skill in terms of TSA. The average of several members offers a better forecast than most of the individual forecasts, as already mentioned, [29–31]. The TSA score for 16 h forecasting of accumulated precipitation in Figure 9(b) indicates that the model with coarser resolution and larger domain size tends to better predict the weaker accumulated precipitation (<10 mm), while E100 with the finest resolution of 0.5 km grid length has better TSA for heavier precipitation (>20 mm). The arithmetic mean can improve TSA score in general. However, the skill for stronger precipitation is always much lower than for weaker precipitation regardless of model configurations.

Figure 10(a) plots the BS and TSA with threshold of 1 mm h^{-1} for hourly precipitation in E210, E211, E220, E221, E100, E310, E440, and their arithmetic mean. TSA clearly shows that forecasts with LHN are superior to the forecasts without LHN during the first few forecast hours. The arithmetic mean is more skillful than single member. However, in terms of BS, the arithmetic mean tends to overpredict the precipitation area. The forecast skill in terms

FIGURE 10: The time evolution of BS and TSA for hourly rainfall against radar observation with thresholds: (a) 1 mm h^{-1} and (b) 5 mm h^{-1} for the Toowoomba/Brisbane flood case.

of BS and TSA becomes lower as the threshold increases to 5 mm h^{-1} (Figure 10(b)). The time evolution of BS and TSA for hourly precipitation against radar observations with thresholds of 1 mm h^{-1} and 5 mm h^{-1} indicates that BS and TSA vary significantly with forecast lead time and the skill with 5 mm h^{-1} threshold is much lower than the skill with 1 mm h^{-1} threshold.

5.2. Fractions Skill Score (FSS). The traditional categorical skill scores provide information about the quality of a forecast on point-by-point comparisons between simulation and observed precipitation but without regard to spatial information and forecast uncertainty. A prediction of a precipitation structure that is correct in terms of amplitude, size, and timing but with small positional errors may be very poorly rated by categorical scores, for example, BS and TSA [28, 32, 33]. To address this problem, several new verification

techniques have been developed specifically for higher spatial and temporal resolution forecasts [34–39]. The Fractions Skill Score (FSS) is a scale-dependent neighbourhood based verification method [6]. To compute the FSS, the hourly precipitation is interpolated onto the 1.5 km spacing grid for the radar observed precipitation analysis. Then, fractions are generated using the neighbourhood approach: for every $1.5 \times 1.5 \text{ km}^2$ pixel, the fraction of surrounding pixels within a given sized square “neighbourhood” or “elementary area” that exceed a particular threshold (e.g., 2 mm in a 1 h period) is computed. As a result, every pixel in the forecast field has a fraction that can be compared to its equivalent pixel in the radar observation field. This is done to both the observed and forecast precipitation fields. The FSS is then computed with

$$\text{FSS} = 1 - \frac{\text{FBS}}{\text{FBS}_0}, \quad (2)$$

FIGURE 11: The time evolution of BS and TSA for hourly rainfall against radar observation with thresholds: (a) 1 mm h^{-1} and (b) 5 mm h^{-1} for the Toowoomba/Brisbane flood case.

where FBS is Fraction Brier Score,

$$\text{FBS} = \frac{1}{N} \sum_{j=1}^N (p_j - o_j)^2 \quad (3)$$

$0 \leq p_j \leq 1$ forecast fractions, $0 \leq o_j \leq 1$ radar fractions,

N is grid points in whole neighbourhood square, and o_j and p_j are the observation and forecast fractions, respectively, at each point j :

$$\text{FBS}_0 = \frac{1}{N} \left(\sum_{j=1}^N p_j^2 + \sum_{j=1}^N o_j^2 \right). \quad (4)$$

FSS has a value that varies between 0 and 1. A forecast with perfect skill has a score of 1 and a score of 0 means zero skills.

