



Research papers

Developing spin-up time framework for WRF extreme precipitation simulations

Ying Liu^{a,*}, Lu Zhuo^b, Dawei Han^a

^a School of Civil, Aerospace and Mechanical Engineering, University of Bristol, Bristol BS8 1TR, UK

^b School of Earth and Environmental Sciences, Cardiff University, Cardiff CF10 3AT, UK



ARTICLE INFO

This manuscript was handled by Andras Barossy, Editor-in-Chief, with the assistance of Sheng Yue, Associate Editor.

Keywords:

WRF spin-up time
OSTI framework
ERAS
IMERG

ABSTRACT

Despite the wide application of the Weather Research and Forecasting (WRF) model in extreme precipitation simulations, there is a lack of consensus and clear guidance on identifying the suitable length of spin-up time. In this study, the WRF model was used to simulate the extreme precipitation events that happened on the 4 November 2015 at Alexandria of the Nile Delta. According to the observation from the Integrated Multi-satellite Retrievals for Global Precipitation Measurement (IMERG), 21 spin-up time experiments of the 3-level-nested domain scenario and 21 spin-up time experiments of the 2-level-nested domain scenario were designed to explore the relationship between model required spin-up time and initial conditions. The simulation performances were evaluated by seven verification metrics and one overall performance score. Here we try to provide guidelines on how to determine the optimal spin-up time through satellite data without too many trial-and-error tests. An Optimal Spin-up Time Identifying (OSTI) framework with possible weather situations and spin-up time determining steps is proposed to help future work. It is found that the occurrences of disturbing weather events strengthen the influence of initial conditions on simulation outputs and increase model requirements for spin-up time lengths. Moreover, this framework is more useful for precipitation events with strong synoptic backgrounds because their simulation performances depend largely on the development of appropriate atmospheric circulations and physical equilibrium states in the model.

1. Introduction

Global and regional climate models (GCMs and RCMs) are becoming established tools for providing reliable information for climate predictions. However, the high-resolution simulations of GCMs are very computationally expensive. In contrast, RCMs have shown their ability to effectively downscale global information and accurately reproduce mesoscale and local features over a limited region with affordable computational costs (Rummukainen, 2010; Jacob et al., 2014; Schewe et al., 2019). The mesoscale atmospheric Weather Research and Forecasting (WRF) model is a typical RCM that numerically solves a set of differential equations describing the climate system over a selected domain after required discretization and parameterization processes. To perform dynamical downscaling, WRF needs initial and boundary conditions (IC and BC) provided by global reanalysis datasets to initialise and drive atmospheric variables through their domain boundaries. Compared to the larger gridded meteorological datasets that poorly represent atmospheric patterns, running dynamical downscaling WRF

with high resolution can help resolve the spatial variability of key variables and surface-atmosphere exchanges, because heterogeneous land fluxes and near-land meteorological parameters are better resolved (Talbot et al. 2012).

It is commonly believed that WRF may forget the IC after a certain execution time. To save computational resources and reduce delivery time, the whole simulation period is usually divided into subperiods with spin-up periods followed by continuous executions (Gómez-Navarro et al., 2011; Jerez et al., 2018). The spin-up time is necessary to avoid inhomogeneities when connecting subperiods as well as decrease the masking effects of IC issues on model results. In WRF extreme precipitation simulations, spin-up time allows the model to adjust from the IC to a state that is consistent with its own numerics and physics and to develop appropriate large-scale circulations (Jankov et al., 2007; Skamarock and Klemp, 2008). Therefore, spin-up time is the required execution time that allows the WRF to reach a physical equilibrium state following the path defined by the BC, while forgetting about the IC (Yang et al., 1995; Giorgi and Mearns, 1999; Denis et al., 2002).

* Corresponding author.

E-mail address: emily.liu@bristol.ac.uk (Y. Liu).

Moreover, until reach the equilibrium state the results can be easily tainted by spin-up-induced biases, the model outputs will not be realistic and must be discarded (Cosgrove et al., 2003).

To prevent instabilities in the WRF model, Jankov et al. (2007) found that at least 12 h spin-up time should be used. And this spin-up time is often regarded as a suitable choice directly without enough verification in many WRF studies (Huva et al., 2012; Dzebre et al., 2019; Afshar et al., 2020). Furthermore, the lengths of spin-up time used in some other studies vary widely, from a minimum of 6 h to a maximum of 1 year (Cha and Wang, 2013; Naabil et al., 2017; Zhuo et al., 2019). In fact, the length of spin-up depends on the quality of initial inputs and the soil conditions because the soil moisture content and latent heat flux will influence precipitation processes (Kleczeck et al., 2014). If the IC is very complex and impressive, the model will take more time to forget it and get away from its masking effects. Therefore, the optimal spin-up time is not fixed and should be different for each event with different IC. In addition, the study by Hwang et al. (2019) found that using multiple hydrometeors as input in the BC can reduce spin-up time and accelerate precipitation initialization. Their sensitive experiments were conducted with different numbers of hydrometeors including specific humidity, specific cloud liquid water, specific cloud ice water, specific rain water and specific snow water. The results show that this method has a good effect on short-range precipitation simulation. However, most weather prediction studies ignore the effect of initial weather conditions and directly use 12 h spin-up time because conducting too much trial-and-error testing before the formal simulation is computationally expensive.

In fact, it is very difficult to determine the WRF optimal spin-up time for different events since it depends on many factors. First of all, the climate system has a large number of components with various response times. For example, soil moisture responds much slower than atmospheric variables to dynamical and thermo-dynamical processes (Skamarock, 2004; De Elia et al., 2002; Doblas-Reyes et al., 2013). The second factor is the size of the simulated domains, since the BC effect decays as we move away from the boundaries (Leduc and Laprise, 2009). Third, WRF has many physical schemes that can be used. The choice of physical scheme often results in differences in the internal variability of simulations (Awan et al., 2011; Evans et al., 2012; Ji et al., 2014). Fourth, the IC of WRF is obtained by spatial interpolating the global datasets to the selected regions. The greater the inconsistencies between IC and model physics, the longer the spin-up time required (Turco et al., 2013). The last factor is the model spin-up would be affected by extreme conditions like extremely hot or cold temperatures, extremely dry air and so on (Day et al., 2014; Seck et al., 2015).

In this study, to solve the lack of consensus on the choice of WRF spin-up time, 42 sensitivity experiments with different IC are evaluated to explore the effects of disturbing weather on required spin-up times. Disturbing weather refers to the other weather event that occurs within the spin-up window, which causes large-scale changes in the synoptic background at the initial time. Moreover, an Optimal Spin-up Time Identifying (OSTI) framework is proposed to provide guidelines for future simulation work. The purpose of this study is to help other researchers to avoid unfavourable IC and determine suitable spin-up times for their WRF precipitation simulations. This study aims to address the following two key research questions:

- Does the occurrence of disturbing weather events weaken or strengthen the influence of initial conditions on simulation outputs?
- Why does WRF's optimal spin-up time vary by event and situation?
- How can a structured selecting framework help to determine suitable WRF spin-up time more efficiently?

The significance of this study is to highlight the importance of spin-up time and show a guiding framework that can determine spin-up time more efficiently before simulation without the need for too much trial-and-error testing. This study is organized as follows: a brief description of the study area and event are provided in Section 2. The datasets,

experimental design, verification metrics, framework planning are illustrated in Section 3. The results and discussions of different experiments are shown in Section 4. The detailed OSTI framework and framework application is explained in Section 5. Finally, the summary and conclusions of this study are presented in Section 6.

2. Study area and event

Alexandria in the Nile delta was selected as the study area. Due to the influence of the Mediterranean climate, this area has long dry summers and short mild winters. In recent decades, this area becomes a vulnerable zone that faces growing pluvial flooding hazards. However, the inadequate coverage of radars and rain gauges makes the development of a hydrometeorological early warning system in Egypt very difficult. WRF model is an effective way to simulate extreme precipitation by downscaling global NWP products to interesting areas, which is very suitable and feasible for countries like Egypt. Our previous study has investigated the sensitivities of various model configurations and obtained some useful findings (Liu et al., 2021), but the hypothesis that regional weather conditions may impact the time required for model initialization has not been fully confirmed. Therefore, further exploration of the relationship between WRF spin-up time and weather conditions through the extreme event in this area would be meaningful for Egypt and other similar regions.

The study event happened on the 4 November 2015, which was a 50-year storm that caused 60% of the city area to be flooded as well as the stagnant water in some low-lying areas remained for more than 15 days. Some places even recorded more than 200 mm of precipitation in 2 h (Zevenbergen et al., 2017). The main extreme precipitation of this event lasted for about 18 h from 06:00 to 24:00 UTC. The centred latitude and longitude of this event are at 31.5°N and 30°E. This extreme precipitation event caused a devastating flood that has been reported 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). The terrain of the study area and the location of the surrounding Red Sea and the Mediterranean Sea are shown in Fig. 1 (b, c). The detailed descriptions of the nested domain and study event locations are presented in Section 3.2. Furthermore, the proposed framework is also applied and verified through three other extreme precipitation events that happened around Hurghada, Egypt (26 and 27 October 2016), Antalya, Turkey (28 December 2013) and Beirut, Lebanon (25 October 2015). These precipitation events have different intensities, distributions and synoptic backgrounds, which is very helpful to explore the generalization ability of the OSTI framework.

3. Materials and methods

3.1. Datasets

3.1.1. Initial and lateral boundary dataset

The European Centre for Medium-Range Weather Forecasts (ECMWF) Global Climate Reanalysis v5 (ERA5) was used to provide the initial and lateral boundary meteorological conditions in the WRF simulations. ERA5 is a newly developed dataset since early 2016 and has replaced the old reanalysis dataset ERA-Interim. This dataset spans the period from 1 January 1950 onwards. ERA5 offers over 240 parameters related to the atmosphere, land, and ocean climate, etc on the surface and vertical levels, which are much more than 100 parameters in ERA-Interim. Furthermore, the spatiotemporal resolution of the ERA5 dataset (hourly 31 km grid spacing analysis fields) is significantly better than the ERA-Interim dataset (6-hourly 80 km grid spacing analysis fields) (Hersbach et al., 2020). Along with the development of this new dataset, there have been some studies using ERA5 as a lateral boundary condition for WRF simulations in Egypt and the surrounding area of the Mediterranean Sea (Duzenli et al., 2021; Liu et al., 2021; Papavasileiou et al., 2022; Ludwig and Hochman, 2022;). The detailed introduction of

