To develop a progressive multimetric configuration optimisation method for WRF simulations of extreme rainfall events over Egypt

Ying Liu[a, *], Yiheng Chen[a], Otto Chen[a], Jiao Wang[a], Lu Zhuo[b, *], Miguel A. Rico-Ramirez[a] and Dawei Han[a]

[a]Department of Civil Engineering, University of Bristol, Bristol, BS8 1TR, UK

[b]Department of Civil and Structural Engineering, University of Sheffield, Sheffield, S1 3JD, UK

*Corresponding author.

E-mail address: emily.liu@bristol.ac.uk (Y. Liu), lu.zhuo@sheffield.ac.uk (L. Zhuo)

Abstract

The Weather Research and Forecasting (WRF) model can help to improve our understanding of analysing and forecasting hydrometeorological disasters. Especially for some regions like the Nile Delta, which faces growing climate hazards but has inadequate in situ rainfall observations. However, identifying an optimal configuration to run the WRF model is often a challenge. In this study, the WRF model was used to simulate extreme rainfall events at high spatial and temporal resolutions centered around Alexandria, in northern Egypt. In particular, a progressive multimetric configuration optimisation (PMCO) method is proposed to identify the possible optimal configurations of WRF in the aspect of domain size, numbers of vertical levels, nesting ratio, spin-up times, and physical parameterization schemes (microphysics, planetary boundary layer, and cumulus), based on 48 specifically designed experiments. The simulation performances are quantified and sorted by the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). All WRF simulations use the ERA5 reanalysis dataset as boundary conditions and the WRF results are verified against the Integrated Multi-satellitE Retrievals for GPM (IMERG). The results show that the rainfall distribution and magnitude are most sensitive to the spin-up time and physical parameterization schemes. It is also observed that the improvement of WRF's reproducibility of rainfall intensity is usually accompanied by a decrease in the reproducibility of rainfall distribution. The best model configuration for the study area comprises of three-level nesting (D01 80x80; D02 112x112; D03 88x88), 58 vertical levels, 1:3:3 downscaling ratio, 48h spin-up time, WRF Single-Moment 6-class microphysics scheme, Mellor-Yamada-Janjic planetary boundary layer scheme, and Grell-Freitas cumulus. The stability of this configuration is also verified with the other three extreme rainfall events over Egypt. The results show that there exists a common WRF configuration set in Egypt that produces the relatively good simulations for extreme rainfall events.

Keywords

WRF rainfall simulation; ERA5; IMERG; PMCO method; Egypt.

1 Introduction

One of the most challenging parts of flood forecasting is the lack of meteorological observations, especially rainfall. In real-time flood forecasting, rainfall needs to be forecasted to extend the flood forecast lead time, which enables the implementation of flood control more promptly (Brath et al., 1988; Cluckie and Han, 2000). Nevertheless, for many catchments in the world, the locations of the rain gauges are too sparse to provide accurate and representative catchment rainfall measurements (Dai et al., 2017) or are non-existent. Therefore, numerical weather prediction (NWP) models are very useful tools for the development of flood forecasting systems because they are able to simulate the atmospheric processes with high spatial and temporal resolutions (Zhuo et al., 2019). However, these models often have a wide range of configuration options available and this diversity brings its own problems at the same time. Because of the occurrence of high-dimensional and nonlinear interactions, it becomes highly complex to identify the best set of physical, dynamical and computational configurations (Nossent et al., 2011). Thus, examining the sensitivity of models to the changes in their configuration options constitutes an essential evaluation work. These sensitivity tests can not only help improve our understanding of how NWP models work, but also help identify which model parameters need to be specified more accurately (Barnsley, 2007). In addition, sensitivity analysis can give modelers useful information about the choice and influence of model configurations. Besides the large number of options, another critical problem for NWP models is that the best configuration combination of one region is not necessarily applicable to others (Krieger et al., 2009). In the past twenty years, many WRF configuration studies to simulate rainfall have been done in regions with rich meteorological/hydrological data, such as Beijing (Di et al., 2015; Chu et al., 2018), southwest England (Liu et al., 2012; Yang et al., 2019) and the United States (Pei et al., 2014; Li et al., 2014), while limited studies have been carried out over Egypt (ElTahan and Magooda, 2017). Due to the lack of surface radars and rain gauges in Egypt, the development of their flood forecasting system would benefit from numerical weather models. If an early flood warning system can be well established, the flood damages and losses could be mitigated while the water resources could be managed for other uses like agricultural and residential water supply. The previous study only evaluated the impact of different microphysics schemes on WRF rainfall simulations over Egypt (ElTahan and Magooda, 2017). There are still many uncertainties in the model configuration that have not been fully explored. Thus, it is meaningful to carry out an extensive sensitivity test and identify optimal WRF configurations to simulate rainfall processes over Egypt.

The Weather Research and Forecasting (WRF) model is one of the most commonly used NWP models. It offers multiple domain and physics configurations that can be combined in a wide range of ways. According to Awan et al. (2011), the dependence of WRF on different domain sizes, horizontal and vertical resolutions, initial and boundary conditions, numerical solvers, terrain and vegetation features, along with assimilation and nudging techniques result in varied results. Previous studies have evaluated the sensitivity of the model to these configurations. Seth and Rojas (2003) demonstrated that simulations of small domain sizes could suppress the feedback from local disturbances on the large-scale general circulation and then easier to benefit from lateral boundary conditions. However, Vannitsem and Chomé (2005) also noted that too small domain sizes would prevent detailed mesoscale processes from being developed in the area of interest. To balance the trade-off of domain sizes, the WRF official guidance (Warner, 2011) recommends that the domain sizes should contain the major features of the regional mesoscale circulation systems, and at least five grid points exist between adjacent nested domains to have sufficient space for relaxation. In WRF, domain size implicitly decides the impacts of terrain and large-scale dynamics, while the horizontal and vertical grid spacings decide the smallest resolution (Goswami et al., 2012). It is reasonable to expect that WRF runs with small grid spacings can produce good results because such simulations could resolve more small-scale features that are contained in the boundary conditions. However, there are many studies showings that WRF runs at relatively high resolutions not necessarily produce accurate outputs (Roberts and Lean, 2008; Kain et al., 2008; Schwartz et al., 2009). Liu et al. (2012) and Knievel et al. (2004) also demonstrated that the performance of WRF to forecast rainfall decreased with the increase of the nesting ratio. A similar conclusion was drawn by Aligo et al. (2009) who concluded that too small vertical grid spacings tend to weaken the WRF rainfall simulation performance. Chu et al. (2018) suggested that WRF rainfall simulations with horizontal and vertical grid spacings of approximately 4 km and less than 1 km (in the troposphere), respectively, may be a suitable compromise between accuracy and computational efficiency. Overall, these domain configurations together affect the range of resolved scales and the nature of dynamical interactions in the model.

Apart from the WRF domain configurations, there are other sources of uncertainty in the rainfall simulations that come from using different physical parameterization schemes. WRF has various physical parameterizations available for microphysics (MP), planetary boundary layer (PBL) and cumulus (CU). Physical parameterization schemes interact non-linearly with each other and with the dynamical core of the model, and these complex relationships make the exploration of uncertainty in rainfall simulation very challenging. Many studies have shown varying model performances when regional climate simulation with different physical parameterizations. According to the study of Flaounas et al. (2011), PBL schemes have the greatest impact on temperature, rainfall amount, humidity vertical distribution while CU schemes strongly affect the dynamics and rainfall variability. They also highlighted that a combination of the Mellor-Yamada-Janjic (MYJ) PBL scheme and Kain-Fritsch (KF) CU scheme was found to produce more realistic temperature, humidity, and the onset of West African monsoon. Another study by Evans et al., (2012) carried out near the southeast coast of Australia pointed out the MYJ PBL scheme also performed well when used with Betts–Miller–Janjić (BMJ) CU scheme. In addition, they suggested the Yonsei University (YSU) PBL scheme, KF CU scheme and Rapid

Radiative Transfer Model for General Circulation Models (RRTMG) radiation schemes should not be used in combination. However, Ji et al. (2013) found that using the YSU PBL scheme with the KF CU scheme could improve rainfall prediction, especially for heavy rainfall. These inconsistent results suggest that no single model configuration is universally suitable for all cases, one study may conflict with another. Therefore, it is useful to explore as many combinations of different physical parameterization schemes as possible to gain more understanding about their suitability in different regions.

In addition, the WRF spin-up time is another important factor affecting the performance of the simulations. In regional modelling, a suitable spin-up time is often required to balance the inconsistencies between the initial and boundary conditions of forcing data and the model simulation results. However, there is still a lack of consensus on the best spin-up time. The spin-up time mainly depends on the domain size and boundary conditions disturbances (Kleczek et al., 2014). Generally, the presence of disturbances causes model performances to decrease with the reduced spin-up time. Jankov et al. (2007) and Skamarock and Klemp (2008) suggest that a minimum of 12 h spin-up time should be used in the mesoscale NWP model. In most previous studies, a spin-up time of 12 h is often regarded as the best choice directly but without enough verification (Chu et al., 2018). As for complex NWP models like WRF, the selection of spin-up time still needs to be further explored in the context of rainfall simulation.

In this study, to fill in the aforementioned knowledge gaps, a progressive multimetric configuration optimisation (PMCO) method is proposed to explore the sensitivities of the WRF model to the domain configurations, physical parameterized configurations and spin-up times. The goal of this study is to derive a skilful configuration set for the WRF model, which can help to simulate extreme hydrometeorological events in Egypt where ground meteorological observations are usually not available. The questions we attempt to answer in this study are as follows:

- What is the optimal set of WRF configurations for domain size, number of vertical levels, horizontal resolution, spin-up time and physical parameterization schemes (MP, PBL, CU) to simulate extreme rainfall in Egypt?
- Which are the most sensitive WRF configurations for rainfall distribution and intensity?
- How to systematically screen ideal configurations by the proposed PMCO method?
- How to adjust this set of optimal configurations in other rainfall event simulations?

The significance of this study is to show a new WRF configuration optimization method that can make further refinement of the model and address new questions for subsequent research. This study is organized as follows: a brief description of the study area and dataset are presented in Section 2. The WRF model and PMCO method (including experimental design and verification metrics) are illustrated and explained in Section 3. The results of different experimental scenarios are shown in Section 4. Finally, the summary and discussions of this study are presented in Section 5.

2 Study area and datasets

2.1 Study area and event

Alexandria city (and its surrounding region) is selected as the study area. It is the second-largest city in

Egypt and one of the most important trading centres in the world. Its largest port hosts approximately 40% of its industry. Alexandria coastline extends on more than 70 km, from the northwest side of the Nile Delta to Mariout Lake in the east. Under the influence of the Mediterranean climate, Alexandria experiences short mild winters (November to February) and long dry summers (March to October) (Zevenbergen et al. 2017). The temperature usually varies from 10 to 17℃ in winter and from 24 to 30℃ in summer. Besides, the winter is wet with 165 mm of average rainfall, whereas the summer is usually dry with about 30 mm of rainfall on average. The average annual rainfall in Alexandria is only 200 mm.

However, a severe storm occurred on 4th November 2015 in Alexandria and this event was selected for the WRF simulations. During this 50-year storm, more than 100 mm in 2 h rainfall was recorded in some places (Zevenbergen et al. 2017). This event led to a devastating flood that has been described as "the worst flooding of Alexandria City over the past decades in terms of the number of people affected and the amount of economic damage" (IHE Delft, 2017). This rainfall event lasted for about 18 h (from 06:00 to 24:00 UTC). According to the satellite observation used for comparison in this study, the event can be divided into two stages that are heavy rainy stage (6:00-15:00 UTC) and moderate rainy stage (15:00-24:00 UTC). The average rainfall over the Alexandria city and its surrounding region is about 30 mm in the first stage and about 15 mm in the second stage. As the existing urban drainage network was not designed to hold such large volumes of water, 60% of the city area was flooded while the stagnant water remained for more than 15 days in some low-lying areas. This huge flood heavily exceeded the pumping capacity of Alexandria city. If an early flood warning system had been well established to predict this event in advance, appropriate measures could have been taken to mitigate flood damages and losses. Furthermore, the performance stability of the optimal configuration set is also verified through three other rainfall events over Alexandria (25th October 2015), Hurghada (27th October 2016) and Cairo (24th April 2018) in Egypt. These verification rainfall events have different intensities and scales, which is very helpful to understand the moderateness of the optimal configuration set. The location relationships between rainfall events and the Nile River are shown in Figure 1.

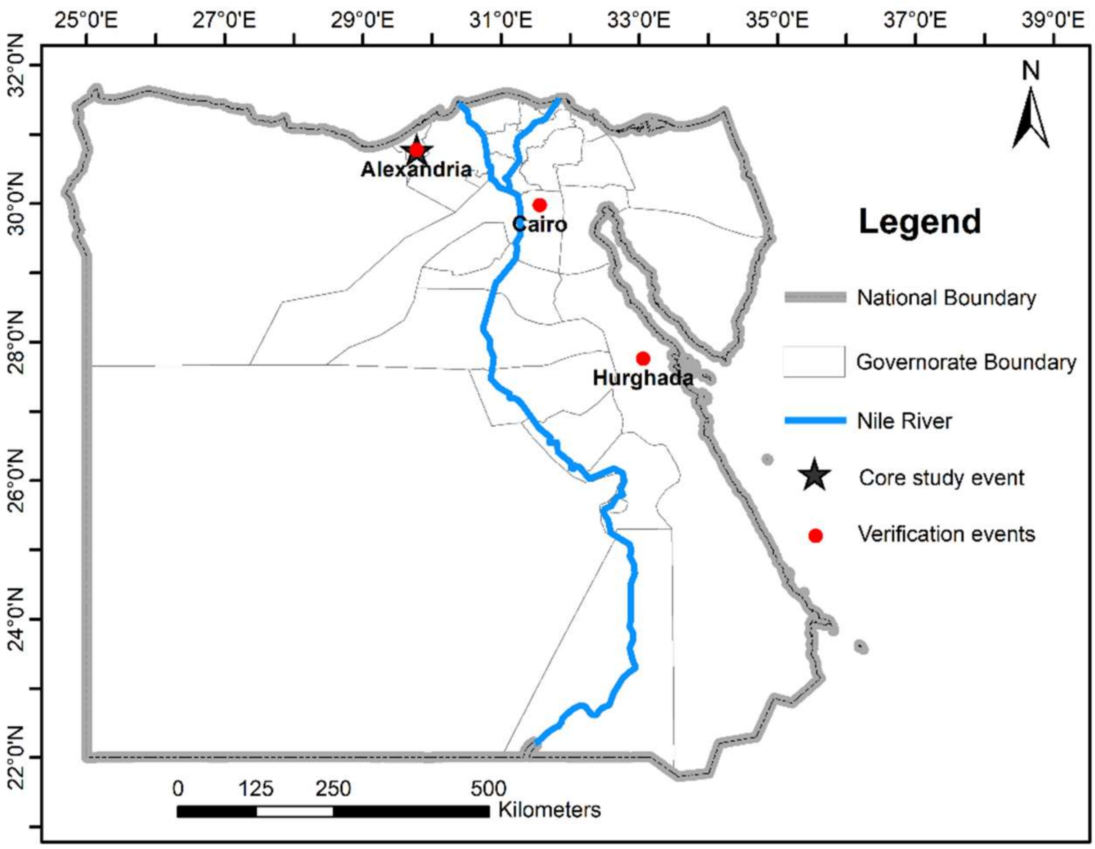

Figure 1. Location of the study events and Nile River in Egypt.