As the sampling square becomes larger, the spatial errors (e.g., misplaced rainbands) in the forecast become less significant; the forecast and radar fractions will become more alike. An unbiased forecast will tend toward a FSS of 1.0 as the sampling square approaches the size of the verification domain. The FSS of hourly precipitation from a few experiments at 3 h (09UTC) and 6 h (12UTC) forecast lead time is plotted in Figure 11. The FSS increases with the neighbourhood size for a given threshold while for a fixed neighbourhood size the FSS decreases as the threshold value increases. According to [6], a characteristic value of $\text{FSS} = 0.5$ represents a hit rate of 0.5 and a CSI of 0.33 in categorical scores when the precipitation area is small comparing with the evaluation domain. So we select $\text{FSS} = 0.5$ as a reference for the length scale estimation. Figure 11(a) shows that the square neighbour length is around 90 km for 10 mm h^{-1} threshold,

FIGURE 12: FSS of hourly rainfall against the verification spatial scales using thresholds 2.5, 5, and 10 mm h^{-1} , respectively, (a) valid at 9 UTC 26 January 2013 and (b) 12 UTC 26 January 2013 for the Ex-tropical cyclone Oswald case.

while for the 5 mm h^{-1} threshold the square neighbour length is around 60 km as $\text{FSS} = 0.5$. Consistent with the TSA verification, the FSS indicates that the model is more skillful at predicting widespread rainfall than predicting localized heavy precipitation. Figure 11(a) shows that the E211 (1.5 km grid length with LHN) has the best FSS for 2.5 mm h^{-1} threshold, which means the LHN improves the QPF of relatively widespread precipitation. For 10 mm h^{-1} threshold, the LHN has negative impact on QPF. Figure 11(a) also shows the FSS increases with model horizontal resolution. The forecast with 0.5 km grid length, the highest resolution in this study, has the best FSS for 5 and 10 mm h^{-1} threshold at smaller neighbourhood square size. Figure 11(b) shows the FSS for the sixth hour forecast. Comparing with Figure 11(a),

it is evident that forecast skill decreases quickly with forecast lead time. Figure 11(b) shows that for the 2.5 and 5.0 mm h^{-1} threshold the E310 has best skill, which means that larger domain can better predict the light to moderate precipitation, while with 10 mm h^{-1} threshold FSS is very small regardless of the model domain size, resolution, and LHN. Figure 12 shows the FSS for the case of extratropical cyclone Oswald. The forecast with 0.5 km grid spacing has better FSS up to 6 h forecast lead time. Figures 11 and 12 indicate that the higher resolution models have the potential to forecast precipitation which is highly localized and episodic rainfall in first few hours of lead time. The results for the third case are quite similar. These verifications indicate that higher horizontal resolution improves the forecast skill of the convective precipitation.

FIGURE 13: Intensity-scale diagrams show FSS of hourly precipitation from E210 with radar observation as truth, (a) 6th hour forecast (valid 1200 UTC 9 January 2011) and (b) 12th hour forecast (valid 18 UTC 9 January 2011); unit is %. Grid value with color shading shows the FSS. The abscissa and the ordinate indicate the threshold value and the spatial scale, respectively.

To demonstrate further the dependence of the FSS on the precipitation intensity and the verification scale or neighbourhood square, the intensity-scale diagrams for the E210 at forecast lead times 6 h and 12 h (12UTC and 18UTC 9th January 2011) are plotted in Figures 13(a) and 13(b). The intensity-scale diagrams demonstrate that the FSS increases with spatial scale and decreases with threshold value and forecast lead time.

Percentile thresholds (e.g., the strongest precipitation of top 50%, 25%, 10%, and 1%, that is, precipitation exceeding the 50th, 75th, 90th, and 99th percentile values) are also used for the FSS computation. Similar to the results of Roberts [40],

the FSS verification with percentile thresholds has shown that the skill is dependent on the precipitation area, which is closely related to precipitation intensity. The model has better skill for weak precipitation.

6. QPF Sensitivity to Model Configurations

6.1. The Intensity and Scale Dependence of Sensitivity. Figure 14 shows the intensity-scale histogram of radar observed and model simulated hourly precipitation. The ordinate shows the number of precipitation grid points (N) against hourly precipitation intensity (abscissa; with 0.05 mm bin

FIGURE 14: Intensity histogram of hourly precipitation with different model configurations. The ordinate indicates the number of precipitation grid points in their 36 h simulations and the abscissa indicates the precipitation intensity. (a) The impacts of domain size and resolution and (b) the impacts of perturbed physics, IC uncertainty, and LHN.