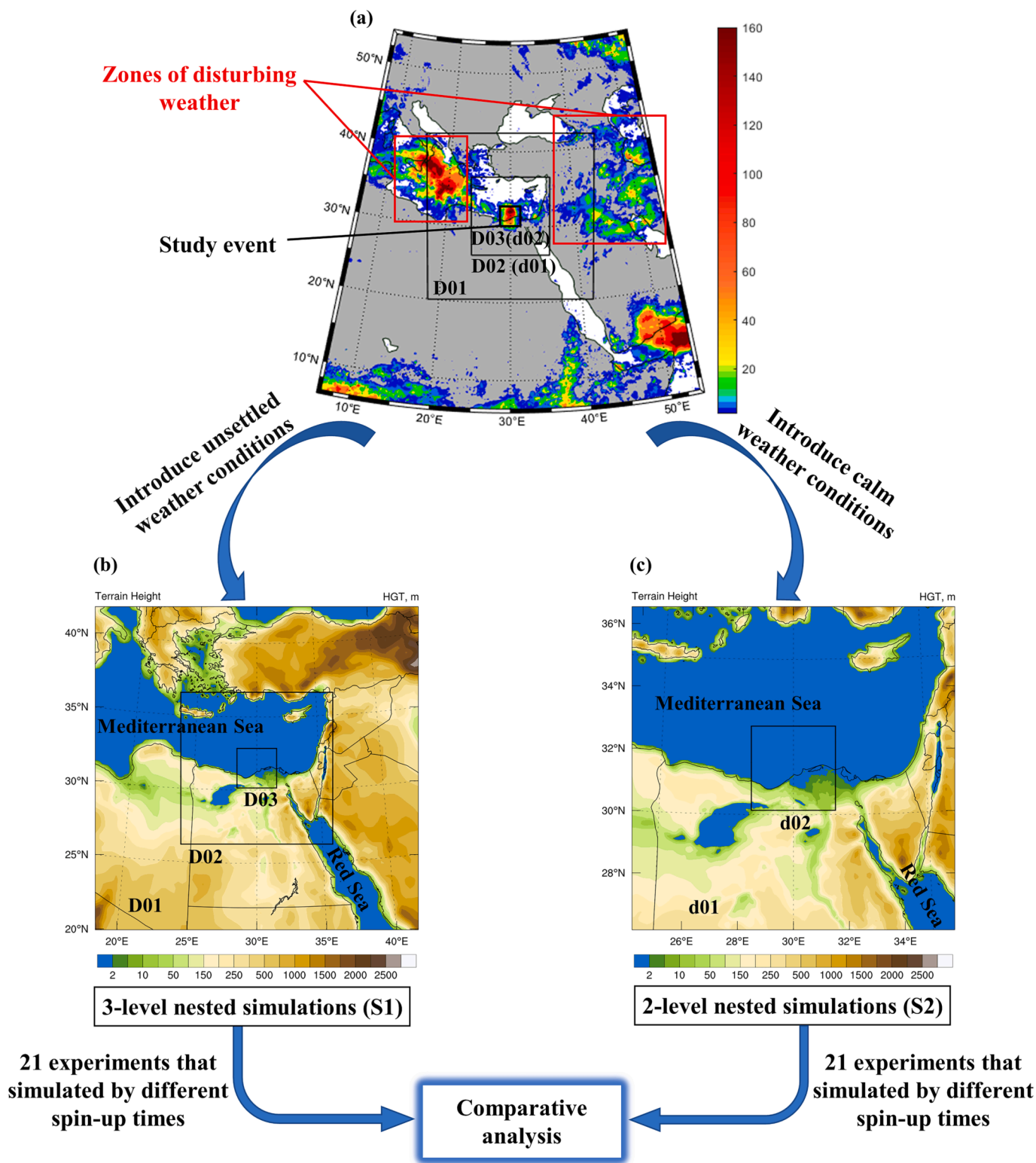


Fig. 1. The flow chart of experimental scenario establishment. (a) Accumulated precipitation (mm) around the study area from 06:00 on 30th Oct to 00:00 on 5th Nov (138 h in total, including the 120 h spin-up period and the 18 h study event duration). (b) The 3-level nested domains used in Scenario 1, which contains the innermost domain (D03), middle domain (D02) and outermost domain (D01). (c) The 2-level nested domains used in Scenario 2, which only contains the innermost domain (d02) and outermost domain (d01). The size and position of d01 and d02 are the same as D02's and D03's, respectively.

the ERA5 reanalysis dataset and its parameters can be found in the ECMWF documentation (<https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation>). The ERA5 reanalysis dataset can be downloaded on the ECMWF website (<https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>).

3.1.2. Observational data

The Integrated Multi-satellite Retrievals for Global Precipitation Measurement (IMERG) version 06B was used for simulation verification. It combines precipitation estimates from geosynchronous infrared observations from geo-IR satellites, all passive microwave sensors of the

Global Precipitation Measurement (GPM) constellation and ground-based measurements from precipitation gauges (Huffman et al., 2015). IMERG provides the quasi-global precipitation estimates with $0.1^\circ \times 0.1^\circ$ spatial resolution and 30-min temporal resolution. According to the differences in latency and accuracy, IMERG has three sequences called Early Run (near real-time low-latency gridded global multi-satellite precipitation estimates, 4 h latency), Late Run (near real-time gridded global multi-satellite precipitation estimates with quasi-Lagrangian time interpolation, 12 h latency), and Final Run (research-quality gridded global multi-satellite precipitation estimates with quasi-Lagrangian time interpolation, gauge data, and climatological adjustment, 3.5 months latency). Particularly, the IMERG Final Run product has a month-to-month climatological correction through the Global Precipitation Climatology Centre (GPCC) precipitation gauge analysis (Hou et al., 2014, Huffman et al., 2015). Since IMERG product's release in early 2015, a substantial number of studies have used and recommended it for various applications, such as analysis of extreme precipitation events (Huang et al., 2019), streamflow simulation (Tang et al., 2016) and flood forecasting (Wang et al., 2017). And a large number of studies have evaluated the performance of IMERG precipitation products at various temporal and spatial scales (e.g., Wang et al., 2023; Navarro et al., 2019, Palomino-Ángel et al., 2019; Prakash et al., 2018; Manz et al., 2017). Considering the insufficient coverage of surface observations (radar and rain gauges) in Egypt, IMERG is a useful alternative dataset to be adopted in this study due to its fine spatial and temporal resolutions, which favour simulation evaluations, and its stable performance, as evidenced in previous studies. In this study, the IMERG Final Run product is used because it generally has advantages over the uncalibrated Early Run and Late Run products. IMERG products can be downloaded on the GPM website (<https://gpm.nasa.gov/data/directory>) and detailed introductions can be found in Huffman et al. (2015).

3.2. Experimental design

The precipitation simulation experiments of this study were conducted by the WRF-ARW version 4.0 (Skamarock et al., 2019). The model configuration settings are listed in Table 1. First, all experiments

Table 1
Summary of physical parameterisations and other configurations used in 3-level and 2-level nested simulations.

Mode configurations	3-level nested scenario (S1)	2-level nested scenario (S2)
Nesting ratio	1:3:3, D01 31.5 km; D02 (d01) 10.5 km; D03 (d02) 3.5 km	
Vertical levels	58 levels	
Microphysics	WRF Single-Moment 6-class scheme (Hong and Lim, 2006)	
Planetary boundary layer	Mellor-Yamada-Janjic scheme (Janjić, 1994)	
Cumulus parameterization	Kain-Fritsch scheme (Kain, 2004)	
Longwave radiation	RRTM scheme (Mlawer et al. 1997)	
Shortwave radiation	RRTM scheme (Mlawer et al. 1997)	
Land surface	Noah land-surface model (Chen and Dudhia; Ek et al. 2003)	
Hydrometeor species from ERA5 input data	Specific humidity, Specific cloud liquid water, Specific cloud ice water, Specific rain water and Specific snow water	
Domain size and levels	3-level nested: D01 80x80 (about 6.19 million km ²), D02 112x112 (about 1.36 million km ²), D03 88x88 (about 0.09 million km ²)	2-level nested: d01 112x112 (about 1.36 million km ²), d02 88x88 (about 0.09 million km ²)
Spin-up time	0–120 h, per 6 hCase 1-Case 21 (C1-C21)	0–120 h, per 6 hcase 1-case 21 (c1-c21)

were simulated using a two-way interactive nested configuration with the 1:3:3 downscaling ratio and 58 vertical levels. The physical parameterization schemes used in all model domains and experiments are the WRF Single-Moment 6-class microphysics scheme (Hong and Lim, 2006), the Mellor-Yamada-Janjic planetary boundary layer scheme (Janjić, 1994), the rapid radiative transfer scheme (RRTM, Mlawer et al. 1997) and the unified Noah land-surface model (Chen and Dudhia, 2001; Ek et al., 2003). All the settings mentioned above are the optimal configurations for this simulation region summarized by Liu et al. (2021). Next, refer to the study by Hwang et al. (2019), all experiments used ERA5 data with additional hydrometeors (all five available hydrometeors) as input conditions to help observe the changes in the model required spin-up times. Therefore, the Kain-Fritsch cumulus scheme (Kain, 2004) with more moisture tendencies (Qc, Qr, Qi and Qs) was used in this study instead of the Grell-Freitas cumulus scheme (Grell and Freitas, 2014) suggested by Liu et al. (2021). The time step used for the lateral boundary condition file is one hour. The history output files of each domain are logged hourly.

To evaluate the effects of synoptic conditions, all 42 experiments were divided into two scenarios: the 3-level nested scenario (S1, Case 1-Case 21) and the 2-level nested scenario (S2, case 1-case 21). All nested domains are centred on the same latitude and longitude (31.5°N, 30°E) and employ Lambert conformal projection (Fig. 1 (b–c)). Fig. 1 (a) shows the IMERG observed accumulated precipitation around the study area from 06:00 on 30th Oct to 00:00 on 5th Nov, which occurs within the 120 h spin-up period and the 18 h event duration. In this study, we refer to the precipitation other than the study event as “disturbing precipitation” for this simulation in order to differentiate. Indicated within the red box are the disturbing precipitations that occurred around the domain boundary during the model spin-up time, while the study event is the central precipitation. Therefore, S1 and S2 were designed as comparison simulations that introduced the unsettled and calm weather conditions, respectively. Taking Case 1 (C1) of S1 as an example, it is a 3-level nested simulation with horizontal grid sizes of 31.5 km, 10 km, and 3.5 km for the outermost domain (D01), middle domain (D02) and innermost domain (D03), respectively (Fig. 1 (b)). The D03 covers the study area of Alexandria and the adjacent areas. Besides, the grid points and domain sizes for D01, D02 and D03 are 80 × 80 (about 6.19 million km²), 112 × 112 (about 1.36 million km²) and 88 × 88 (about 0.09 million km²) respectively. In comparison with C1, case 1 (c1) of S2 is a 2-level nested simulation that only has the outermost domain (d01) and innermost domain (d02) with horizontal grid sizes of 10 km and 3.5 km, respectively (Fig. 1 (c)). The horizontal sizes, grid points and domain sizes for d01, d02 are the same as D02 and D03. The study event happened in the centre of D03/d02. Hence, the only difference between S1 experiments and S2 experiments is the presence or absence of the largest domain (D01). According to the IMERG observations, it is assumed that the D01 of S1 experiments have more complex synoptic conditions than the d01 of S2 experiments. After determining the domain and physical configurations, Case 1-Case 21 (C1-C21) of S1 and case 1-case 21 (c1-c21) of S2 were simulated with different spin-up times. Their spin-up times increase from 0 to 120 h with a 6-hour time increment, which makes these experiments introduce different IC at the beginning times. By conducting the above comparative experiments, the relationships between synoptic conditions as well as lengths of spin-up times and final target simulation performances would be drawn.