2.2 Datasets

2.2.1 ERA 5 reanalysis dataset

The ERA5 reanalysis dataset was used to initialize the surface and meteorological fields of the WRF model. ERA5 is a newly developed dataset since early 2016 and covers the period from 1950 to the present. This new reanalysis has replaced the ERA-Interim reanalysis started in 2006 and spans the period from 1 January 1979 to 31 August 2019. The new version of ERA5 has a fine spatial resolution (31 km grid spacing compared with 79 km grid spacing for ERA-Interim) and a high temporal resolution (hourly analysis fields compared with 6-hourly for ERA-Interim). In addition, ERA5 contains over 240 parameters on surface and single level alone, which are much more than the 100 parameters in ERA-Interim. These parameters are related to the atmosphere, land, and ocean climate, etc. The reanalysis dataset is available on the European Centre for Medium-Range Weather Forecasts (ECMWF) website (https://www.ecmwf.int/en/forecasts/datasets/browse-reanalysis-datasets). The detailed changes from ERA-Interim to ERA5 can be found in the ECMWF knowledge document (https://confluence.ecmwf.int/pages/viewpage.action?pageId=74764925).

2.2.2 Integrated Multi-satellitE Retrievals for GPM (IMERG) dataset

The IMERG version 06B rainfall product was used for model verification. The IMERG level 3 multisatellite precipitation product combines precipitation estimates from all passive microwave sensors of the GPM constellation, geosynchronous infrared observations from geo-IR satellites, and ground-based measurements from precipitation gauges (Huffman et al. 2019). IMERG provides the quasi-global rainfall estimates from 60°S to 60°N with 0.1°×0.1° gridded resolution and 30 min time interval. IMERG has three product sequences called early, late, and final runs, which with different latency and accuracy. However, by comparing the total rainfalls of the study event in the early and final products, it is found that their distribution and intensity are very close. The reason may be because the small number of rain gauges in Egypt could not provide sufficient adjustments for early run estimates. In addition, the final run product has longer latency (4 months) than the early run product (3.5 hours). Therefore, model performance verifications rely on the early run dataset in this study, which uses the information of geosynchronous infrared observations with a fine time scale to fill the gaps of microwave overpasses coverage (Joyce et al. 2004). This dataset is available on the Global Precipitation Measurement website (https://gpm.nasa.gov/data-access/downloads/gpm) and detailed descriptions can be found in Huffman et al. (2019).

3 WRF model and PMCO method

3.1 WRF model

The model chosen to conduct the rainfall simulation in this study is WRF-ARW version 4.0, the latest generation of the mesoscale NWP models developed by the National Centre for Atmospheric Research (NCAR). WRF-ARW is a compressible, nonhydrostatic, meteorological model with advanced dynamic, physics, software framework and data assimilation system. Its dynamic solver employs Eulerian equations and has a run-time hydrostatic option. As for discretization, WRF uses Arakawa C-grid staggering for the horizontal grid and second- or third- order Runge-Kutta scheme for time integration. The model can conduct one-way interactive, two-way interactive and moving nests with multiple levels and ratios. Besides, it contains nudging capabilities that can be applied on the grid, spectral and observation. Detailed dynamic and physics description of WRF-ARW version 4 can be found in Skamarock et al. (2019). While it received the official support from NCAR, WRF has become a real community model through the long-term contributions from the global user base. Thanks to these, WRF has grown to offer portability capabilities for a range of earth system prediction applications, such as WRF-Chem, WRF-Hydro, and WRF-Fire system.

3.2 Experimental WRF simulations

The proposed sensitivity test is designed as a progressive process to screen the optimal domain configurations, spin-up times, and physical parameterization schemes with enhanced WRF performances. The whole test is divided into two main sections. As shown in Table 1, the first section contains four scenarios that evaluate domain size (S1), vertical levels (S2), nesting ratio (S3) and spin-up time (S4) respectively. In this section, considering the study area features and high spatial resolution of the simulations (less than 5 km), all cases adopt the same physical configurations (mp_physics= Thompson; bl_pbl_physics=YSU; cu_physics=GF) to see the sensitivity of WRF to the above four configurations (Sikder et al., 2016; Srivastava and Bran, 2018). Table 2 shows the second section that investigates the impact of different combinations of physical parameterization schemes (S5), including MP, PBL and CU schemes. The optimal domain configuration and spin-up time filter out in the first section will be applied to the second one. Through these two major experimental parts, a comprehensive suitable WRF configuration plan will be obtained.

Firstly, all cases adopt three levels of two-way nested domains in order to sufficiently improve the horizontal resolution and explicitly resolve the convective-scale processes. Odd nesting ratios (1:3:3; 1:5:5; 1:7:7) are applied to reduce the initial error brought from interpolating the initial fields to the Arakawa grids (Wang et al., 2019). Besides, all nested domains are centred on the same latitude and longitude (31.5°N, 30°E) and all simulations employ Lambert conformal projection (Figure 2 (a-c)). Taking Case 1 as an example, to ensure high horizontal resolution and better application of ERA5 data, the horizontal grid size of the outermost domain (D01) is set to 31.5 km. The largest domain (D01) contains all the main perturbed synoptic features covering the study area. According to the nesting ratios, the middle domain (D02) is the child of D01 with the horizontal grid size of 10.5 km while the smallest domain (D03) is the child of D02 with the horizontal grid size of 3.5 km. The innermost domain (D03) covers the study area of Alexandria and the adjacent areas. The domain sizes depend on the number of grid points. In Case 1, the grid points and domain sizes for D01, D02 and D03 are 80x80 (about 6.19 million km^2), 112x112 (about 1.36 million km^2) and 88x88 (about 0.09 million km^2) respectively. There are different number of vertical levels depending on the experiment (see Table 1) and the top-level pressure is 5,000 Pa. Because some studies (Jankov et al., 2007; Skamarock and Klemp, 2008) suggest a minimum of 12 h spin-up time should be used before rainfall while there was some light rain over the study domain in the 6 hours before the extreme rainfall event, a spin-up time of 18 hours is chosen to warm up the WRF simulations of the first three scenarios. The model outputs are logged hourly for each domain. The lateral boundary conditions are updated every hour using ERA5.

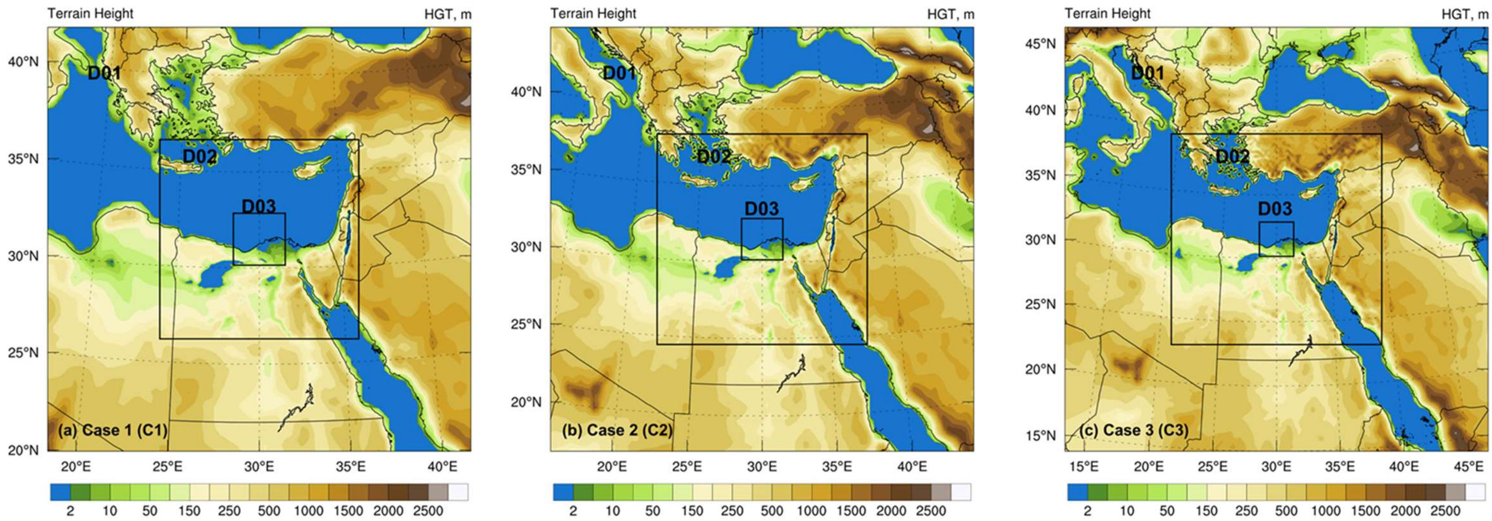

Figure 2. (a)(b)(c) Three different nested domain configurations used in Case 1, Case 2 and Case 3 respectively.

As shown in Table 1, four scenarios are designed to explore the most ideal domain configuration options and spin-up time. S1 includes three cases (C1-C3) that adopt the WRF configurations mentioned above and focus on evaluating the impact of domain size on model performance. To verify whether the domain size assigned in C1 is large enough to develop small-scale atmospheric features, C2 and C3 are devised as comparative experiments. As seen in Figure 2 (a-c), D03 of C1-C3 are the same, while the sizes of D02 and D01 in C2 and C3 are larger than C1. Besides, three nested domains are more than five grid points away from each other to allow for sufficient relaxation. Following this, S2 aims to investigate whether the model run with a higher vertical resolution could get better performance. The optimal domain size found in S1 is directly applied in S2. In scenario 2, the first experiment is the optimal experiment identified in S1 (OS1) and then followed by five comparative experiments (C4-C8). Since the vertical levels should be set at least 34 to reach the required top-level pressure of 5000 Pa, the vertical levels of S2 experiments start from 34 and up to 64. Among them, Case 4 is designed to use the same model level (38 vertical levels) as the ERA5 dataset for comparison. All cases in S2 met the requirement of a grid spacing of less than 1 km in the troposphere. Then comes scenario 3 (S3), the optimal case in S2 (OS2) with nesting ratio of 1:3:3 is compared with another two experiments with increased nesting ratio of 1:5:5 (C9) and 1:7:7 (C10). The grid spacing of D01 is 31.5 km in all S3 experiments, while the grid spacings of D02 and D03 are 10.5 km and 3.5 km (OS2), 6.3 km and 1.26 km (C9), 4.5 km and 0.643 km (C10) respectively. In addition, the grid points of C9 and C10 are increased correspondingly to keep their domain size similar to OS1 (Table 1). Since the number of grid points minus 1 should be an integer multiple of the nesting ratios, it is hard to make the domain size exactly the same when using the different nesting ratios. But their sizes are designed to be as equal as possible in the model. This scenario (S3) adopts the best domain size and vertical resolution configuration found in S1 and S2, as well as examines the change of WRF performance using different domain nesting ratios (horizontal resolutions). After minimizing the uncertainties introduced by domain

Table 1. Experiment categories with different domain configurations (domain size, vertical levels and nesting ratio) and spin-up times.

Scenario	Experiment number	Domain size (grid points)	Vertical levels (vertical resolution)	Nesting ratio (horizontal resolution)	Spin-up time
Domain size (S1)	Case 1 (C1)	D01 80x80; D02 112x112; D03 88x88	34	D01 31.5 km; D02 10.5 km; D03 3.5 km; (1:3:3)	18h
	Case 2 (C2)	D01 100x100; D02 148x148; D03 88x88	as C1	as C1	as C1
	Case 3 (C3)	D01 120x120; D02 178x178; D03 88x88	as C1	as C1	as C1
Vertical levels (S2)	Optimal case in S1 (OS1)	as OS1	34	as C1	as C1
	Case 4 (C4)	as OS1	38 (model level)	as C1	as C1
	Case 5 (C5)	as OS1	44	as C1	as C1
	Case 6 (C6)	as OS1	53	as C1	as C1
	Case 7 (C7)	as OS1	58	as C1	as C1
	Case 8 (C8)	as OS1	64	as C1	as C1
Nesting ratio (S3)	Optimal case in S2 (OS2)	as OS1 (D01 80x80; D02 112x112; D03 88x88)	as OS2	D01 31.5 km; D02 10.5 km; D03 3.5 km; (1:3:3)	as C1
	Case 9 (C9)	Determined by the domain size of OS1 (D01 80x80; D02 186x186; D03 246x246)	as OS2	D01 31.5 km; D02 6.3 km; D03 1.26 km; (1:5:5)	as C1
	Case 10 (C10)	Determined by the domain size of OS1 (D01 80x80; D02 260x260; D03 477x477)	as OS2	D01 31.5 km; D02 4.5 km; D03 0.643 km; (1:7:7)	as C1
Spin-up time (S4)	Case 11 - Case 13 (C11-C13)	as OS1	as OS2	as OS3	0-12h, per 6h
	Optimal case in S3 (OS3)	as OS1	as OS2	as OS3	18h
	Case 14 - Case 31 (C14-C31)	as OS1	as OS2	as OS3	24-126h, per 6h

All these cases use the physics configurations: mp_physics= Thompson; bl_pbl_physics=YSU; cu_physics=GF.

configuration options through the above scenarios (S1-S3), S4 is devised to identify a likely optimal range of spin-up time. This scenario includes the optimal experiment in S3 (OS3) that ran with 18 h spin-up time, and 21 comparative experiments (C11-C31). From C11 to C13, their spin-up times increase from 0 to 12 h with a 6-hour time step. The spin-up times of the remaining comparative experiments (C14-C31) increase from 24 to 126 with a 6-hour time step. All ideal domain configuration options identify in S1-S3 are used in these spin-up time experiments.