size as increment). The hourly accumulated precipitation from each 36 h forecasts is used for the sampling of these precipitation grid points. To keep the model precipitation statistically consistent with the radar observations, the hourly accumulated model precipitation is interpolated to the radar observation grid points with bilinear interpolation approach. The exponential decrease of N with increasing precipitation intensity in Figure 14 indicates that there is a log-linear relationship existing between N and precipitation intensity. Since each grid point represents an area of Δx^2 , assuming the precipitation area is circular in shape, $2\sqrt{N/\pi}\Delta x$ approximately represents the diameter of the precipitation area, and hence $2\sqrt{N/\pi}\Delta x$ can be taken as a length scale of precipitation area approximately. Larger N and weak intensity usually represent wide spread precipitation area while smaller N and heavy intensity imply localized small scale precipitation. The increasing spread of model histograms with rainfall intensity in Figure 14 indicates that the heavier precipitation ($>20 \text{ mm h}^{-1}$) is more sensitive to model configurations than the lighter precipitation ($<10 \text{ mm h}^{-1}$). The histograms also show that the model precipitation in general is more widely spread than radar observation. This result is consistent with the BS verification shown in previous section. The better agreement between the histograms with perturbed physics (E210RP1 or E210RP2) and radar observation indicates that the perturbed physics improves the intensity-scale statistics, particularly for the heavy precipitation, while the LHN degrades the statistical property of heavy precipitation (Figure 14(b)).

The dependence of sensitivity on the precipitation intensity is further illustrated by using the hourly precipitation data collected from the 36 h forecasts of E210, E211, E210RP1, E210RP2, E310, E110, E220, E221, E320, E340, E440,

E210ICR01, and E210ICR02. The variations of their arithmetic mean, the maximum, the minimum, and the variance of each member against the precipitation intensity of E210 are shown in Figure 15. The variance for strong precipitation ($>20 \text{ mm h}^{-1}$) is much larger than that of weak precipitation ($<10 \text{ mm h}^{-1}$) implying that stronger precipitation has larger sensitivity. In other words, the heaviest precipitation has larger uncertainty or error.

6.2. Skill Score Disparity and Quantitative Sensitivity. To quantify the relative importance of each model configuration on the QPF, we need to choose a norm or metric by which the sensitivity is measured. Usually skill scores are designed to quantitatively measure or characterize the accuracy of model forecast by measuring the “distance” between model forecast and the observation (e.g., Section 5). Intuitively the BS and TSA “skill scores” of different experiment with control run (E210 here) as “truth” can be used to measure the sensitivity of model precipitation to the model configurations quantitatively. To illustrate the relevance of different sources of uncertainty at different lead time, the time evolution of BS and TSA of different experiments with E210 as truth is shown in Figure 16. As expected, the E210 against E210 is perfect with BS = 1 and TSA = 1. While the BS of different experiment is not too far from 1 (Figure 16(c)), the TSA is a suitable metric to show the impact of various model configurations on precipitation. The experiments with LHN (E211, E221, and E321) have the smallest TSA in Figures 16(a) and 16(b), that is, the largest disparity from E210 during 1 h and 6 h, while the large TSA of E110, E310, E220, and E320 indicates the smallest disparities from E210 during this time period. The domain size has little significant impact and the perturbed physics and initial field perturbations have less impact on the TSA

FIGURE 15: The maximum, minimum, arithmetic mean, and variance of hourly precipitation against precipitation intensity in E210 (CTL).

than LHN before 6 h lead time. These TSA characteristics indicate that LHN has the strongest impact on precipitation forecasting in the initial stage. The moderate disparity of E100 from E210 in Figure 16 indicates that the higher resolution has moderate impact on precipitation before 6 h lead time. After 6 h lead time, the TSA of E211 becomes the largest, while the impacts of other experiments become more significant. TSA of E440, E100, E320, and E310 decreases to small values around 0.2–0.3. The TSA evolution indicates that the impact of LHN decreases while the impact of model resolution and domain size becomes more significant after 6 h forecast lead time.