3.3. Verification method

To show the spatiotemporal performance changes of these experiments, this study uses seven error metrics and one overall performance score to evaluate WRF simulations with respect to the IMERG observations. Seven error metrics include the Probability of Detection (POD), the False Alarm Ratio (FAR), the Critical Success Index (CSI), the Frequency Bias Index (FBI), the Root Mean Square Error (RMSE), the Mean Bias Error (MBE) and the Standard Deviation (SD). The first four metrics

are spatial metrics, they represent the probability of detecting precipitation and generating false precipitation, the critical performance, and the tendency to overestimate or underestimate precipitation, respectively. The calculation of these metrics is based on the rainfall contingency table, as shown in Table 2. RR, NR, RN and NN represent the grid numbers of hits, misses, false alarms and correct negatives, respectively. Then the metric values can be obtained through Equations (1)-(4). In these equations, i and N refer to each time step and the total time step of the simulation run. Considering different thresholds may help to further investigate the simulation accuracy of extreme precipitation, the precipitation above 0.1 mm is used to calculate *POD*, *FBI*, *CSI* and *FAR* in this study. The other three metrics are temporal metrics that indicate the average magnitude of error between simulations and observations, the average bias of cumulative error, and the magnitude of random error, respectively. The ranges and the ideal value of these error metrics are shown in Table 3. A detailed description of the seven metrics can be found in Liu et al. (2012). In this study, they are calculated by interpolating WRF simulations to the IMERG Final Run observation grid at a 3-hour time step in D03/d02.

$$POD = \frac{1}{N} \sum_{i=1}^N \frac{RR_i}{RR_i + NR_i} \tag{1}$$

$$FBI = \frac{1}{N} \sum_{i=1}^N \frac{RR_i + RN_i}{RR_i + NR_i} \tag{2}$$

$$FAR = \frac{1}{N} \sum_{i=1}^N \frac{RN_i}{RR_i + RN_i} \tag{3}$$

$$CSI = \frac{1}{N} \sum_{i=1}^N \frac{RR_i}{RR_i + RN_i + NR_i} \tag{4}$$

Since every metric show different performance characteristics in spatial or temporal dimensions, it is difficult to identify how the overall simulation performance changes with spin-up times. The Relative Closeness Value of Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS RCV) proposed by Hwang and Yoon (1981) and further developed by Liu et al. (2021) is applied to evaluate the overall performance of each experiment in this study. In order to have a uniform score with a value of 0 representing the worst and 1 representing the best, all seven metric values are scaled to a range from 0 to 1 and assigned equal weights to calculate the TOPSIS RCV. So the possible range of TOPSIS RCV is 0–1 and the perfect value is 1. In this way, TOPSIS RCV could easily show the total performance of WRF simulations under different weather conditions. The detailed calculation method for uniform score could be found in Liu et al. (2021).

3.4. Framework planning

The aim of this study is to take the shape of a WRF optimal spin-up time identifying framework. To achieve this goal, the following work was conducted. First, the proposed framework is inspired by the above-mentioned different nested-level simulations over Alexandria, Egypt. These 42 extreme precipitation simulations revealed the relationship between regional model IC and optimal spin-up time. In Section 4.1, the impact of unsettled weather on the study event and the thermodynamic and dynamic processes during different spin-up times are discussed. Then, seven spatiotemporal performance metrics and the overall

Table 2
Contingency table of the WRF simulation against observation.

Simulation/Observations	Rain	No Rain
Rain	RR (hits)	RN (false alarms)
No Rain	NR (misses)	NN (correct negatives)

Table 3
The ranges and ideal values of seven verification metrics.

Verification metrics	Ranges	Ideal values
Probability Of Detection (<i>POD</i>)	0–1	1
False Alarm Ratio (<i>FAR</i>)	0–1	0
Critical Success Index (<i>CSI</i>)	0–1	1
Frequency Bias Index (<i>FBI</i>)	0–∞	1
Root Mean Square Error (<i>RMSE</i>)	0–∞	0
Mean Bias Error (<i>MBE</i>)	–∞–∞	0
Standard Deviation (<i>SD</i>)	0–∞	0

performance score TOPSIS RCV were used to further evaluate the performance of precipitation simulations and determine the optimal spin-up time. By analysing the variation of these metrics, the concept of three standard periods (critical period, minimum period and adequate period) of model spin-up time is introduced in Section 4.2. Based on the knowledge of unsettled weather influences and three standard periods, a simple optimal spin-up time identifying framework that considers the different situations the WRF simulation set-up will meet and the corresponding solutions could use is designed in Section 5.1. The possible weather situations were classified regarding the time between WRF target precipitation and disturbing weather. At the same time, according to the critical, minimum and adequate periods of the WRF model in the study area, determine the recommended, unrecommended, unnecessary and cautious model spin-up periods. Finally, in Section 5.2, three additional precipitation events around the Mediterranean Sea were used to test the proposed framework and show how to apply this framework. These test events have different precipitation distributions, precipitation intensities and synoptic backgrounds, which is very helpful in understanding the generalization ability of this framework. However, further research is needed to refine the details of the framework, particularly the spin-up critical period, minimum period, and adequate period, which may vary in different regions with different geographical features. A comprehensive global list of these standard periods would be very useful for the OSTI framework to be generalised and applied to other events around the world.

4. Results and discussions

4.1. Results of simulated precipitation patterns under different weather conditions

The accumulated precipitation during the study event (18 h) of the 3-level nested simulations scenario and the 2-level nested simulations scenario are shown in Fig. 2 and Fig. 3. The first subfigure (i.e., Fig. 2 (a) and Fig. 3 (a)) present the IMERG observation as the validation for WRF experiments. The other eleven subfigures (i.e., Fig. 2 (b-l) and Fig. 3 (b-l)) display the WRF results of simulations run with spin-up times of 0 h (C1 and c1), 12 h (C3 and c3), 24 h (C5 and c5), 36 h (C7 and c7), 48 h (C9 and c9), 60 h (C11 and c11), 72 h (C13 and c13), 84 h (C15 and c15), 96 h (C17 and c17), 108 h (C19 and c19) and 120 h (C21 and c21), respectively. They are plotted in the D02 domain of S1 and d01 domain of S2, which have the same size and for best visual comparisons. Next, the disturbing precipitations (precipitation rates estimated by IMERG) around the study area at the start times of WRF experiments are illustrated in Fig. 4. Fig. 4 (a) is the total precipitation from 06:00 on 30th Oct to 00:00 on 5th Nov, which accumulated in the maximum 120 h spin-up period and 18 h study event duration. Fig. 4 (b-l) correspond to the disturbing precipitation rate at different time points happened in WRF experiments of S1 and S2. These different onset time points also indicate that different lengths of spin-up time were used in the WRF experiments. For example, WRF simulations (C21 and c21) that run with 120 h spin-up time spin up the model from 06:00 on Oct 30 to 06:00 on Nov 4 and then simulated the 18 h study event from 06:00 on Nov 4 to 00:00 on Nov 5. Finally, the synoptic weather patterns in the D01 domains at different start times of WRF experiments are displayed in Fig. 5.

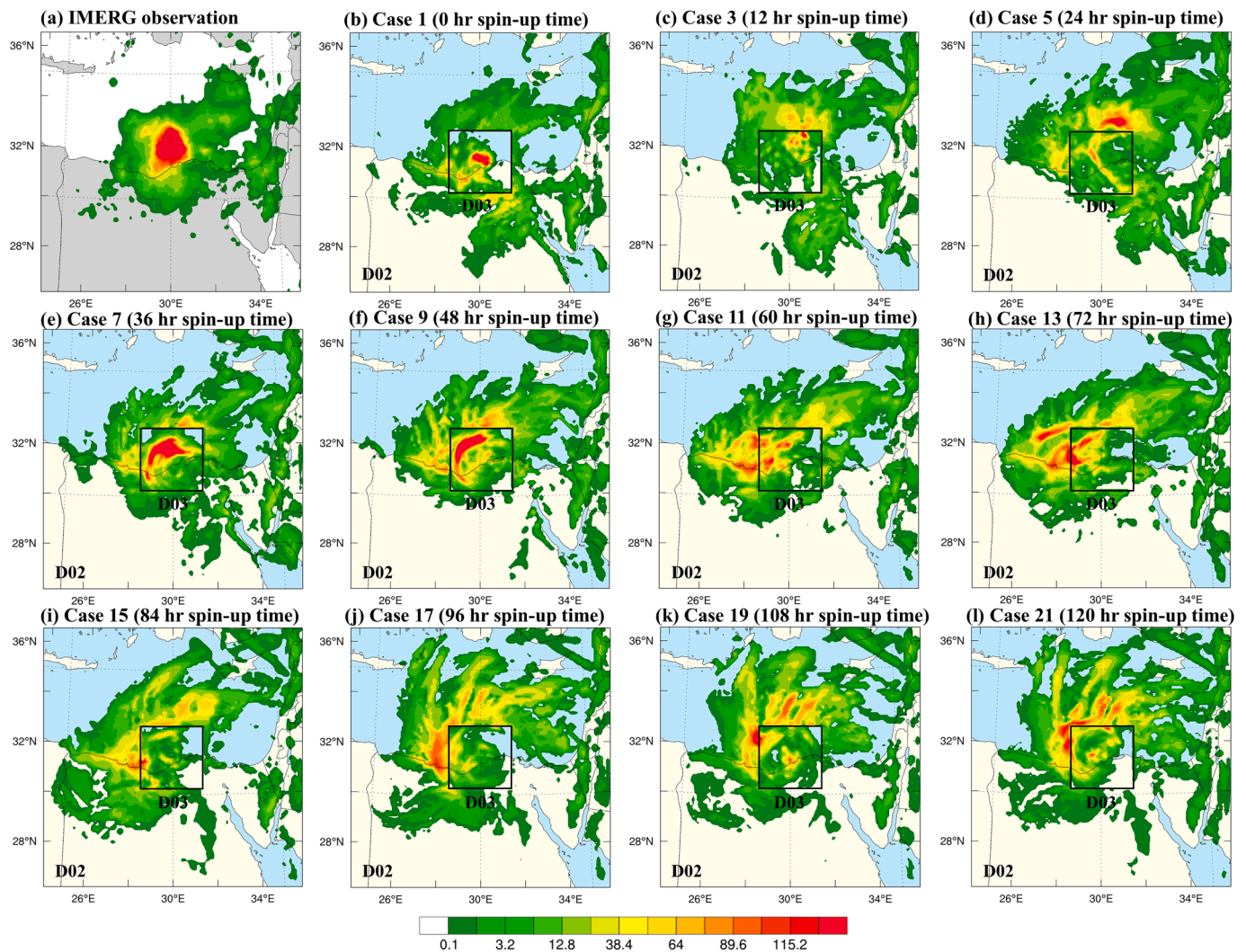


Fig. 2. Comparison of accumulated precipitation for study event (18 h) in the D02 domains of 3-level nested simulations with different spin-up times. (Scenario 1). (a) The IMERG observed precipitation. (b–l) WRF simulation results with spin-up times of 0 h (C1), 12 h (C3), 24 h (C5), 36 h (C7), 48 h (C9), 60 h (C11), 72 h (C13), 84 h (C15), 96 h (C17), 108 h (C19) and 120 h (C21), respectively.

D01 domain is the largest domain that contains the area of D02 (d01) and D03 (d02). For better viewing, domain boundaries are not drawn in Fig. 5. Fig. 5 (a) present surface temperature and sea-level pressure averaged over the same 138 h as in Fig. 4 (a). Fig. 5 (b–l) show geopotential heights, wind speeds and relative humidity at 500-hPa, which also correspond to the different IC at different time points used in WRF experiments of S1 and S2. Fig. 4 and Fig. 5 will help to understand the weather conditions around the domain boundaries for the entire study period.