After the first section, Table 2 demonstrates how to cross-combine MP, PBL and CU parameterization schemes in the S5. In this scenario, available MP schemes contain Thompson (mp_physics=8), WRF Single-Moment 5-class (WSM5, mp_physics=4) and WRF Single-Moment 6-class (WSM6, mp_physics=6). As for the PBL schemes, this part uses YSU (bl_pbl_physics=1) and MYJ (bl_pbl_physics=2) mentioned in the introduction chapter. Moreover, the three types of CU schemes are Grell-Freitas (GF, cu_physics=3), BMJ (cu_physics=2) and KF (cu_physics=1). Apart from the optimal case in S4 (OS4) that using the Thompson MP scheme, YSU PBL scheme and GF CU scheme, there are other 17 comparative experiments (C32-C48) using different arrangements of three physical parameterization schemes. By conducting the above two main experiment sections, the ideal options for the domain configuration, spin-up time and physical schemes would be drawn.

The order of these five scenarios is determined after fully considering the calculation efficiency of the PMCO method and the sensitivities of rainfall simulations to different configurations. Some test cases have been conducted before the whole optimization process to understand the impacts of different configurations on rainfall simulations. The evaluation starts with domain configuration (S1-S3) and

Table 2. Experiment categories with different physical parameterization schemes (MP, PBLand CU).

Scenario	Experiment number	Microphysics (MP)	Planetary Boundary Layer (PBL)	Cumulus (CU)
Physical parameterization schemes (S5)	Optimal case in S4 (OS4)	8: Thompson (Thompson et al., 2008)	1: YSU (Hong et al, 2006)	3: GF (Grell et al., 2013)
	Case 32 (C32)	8: Thompson	1: YSU	2: BMJ (Janjic, 1994, 2000)
	Case 33(C33)	8: Thompson	1: YSU	1: KF (Kain, 2004)
	Case 34(C34)	8: Thompson	2: MYJ (Janjic, 1994)	3: GF
	Case 35 (C35)	8: Thompson	2: MYJ	2: BMJ
	Case 36 (C36)	8: Thompson	2: MYJ	1: KF
	Case 37 (C37)	4: WSM5 (Hong, Dudhia and Chen, 2004)	1: YSU	3: GF
	Case 38 (C38)	4: WSM5	1: YSU	2: BMJ
	Case 39 (C39)	4: WSM5	1: YSU	1: KF
	Case 40 (C40)	4: WSM5	2: MYJ	3: GF
	Case 41 (C41)	4: WSM5	2: MYJ	2: BMJ
	Case 42 (C42)	4: WSM5	2: MYJ	1: KF
	Case 43 (C43)	6: WSM6 (Hong and Lim, 2006,)	1: YSU	3: GF
	Case 44(C44)	6: WSM6	1: YSU	2: BMJ
	Case 45 (C45)	6: WSM6	1: YSU	1: KF
	Case 46 (C46)	6: WSM6	2: MYJ	3: GF
	Case 47 (C47)	6: WSM6	2: MYJ	2: BMJ
	Case 48 (C48)	6: WSM6	2: MYJ	1: KF

All these cases use the optimal domain configuration and spin-up time found in S1-S4.

follows by spin-up time (S4) and physical parameterization configuration (S5), during which the computational demand of configuration scenarios and the sensitivities of rainfall simulation increase gradually. This order ensures maximum simplicity and stability throughout the optimisation process. Determining domain and spin-up time configuration priority also helps to avoid the influence of poor-quality boundary conditions on subsequent simulations. Besides, the relatively satisfied physical parameterization combination and spin-up time found in the test cases are used as the initial configuration to reduce the impacts of optimization order. The above is the experimental construction of the PMCO method in this study.

3.3 Verification metrics

To cover both spatial and temporal model performances, this study uses seven error metrics proposed by Liu et al. (2012) to evaluate WRF simulation performances with respect to the IMERG observations. On the one hand, four categorical metrics are employed for spatial verification. These metrics include the probability of detection (*POD*), the false alarm ratio (*FAR*), the critical success index (*CSI*) and the frequency bias index (*FBI*). The *POD* and *FAR* represent the probability of detecting rainfall and the probability of false rainfall generated by model simulations. The *CSI* not only shows the probability of rainfall detection but also critical performance, which rewards 'hits' and penalizes both 'misses' and 'false alarms'. At the same time, the *FBI* indicates whether WRF has the tendency to overestimate ($FBI>1$) or underestimate ($FBI<1$) rainfall. But *FBI* does not measure how well the simulation corresponds to the observation. The ideal score for *POD, FAR, CSI* and *FBI* are 1, 0, 1 and 1, respectively. On the other hand, three continuous metrics including the root mean square error (*RMSE*), the mean bias error (*MBE*) and the standard deviation (*SD*) are used for temporal verification. The *RMSE* indicates the average magnitude of error between simulations and observations without showing the bias, while *MBE* indicates the average bias of cumulative error but not corresponds to simulations and observations. The *SD* shows the variation of the simulation error about the *MBE* that reflects the magnitude of random error but without error direction.

All these metrics are calculated by interpolating WRF simulations to the IMERG observation grid at a 3-hour time step in D03. Each verification metric represents the different performance characteristics in spatial or temporal dimensions. Since it is difficult to identify the best case based on these seven different metrics, to uniformly quantify the results of seven metrics, a multimetric decision making analysis called Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) (Hwang and Yoon, 1981) is conducted to obtain the likely best WRF configuration set from 48 cases. Moreover, the uniform performance score is applied to compare the results of different configurations. The further introduction of the multimetric decision making analysis method and the uniform performance score is presented as follows.

TOPSIS determines the best alternative according to the shortest and longest geometric distance from the positive ideal solution and the negative ideal solution, respectively (Assari et al., 2012). TOPSIS method was originally developed by Hwang and Yoon (1981) and further extended and used in numerous researches such as Boran et al. (2009), Sikder et al. (2016) and Goodarzi et al. (2019). In the TOPSIS method, the TOPSIS Relative Closeness Value (*TOPSIS RCV*) is a relative value that determines the best performance based on given metrics. In this study, *TOPSIS RCV* is used as a uniform score to compare overall WRF performances across various configurations. Firstly, seven error metrics are rescaled to calculate the uniform score. Table 3 shows the conversion method between the original error metrics and the rescaled metrics, where subscript "r" means "rescaled". When rescaling the metric values, all the 48 cases used the same max thresholds (FBI_{MAX}= 2, $RMSE_{MAX}$= 12, MBE_{MAX}= 12 (If MBE>0), MBE_{MAX}= -12 (If MBE<0) and SD_{MAX} = 6) and corresponding calculation formulas. The threshold of FBI was set to facilitate rescaling and other thresholds are set according to the rainfall intensity of the study event. If the performance of some simulations exceeds the threshold, they will be removed directly in the configuration experiment. Because in TOPSIS, each metric should have the norm of "high is better" or "low is better." To meet the requirements, seven metrics values are adjusted to range from 0 to 1, where 0 represents the worst performance and 1 is for the best performance. Therefore, the *POD*, *CSI*, rescaled *FBI,* and rescaled *MBE* are "high is better" while *FAR*, *RMSE* and *SD* are "low is better" metrics in this study. Then, all error metrics are assigned equal weight to calculate *TOPSIS RCV* (Equation 1) as the uniform score. The purpose of deriving a uniform score is to allow a convenient and multiscale simulation quality assessment for numerous WRF configurations. In this way, the higher *TOPSIS RCV* means closer to the observation or better simulation performance.

Table 3. Conversion between original and rescaled error metrics.

Original and rescaled error metrics	Original range	Original perfect score	Rescaled thresholds	Rescaled range	Rescaled perfect score
$PODr = POD$	0-1	1	N/A	0-1	1
$FARr = 1 - FAR$	0-1	0	N/A	0-1	1
$CSIr = CSI$	0-1	1	N/A	0-1	1
If $FBI>1$: $FBIr = FBI_{MAX} - FBI$ If $FBI<=1$: $FBIr = FBI$	0-∞	1	+2 max	0-1	1
$RMSEr = (1 - RMSE/RMSE_{MAX})$	0-∞	0	+12 max	0-1	1
If $MBE>0$: $MBEr = 1 - MBE/MBE_{MAX}$ If $MBE<0$: $MBEr = 1 - MBE/-MBE_{MAX}$	-∞-∞	0	-12 to +12	0-1	1
$SDr = (1 - SD/SD_{MAX})$	0-∞	0	+6 max	0-1	1

$$TOPSIS\ RCV = \frac{PODr + FARr + CSIr + FBIr + RMSEr + MBEr + S}{7} \quad (1)$$

4 Results and discussions

In every scenario, seven metrics are compared between the cases, which are calculated at the same period and domain (D03). For the first three scenarios (S1-S3) (i.e., Figure 3-5), the results of each scenario are presented in six subfigures. The first four subfigures (i.e., Figure 3-5 (a-d)) show the values of spatial verification metrics (*POD*, *FAR*, *CSI* and *FBI*) considered over six evaluated sub-periods (6:00-9:00, 9:00-12:00, 12:00-15:00, 15:00-18:00, 18:00-21:00 and 21:00-24:00 UTC on 4th November 2015). The last two subfigures (i.e., Figure 3-5 (e, f)) show the original and rescaled values of all metrics that calculated over the whole event duration (from 6:00 to 24:00). Next is the spin-up time scenario (S4) containing two broken-line graphs. One broken-line graph (Figure 6 (a)) illustrates the changes of the rescaled metric values for 22 simulation cases that run with different spin-up times. Another (Figure 6 (b)) displays the variations of uniform scores (*TOPSIS RCV*) for model simulation results, which also indicates the change in WRF performance with spin-up time. All the metrics in S4 are calculated over the entire event duration. The last scenario (S5) uses three subfigures (Figure 7 (a-c)) to demonstrate the impacts of three physical parameterization schemes on the WRF model performance. After the whole configuration screening process, the total performance ranking of all 48 experiments is displayed in Table 4. Then, the cumulative rainfall maps and rainfall series of the five optimal cases are compared with the IMERG observations (Figure 8, 9). The original and rescaled metric values of the five optimal simulations are shown in Table 5. Finally, the cumulative rainfall maps of three verification events are plotted to show the reproducibility of the recommended configuration in Egypt (Figure 10). Seven performance metric values of the three verification simulations are shown in Table 6.

4.1 Results of the domain size scenario

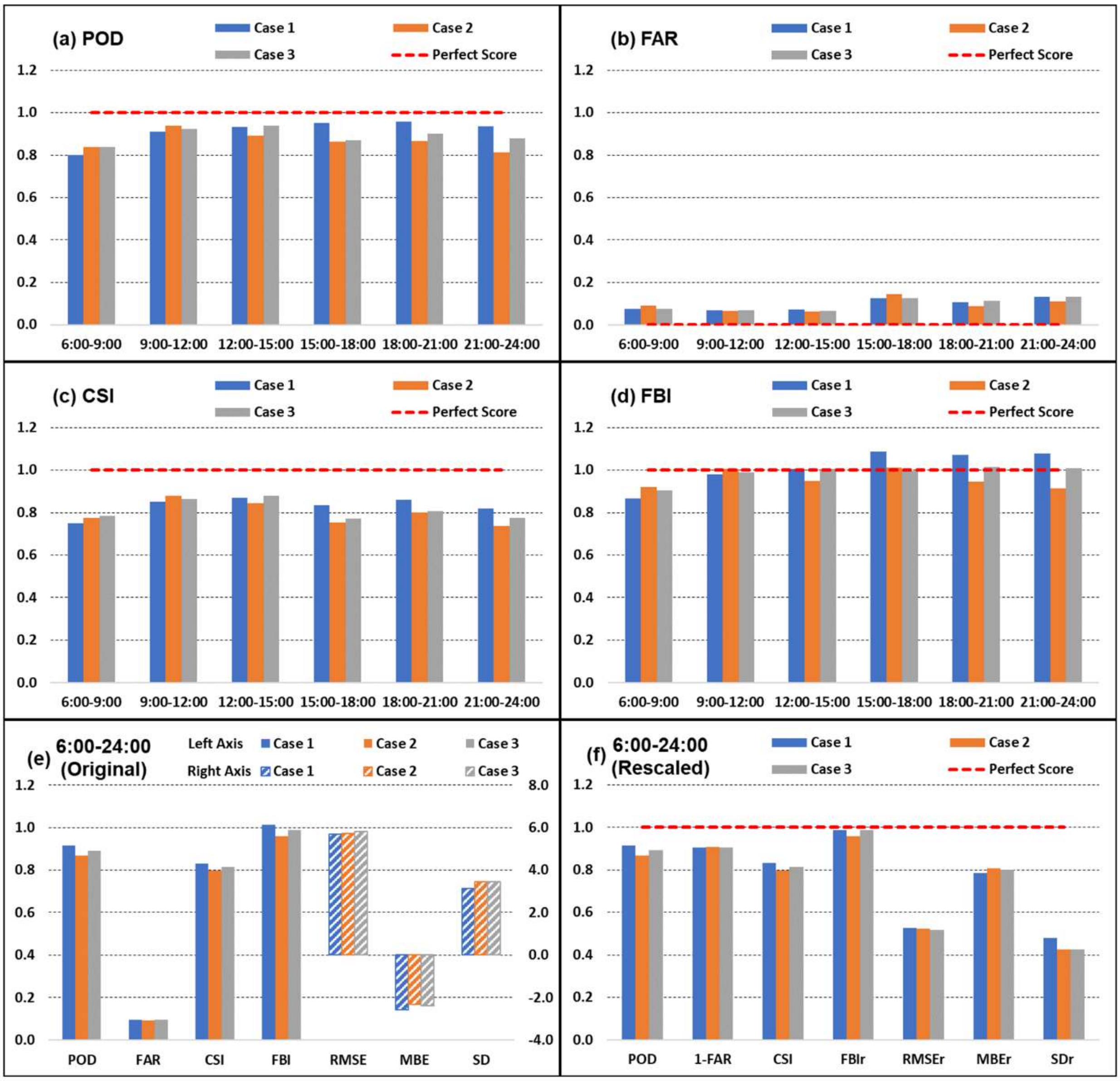

Figure 3. Original and rescaled values of verification metrics for S1. C1 incorporates the smallest nested domains, while C2 and C3 are the intermediate-sized nested domain and largest nested domain. (a)(b)(c)(d) The four spatial metrics calculated over different durations (every three hours) start from 06:00 UTC on 4 November 2015 for the innermost domain. (e) The original values of all metrics calculated over the whole rainfall event duration. (f) The rescaled values of all metrics based on the Table.3 conversion method.