To further demonstrate the dependence of model parameter sensitivity on the precipitation intensity, the BS and TSA “skill scores” of 16 h accumulated precipitation against the precipitation intensity from various model configurations and model parameters using E210 as “truth” which is shown in Figure 17. The increment of disparity along with the precipitation intensity indicates that the heavier precipitation has larger sensitivity and uncertainty than weaker precipitation to model configurations and model parameters. Figure 17 illustrates that impacts of domain size and model resolution are larger than that of random physics and IC uncertainty for the 16 h precipitation accumulation. This result suggests that *the forecast precipitation is much more affected by model domain size and resolution than by IC uncertainty and model physics in the longer time forecast*. The same analysis for the other two cases (not shown here) suggests that their precipitation forecasts have similar properties, though the resulting sensitivity exhibits case-to-case variability.

7. Summary with Discussion

The main purpose of this study is to investigate the sensitivity of heavy precipitation to model configurations. The impacts of domain size, horizontal grid spacing combining with perturbed physics, and LHN, and IC uncertainty on QPF are evaluated through a series of convection-permitting simulations of three heavy precipitation cases with various configurations of ACCESS model. The aim is to better understand the model configurations on heavy precipitation forecast. The categorical statistical metrics of BS, TSA, and the scale-dependent neighbourhood based skill score (FSS) have been employed to verify the forecasts with compiled radar observational data. The ACCESS model with coarser resolution and larger domain provides slightly better forecasts for wide spread precipitation, whereas the model with higher resolution has better skill for more localized heavy rainfall. The LHN can shorten the spin-up time of convective systems and improve the QPF in the first few forecast hours. The perturbed physics improves the statistical relationship between precipitation intensity and precipitation scale. However, we should say these results are yet to be further investigated with more study cases.

The sensitivity of rainfall to the model configurations is estimated by the “ensemble” spread and cross-verification using the “skill scores” of different experiments with the control run used as “truth.” The variance for strong precipitation ($>20 \text{ mm h}^{-1}$) is much larger than that of weak precipitation ($<10 \text{ mm h}^{-1}$) implying that stronger precipitation has larger sensitivity and uncertainty. A detailed assessment of the relevance of different aspects of model configuration at different forecast ranges is also conducted. The time evolution of TSA indicates that the LHN is taking effect only in the initial stage of the prediction while the domain size, that is, the boundary conditions, has a relevant impact after 6 h.

The uncertainty analyses have traditionally focused on model physics and IC errors. The model uncertainty associated with domain size and model resolution is rarely considered. The QPF disparities illustrated by the dispersion of the skill scores show that the impact of model domain size and horizontal resolution on QPF is quite large comparing with the perturbed physics and IC uncertainty after 12 h lead time. To make substantial progress with QPF, we need not only to improve the IC through better data assimilation, but also to reduce the model error and LBC error associated with horizontal resolution and limited domain size. Our tests show that the arithmetical mean precipitation of several forecasts with different resolutions and domain sizes is in general superior to single forecasts in terms of skill scores. These examples indicate that in designing future convection-permitting ensemble prediction systems, besides the uncertainties associated with ICs and model physics, the uncertainties associated with model resolution and domain size should be taken into consideration as model uncertainties.

Due to the limit of computer resources, only three cases are investigated in this study. Testing with more cases is planned as the convection-permitting ACCESS model is used for a real time forecast in Australia.

FIGURE 16: The TSA and BS of hourly rainfalls from different model configurations from 7 UTC 9 January 2011 to 12 UTC 10 January 2011 using E210 as “truth.” (a) Shows TSA of E210 (CTL), E210RP1, E210RP2, E210ICR1, E210ICR2, and ENSM, (b) shows TSA of E210, E110, E100, E211, E221, E220, E310, E320, E321, E340, and E440, and (c) shows BS of all experiments.

FIGURE 17: BS (a) and TSA (b) for 16-hour rainfall accumulations from several experiments with E210 as “truth.”

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors thank Dr. Noel E. Davidson, Dr. David Smith, and Dr. Hongyan Zhu for their careful perusal of this paper. This study was supported by Strategic Radar Enhancement Project (SREP) of Australian Bureau of Meteorology.