As shown in Fig. 2 (b–f), five experiments C1, C3, C5, C7 and C9 of S1 show the WRF simulation improves significantly with the increases in spin-up time and gradually stabilizes around the 36 h and 48 h. It is because the suitable spin-up times can help the model to adjust its own numerics and develop appropriate circulations (Jankov et al., 2007; Skamarock and Klemp, 2008). However, the simulation performances of C11, C13, C15, C17, C19 and C21 (Fig. 2 (g–l)) are decreased instead when the spin-up time is extended over 48 h. In particular Fig. 2(i–l), the simulated precipitations become very unreasonably scattered toward the domain boundary instead of accumulating within D03. The potential reason for it could be that the model struggled to solve the inconsistency in its numerics and physics under unsettled weather conditions. In contrast, the improvements between the nine experiments c1, c3, c5, c7, c9, c11, c13, c15, c17, c19 and c21 of S2 are better and in line with our

common understanding of model spin-up time. That is model performance grows with the increases in spin-up time (Fig. 3 (b–f)) and precipitation patterns become stable after reaching a certain level (Fig. 3 (g–l)). Even though the 11 experiments of S1 and 11 experiments of S2 are simulated under the same 18-hour study event precipitation, their performance varies widely.

When considering the impacts from disturbing precipitation and synoptic backgrounds at different start times displayed in Fig. 4 and Fig. 5, the differences in simulation outputs between S1 and S2 experiments seem to be explained. Thermodynamically, the study event was driven by the convergence of moist air from the Mediterranean Sea and colder air from the surrounding land. The collision of these air masses created a zone of instability in the atmosphere, which allowed for the development of the intense storm. This storm was intensified by the upward vertical motion of air and the release of heat and energy through condensation, which finally led to heavy precipitation and a flash flood in Alexandria, Egypt. As shown in Fig. 5 (a), the Mediterranean Sea surface is very warm and averages 22 °C during the whole spin-up period while the land north of the sea averages only 10 °C and the land south of the sea averages 16 °C. These large temperature gaps between ocean and land promoted the formation of extreme precipitation in both S1 and S2 simulations. But due to the occurrence of disturbing precipitation and different domain sizes of S1 and S2 scenarios, the atmospheric

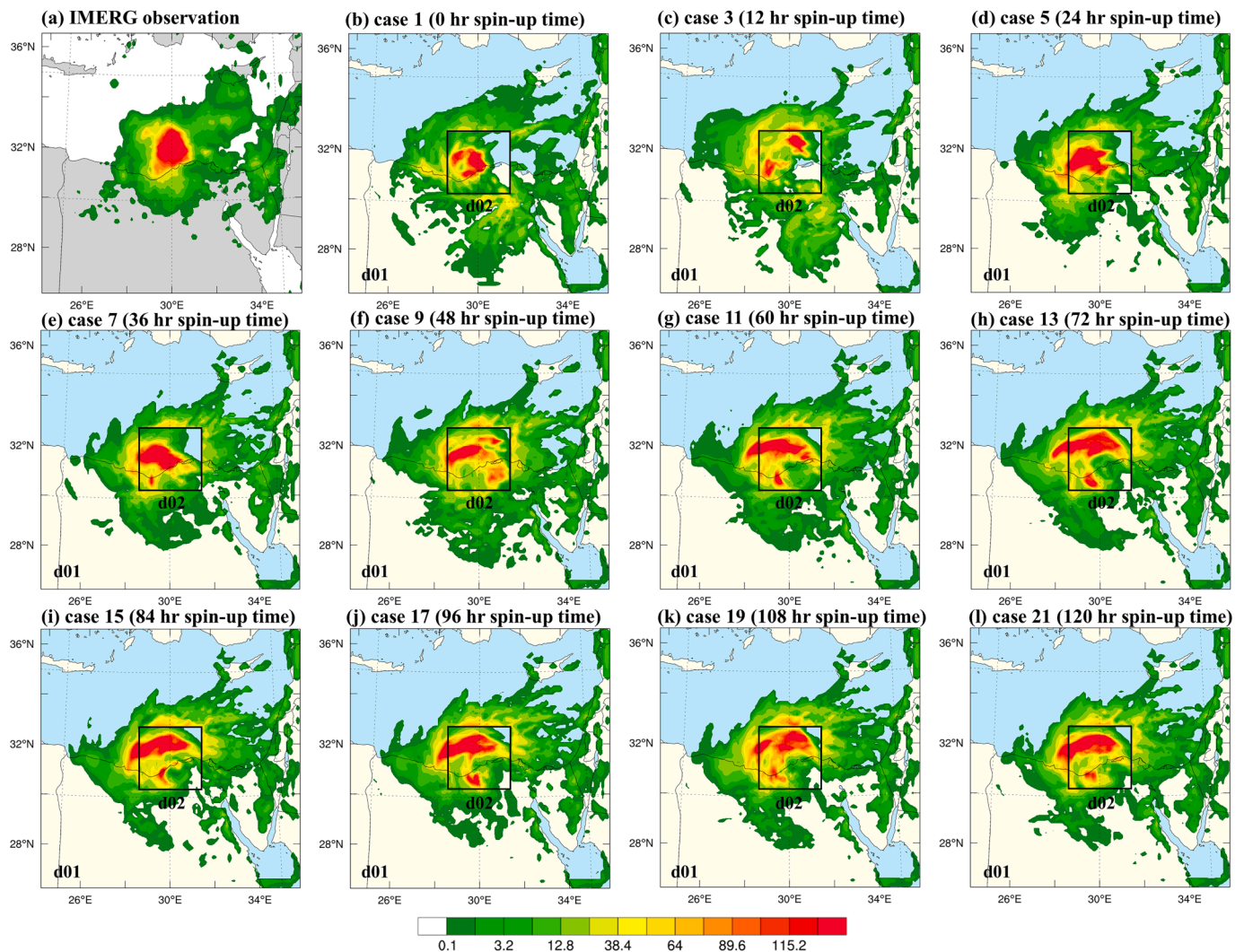


Fig. 3. Comparison of accumulated precipitation for study event (18 h) in the d01 domains of 2-level nested simulations with different spin-up times (Scenario 2). (a) The IMERG observed precipitation. (b–l) WRF simulation results with spin-up times of 0 h (c1), 12 h (c3), 24 h (c5), 36 h (c7), 48 h (c9), 60 h (c11), 72 h (c13), 84 h (c15), 96 h (c17), 108 h (c19) and 120 h (c21), respectively.

humidities in the initial fields of 42 experiments are very different. For S1 experiments from 06:00 on Oct 30 to 06:00 on Nov 2 (Fig. 4 (f–l)), there is disturbing precipitation occurring in the area between the boundaries of D01 and D02 (d01) during the spin-up period and its intensity is very high (up to 25 mm/h). According to the study of Rago et al. (2017) and the news from FloodList (2015), the disturbing precipitation severely affected southern Italy regions with more than 100 mm of precipitation in 24 h. Therefore, the impact and intensity of the Italy disturbing precipitation event are no less than the Alexandria study event. With the dissipation of this disturbing precipitation, the relative humidity at the height of 500 hPa gradually decreases and the remaining moist air moves to the centre of the domains (Fig. 5 (f–l)). However, the areas where the disturbing precipitation was formed and the air humidity fluctuations it caused around the domain boundary were not included in the S2 experiment. Dynamically, the Alexandria event was driven by the low-pressure system over the eastern Mediterranean Sea. The low-pressure system caused strong winds to develop in the lower atmosphere, which helped to transport the moist air into Alexandria and further lead to strong atmospheric instability and thunderstorms prevalent. Fig. 5 (b–d) displays an obvious low-pressure system starting 24 h before the study event, which is characterized by a counterclockwise flow of air around a central region of low pressure. It is also accompanied by cloudiness, strong winds and heavy precipitation, which are

often referred to as cyclones. But when the spin-up time is longer (Fig. 5 (f–l)), the disturbing precipitation also formed a small cyclone over the left domain boundary D01 of S1 experiments. Overall, the research event was triggered by a low-pressure system, then began to precipitate at 06:00 on Nov 4 and developed to the strongest 4 h later. It has a relatively strong synoptic background and shows large-scale atmospheric circulation patterns (supporting files 1 and 2 from NASA worldview, <https://worldview.earthdata.nasa.gov/>). The reason for the difference in the performance of the S1 and S2 experiments is that the model spends a lot of effort on forgetting complex IC that does not contribute much to the formation of the target event, such as the high humidity and cyclone near the domain boundary caused by the disturbing event. This not only delays the model to following the path indicated by BC but also allows masking effects in the model outputs due to IC issues. Besides, suppose the spin-up time is unsuitable. In that case, the model can not fully reach its physical equilibrium state and the model outputs for the target event also could be adversely affected.

As described in the experimental design of Section 3.2, the S2 experiments (c1–c21) remove the outermost domain of the S1 experiments (C1–C21). The location of the middle domain (D02 and d01) is also designed referencing the satellite-observed precipitation distributions, to avoid introducing complex weather conditions into S2 experiments (Fig. 4 (a)). This design aims to figure out whether the occurrence of

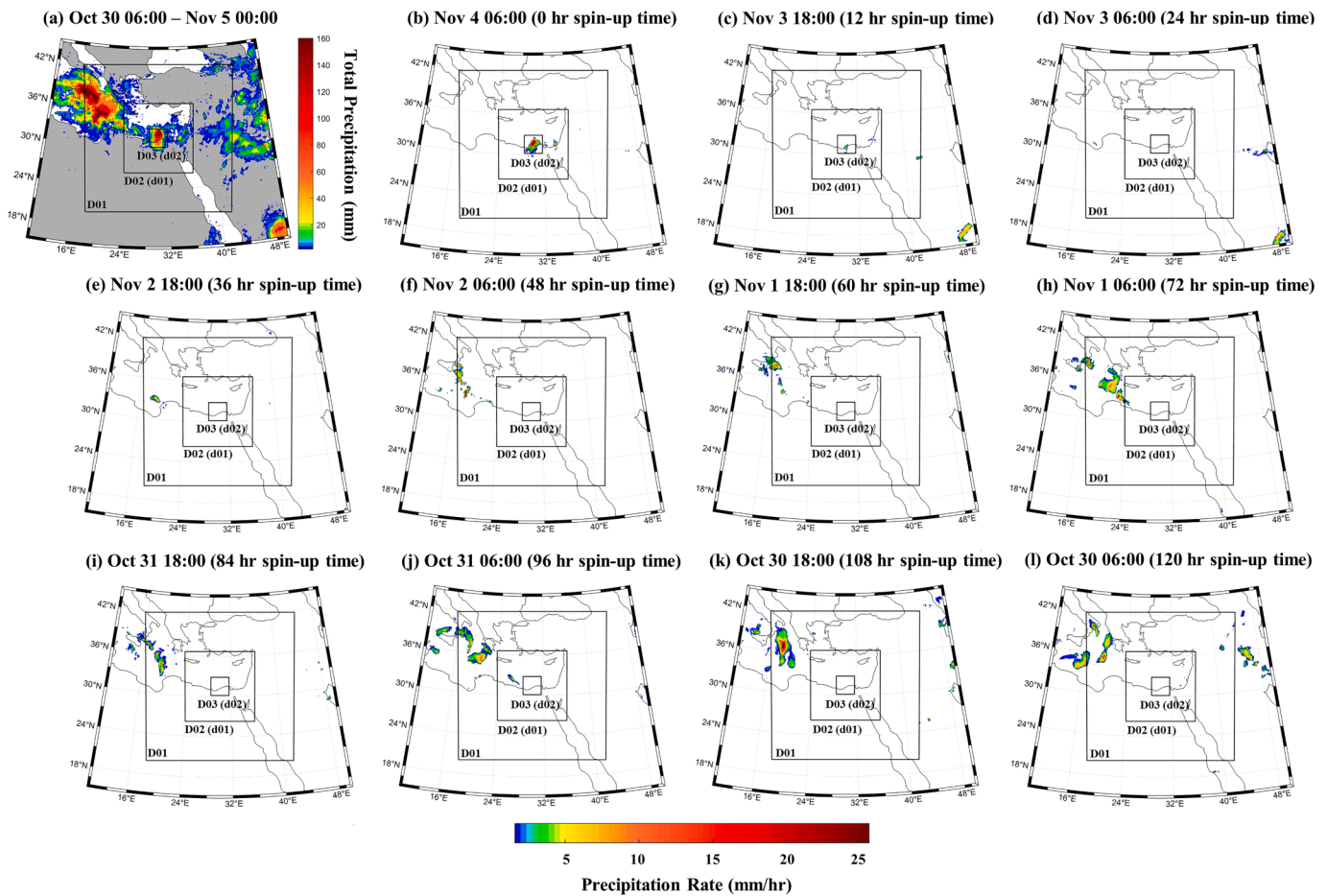


Fig. 4. Comparison of disturbing precipitations (based on IMERG observation data) around the nested domains. (a) shows total precipitation over the 120 h spin-up period and the 18 h study event duration. (b-l) show the precipitation rate at the different start points before the study event.