As shown in Figure 3 (a-d), three experiments of S1 show clearly different results in spatial metrics at the early (6:00-15:00) and later (15:00-24:00) evaluation stages. The most obvious differences are detected in the *FAR* that is initially stabilized at 0.07 but all rise to around 0.12 later, which represents the false alarms simulated by the model that increased enormously. The same variation is also found in the *FBI* that most experiments underestimate (*FBI*<1) rainfall occurrences at the first 9 hours whereas overestimate (*FBI*>1) rainfall occurrences in the following period. Furthermore, the *POD* and *CSI* of C2 and C3 significantly decrease in the later stage. However, the *POD* of C1 approximately increases by 0.16 as well as *CSI* by 0.1. The reason for these changes is that the heavy rainfall is mainly concentrated in the first 9 hours and became moderate in the later. Thus, the WRF simulations are less likely to misreport and overestimate rainfall occurrences at the early heavy rainy stage. On the other hand, the spatial association between simulations and observations improves in C1 but deteriorates in C2 and C3 which could be due to the role of the updated boundary conditions in modifying the local model solutions to approach the real atmospheric circulation conditions (Seth and Rojas, 2003). In addition, the domain size of every experiment and rainfall scale of every period can also lead to different result trends.

Comparison of *TOPSIS RCVs* on the whole event scale shows that C1 (Score: 0.775) is a relatively better experiment than C2 (Score: 0.755) and C3 (Score: 0.762). Although C2 and C3 perform slightly better than C1 in terms of *POD, CSI,* and *FBI* at the first two sub-periods (6:00-9:00 and 9:00-12:00), the superiority of C1 is more obvious in the remaining stages. It is because the small domain of C1 uses boundary conditions more efficiently in modifying the false disturbance generated by the local model run. Moreover, C1 performs better in two of rainfall amount estimate metrics (*RMSE* and *SD*) than other experiments. Finally, it achieves the highest score of the five metrics (*POD, CSI, FBIr, RMSEr* and *SDr*) in S1 (Figure 3 (e, f)). These results indirectly indicate that small size domains are more likely to benefit from updated boundary conditions. Small domains are also helpful to simulate rainfall amount stably and accurately. But when simulating large-scale heavy rainfall, like the early evaluation stage of this event, it is easier for large domains to capture the correct hits and spatial patterns of rainfall. Overall, C1 is chosen as the OS1 from both statistical and physical perspectives.

4.2 Results of the vertical levels scenario

According to the S1 analysed results, C1 (OS1) is selected as the starting experiment in S2. As shown in Figure 4 (e, f), unlike the obvious superiority of C1 indicated in S1, the differences in rainfall-related metrics are not apparent between S2 experiments with different vertical levels. But there still exist differences between heavy rainy stage (6:00-15:00) and moderate rainy stage (15:00-24:00) in *FAR* and *FBI* (Figure 4 (b, d)). Similar to Scenario 1, *FAR* stays at 0.07 in the early stage and rise to around 0.11 in the later stage. At the same time, the *FBI* of the most experiments show that WRF underestimates heavy rainfall and overestimates moderate rainfall. However, the gaps in *POD* and *CSI* between different rainy stages became smaller in the most of experiments (OS1, C4, C5 and C7) of S2 (Figure-4(a, c)). S1 filters C2 and C3 and remains C1, which has stable performances in any rainy stages, as the next starting experiment, so the values of *POD* and *CSI* in S2 seem to be more stable than before. The small domain configuration of C1 also helps the follow-on experiments which benefit from the updated boundary conditions with better spatial simulations.

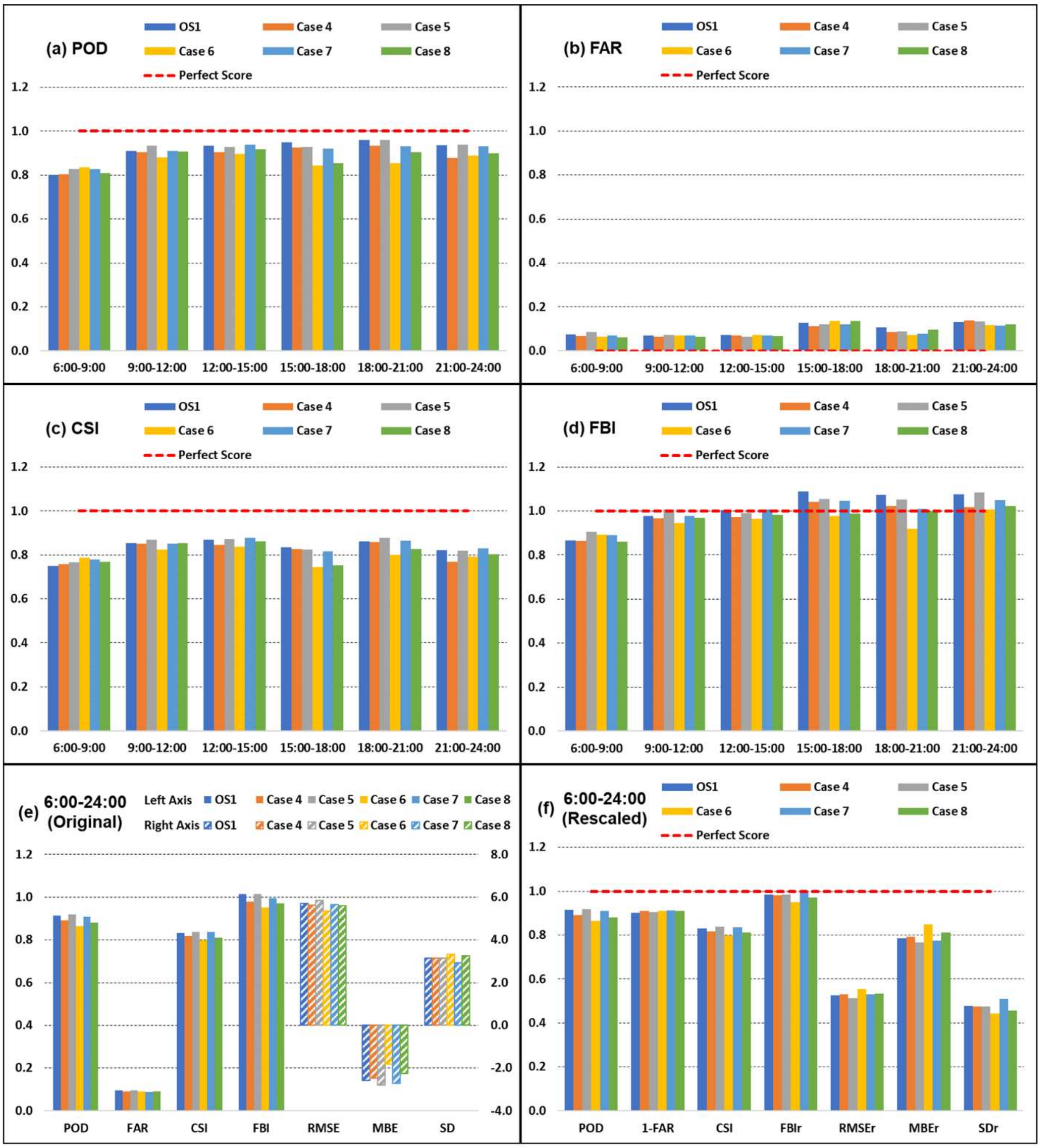

Figure 4. Similar to Figure 3, but for the experiments in S2 with different vertical levels. OS1, C4, C5, C6, C7 and C8 forced by the ERA5 pressure-level data with 34, 38 (WRF model level), 44, 53, 58 and 64 vertical levels, respectively.

Based on the *TOPSIS RCVs* of S2 experiments, C7(Score: 0.782, 58 vertical levels) displays the greatest overall skill. Although C7 does not have the best score in every sub-period, it has stable performance for the spatial metrics and not too poor for the temporal metrics. Comparing C4 (Score: 0.771, the same

38 vertical levels as ERA5) and C5 (Score: 0.772, 44 vertical levels) indicates that the increase of model vertical resolution could enhance WRF's ability to solve small-scale physical processes and improve the spatial correlation (*POD* and *CSI*) of simulated rainfall. However, with a further increase of the vertical resolution in C6 (Score: 0.767, 53 vertical levels) and C8 (Score: 0.768, 64 vertical levels), the scores of all the spatial metrics significantly decline while the accuracy of rainfall amount (*RMSE* and *MBE*) greatly increase. It may be because the progressive reductions in the vertical grid spacing weaken the kinetic energy that favours precipitation and further impacts the scale and intensity of rainfall systems (Sun et al., 2013; Bryan et al., 2003). In contrast, C7 employs a relatively appropriate grid spacing to balance the two aspects of excessive propagation of surface interference and accurate capture of small-scale physical processes. Therefore, C7 gets the better kinetic energy for rainfall and shows better consistency with the observations in terms of the amount and distribution of rainfall. On the other hand, comparing C5, C6, C8 with OS1 (Score: 0.775, 34 vertical levels) shows that, although at a lower vertical resolution, the OS1 overall score is better than the other three cases with a higher resolution. This means that the initial errors introduced by the interpolation process could also cause performance degradation. Overall, C7 shows the best agreement with observations in this scenario.

4.3 Results of the nesting ratio scenario

In the light of the S2 analysed results, C7 (OS2) is chosen as the starting experiment in S3. As Figure 5 (e, f) and Figure 4 (e, f) show, the spatial and temporal metrics are more sensitive to the change of horizontal resolutions than to the variation of vertical resolutions. Especially for *FAR* and *FBI*, the modelling skills of the S3 experiments display different performance trends in heavy and moderate rainy stage as that in the S1 and S2 experiments. Over the evaluated periods, C9 (horizontal grid spacing

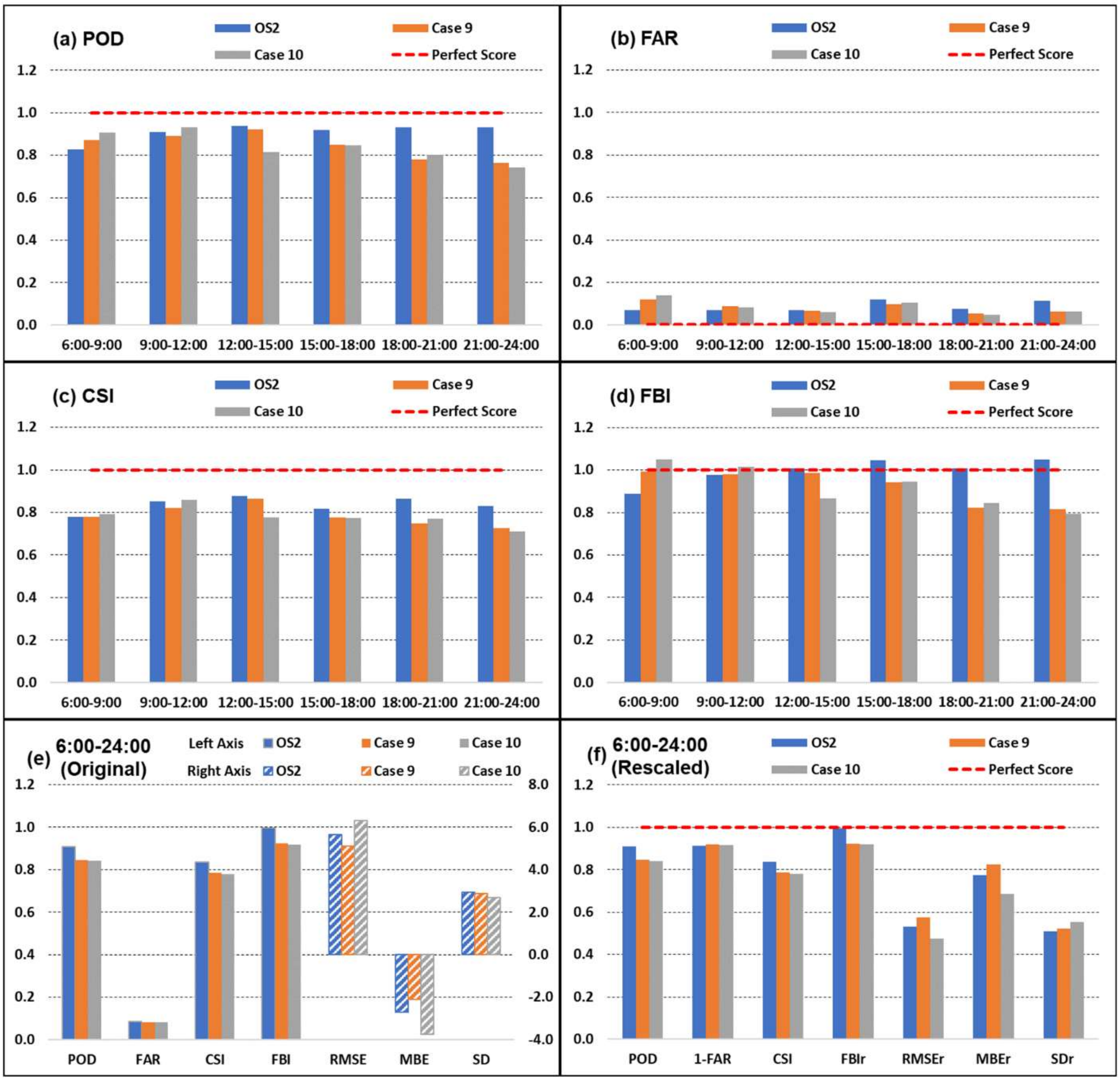

Figure 5. Similar to Figure 3, but for the experiments in S3 with different horizontal resolutions. OS2 has a nesting ratio of 1:3:3 with the horizontal grid spacing of 31.5, 10.5, and 3.5 km, while C9 and C10 have the same largest horizontal grid spacing (31.5 km) with downscaling ratios of 1:5:5 and 1:7:7, respectively. The innermost domain grid spacing is 1.26 km in C9 and 0.643 km in C10.

of 1.26 km) and C10 (horizontal grid spacing of 0.643 km) show either more or less false alarm than OS2 (horizontal grid spacing of 3.5 km) and have no obvious difference between two kinds of rainfall stages. C9 and C10 also present a tendency to underestimate (*FBI*<1) rainfall-scale in most of the periods. However, the *POD* and *CSI* of C9 and C10 extremely decrease when rainfall scale became less. Therefore, WRF simulations with smaller horizontal grid spacings could also lead to poorer results. In theory, this kind of deterioration is due to the biases from the initial and boundary conditions or accumulated errors caused by imperfect model physics, and the chaotic nature of NWP systems also exaggerates them (Liu et al., 2012). As the amount and scale of rainfall became smaller, these errors seem to be more pronounced.