References

- [1] M. L. Weisman, W. C. Skamarock, and J. B. Klemp, "The resolution dependence of explicitly modeled convective systems," *Monthly Weather Review*, vol. 125, no. 4, pp. 527–548, 1997.
- [2] J. Done, C. A. Davis, and M. Weisman, "The next generation of NWP: explicit forecasts of convection using the weather research and forecasting (WRF) model," *Atmospheric Science Letters*, vol. 5, no. 6, pp. 110–117, 2004.
- [3] H. W. Lean, P. A. Clark, M. Dixon et al., "Characteristics of high-resolution versions of the Met Office unified model for forecasting convection over the United Kingdom," *Monthly Weather Review*, vol. 136, no. 9, pp. 3408–3424, 2008.
- [4] C. S. Schwartz, J. S. Kain, S. J. Weiss et al., "Next-day convection-allowing WRF model guidance: a second look at 2-km versus 4-km grid spacing," *Monthly Weather Review*, vol. 137, no. 10, pp. 3351–3372, 2009.
- [5] E. Richard, A. Buzzi, and G. Zängl, "Quantitative precipitation forecasting in the Alps: the advances achieved by the Mesoscale Alpine Programme," *Quarterly Journal of the Royal Meteorological Society*, vol. 133, no. 625, pp. 831–846, 2007.
- [6] N. M. Roberts and H. W. Lean, "Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events," *Monthly Weather Review*, vol. 136, no. 1, pp. 78–97, 2008.
- [7] T. Weusthoff, F. Ament, M. Arpagaus, and M. W. Rotach, "Assessing the benefits of convection-permitting models by neighborhood verification: examples from MAP D-PHASE," *Monthly Weather Review*, vol. 138, no. 9, pp. 3418–3433, 2010.
- [8] W. C. Skamarock, "Evaluating mesoscale NWP models using kinetic energy spectra," *Monthly Weather Review*, vol. 132, no. 12, pp. 3019–3032, 2004.
- [9] D. Harris, E. Foufoula-Georgiou, K. K. Droegemeier, and J. J. Levit, "Multiscale statistical properties of a high-resolution precipitation forecast," *Journal of Hydrometeorology*, vol. 2, no. 4, pp. 406–418, 2001.
- [10] D. E. Stevens, A. S. Ackerman, and C. S. Bretherton, "Effects of domain size and numerical resolution on the simulation of shallow cumulus convection," *Journal of the Atmospheric Sciences*, vol. 59, no. 23, pp. 3285–3301, 2002.
- [11] R. Laprise, R. de Elia, D. Caya et al., "Challenging some tenets of regional climate modelling," *Meteorology and Atmospheric Physics*, vol. 100, pp. 3–22, 2008.
- [12] M. Leduc and R. Laprise, "Regional climate model sensitivity to domain size," *Climate Dynamics*, vol. 32, pp. 833–854, 2009.
- [13] P. Goswami, H. Shivappa, and S. Goud, "Comparative analysis of the role of domain size, horizontal resolution and initial conditions in the simulation of tropical heavy rainfall events," *Meteorological Applications*, vol. 19, no. 2, pp. 170–178, 2012.
- [14] National Climate Centre (NCC). Bureau of Meteorology, *Special Climate Statement 24*, 2011, <http://www.bom.gov.au/climate/current/statements/scs24b.pdf>.
- [15] K. Puri, Y. Xiao, X. Sun et al., "Preliminary results from numerical weather prediction implementation of ACCESS," *CAWCR Research Letters*, vol. 5, pp. 15–22, 2010.
- [16] T. Davies, M. J. P. Cullen, A. J. Malcolm et al., "A new dynamical core for the Met Office's global and regional modelling of the atmosphere," *Quarterly Journal of the Royal Meteorological Society*, vol. 131, no. 608, pp. 1759–1782, 2005.
- [17] D. Gregory and P. R. Rowntree, "A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure," *Monthly Weather Review*, vol. 118, no. 7, pp. 1483–1506, 1990.
- [18] A. P. Lock, A. R. Brown, M. R. Bush, G. M. Martin, and R. N. B. Smith, "A new boundary layer mixing scheme. Part I: scheme description and single-column model tests," *Monthly Weather Review*, vol. 128, no. 9, pp. 3187–3199, 2000.
- [19] R. L. H. Essery, M. J. Best, and P. M. Cox, "MOSES 2.2 technical documentation," Hadley Centre Technical Note 30, Met Office, Bracknell, UK, 2001.
- [20] J. M. Edwards and A. Slingo, "Studies with a flexible new radiation code. I: choosing a configuration for a large-scale model," *Quarterly Journal of the Royal Meteorological Society*, vol. 122, no. 531, pp. 689–719, 1996.
- [21] D. R. Wilson and S. P. Ballard, "A microphysically based precipitation scheme for the UK Meteorological Office Unified Model," *Quarterly Journal of the Royal Meteorological Society*, vol. 125, no. 557, pp. 1607–1636, 1999.
- [22] F. Rawlins, S. P. Ballard, K. J. Bovis et al., "The met office global four-dimensional variational data assimilation scheme," *Quarterly Journal of the Royal Meteorological Society*, vol. 133, no. 623, pp. 347–362, 2007.
- [23] J. Dorte and A. W. Seed, "Spatio-temporal characteristics of blended radar/gauge data and the estimation of design rainfalls," in *Hydrology and Water Resources Symposium 2014*, pp. 326–333, ACT: Engineers Australia, Barton, Australia, 2014.
- [24] C. D. Jones and B. Macpherson, "A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model," *Meteorological Applications*, vol. 4, no. 3, pp. 269–277, 1997.
- [25] J. W.-B. Lin and J. D. Neelin, "Influence of a stochastic moist convective parameterization on tropical climate variability," *Geophysical Research Letters*, vol. 27, no. 22, pp. 3691–3694, 2000.
- [26] R. B. Stull, *An Introduction to Boundary Layer Meteorology*, Kluwer, 1988.
- [27] D. S. Wilks, *Statistical Methods in the Atmospheric Sciences: An Introduction*, Academic Press, 2nd edition, 2006.
- [28] M. E. Baldwin and J. S. Kain, "Sensitivity of several performance measures to displacement error, bias, and event frequency," *Weather and Forecasting*, vol. 21, no. 4, pp. 636–648, 2006.
- [29] M. S. Tracton and E. Kalnay, "Operational ensemble prediction at the National Meteorological Center: practical aspects," *Weather & Forecasting*, vol. 8, no. 3, pp. 379–398, 1993.
- [30] R. Mureau, F. Molteni, and T. N. Palmer, "Ensemble prediction using dynamically conditioned perturbations," *Quarterly Journal—Royal Meteorological Society*, vol. 119, no. 510, pp. 299–323, 1993.
- [31] P. L. Houtekamer, L. Lefavre, J. Derome, H. Ritchie, and H. L. Mitchell, "A system simulation approach to ensemble