disturbing weather events weakens or strengthens the influence of IC on simulation outputs and explores a method that can determine the suitable spin-up time in advance. The results of the S1 and S2 experiments demonstrate it is feasible to determine the optimal model spin-up start points based on observation data. For example, when WRF users check observations and find there is disturbing precipitation that happened between 120 h and 48 h before the study event (Fig. 4 (f-l)), they could choose to start simulations within 48 h (Fig. 4 (b-e)) and get better performance. Alternatively, they can directly avoid unsettled weather conditions by controlling the size and location of the simulation domains as demonstrated in the S2 experiments. In addition, the large performance difference between S1 and S2 experiments illustrates that it is not always the longer spin-up time the better simulations. The lengths of the model required spin-up time should be determined by the BC and IC during the spin-up period. And the effect of spin-up time and weather conditions on the extreme precipitation simulation performance is non-negligible. Thus, it is very important to have clear guidance or framework to help more WRF users identify optimal spin-up time systemically.

4.2. Results of verification metrics and overall performance score

Apart from comparing precipitation patterns, the changes in seven verification metrics and overall performance score for 42 experiments are displayed in Fig. 6 (a-h). These experiments are sorted by the length of spin-up time and present the performance of individual metrics over the whole study event duration. As shown in Fig. 6 (a-c), three spatial metrics *POD*, *FBI*, and *CSI* of S1 experiments show similar variations at different spin-up times. These metrics all reach the highest at 36 to 48 h and drop significantly after that. But these three metrics exhibit

fluctuations, which implies simulation performances are not very stable. In contrast, the *POD*, *FBI*, and *CSI* of S2 experiments gradually increase to the top around 48 h and maintain a nice performance in the remaining. Furthermore, the maximum value of these three spatial metrics for the S2 experiments is about 0.1 higher than that for the S1. Another spatial metric *FAR* shows less sensitivity to spin-up time than other metrics. As illustrated in Fig. 6 (d), the *FAR* of S2 experiments remained at a good level of around 0.1, while the *FAR* of S1 experiments has a small raise from 60 h to 120 h. On the other hand, the temporal metrics *RMSE* and *MBE* of S1 experiments both achieve the best values (approximately 4.5 and -2) closest to the ideal value of 0 near the 48 h but become worse after that (Fig. 6 (e, f)). In contrast to S1, the *RMSE* and *MBE* of S2 experiments reach good values (approximately 3 and 0) around 24 h and keep the performance thereafter. As for *SD*, it has similar evident improvements in the first 48 h for both S1 and S2 experiments (Fig. 6 (g)). But S1 experiments still exist small fluctuations from 3 to 2 after 48 h, unlike S2 which always remains at around 2. Finally, the overall performance score *TOPSIS RCV* trend is generally consistent with the trends of the seven metrics summarized above. The uniform score method is also convenient to evaluate the impacts of spin-up time and weather conditions on simulation performances. As shown in Fig. 6 (h), the overall performance of S2 experiments experiences a large increase in the first 48 h and maintains at 0.727 while the overall performance of S1 is less stable which can reach the highest score of about 0.638 and the lowest of about 0.375.

Linking the seven metric results with *TOPSIS RCV*, it can be found that 12 h, 24 h and 48 h are three important time points for precipitation simulation with good IC and BC (S2 experiments) in this study area. Because these three points correspond to the rapid adjustment period

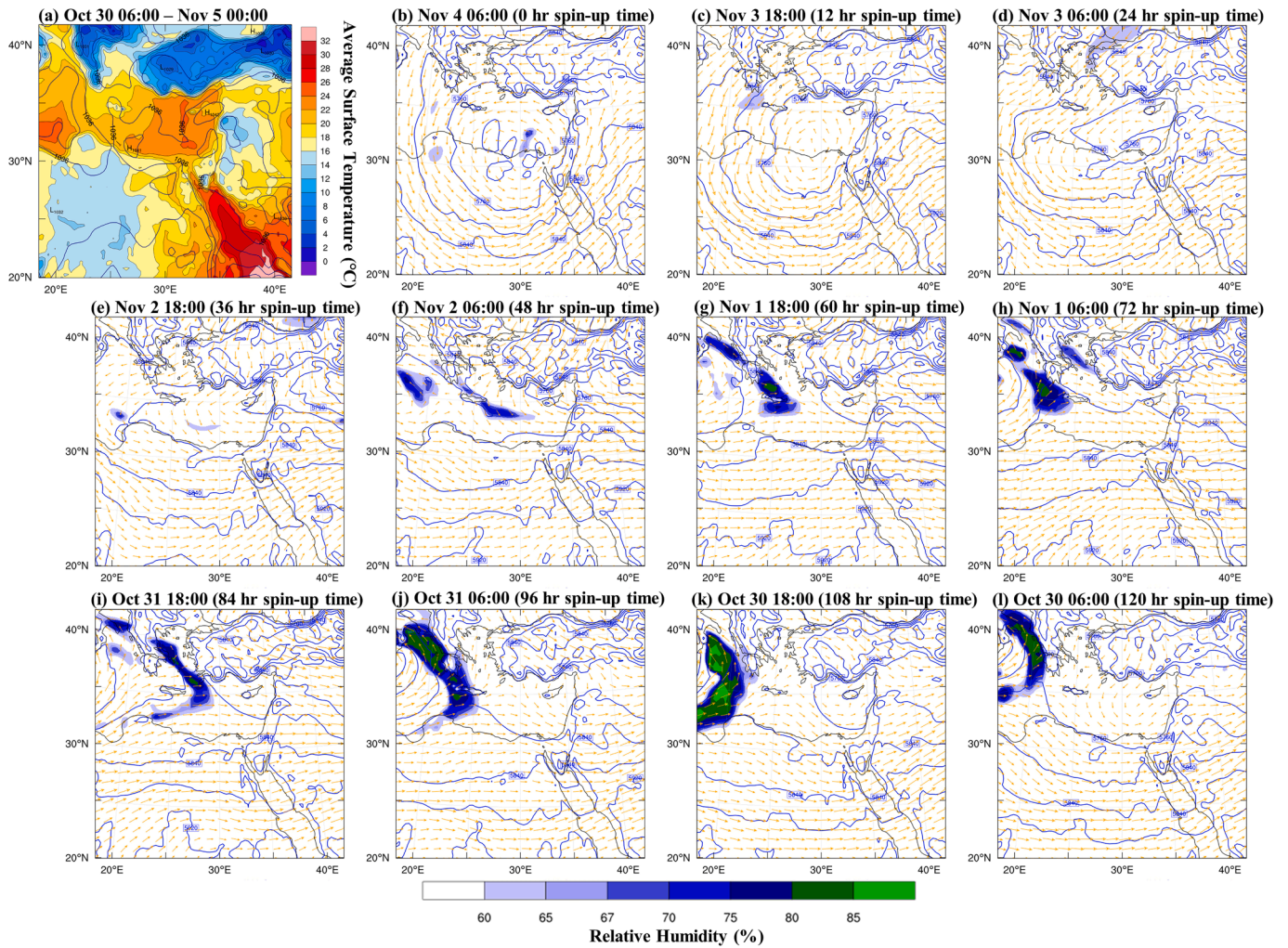


Fig. 5. Comparison of synoptic weather patterns in the D01 domains. (a) shows surface temperature and sea-level pressure averaged over the 120 h spin-up period and the 18 h study event duration. (b-l) show geo-potential heights (blue lines), wind speeds (arrow) and relative humidity (shaded) at 500-hPa at the different time points before the study event.

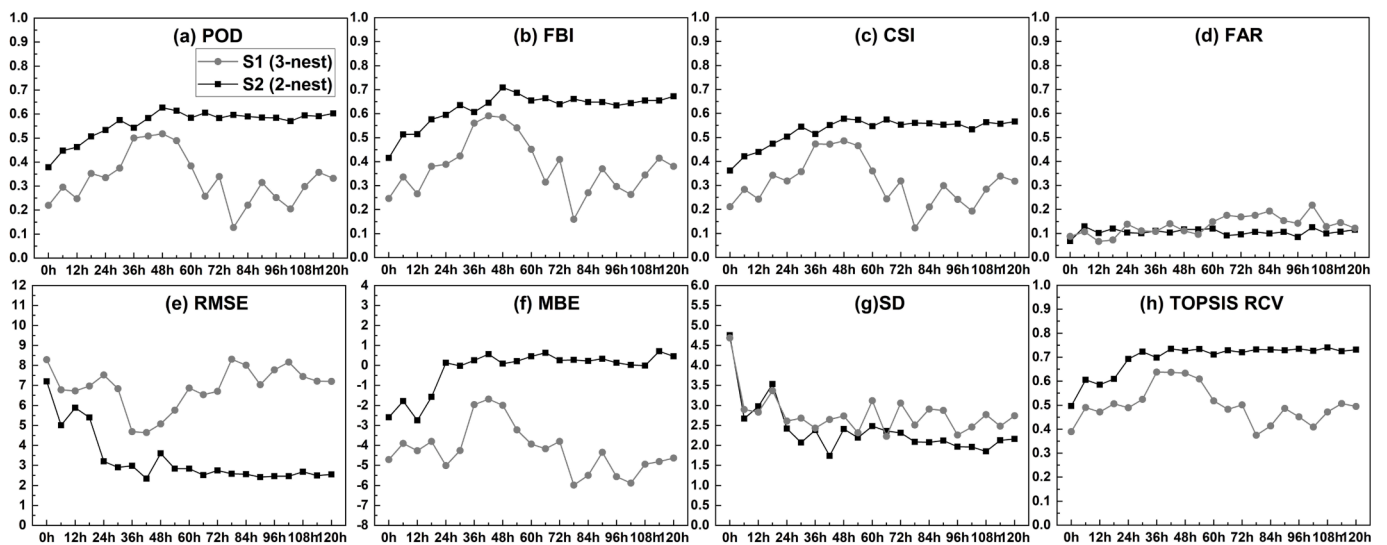


Fig. 6. The changes of seven verification metrics and overall performance score (*TOPSIS RCV*) with the increase of spin-up times. The grey lines represent the 3-level nested simulations while the black lines represent the 2-level nested simulations.