According to the *TOPSIS RCVs* of the horizontal resolution experiments, OS2 (Score: 0.782) is still the optimal experiment in S3 compares with C9 (Score: 0.770) and C10(Score: 0.739). As shown in Figure 5 (a, c), OS2 tends to produce a more accurate spatial pattern than C9 and C10, particularly during the less rainy stage. Besides, OS2 displays a relative perfect tendency ($FBI \approx 1$) to estimate rainfall occurrences (Figure 5 (e)). However, in terms of rainfall amount estimations, C9 presents a significant advantage in *RMSE* and *MBE*. This phenomenon is due to the WRF microphysics scheme which resolves more small-scale features that do not contain in the boundary conditions through higher horizontal resolution. But at the same time, the more external biases and more model accumulated errors mentioned before also reduce the spatial correlation of the simulations. Thus, considering the spatiotemporal accuracy and computational efficiency of WRF simulations, OS2 with the nesting ratio of 1:3:3 is also verified as the optimal experiment of S3.

4.4 Results of the spin-up time scenario

As mentioned above, C7 (OS3) is adopted as the starting experiment in S4. To reduce the influence of chaotic nature on simulations and extend model lead time, this scenario explores 18 different spin-up times for WRF to balance the inconsistencies between boundary conditions and simulation results. Due to the length of spin-up time mostly relies on the domain size and boundary condition disturbance, the

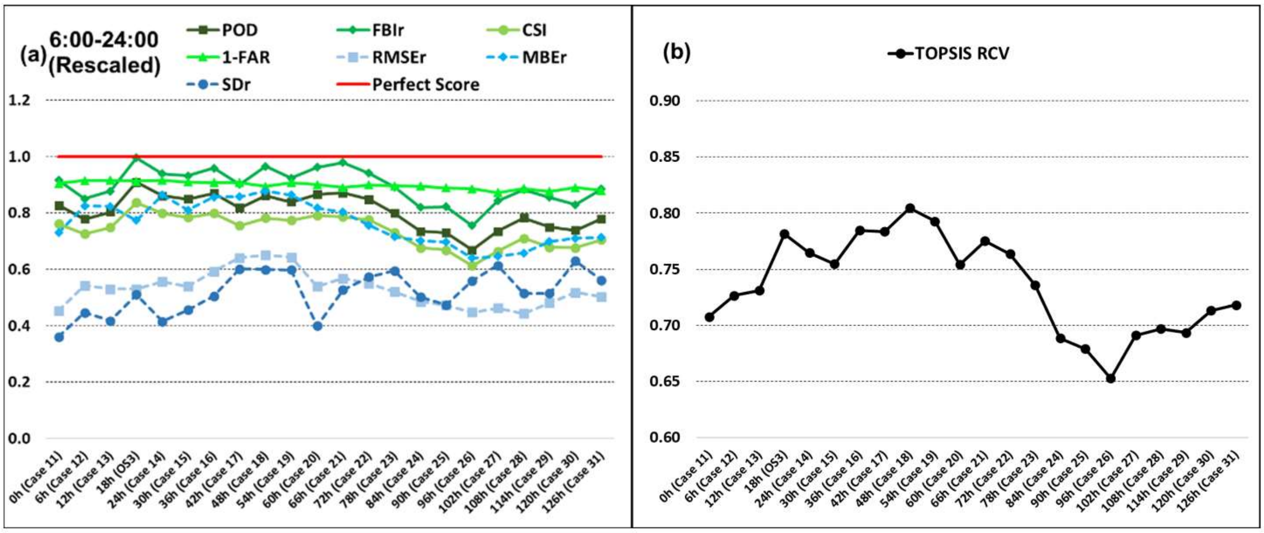

Figure 6. Rescaled verification metrics and *TOPSIS RCV* (uniform score) for the experiments in S4 with different spin-up times. C11 to C13 employs the spin-up time increased from 0 to 12 h by every 6 h. OS3 used a spin-up time of 18 h. From C14 to C31, the spin-up time is grown from 24 to 126 h by every 6 h. (a) shows rescaled values of all verification metrics calculated over the whole rainfall event duration for the innermost domain. (b) shows *TOPSIS RCV* of experiments in S4 calculated by rescaled verification metrics using Formula 1.

spin-up time scenario is placed after the domain configuration scenarios. Unlike the previous scenarios, S4 experiments are sorted by the length of spin-up time and present the performance of individual metrics over the whole event duration. As shown in Figure 6 (a), the performances of rainfall simulations evidently vary with spin-up time. Three spatial metrics *POD, CSI* and *FBI* show similar variations for different spin-up time, while *FAR* is found to display less sensitivity to spin-up time. Furthermore, although there are some small fluctuations, the four spatial metrics maintain a good performance between 18 h and 66 h. As for the temporal metrics, *RMSE, MBE* and *SD* have an evident rise near 48 h. But all these rainfall-related metrics have an obvious decline after 66 h and reach the lowest around 96h. Comparing the variation of the weather conditions during the spin-up period, it is found that a small-scale rainfall occurred near the outmost domain boundary from 30th October 2015. This rainfall became heavier near the start time of the 96 h spin-up time experiment, and it gradually disappeared around the start time of 48 h spin-up time experiment. The water vapour mixing ratio and cloud water mixing ratio at these two times are also significantly different. Therefore, it is reasonable to speculate that WRF runs with longer spin-up times can produce simulations with better spatiotemporal correlation when the model initial conditions are clear and calm. But when the start time is during unstable conditions which would introduce large boundary disturbances, WRF performance could be poor and even need more time to balance the inconsistencies. The small fluctuations among adjacent experiments could be due to the change of initial and boundary conditions, such as water vapour amounts and temperature at the beginning of the simulations.

Comparing the *TOPSIS RCVs* of S4 experiments also indicates that C18 (Score: 0.805, 48 h spin-up time) has the best performance, while C26 (Score: 0.653, 96 h spin-up time) has the worst performance. This overall trend is consistent with the trend summarized above. Moreover, as the optimal experiment of the previous scenario, the overall performance of OS3 (Score: 0.782, spin-up time 18 h) ranks the fifth among the 18 experiments in this scenario. But the spatial correlation performance (*POD, FAR, CSI* and *FBI*) of OS3 ranks first, with an average score up to 0.914. After analysing the weather variables at the start time of OS3, it is found that there is no rainfall and other weather disturbances in the whole simulated domain, which may help WRF simulate rainfall distribution more accurately. Moreover, the water vapour in the domain is also at a relatively low level, and the state of the various hydrometeors (e.g. cloud water, rain water and ice water) is also shown as the constant field. Since the various water contents decide the maximum possible rainfall amount, the *RMSE* and *MBE* of OS3 are worse than other experiments. On the other hand, C11, C12 and C13 simulated with the short spin-up times are not ideal in terms of rainfall distribution and rainfall amount. Thus, it is concluded that the WRF simulations must have a spin-up period of at least 12-18 h before the start of the event. Overall, C18 with the 48 h spin-up time is regarded as the optimal experiment of S4.

4.5 Results of the physical parameterization schemes scenario

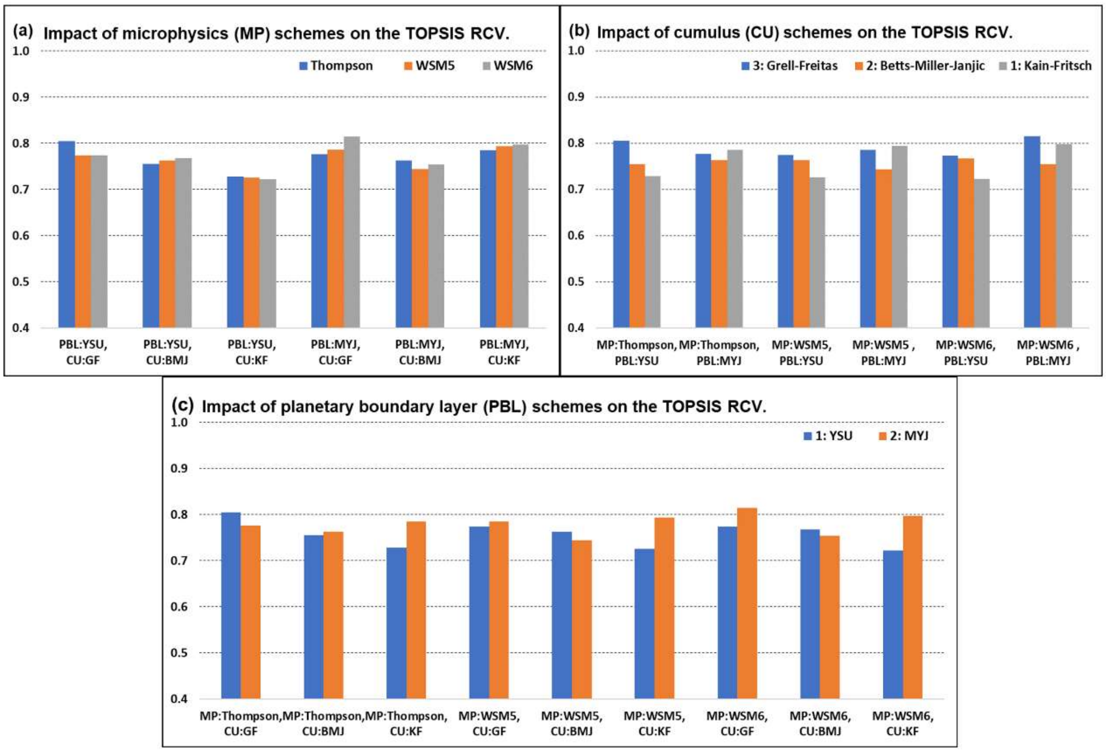

Figure 7. *TOPSIS RCVs* of the experiments in S5 with different physical parameterization schemes. (a) The impact of MP schemes on *TOPSIS RCV*, in which the effects of combining three MP schemes (Thompson, WSM5 and WSM6) with different CU schemes and PBL schemes are compared. (b) Similar to (a) but shows the impact of CU schemes on *TOPSIS RCV*, in which three CU schemes (GF, BMJ and KF) are compared. (c) Similar to (a) but shows the impact of PBL schemes on *TOPSIS RCV*, in which two PBL schemes (YSU and MYJ) are compared.

After the domain size and spin-up time configuration evaluations, C18 (OS4) is selected as the starting experiment to further explore the impact of MP, CU and PBL schemes on WRF performance. As Figure 7 (a-c) shows, WRF performance is less sensitive to MP schemes than CU and PBL schemes. There is no significant performance gap between simulations using the same CU and PBL schemes but different MP schemes, in which the maximum score difference of 0.038 occurs between Case 34 (MP: Thompson, PBL: MYJ and CU: GF) and Case 46 (MP: WSM6, PBL: MYJ and CU: GF). As for CU schemes, its maximum score difference of 0.076 happens between C33 (MP: Thompson, PBL: YSU and CU: KF) and OS4 (MP: Thompson, PBL: YSU and CU: GF). Moreover, the maximum score difference for PBL schemes is 0.075 that appears between C45 (MP: WSM6, PBL: YSU and CU: KF) and C48 (MP: WSM6, PBL: MYJ and CU: KF). These results are due to CU schemes have large influences on the rainfall dynamics and variability, while PBL schemes strongly affect temperature, humidity distribution and rainfall amount (Flaounas et al., 2011). On the other hand, as displayed in Figure 7 (b), BMJ CU scheme seems to be less superior than KF CU scheme and GF CU scheme when combined with various microphysics and planetary boundary layer schemes. The best BMJ CU case (C44, Score: 0.768) is ranked only 10th in terms of overall performance in all physical parameter scheme experiments, while the best KF CU case (C48, Score: 0.797) and the best GF CU case (C46, Score: 0.815) are ranked 3rd and 1st respectively (Table 4). Sometimes, the microphysics should be able to reproduce the convective precipitation in high resolutions and the use of CU scheme is not necessary. In such a case, the computation time will increase, and the rainfall amount may be overestimated (Zheng et al, 2016). By comparing the biases of S5 cases, it can be found that the differences in *RMSE* and *MBE* between GF CU and KF CU are small, while differences for BMJ CU are obviously larger. Thus, the BMJ CU may not be very suitable for this high resolution (3.5 km) application. The lower overall performances of BMJ CU cases are also due to this reason. Besides, the evaluation results of the 6 experiments in Figure 7 (c) show that the MYJ PBL scheme performs much better than the YSU PBL scheme when combined with the KF CU scheme. Thus, it is concluded that the WRF simulation in the Nile Delta and its surrounding regions should avoid using YSU PBL and KF CU together.