- prediction,” *Monthly Weather Review*, vol. 124, no. 6, pp. 1225–1242, 1996.
- [32] E. E. Ebert, “Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework,” *Meteorological Applications*, vol. 15, no. 1, pp. 51–64, 2008.
 - [33] E. E. Ebert, “Neighborhood verification of high resolution precipitation forecasts,” *Weather and Forecasting*, vol. 24, pp. 1498–1510, 2009.
 - [34] D. Ahijevych, E. Gilleland, B. G. Brown, and E. E. Ebert, “Application of spatial verification methods to idealized and NWP-gridded precipitation forecasts,” *Weather and Forecasting*, vol. 24, no. 6, pp. 1485–1497, 2009.
 - [35] E. E. Ebert, U. Damrath, W. Wergen, and M. E. Baldwin, “The WGENE assessment of short-term quantitative precipitation forecasts,” *Bulletin of the American Meteorological Society*, vol. 84, no. 4, pp. 481–492, 2003.
 - [36] E. E. Ebert and J. L. McBride, “Verification of precipitation in weather systems: determination of systematic errors,” *Journal of Hydrology*, vol. 239, no. 1–4, pp. 179–202, 2000.
 - [37] B. Casati, G. Ross, and D. B. Stephenson, “A new intensity-scale approach for the verification of spatial precipitation forecasts,” *Meteorological Applications*, vol. 11, no. 2, pp. 141–154, 2004.
 - [38] B. Casati, L. J. Wilson, D. B. Stephenson et al., “Forecast verification: current status and future directions,” *Meteorological Applications*, vol. 15, no. 1, pp. 3–18, 2008.
 - [39] C. Keil and G. C. Craig, “A displacement and amplitude score employing an optical flow technique,” *Weather and Forecasting*, vol. 24, no. 5, pp. 1297–1308, 2009.
 - [40] N. Roberts, “Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model,” *Meteorological Applications*, vol. 15, no. 1, pp. 163–169, 2008.