(0–12 h), steady growth period (12–24 h) and slow improvement period (24–48 h) of WRF simulation performance, respectively. First of all, the *TOPSIS RCV* of S2 increased significantly from 0.497 to 0.586 in 12 h spin-up time because the WRF model roughly adjusted numerics and physics using IC and BC and made basic progress on precipitation formation. During this period, the values of many metrics such as *RMSE*, *MBE*, and *SD* fluctuated and were far from the ideal value of 0. Therefore, 12 h could be seen as the critical period for this study event. But the WRF users are not recommended use spin-up times lower than the critical period to simulate precipitation as the model performance is very unstable and poor. Following this, the *TOPSIS RCV* of S2 grew greatly to 0.692 at 24 h spin-up time as the appropriate atmospheric circulations were developed. As shown by the black lines in Fig. 6, all metrics except *SD* became better without fluctuations. Besides, the current *TOPSIS RCV* is already very close to the performance peak value that can be achieved by the model under the influence of spin time. So 24 h could be seen as the minimum period to get good simulations for this study event, which is a suitable compromise between accuracy and computational efficiency. Finally, the *TOPSIS RCV* of S2 raise slowly and peaked at 0.727 at 48 h spin-up time when WRF made further refinement. Accordingly, 48 h could be seen as the adequate period for WRF to reach the best performance in this study event. The spin-up times longer than the adequate period would be unnecessary, as precipitation simulation cannot be further improved which would just increase unnecessary computational costs. If IC are ideal enough, *TOPSIS RCV* could be higher after this adequate spin-up period. But unsettled boundary conditions can still prevent further improvement (like the experiments of S1, grey lines in Fig. 6). Overall, the critical period, minimum period and adequate period are important for WRF users to determine the spin-up time in different circumstances. But the three periods found in this study event (12, 24 and 48 h) could vary for other events as it depends on the geographical features of the study area. Besides, comparing the S1 and S2 simulations in this study (with additional hydrometeors as input conditions) with the normal simulations (without additional hydrometeors as input conditions), it is found that the method of Hwang et al. (2019) does improve the performance of very short-range simulations. But Hwang et al. (2019)'s method can't compensate for the effects of unsettled weather conditions. When WRF users conduct simulations, the model's spin-up time should not be fixed and need to be adjusted reasonably by some steps. Section 5 introduces the proposed framework with several possible weather situations and processing steps for how to identify optimal spin-up time for different extreme precipitation events. The role of the critical period, minimum period, adequate period and satellite observations will also be reflected in this framework.

5. OSTI framework

5.1. Proposed framework

The OSTI framework provides a simple spin-up time identifying methodology that incorporates four steps and four possible weather situations, which aims to improve the efficiency of WRF short-term precipitation simulation. By judging which possible weather situation the proposed simulation is, the users could obtain the recommended spin-up time or determine whether the proposed domain configuration is feasible. Besides, some performance improvement methods suitable for specific situations were also introduced. The detailed processing steps and the situation explanations are shown below:

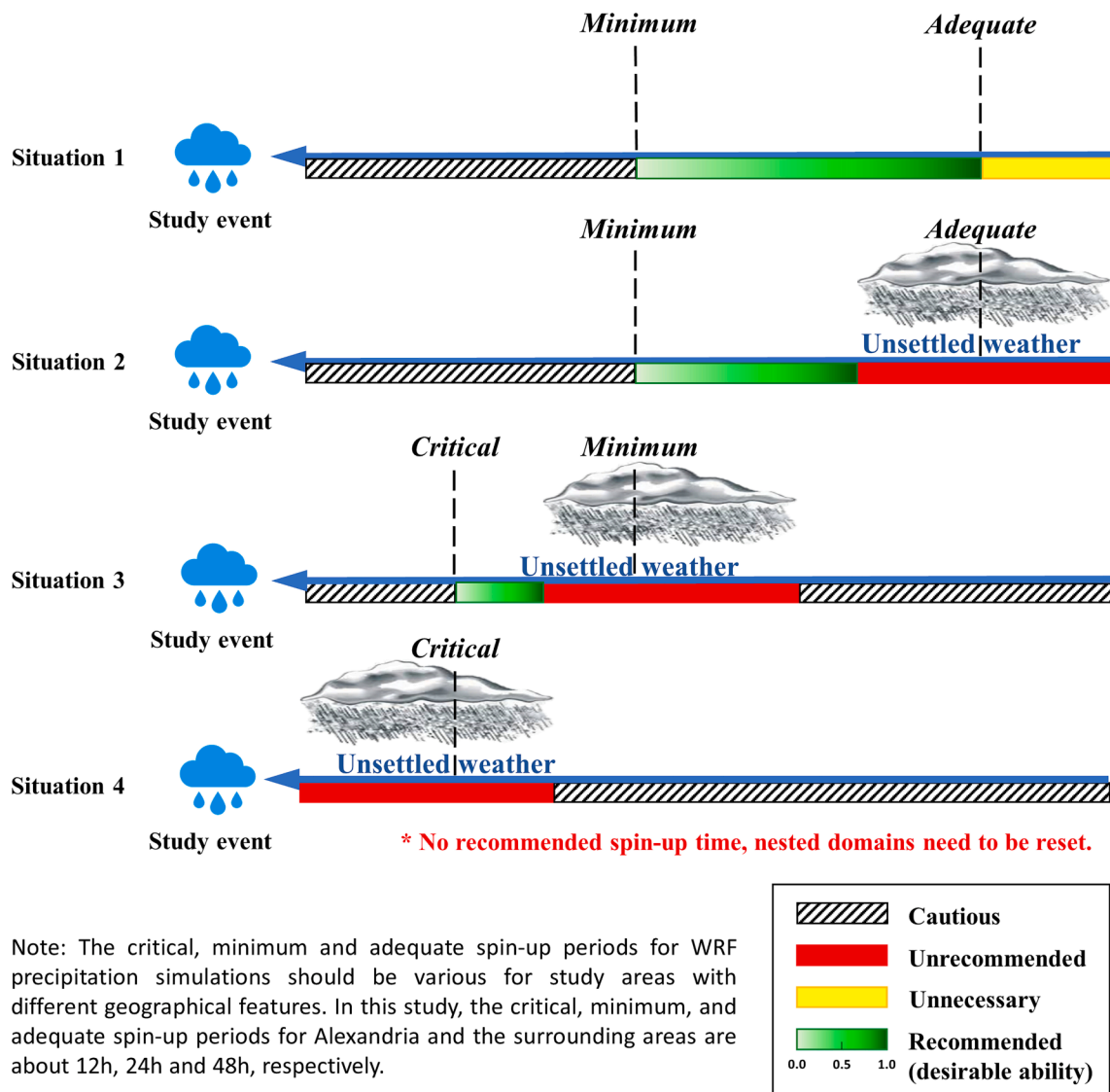
Step 1. Setting appropriate nested domain locations and sizes according to study event and WRF user guidance. It is recommended that the proposed domain should contain major mesoscale circulation features and have at least five grid points between adjacent nested domains for sufficient relaxation space (Warner, 2011).

Step 2. Checking the weather conditions around the proposed domain boundaries in the period prior to the study event. Observation data such as satellite and radar can be used. In this study, precipitation

as one of the most obvious signals is used to determine the area with unsettled weather conditions. But in the future study, more weather elements such as humidity, air temperature and pressure, wind speed and direction could also be employed as the indicator of input initial weather conditions.

Step 3. Find out which of the following situation best describes the occurrence of unsettled weather conditions prior to the study event around the domain boundaries. As mentioned in Section 4.2, the critical period, minimum period and adequate period are important standards of spin-up time for WRF model initialization and advancement. In Fig. 7, these periods are illustrated as the temporal distance from three simulation start points ("Critical", "Minimum" and "Adequate") to the study event, respectively. In addition, the temporal distance from the Study event To the Unsettled weather conditions period is defined as STU. The relationships between STU and the three standard periods are classified in the four possible weather situations in Fig. 7. For example, if an extreme precipitation event begins on October 2 at 10:00 while the unsettled weather conditions occur on October 1 at 08:00, the STU should be 26 h. And if we assume that the example event occurs at the same location as the study event and use the same model configurations, then the critical, minimum, and adequate spin-up periods are 12 h, 24 h and 48 h, respectively. Since minimum period (24 h) \leq STU (26 h) $<$ adequate period (48 h), it is classified as Situation 2 and the recommended spin-up time is between 24 h to just prior to the disturbing event. The four possible weather situations and different types of spin-up time ranges are explained below:

- **Situation 1:** $STU \geq$ adequate period, the proposed nested domains are ideal and can conduct WRF extreme precipitation simulation by a flexible spin-up time. The WRF model is recommended to start to spin up within the range between the minimum point and the adequate point (green bar). The spin-up time closest to the minimum period is with the least desirable ability (less recommended), while the spin-up time closest to the adequate period is with the most desirable ability (more recommended). For the spin-up times beyond the adequate period (yellow bar), they are considered unnecessary because of the unnecessary computational cost. On the contrary, the spin-up times less than the minimum period (striped bar) should be used cautiously as the atmospheric circulations may have not been developed thoroughly. If so, WRF users should check the precipitation spatial distribution of simulations with observations and pay more attention to adjusting the configuration.
- **Situation 2:** Minimum period \leq STU $<$ adequate period, the proposed domains would be affected by unsettled weather conditions when using the adequate spin-up time. If times and conditions allow, the spin-up time is recommended to be larger than the minimum period and further extended to get the best performance (green bar) but not extended to the unsettled weather period. The unsettled weather period (red bar) is definitively unrecommended for WRF initialization, which will introduce a large number of initial inconsistencies and make the model hard to resolve. The spin-up time usage in the range of the striped bar in Situation 2 is the same as in Situation 1.
- **Situation 3:** Critical period \leq STU $<$ minimum period. The explanations for the left striped bar and red bar of Situation 3 are similar to those above. The spin-up period is recommended to be as long as possible but not overlap with the unsettled weather period (green bar). Because the unsettled weather period is close to the study event, WRF can only complete rough adjustments during the critical spin-up time, the proposed domains are not very ideal and other model support methods would be required. For instance, users can use additional hydrometeor species as input parameters to help the model to achieve balance faster (Hwang et al., 2019). Alternatively, users can also use a much longer spin-up time, that is beyond the STU, i.e., the right striped bar. Such a way will allow the model to get stable IC and allow more time to balance later disturbance's introduction. However, using this cautiously because the simulations with



Note: The critical, minimum and adequate spin-up periods for WRF precipitation simulations should be various for study areas with different geographical features. In this study, the critical, minimum, and adequate spin-up periods for Alexandria and the surrounding areas are about 12h, 24h and 48h, respectively.

Fig. 7. Possible weather situations for WRF rainfall extreme simulations as well as the Cautious, Unrecommended, Unnecessary and Recommended ranges for model spin-up.

very long spin-up time may exhibit a certain level of freedom and chaotic behaviour that drift away from the forcing data (Alexandru et al., 2009). If simulation performance is poor, data assimilation and nudging would be support methods. Otherwise, go back to step 1 to set new nested domains to avoid unsettled weather conditions and perform the above steps again.

- **Situation 4:** $STU < \text{Critical Period}$, the proposed domains are not suitable and simulation performance would be severely affected by the IC. No recommended spin-up time for Situation 4. Go back to step 1 to reduce or enlarge the domain size to avoid unsettled weather conditions near the boundaries and perform these steps again.