Table 4. Total performance (*TOPSIS RCV*) ranking of all experiments.

Scenario	Rank in scenario	Rank in all experiments	Experiment number	*TOPSIS RCV*
Domain size (S1)	1	12	Case 1	0.775
	2	26	Case 3	0.762
	3	27	Case 2	0.755
Vertical levels (S2)	1	10	Case 7	0.782
	2	12	Case 1 (OS1)	0.775
	3	16	Case 5	0.772
	4	17	Case 4	0.771
	5	19	Case 8	0.768
	6	21	Case 6	0.767
Nesting ratio (S3)	1	10	Case 7 (OS2)	0.782
	2	18	Case 9	0.770
	3	33	Case 10	0.739
Spin-up time (S4)	1	2	Case 18	0.805
	2	5	Case 19	0.793
	3	8	Case 16	0.785
	4	9	Case 17	0.783
	5	10	Case 7 (OS3)	0.782
	6	12	Case 21	0.775
	7	22	Case 14	0.765
	8	23	Case 22	0.764
	9	27	Case 15	0.755
	10	31	Case 20	0.754
	11	34	Case 23	0.736
	12	35	Case 13	0.731
	13	37	Case 12	0.727
	14	40	Case 31	0.718
	15	41	Case 30	0.713
	16	42	Case 11	0.708
	17	43	Case 28	0.697
	18	44	Case 29	0.694
	19	45	Case 27	0.691
	20	46	Case 24	0.688
	21	47	Case 25	0.679
	22	48	Case 26	0.653
Physical parameterization schemes (S5)	1	1	Case 46 (OS5)	0.815
	2	2	Case 18 (OS4)	0.805
	3	3	Case 48	0.797
	4	4	Case 42	0.794
	5	6	Case 40	0.786
	5	6	Case 36	0.786
	7	11	Case 34	0.777
	8	14	Case 37	0.774
	8	14	Case 43	0.774
	10	19	Case 44	0.768
	11	24	Case 35	0.763
	11	24	Case 38	0.763
	13	27	Case 32	0.755
	13	27	Case 47	0.755
	15	32	Case 41	0.744
	16	36	Case 33	0.729
	17	38	Case 39	0.726
	18	39	Case 45	0.722

Based on the *TOPSIS RCVs* of S5 experiments, the most recommended scheme combination is C46 (Score: 0.815, MP: WSM6, PBL: MYJ and CU: GF). Moreover, OS4 (Score:0.805, MP: Thompson, PBL: YSU and CU: GF) is also worth recommending. It has a good spatial correlation (*POD=0.862, FAR=0.106, CSI=0.783 and FBI=0.965*) while the ability to estimate rainfall amount is slightly weaker than C46. The scores of the other 16 experiments range from 0.722 (C45) to 0.797 (C48), which demonstrate the strong influence of physical parameterization schemes on WRF performance. After conducting screening experiments of the five scenarios, C46 (OS5) is identified as the experiment that best reproduces the Alexandria extreme rainfall event with the optimal set of domain-related configuration, the ideal length of spin-up time, and the best combination of the physical parameterization schemes. Furthermore, the total performance ranking of all 48 experiments is shown in Table 4 to help understand performance improvements of WRF simulations in the PMCO method.

4.6 Comparison of the optimal simulations in each scenario

To show the importance of model configuration checking, the rainfall maps of five optimal simulations and IMERG observation are compared in Figure 8 (a-e). After five steps of screening, the rainfall amount and its distribution of OS5 are the closest to the IMERG, which has a great improvement in rainfall amount estimation comparing with other optimal experiments. Besides, the rainfall patterns of OS4 and OS5 are more similar to IMERG, while OS1 and OS2 (OS3) are much more dispersed than IMERG. Therefore, the model spin-up time and physical parameterization schemes have a greater impact on simulation performances than domain configuration options. The hourly rainfall changes of WRF optimal simulations and satellite observation are also shown in Figure 9. These rainfall changes are calculated over the D03 (about 0.09 million km^2). It can be found that OS5 is more consistent with the IMERG observation than other simulations, which has an obvious change from the heavy rainy stage (6:00-15:00) to the moderate rainy stage (15:00-24:00). However, rainfall pattern rotation and convection-related mesoscale wind rotation are found in the 2D and 3D animations of the WRF rainfall simulation (included in the "supporting file"), which is not obvious in the IMERG observation. This could be because the satellite didn't sufficiently capture the spatial variability and complex internal structure of the storm at the spatial resolution of about 10 km. It could also be due to the mixture of the higher and lower quality input data from multiple satellite sensors intensified the inconsistency of the IMERG data. Therefore, the WRF simulation is valuable in capturing the rainfall movement that is not easily observed in the IMERG data.

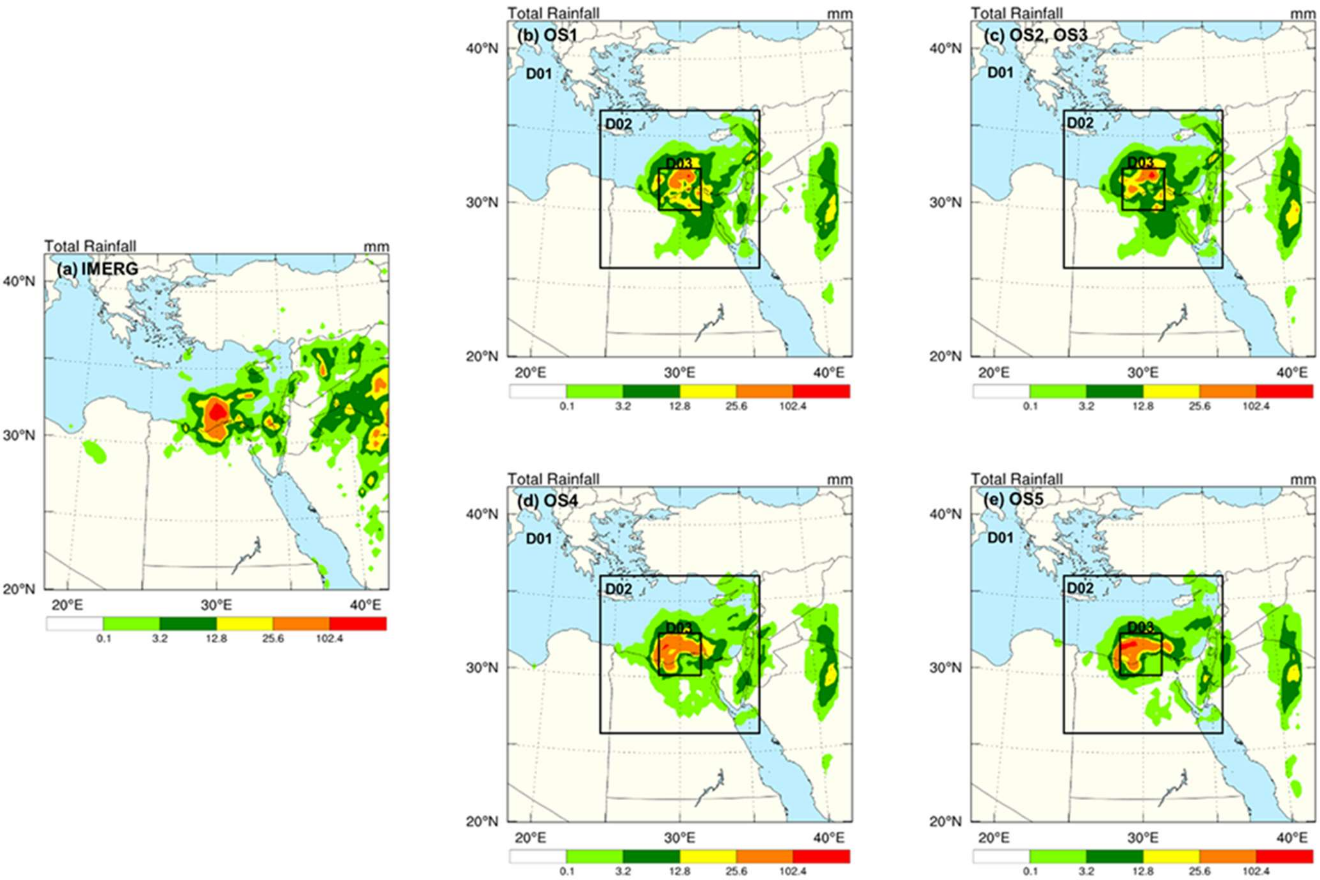

Figure 8. Rainfall maps of satellite observation and WRF simulations drawn over the whole rainfall event from 06:00 UTC to 24:00 UTC on 4 November 2015 for the outermost domain. (a) shows the rainfall map drawn by IMERG data. (b)(c)(d)(e) show the rainfall maps of optimal experiments of every scenario. Among them, the rainfall maps of OS2 and OS3 are both (c). (b)(d)(e) are the rainfall maps of OS1, OS4 and OS5 respectively.

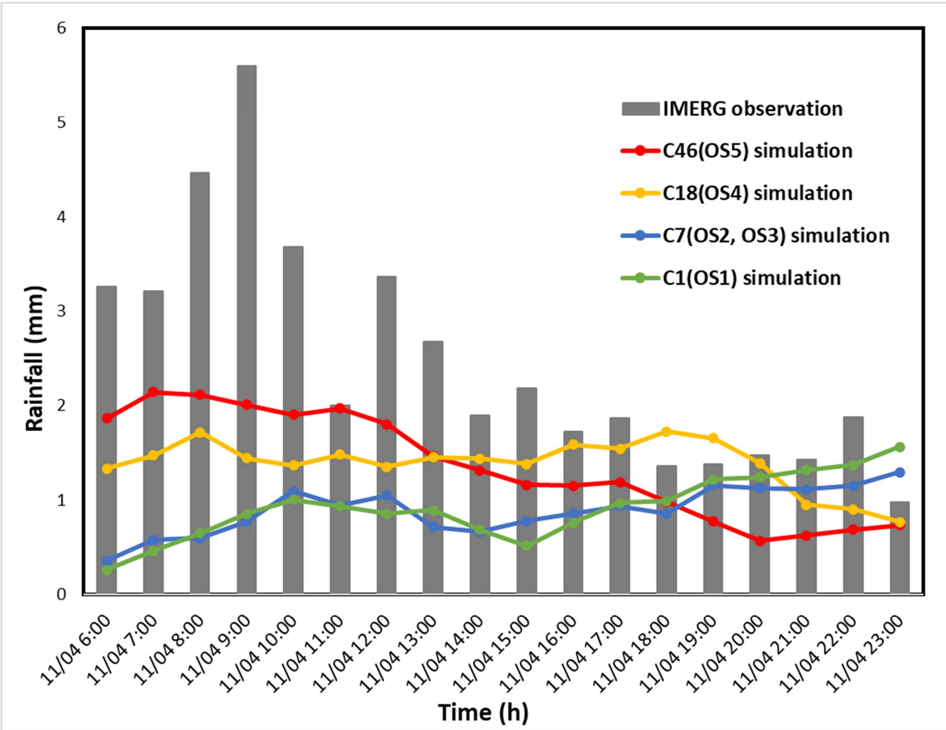

Figure 9. Hourly rainfall series of satellite observation and WRF optimal simulations over the innermost domain (D03) from 06:00 UTC to 24:00 UTC on 4 November 2015. The observed rainfall can be roughly divided into two stages that are heavy rainy stage (6:00-15:00 UTC) and moderate rainy stage (15:00-24:00 UTC).

Apart from the rainfall maps and rainfall series, the performances of the optimal experiments are quantified by seven verification metrics to show their integrated improvements. Table 5 shows the original and rescaled verification metric values calculated over the total evaluation duration of 18 h. Besides, *TOPSIS RCVs* are used to show the extent of total performance improvements between the simulations of different model configurations for the same rainfall event. First, the improvement of OS2 (OS3) contrasted with OS1 is mainly reflected in *FAR*, *FBI* and *SD*. This improvement is because the increase of grid resolution can capture small-scale processes more accurately. Next, when compared with OS3, OS4 performs much better on the temporal metrics *RMSE*, *MBE* and *SD*, but slightly decreases on the spatial metrics. It is related to the different boundary conditions and disturbances at the different model beginning times. Due to the limited kinetic energy for rainfall, the increased ability to simulate rainfall intensity could be accompanied by the decreased ability to simulate rainfall distribution. Finally, OS5 further improves *RMSE, MBE*, *SD* and even *FAR* after employing a better combination of physical parameterization schemes. In summary, *RMSEr* increases from 0.525 in OS1 to 0.699 in OS5, *MBEr* increases from 0.785 to 0.863, *SDr* increases from 0.478 to 0.757, and *FAR* decreases from 0.097 to 0.089. Moreover, the simulation total performance *TOPSIS RCV* also increases notably from 0.775 to 0.815. Although not all the verification metric values improve, the improvements in *TOPSIS RCVs* and rainfall maps all reflect a significant increase in WRF skill after conducting the re-evaluation process.

Table 5. Original metric values and rescaled metric values of the optimal experiments of every scenario.