Step 4. Conducting the WRF simulations with the recommended spin-up times. Observe whether simulated precipitation is accumulated reasonably or scatters toward the domain boundary as shown in Fig. 2 (j-l). If scatters are present, adjust other configurations in WRF, such as vertical and grid resolution, nesting ratio, and physical parameterization scheme to refine the simulation.

5.2. Framework application

To show the generalization ability of the proposed OSTI framework,

we have carried out simulations of three additional events around the Mediterranean Sea. Considering their simulation areas have the same geographical features and similar domain sizes as the Alexandria study event, the optimal spin-up time is determined by the same three standard periods (12 h, 24 h and 48 h) found before. Besides, because the most important part of the OSTI framework is to determine which situation the precipitation event belongs to, the example events with different STUs, synoptic backgrounds, precipitation distribution and intensity were chosen to demonstrate. Firstly, a precipitation event that occurred over Hurghada, Egypt on 26–27 October 2016 is an example of Situation 1 (Example 1) in the OSTI framework. Based on the weather maps from the NASA worldview (supporting files 3 and 4), the Red Sea and its surrounding area were calm, windless and rainless for several days prior to Example 1. This event happened over a narrow inland sea between the Arabian Peninsula and Africa, which has a relatively weak synoptic background than other events. Due to $STU \geq \text{adequate period}$, the most recommended spin-up time would be 48 h and longer times are unnecessary for performance improvement. Fig. 8 (a-c) show the accumulated precipitations in two days from IMERG observation, WRF simulation with 48 h recommended spin-up time (also adequate period) and WRF simulation with 60 h unnecessary spin-up time, respectively. It can be found that the intensity of precipitation in the centre is stable,

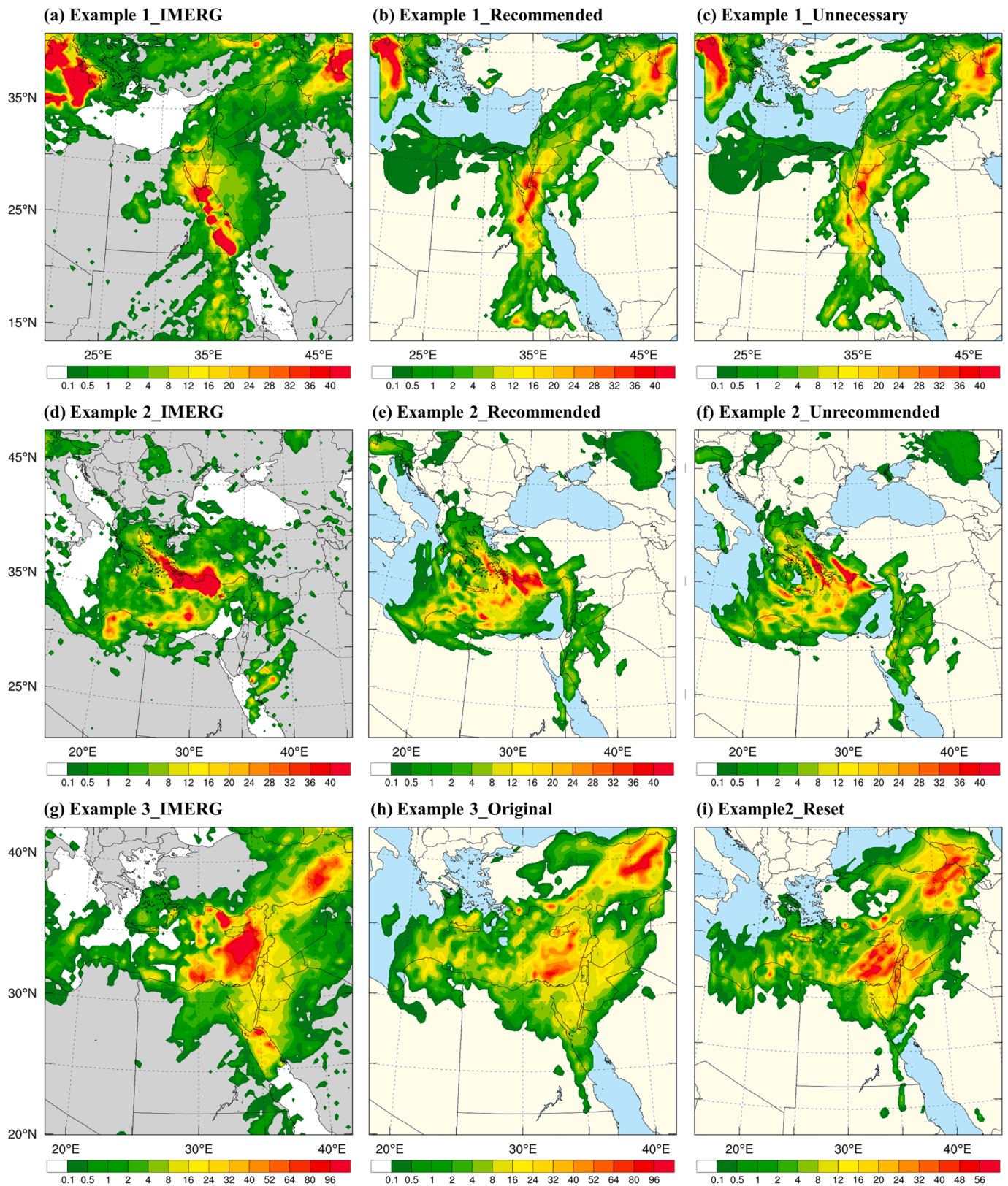


Fig. 8. Comparison of accumulated precipitation for three example events. (a) The IMERG observed precipitation for example 1 that happened around Hurghada, Egypt on the 26th and 27th of October 2016. (b-c) WRF precipitation results for example 1 simulated using the recommended spin-up time (48 h) and unnecessary spin-up time (60 h). (d) The IMERG observed precipitation for example 2 that happened around Antalya, Turkey on 12th December 2013. (e-f) WRF precipitation results for example 2 simulated using the recommended spin-up time (12 h) and unrecommended spin-up time (24 h). (g) The IMERG observed precipitation for example 3 that happened around Beirut, Lebanon on 4th November 2015. (h-i) WRF precipitation results for example 3 simulated using the same spin-up time (48 h) but different sized domains. The original and reset domain sizes are 80x80 (about 6.19 million km²) and 100x100 (about 9.73 million km²), respectively.

and there is not much improvement even with a longer execution time. For Situation 2 in the OSTI framework, the above study event over Alexandria, Egypt is a good example. The results of S1 experiments illustrate that WRF with the 40 h spin-up time has the best performance, while WRF simulations have poor performances when the disturbing weather includes in the IC (Fig. 6 (h)). Following this, an event that happened in Antalya, Turkey on 28 December 2013 is an example of Situation 3 (Example 2). The synoptic background of this event is relatively strong, and we can see obvious strong wind and atmospheric circulation moving from the left side of the Mediterranean Sea to the right (supporting files 5). In addition, the disturbing precipitation during the spin-up period is not heavy but exists close to the critical time of the Example 2 event (supporting files 6). In this case, critical period \leq STU $<$ minimum period, the WRF simulations with 12 h recommended and 24 h unrecommended spin-up times are conducted, respectively. It should be noted that there are certain ranges of recommended and unrecommended spin-up times, and 12 h and 24 h are selected only as demonstrations. As demonstrated in Fig. 8 (d-f), the precipitation distribution and intensity of recommended simulation are much better than the unrecommended when compared with IMERG observation. However, Situation 3 is the least frequent among all possible weather scenarios of WRF simulations. Because it requires that the disturbing precipitation occurring around the boundary is not too strong and ends before the critical period of the target event, to ensure that the performance of the WRF simulation will not be greatly affected. Furthermore, since the recommended spin-up time of this type of precipitation event is less than the minimum period, the simulation performance of these events will be more limited than others. Finally, Situation 4 of the OSTI framework is illustrated by taking the extreme precipitation event (Example 3) that occurred near Beirut, Lebanon on 25 October 2015 as an example. This event has the strongest weather background, the largest scale and the highest precipitation intensity among all the events in this study (supporting files 7 and 8). Significant large-scale atmospheric circulations occurred over the Mediterranean Sea and surrounding lands. In this case, based on the framework STU $<$ Critical Period, it is hard to get satisfied simulations by using the original small domains, and it is recommended to reset domain configurations. As shown in Fig. 8 (g-i), even though they were simulated with the same 48 h spin-up time, the simulations of the larger domain exhibited a significant improvement in reproducing the centre's extensive extreme precipitation. And this improvement is difficult to achieve by changing the length of model spin-up times. In sum, the OSTI framework can indeed help to decide the optimal spin time for WRF simulations before performing the simulations. According to time ranges indicated by the OSTI framework, the WRF precipitation results simulated by recommended, unrecommended and unnecessary spin-up times are basically in line with the expectations. Besides, this framework is more useful for precipitation events with strong synoptic backgrounds because their simulation performances depend largely on the development of appropriate atmospheric circulations and physical equilibrium states in the model. Such as the Alexandria study event, Antalya example 2 event and Beirut example 3 event driven by the large-scale movement of air together with ocean circulation, the simulated precipitation intensity and distribution vary greatly under different initial conditions and spin-up times. In contrast, the precipitation scale of Hurghada example 1 event with a relatively weak synoptic background is small and shows less sensitivity to the spin-up times than others.

6. Summary and conclusions

This study focuses on exploring the relationship between the WRF model required spin-up times and weather conditions. In particular, 21 3-level nested experiments and 21 2-level nested experiments for the extreme precipitation event that occurred on the 4 November 2015 in Alexandria are conducted for comparative analysis. Firstly, the simulated precipitation distributions of S1 experiments with unsettled

weather conditions are compared to the S2 experiments with calm weather conditions. These results are also analysed and linked to the time and location of disturbing precipitation. Then, to quantify the WRF performances under different spin-up times, seven error metrics *POD*, *FBI*, *CSI*, *FAR*, *RMSE*, *MBE*, *SD*, and an overall performance score *TOPSIS RCV* are calculated to show the differences between S1 and S2 experiments. Finally, the optimal spin-up time identifying (OSTI) framework is proposed to help WRF users understand and choose a suitable spin-up time for each precipitation event. The application and verification of this framework are also presented by three other extreme precipitation events around the Mediterranean Sea. The whole study aims to highlight the relationship between spin-up times and weather conditions and give clear guidelines on how to determine optimal spin-up time without too much trial-and-error testing.

By evaluating the results of the S1 and S2 experiments, it is demonstrated that even simulated with the same spin-on time, the spatiotemporal and overall performance of WRF simulations varies greatly under the influence of different IC. Specifically, the greatest differences of *POD*, *FBI*, *CSI* and *FAR* are about 0.5 (78 h spin-up time, C14 compared to c14), 0.5 (78 h spin-up time, C14 compared to c14), 0.45 (78 h spin-up time, C14 and c14), and 0.1 (102 h spin-up time, C18 and c18), respectively. Moreover, the *RMSE* is improved by up to 70.5%, and the overall performance score *TOPSIS RCV* improves from about 0.375 to 0.732 at most. The timing of these large differences in performance is all related to the timing of the Italy extreme precipitation event occurring around the domain boundaries. Since the timing and location of complex weather are uncertain, the optimal spin-up time for WRF should vary by study event and domain configuration. On the other hand, other factors could also have contributed to the simulation performance of S1 being worse than S2, such as model internal variability, which is well known to increase with domain size (Giorgi and Bi, 2000; Alexandru et al., 2007). But comparing the 21 S1 simulations with same domain configurations, the model performance can change significantly in very short times, which exceeds the level of internal model variability. Thus, the main reason for the model performance degradation in the larger domain simulations with long spin-up times should still be due to synoptic conditions.