Original values	***POD***	***FAR***	***CSI***	***FBI***	***RMSE***	***MBE***	***SD***	**Calculation time step**
Case 1 (OS1)	0.915	0.097	0.832	1.015	5.698	-2.575	3.135	3 h in D03
Case 7 (OS2, OS3)	0.910	0.086	0.837	0.996	5.642	-2.705	2.940	3 h in D03
Case 18 (OS4)	0.862	0.106	0.783	0.965	4.194	-1.471	2.404	3 h in D03
Case 46 (OS5)	0.818	0.089	0.759	0.898	3.607	-1.639	1.461	3 h in D03
Rescaled values	***POD***	***1-FAR***	***CSI***	FBI_r	$RMSE_r$	MBE_r	SD_r	***TOPSIS RCV***
Case 1 (OS1)	0.915	0.903	0.832	0.985	0.525	0.785	0.478	0.775
Case 7 (OS2, OS3)	0.910	0.914	0.837	0.996	0.530	0.775	0.510	0.782
Case 18 (OS4)	0.862	0.894	0.783	0.965	0.651	0.877	0.600	0.805
Case 46 (OS5)	0.818	0.911	0.759	0.898	0.699	0.863	0.757	0.815

Note. TOPSIS RCV is used here to show the improvements between simulations of different model configurations for the same rainfall event.

4.7 Reproducibility over Egypt

To investigate the reproducibility of the recommended configurations, three verification simulations are established in Egypt. The first verification event also occurred in Alexandria and it uses the exact same configuration as the core study event. The simulation domains of the other two verification events are moved to the locations centred on Hurghada (latitude 28°N, longitude 34°E) and Cairo (latitude 29.5°N, longitude 31°E) while the other configurations remain unchanged. The rainfall intensity and scale of these three verification events are various in order to better compare with the IMERG observations and verify the stability of the recommended configuration set.

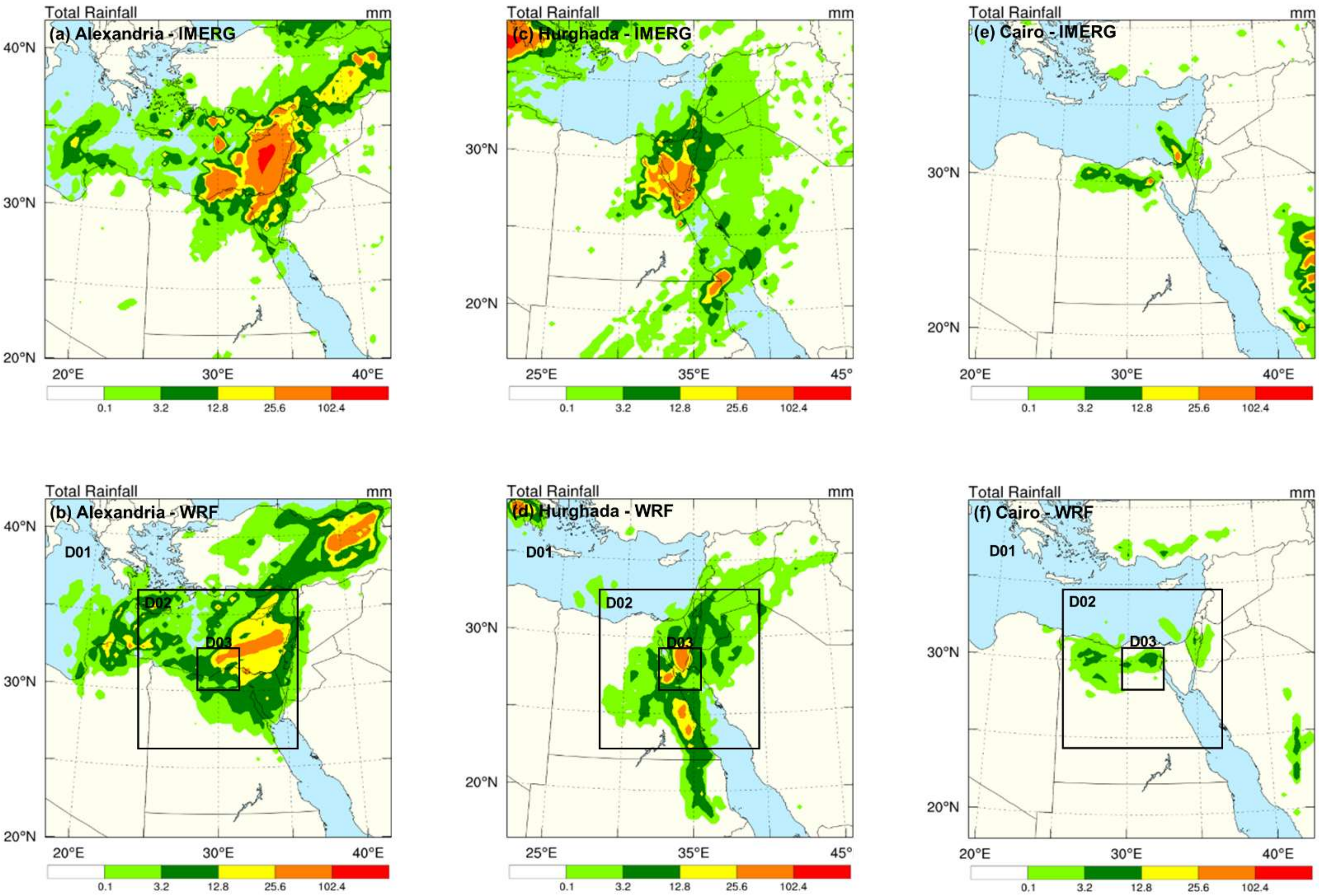

Figure 10. Rainfall maps for different rainfall events drawn by satellite observation (upper figures) and WRF simulations (lower figures). (a) and (b) show the rainfall maps for another Alexandrian rainfall event (Event 1) occurred from 00:00 UTC to 12:00 UTC on 25 October 2015 over their outermost domain, which are plotted by IMERG data and WRF simulation (applied the optimal configuration set found in the core study event), respectively. Similarly, (c) and (d) show the rainfall maps for the Hurghada rainfall event (Event 2) occurred from 08:00 UTC to 20:00 UTC on 27 October 2016. (e) and (f) show the rainfall maps for Cairo rainfall event (Event 3) occurred from 12:00 UTC to 24:00 UTC on 24 April 2018.

As shown in Figure 10, rainfall locations and ranges of the three verification events simulated using the recommended configurations are very similar to the IMERG observations, which is a satisfying result for the NWP model simulations. The simulated rainfall distributions of Alexandria and Cairo events agree well with the observations. Although the Hurghada event had some light rain poorly captured, the rainfall in the central target city is well presented. In addition to the rainfall maps, the original metric values of the three verification events are shown in Table 6. They are also calculated between WRF and IMERG for every 3 h, which is the same as the core event study. Because these three rainfall events have different scales and intensities, different thresholds need to be used to rescale their simulation verification metrics and show their performances. But if different thresholds are used to calculate *TOPSIS RCV*, the comparison between the simulations of different verification events will not be objective. Therefore, the *TOPSIS RCV* of the PMCO method is not used here for comparison, but this method is still useful for re-evaluating the simulations of different configurations for the same event. According to Table 6, the performances of the first two large-scale extreme rainfall simulations are very good in both spatial and temporal dimensions. Although the scales and intensities of Event 1 and Event 2 are larger than the core study event, the recommended configurations still give the simulations stable performances. But the spatial performances of the last small-scale rainfall simulation seem to be poor. It may be due to the Saudi Arabia rainfall event near the domain boundary (the lower right corner of Figure 10 (e, f)) which strongly influenced the simulated small-scale rainfall. Besides, the poor spatial results could also be due to the time step is too small to reflect the real performance for small rainfall simulation. If calculated by the accumulated rainfall (12 h), the *POD* of verification Event 3 is 0.943, *FAR* is 0.246, *CSI* is 0.721, and *FBI* is 1.251. For such small-scale rainfall, the results of accumulated rainfall seem more representative than the results calculated by the time step of 3 h.

Table 6. Original metric values of the three verification rainfall events.

Original values	*POD*	*FAR*	*CSI*	*FBI*	*RMSE*	*MBE*	*SD*	Calculation time step
Alexandria (Event 1)	0.866	0.137	0.766	1.005	4.161	-3.575	1.875	3 h in D03
Hurghada (Event 2)	0.842	0.092	0.775	0.933	5.631	-3.947	3.944	3 h in D03
Cairo (Event 3)	0.490	0.645	0.220	2.992	0.584	-0.361	0.436	3 h in D03

Note. TOPSIS RCV is not used here to compare the overall performance because the scale and intensity of verification events are different.

Although the good stability of the recommended configurations is proved in other extreme rainfall events, these rainfall simulations could be further improved if two adjustments can be made. Firstly, it is better to increase the number of grids and expand the simulated domain for the large-scale rainfall in Alexandria and Hurghada verification events. Because their rainfall affects a wider area than the core study event and even reach the boundary of D01, which is unfavourable for the rainfall simulation inside the domain. By contrast, the domain size for the small-scale rainfall simulation in Cairo could be reduced to avoid the disturbances from the Saudi Arabia rainfall event near the boundary. The quality of rainfall simulation is very sensitive to the input conditions such as soil moisture and the latent heat flux (Kleczek et al.,2014; Bonekamp et al.,2018; Lu et al., 2020). Thus, the domain size and location of simulation should be considered according to the specific conditions of each event. It is also worth mentioning that the rainfall near the boundary is easily underestimated, such as the upper left corner of Figure 10 (c, d) and the lower right corner of Figure 10 (e, f). Second, according to the previous hypothesis, the spin-up time may need to be reconsidered to obtain the best starting point for different rainfall simulations. It can help WRF to simulate more accurate rainfall intensity. On the whole, the optimal configuration set found in this study shows relatively stable performance in extreme rainfall events over Egypt. The recommended configurations can also be used as a basis for modification according to different rainfall conditions. Therefore, it can be used as a common set for extreme rainfall simulations over Egypt or a reference set for other simulations beyond Egypt.

5 Summary and conclusions

This study conducts a set of evaluation tests by the PMCO method to explore the effects of domain size, numbers of vertical levels, nesting ratio, spin-up times, and physical parameterization schemes on WRF simulation for the extreme rainfall event on 4th November 2015 in Alexandria, Egypt. It contains five scenarios and 48 sub-experiments. Their initial conditions are provided by the ERA5 reanalysis datasets and WRF static geographical datasets, while the simulation results are verified by the IMERG rainfall product. To help quantify simulation skills and screen the optimal configuration, four rainfall distribution error metrics and three rainfall amount error metrics are calculated at different time periods and summarized as *TOPSIS RCV* to show the overall performance of each experiment. Then the values of seven error metrics and *TOPSIS RCVs* under different conditions are compared as the basis for the subsequent evaluation process of optimal configuration set. Finally, the optimal configuration set is applied in the simulations of other three rainfall events to verify its stability and efficiency. The entire test aims to identify the likely optimal configuration set of WRF to reproduce the extreme rainfall event in Egypt as well as to find key performance relationships to help improve the simulation of other rainfall events.

Comparing the first optimal case C1 (OS1) and finally recommended setting C46 (OS5) shows that the rainfall-related verification metrics and *TOPSIS RCV* are significantly improved by the PMCO method. In particular, *FAR* decreases from 0.097 to 0.089, *RMSE* reduces by 36.84%, *MBE* reduces by 36.28%, and *SD* reduces by 53.45%. Therefore, re-evaluating various WRF configurations could bring substantial benefits for the studies of extreme rainfall simulation. Besides, the ideal optimal configurations found in this study also have good performances in the other three rainfall events that occurred at Alexandria, Hurghada and Cairo, which manifests the importance and practicality of the configuration assessment study. In summary, the most recommended configurations for Egypt comprises of three-level nested domains (D01 80x80; D02 112x112; D03 88x88), 58 vertical levels, 1:3:3 horizontal nesting ratio (31.5, 10.5 and 3.5km), 48h spin-up time, WSM6 microphysics scheme, MYJ planetary boundary layer scheme and GF cumulus convection scheme.

The previous analyses reveal that the performance of WRF simulations varies greatly with different sets of configurations. Therefore, it is worthwhile to explore their relationships on the spatial and temporal scale in order to improve the reproducibility of extreme rainfall. Analysis results reveal the following.

1. The appropriate small domain is more effective in using the updated boundary conditions and modifying model running errors. But domain sizes should not be fixed and need to be adjusted according to the scale and intensity of individual rainfall events.
2. Increasing vertical grid resolution could enhance the development of small-scale physical processes and obtain better rainfall distribution. However, considering the effect of amplifying surface disturbances, a relatively suitable number of vertical levels should be between 50 and 60 when top-level pressure is set as 5000 Pa. This optimal vertical level number is consistent with the suggestion of Chu et al. (2018).
3. The horizontal nesting ratio of 1:3:3 (the grid size of 31.5, 10.5 and 3.5 km) is sufficiently powerful for accurate rainfall simulations, which helps to improve computational efficiency and reduce model running perturbations. According to many previous studies, about three-fifths of them (Sikder et al., 2016, Liu et al., 2012, Wilson and Barros, 2015, Zhang at al., 2019, Goodarzi et al., 2019) employed 3-4 km horizontal grid spacing and obtained good rainfall simulations.
4. WRF simulations with insufficient spin-up periods (less than 12 h) are difficult to balance the inconsistencies between boundary conditions and simulation results, which lead to poor

performance in capturing extreme rainfall. Moreover, there is an interesting finding that the regional weather conditions (especially near the domain boundary) at the start of simulations may impact the time required for model initialization, which has not been seen in other studies. If this hypothesis is true, the optimal length of the model spin-up time may vary for every specific event. A further exploration of this will be conducted in our future studies. But a minimum of 12 h spin-up time is still recommended for each WRF simulations.

5. Some physical parameterization schemes and combinations are found inadequate to simulate rainfall over Egypt, such as the BMJ CU scheme and the combination of YSU PBL scheme and KF CU scheme. Different from some relevant studies (Flaounas et al., 2011, Evans et al., 2012), it is found that the combination of MYJ PBL scheme and GF CU scheme could improve rainfall simulation over Egypt. The selection and combination of physical schemes have large impacts on temperature, humidity distribution, rainfall dynamics, and rainfall variability. So the adaptability of physical scheme combination for the study area should be clearly understood and considered before conducting rainfall simulations.
6. The rainfall distribution and magnitude were more sensitive to the changes in spin-up time and physical parameterization schemes than the three domain configuration options. And the change of vertical resolution has the smallest impact on WRF performance among all these configurations. Besides, it is observed that the improvement of WRF's reproducibility of rainfall intensity is usually accompanied by a decrease in the reproducibility of rainfall distribution.
7. The identification of ideal simulation according to just one type of metrics or only one evaluation period may result in limited conclusions. The multimetric decision making analysis method is a good choice that can summarize various metrics related to rainfall distribution and rainfall

intensity into an overall performance for comparisons. This study details how to use the PMCO method to screen the optimal WRF configuration for rainfall simulation. In future, this evaluation method should be applied to more case studies to help assess the uncertainties of WRF modelling.

8. The moderateness of the recommended model configurations should be considered and adjusted before conducting rainfall simulations and designing hydrometeorological early warning systems. For example, as suggests in Section 4.7, the size of simulated domains and the number of horizontal grids should be modified according to the scale of rainfall. The recommended configuration summarized in this study could be used as a common set over Egypt and a reference set for simulations in other regions.

The Nile Delta in Egypt is a vulnerable zone that faces growing pluvial flooding hazards in recent decades, while inadequate coverage of in-situ rainfall observations (radars and rain gauges) makes the development of a hydrometeorological early warning system very difficult. WRF has been proven to be an effective way to simulate weather events by downscaling global NWP products to the interested areas, which is very suitable and feasible for countries like Egypt. In this study, the sensitivities and uncertainties of various configurations are extensively explored through the PMCO method. However, there is a limitation of this study that the PMCO method may get different results when using different study events. Thus, it is very important to select a representative event of the study area to obtain the optimal configuration. If extreme rainfall with different characteristics can be classified systematically and the PMCO method can be conducted to each classification, then such a comprehensive configuration list may be better for the subsequent applications. However, it is best to conduct this further study in an area with a large number of recorded extreme rainfall events. At this stage, as the research on WRF simulation in Egypt is very limited, we hope our study could just provide a useful guide and fill the knowledge gaps for further studies. Another limitation is the relationship between spin-up time and simulation performance has not been fully demonstrated due to the lack of more accurate verification data like radars and rain gauges. Further work will aim to perfect this issue. Overall, the knowledge gains in this study provides a useful foundation for developing a flood early warning system by linking WRF (with the global forecast data) with WRF-Hydro (to convert rainfall into floods).

CRediT authorship contribution statement

Ying Liu: Methodology, Formal analysis, Validation, Visualization, Software, Funding acquisition, Writing - original draft, Writing - review & editing. **Yiheng Chen:** Methodology, Visualization. **Otto Chen:** Data curation, Funding acquisition. **Jiao Wang:** Software, Visualization. **Lu Zhuo:** Conceptualization, Methodology, Writing - review & editing. **Miguel A. Rico-Ramirez:** Data curation, Supervision, Writing - review & editing. **Dawei Han:** Conceptualization, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study is supported by the Institutional Links grant [332430681] under the Newton-Mosharafa Fund. The grant is funded by the UK Department of Business, Energy and Industrial Strategy (BEIS) and

Egypt Science and Technology Development Fund (STDF) and delivered by the British Council. The first author gratefully acknowledges the China Scholarship Council (CSC No, 201908310086) and University of Bristol joint-funded PhD Scholarship. The authors acknowledge the Advanced Computing Research Centre at the University of Bristol for providing access to the High-Performance Computing system BlueCrystal.

References

Awan, N.K., Truhetz, H. and Gobiet, A., 2011. Parameterization-induced error characteristics of MM5 and WRF operated in climate mode over the Alpine region: an ensemble-based analysis. Journal of Climate, 24(12), pp.3107-3123.

Aligo, E.A., Gallus Jr, W.A. and Segal, M., 2009. On the impact of WRF model vertical grid resolution on Midwest summer rainfall forecasts. Weather and forecasting, 24(2), pp.575-594.

Assari, A., Mahesh, T. and Assari, E., 2012. Role of public participation in sustainability of historical city: usage of TOPSIS method. Indian Journal of Science and Technology, 5(3), pp.2289-2294.

Brath, A., Burlando, P. and Rosso, R., 1988. Sensitivity analysis of real-time flood forecasting to on-line rainfall predictions. In Selected Papers from the Workshop on Natural Disasters in European-Mediterranean Countries, Perugia, Italy (pp. 469-488).

Barnsley, M.J., 2007. Environmental modeling: A practical introduction. CRC Press.

Boran, F.E., Genç, S., Kurt, M. and Akay, D., 2009. A multi-criteria intuitionistic fuzzy group decision making for supplier selection with TOPSIS method. Expert Systems with Applications, 36(8), pp.11363-11368.

Bryan, G.H., Wyngaard, J.C. and Fritsch, J.M., 2003. Resolution requirements for the simulation of deep moist convection. Monthly Weather Review, 131(10), pp.2394-2416.

Bonekamp, P.N.J., Collier, E. and Immerzeel, W.W., 2018. The impact of spatial resolution, land use, and spinup time on resolving spatial precipitation patterns in the Himalayas. Journal of Hydrometeorology, 19(10), pp.1565-1581.

Cluckie, I.D. and Han, D., 2000. Fluvial flood forecasting. Water and Environment Journal, 14(4), pp.270-276.

Chu, Q., Xu, Z., Chen, Y. and Han, D., 2018. Evaluation of the ability of the Weather Research and Forecasting model to reproduce a sub-daily extreme rainfall event in Beijing, China using different domain configurations and spin-up times. Hydrology and Earth System Sciences, 22(6), p.3391.

Dai, Q., Bray, M., Zhuo, L., Islam, T. and Han, D., 2017. A scheme for rain gauge network design based on remotely sensed rainfall measurements. Journal of Hydrometeorology, 18(2), pp.363-379.

Di, Z., Duan, Q., Gong, W., Wang, C., Gan, Y., Quan, J., Li, J., Miao, C., Ye, A. and Tong, C., 2015. Assessing WRF model parameter sensitivity: A case study with 5 day summer precipitation forecasting in the Greater Beijing Area. Geophysical Research Letters, 42(2), pp.579-587.

ElTahan, M. and Magooda, M., 2017. Evaluation of different WRF microphysics schemes: severe rainfall over Egypt case study. arXiv preprint arXiv:1711.04163.

Evans, J.P., Ekström, M. and Ji, F., 2012. Evaluating the performance of a WRF physics ensemble over South-East Australia. Climate Dynamics, 39(6), pp.1241-1258.

Flaounas, E., Bastin, S. and Janicot, S., 2011. Regional climate modelling of the 2006 West African monsoon: sensitivity to convection and planetary boundary layer parameterisation using WRF. Climate Dynamics, 36(5-6), pp.1083-1105.

Goswami, P., Shivappa, H. and Goud, S., 2012. Comparative analysis of the role of domain size, horizontal resolution and initial conditions in the simulation of tropical heavy rainfall events. Meteorological Applications, 19(2), pp.170-178.

Goodarzi, L., Banihabib, M.E. and Roozbahani, A., 2019. A decision-making model for flood warning system based on ensemble forecasts. Journal of Hydrology, 573, pp.207-219.

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E.J., Sorooshian, S., Tan, J. and Xie, P., 2019. NASA Global Precipitation Measurement Integrated Multi-Satellite Retrievals for GPM (IMERG). Algorithm Theoretical Basis Doc., version 6, 34 pp., https://docserver.gesdisc.eosdis.nasa.gov/public/project/GPM/IMERG_ATBD_V06.pdf.

Hwang, C.L. and Yoon, K., 1981. Methods for multiple attribute decision making. In Multiple attribute decision making (pp. 58-191). Springer, Berlin, Heidelberg.

IHE Delft, 2017. Improving flood resilience of Alexandria, Egypt, https://www.un-ihe.org/news/improving-flood-resilience-alexandria-egypt.

Ji, F., Ekström, M., Evans, J.P. and Teng, J., 2013. Evaluating rainfall patterns using physics scheme ensembles from a regional atmospheric model. Theoretical and applied climatology, 115(1-2), pp.297-304.

Jankov, I., Gallus Jr, W.A., Segal, M. and Koch, S.E., 2007. Influence of initial conditions on the WRF–ARW model QPF response to physical parameterization changes. Weather and forecasting, 22(3), pp.501-519.

Joyce, R.J., Janowiak, J.E., Arkin, P.A. and Xie, P., 2004. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. Journal of hydrometeorology, 5(3), pp.487-503.

Krieger, J.R., Zhang, J., Atkinson, D.E., Zhang, X. and Shulski, M.D., 2009. P1. 2 Sensitivity of WRF model forecasts to different physical parameterizations in the beaufort sea region. In The Eighth Conference on Coastal Atmospheric and Oceanic Prediction and Processes.

Kain, J.S., Weiss, S.J., Bright, D.R., Baldwin, M.E., Levit, J.J., Carbin, G.W., Schwartz, C.S., Weisman, M.L., Droegemeier, K.K., Weber, D.B. and Thomas, K.W., 2008. Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. Weather and Forecasting, 23(5), pp.931-952.

Knievel, J.C., Ahijevych, D.A. and Manning, K.W., 2004. Using temporal modes of rainfall to evaluate the performance of a numerical weather prediction model. Monthly weather review, 132(12), pp.2995-3009.

Kleczek, M.A., Steeneveld, G.J. and Holtslag, A.A., 2014. Evaluation of the weather research and forecasting mesoscale model for GABLS3: impact of boundary-layer schemes, boundary conditions and spin-up. Boundary-layer meteorology, 152(2), pp.213-243.

Liu, J., Bray, M. and Han, D., 2012. Sensitivity of the Weather Research and Forecasting (WRF) model to downscaling ratios and storm types in rainfall simulation. Hydrological Processes, 26(20), pp.3012-3031.

Li, L., Li, W. and Jin, J., 2014. Improvements in WRF simulation skills of southeastern United States summer rainfall: physical parameterization and horizontal resolution. Climate dynamics, 43(7-8), pp.2077-2091.

Lu, L., Pontoppidan, M., Sobolowski, S. and Senatore, A., 2020. The impact of initial conditions on convection-permitting simulations of a flood event over complex mountainous terrain. Hydrology and Earth System Sciences, 24(2), pp.771-791.

Nossent, J., Elsen, P. and Bauwens, W., 2011. Sobol'sensitivity analysis of a complex environmental model. Environmental Modelling & Software, 26(12), pp.1515-1525.

Pei, L., Moore, N., Zhong, S., Luo, L., Hyndman, D.W., Heilman, W.E. and Gao, Z., 2014. WRF model sensitivity to land surface model and cumulus parameterization under short-term climate extremes over the southern Great Plains of the United States. Journal of Climate, 27(20), pp.7703-7724.

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

Skamarock, C., Bogumiła Klemp, Jimy Dudhia, Olivia Gill, Zhiquan Liu, Judith Berner, Wei Wang, Georgia Powers, Gerald Duda, Dale Barker and Xiang-Yu Huang, 2019. A Description of the Advanced Research WRF Model Version 4.

Seth, A. and Rojas, M., 2003. Simulation and sensitivity in a nested modeling system for South America. Part I: Reanalyses boundary forcing. Journal of Climate, 16(15), pp.2437-2453.

Schwartz, C.S., Kain, J.S., Weiss, S.J., Xue, M., Bright, D.R., Kong, F., Thomas, K.W., Levit, J.J. and Coniglio, M.C., 2009. Next-day convection-allowing WRF model guidance: A second look at 2-km versus 4-km grid spacing. Monthly Weather Review, 137(10), pp.3351-3372.

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

Sikder, S. and Hossain, F., 2016. Assessment of the weather research and forecasting model generalized parameterization schemes for advancement of precipitation forecasting in monsoon-driven river basins. Journal of Advances in Modeling Earth Systems, 8(3), pp.1210-1228.

Srivastava, R. and Bran, S.H., 2018. Impact of dynamical and microphysical schemes on black carbon prediction in a regional climate model over India. Environmental Science and Pollution Research, 25(15), pp.14844-14855.

Sun, Y., Yi, L., Zhong, Z., Hu, Y. and Ha, Y., 2013. Dependence of model convergence on horizontal resolution and convective parameterization in simulations of a tropical cyclone at gray-zone resolutions. Journal of Geophysical Research: Atmospheres, 118(14), pp.7715-7732.

Vannitsem, S. and Chomé, F., 2005. One-way nested regional climate simulations and domain size. Journal of climate, 18(1), pp.229-233.

Warner, T.T., 2011. Quality assurance in atmospheric modeling. Bulletin of the American Meteorological Society, 92(12), pp.1601-1610.

Wang, W., Bruyère, C., Duda, M., Dudhia, J., Gill, D., Kavulich, M., Werner, K., Chen, M., Lin, H.C., Michalakes, J., Rizvi, S., Zhang, X., Berner, J., Munoz-Esparza, D., Reen, B., Ha, S. and Fossell, K., 2019. WRF-ARW version 4 modeling system User's guide, National Centre for Atmospheric Research.

Wilson, A.M. and Barros, A.P., 2015. Landform controls on low level moisture convergence and the diurnal cycle of warm season orographic rainfall in the Southern Appalachians. Journal of Hydrology, 531, pp.475-493.

Yang, Q., Dai, Q., Han, D., Chen, Y. and Zhang, S., 2019. Sensitivity analysis of raindrop size distribution parameterizations in WRF rainfall simulation. Atmospheric Research, 228, pp.1-13.

Zheng, Y., Alapaty, K., Herwehe, J.A., Del Genio, A.D. and 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), pp.833-860.

Zhuo, L., Dai, Q., Han, D., Chen, N. and Zhao, B., 2019. Assessment of simulated soil moisture from

WRF Noah, Noah-MP, and CLM land surface schemes for landslide hazard application. Hydrology and Earth System Sciences, 23(10), pp.4199-4218.

Zevenbergen, C., Bhattacharya, B., Wahaab, R.A., Elbarki, W.A.I., Busker, T. and Rodriguez, C.S., 2017. In the aftermath of the October 2015 Alexandria Flood Challenges of an Arab city to deal with extreme rainfall storms. Natural hazards, 86(2), pp.901-917.

Zhang, H., Wu, C., Chen, W. and Huang, G., 2019. Effect of urban expansion on summer rainfall in the Pearl River Delta, South China. Journal of Hydrology, 568, pp.747-757.