Overall, the WRF spin-up time should not be too short as it is hard to develop the appropriate atmospheric circulations, but not the longer the better either because it depends on the quality of IC. A solution for this problem is to avoid unsettled weather conditions by using observations as guidance (radar, satellite, etc.), as well as determine the appropriate spin-up time based on the critical period, minimum period and adequate period for the study area (by referring to previous literature carried out in the same areas). Of course, there do exist some other methods for WRF to handle complex weather, like large-scale nudging. But the nudging method also has some problems such as how to choose the nudging strength to preserve internal model variability and keep the error growth under control. Besides, nudging could not yield perfect results because it is still unable to completely overcome the physical and dynamical deficiencies and inconsistencies in the WRF model (Bowden et al., 2012; Tang et al., 2017). Therefore, in some cases, users may consider adopting the proposed OSTI framework to avoid unnecessary adverse effects from complex weather conditions and improve modelling efficiency.

In this study, the sensitivities of WRF simulations to spin-up times and synoptic conditions are investigated, and the framework that estimates the optimal spin-up time through satellite observations is summarized. However, a limitation of this study is that it only investigated the effect of disturbing precipitation on simulation results. In fact, unsettled weather conditions can be of many kinds, while precipitation is one of the easiest to identify. In the future, more weather elements such as humidity, air temperature and pressure, and wind speed could also be explored to see whether they are helpful for optimal spin-up time identification. In addition, this study only considers the information provided by the observation data, because this way is the easiest that

users only need to browse the records on satellite websites. But whether it is possible to obtain more information from the ERA5 reanalysis data and determine the optimal configuration has not been done in this study. Considering the study in this field is very limited and there are no clear guidelines, we hope our study could provide a new idea to solve the lack of consensus on the choice of WRF spin-up time. Moreover, promote the establishment of WRF spin-up time guidelines for the community users. In summary, the results of this study will provide a useful reference for future WRF applications in extreme precipitation simulations.

CRedit authorship contribution statement

Ying Liu: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Lu Zhuo:** Conceptualization, Supervision, Writing – review & editing. **Dawei Han:** Conceptualization, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Ying Liu reports financial support was provided by China Scholarship Council.

Acknowledgements

The first author is grateful to the China Scholarship Council and the University of Bristol for providing the joint scholarship (CSC No, 201908310086) for her PhD study at the University of Bristol. All authors acknowledge the dataset developers from NASA and ECMWF for producing and sharing the datasets employed in this study. We would also like to acknowledge the Advanced Computing Research Centre at the University of Bristol for providing access to the High-Performance Computing system BlueCrystal.

References

- Afshar, M.A., Azadi, M., Rezazadeh, M., 2020. Uncertainty reduction in quantitative precipitation prediction by tuning of Kain-Fritsch scheme input parameters in the WRF model using the simulated annealing optimization method. *Meteorol. Appl.* 27 (4), e1919.
- Alexandru, A., de Elia, R., Laprise, R., 2007. Internal variability in regional climate downscaling at the seasonal scale. *Mon. Weather Rev.* 135 (9), 3221–3238.
- Alexandru, A., De Elia, R., Laprise, R., Separovic, L., Biner, S., 2009. Sensitivity study of regional climate model simulations to large-scale nudging parameters. *Mon. Weather Rev.* 137 (5), 1666–1686.
- Awan, N.K., Truhetz, H., Gobiet, A., 2011. Parameterization-induced error characteristics of MM5 and WRF operated in climate mode over the Alpine region: an ensemble-based analysis. *J. Clim.* 24 (12), 3107–3123.
- Bowden, J.H., Otte, T.L., Nolte, C.G., Otte, M.J., 2012. Examining interior grid nudging techniques using two-way nesting in the WRF model for regional climate modeling. *J. Clim.* 25 (8), 2805–2823.
- Cha, D.H., Wang, Y., 2013. A dynamical initialization scheme for real-time forecasts of tropical cyclones using the WRF model. *Mon. Weather Rev.* 141 (3), 964–986.
- Chen, F., Dudhia, J., 2001. Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Weather Rev.* 129 (4), 569–585.
- Cosgrove, B.A., Lohmann, D., Mitchell, K.E., Houser, P.R., Wood, E.F., Schaake, J.C., Robock, A., Sheffield, J., Duan, Q., Luo, L., Higgins, R.W., Pinker, R.T., Tarpley, J.D., 2003. Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS). *J. Geophys. Res. Atmos.* 108 (D22).
- Day, J.J., Tietsche, S., Hawkins, E., 2014. Pan-Arctic and regional sea ice predictability: Initialization month dependence. *J. Clim.* 27 (12), 4371–4390.
- De Elia, R., Laprise, R., Denis, B., 2002. Forecasting skill limits of nested, limited-area models: A perfect-model approach. *Mon. Weather Rev.* 130 (8), 2006–2023.
- IHE Delft, 2017. Improving flood resilience of Alexandria, Egypt, <https://www.un-ihe.org/news/improving-flood-resilience-alexandria-egypt>.
- Denis, B., Laprise, R., Caya, D., Côté, J., 2002. Downscaling ability of one-way nested regional climate models: the Big-Brother Experiment. *Clim. Dyn.* 18 (8), 627–646.
- Doblas-Reyes, F.J., Andreu-Burillo, I., Chikamoto, Y., García-Serrano, J., Guemas, V., Kimoto, M., Mochizuki, T., Rodríguez, L.R.L., Van Oldenborgh, G.J., 2013. Initialized near-term regional climate change prediction. *Nat. Commun.* 4 (1), 1715.
- Duzenli, E., Yucel, I., Pilatin, H., Yilmaz, M.T., 2021. Evaluating the performance of a WRF initial and physics ensemble over Eastern Black Sea and Mediterranean regions in Turkey. *Atmos. Res.* 248, 105184.
- Dzebre, D.E.K., Acheampong, A.A., Ampofo, J., Adaramola, M.S., 2019. A sensitivity study of Surface Wind simulations over Coastal Ghana to selected Time Control and Nudging options in the Weather Research and Forecasting Model. *Heliyon* 5 (3), e01385.
- Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., Tarpley, J.D., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res. Atmos.* 108 (D22).
- Evans, J.P., Ekström, M., Ji, F., 2012. Evaluating the performance of a WRF physics ensemble over South-East Australia. *Clim. Dyn.* 39 (6), 1241–1258.
- FloodList, 2015. Southern Italy Hit by Deadly Floods After 2 Days of Heavy Rain. <https://floodlist.com/europe/southern-italy-calabria-sicily>.
- Giorgi, F., Bi, X., 2000. A study of internal variability of a regional climate model. *J. Geophys. Res. Atmos.* 105 (D24), 29503–29521.
- Giorgi, F., Mearns, L.O., 1999. Introduction to special section: Regional climate modeling revisited. *J. Geophys. Res. Atmos.* 104 (D6), 6335–6352.
- Gómez-Navarro, J.J., Montávez, J.P., Jerez, S., Jiménez-Guerrero, P., Lorente-Plazas, R., González-Rouco, J.F., Zorita, E., 2011. A regional climate simulation over the Iberian Peninsula for the last millennium. *Clim. Past* 7 (2), 451–472.
- Grell, G.A., Freitas, S.R., 2014. A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmos. Chem. Phys.* 14 (10), 5233–5250.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., 2020. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* 146 (730), 1999–2049.
- Hong, S.Y., Lim, J.O.J., 2006. The WRF single-moment 6-class microphysics scheme (WSM6). *Asia-Pac. J. Atmos. Sci.* 42 (2), 129–151.
- Hou, A.Y., Kakar, R.K., Neeck, S., Azarbarzin, A.A., Kummerow, C.D., Kojima, M., Oki, R., Nakamura, K., Iguchi, T., 2014. The global precipitation measurement mission. *Bull. Am. Meteorol. Soc.* 95 (5), 701–722.
- Huang, C., Hu, J., Chen, S., Zhang, A., Liang, Z., Tong, X., Xiao, L., Min, C., Zhang, Z., 2019. How well can IMERG products capture typhoon extreme precipitation events over southern China? *Remote Sens. (Basel)* 11 (1), 70.
- Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., Yoo, S.H., 2015. NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). Algorithm Theoretical Basis Document (ATBD) Version 4, 26.
- Huva, R., Dargaville, R., Caine, S., 2012. Prototype large-scale renewable energy system optimisation for Victoria, Australia. *Energy* 41 (1), 326–334.
- Hwang, S.O., Park, J., Kim, H.M., 2019. Effect of hydrometeor species on very-short-range simulations of precipitation using ERA5. *Atmos. Res.* 218, 245–256.
- Hwang, C.L., Yoon, K., 1981. Methods for multiple attribute decision making. In: *Multiple attribute decision making*. Springer, Berlin, Heidelberg, pp. 58–191.
- Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O.B., Bouwer, L.M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., 2014. EURO-CORDEX: new high-resolution climate change projections for European impact research. *Reg. Environ. Chang.* 14, 563–578.
- Janjić, Z.I., 1994. The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Weather Rev.* 122 (5), 927–945.
- Jankov, I., Gallus Jr, W.A., Segal, M., Koch, S.E., 2007. Influence of initial conditions on the WRF-ARW model QPF response to physical parameterization changes. *Weather Forecast.* 22 (3), 501–519.
- Jerez, S., López-Romero, J.M., Turco, M., Jiménez-Guerrero, P., Vautard, R., Montávez, J.P., 2018. Impact of evolving greenhouse gas forcing on the warming signal in regional climate model experiments. *Nat. Commun.* 9 (1), 1304.
- Ji, F., Ekström, M., Evans, J.P., Teng, J., 2014. Evaluating precipitation patterns using physics scheme ensembles from a regional atmospheric model. *Theor. Appl. Climatol.* 115 (1), 297–304.
- Kain, J.S., 2004. The Kain-Fritsch convective parameterization: an update. *J. Appl. Meteorol.* 43 (1), 170–181.
- Kleczek, M.A., Steeneveld, G.J., 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. *Bound.-Lay. Meteorol.* 152 (2), 213–243.
- Leduc, M., Laprise, R., 2009. Regional climate model sensitivity to domain size. *Clim. Dyn.* 32, 833–854.
- Liu, J., Bray, M., Han, D., 2012. Sensitivity of the Weather Research and Forecasting (WRF) model to downscaling ratios and storm types in rainfall simulation. *Hydrol. Process.* 26 (20), 3012–3031.
- Liu, Y., Chen, Y., Chen, O., Wang, J., Zhuo, L., Rico-Ramirez, M.A., Han, D., 2021. To develop a progressive multimetric configuration optimisation method for WRF simulations of extreme precipitation events over Egypt. *J. Hydrol.* 598, 126237.
- Ludwig, P., Hochman, A., 2022. Last glacial maximum hydro-climate and cyclone characteristics in the Levant: a regional modelling perspective. *Environ. Res. Lett.* 17 (1), 014053.
- Manz, B., Páez-Bimos, S., Horna, N., Buytaert, W., Ochoa-Tocachi, B., Lavado-Casimiro, W., Willems, B., 2017. Comparative ground validation of IMERG and TMPA at variable spatiotemporal scales in the tropical Andes. *J. Hydrometeorol.* 18 (9), 2469–2489.
- Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J., Clough, S.A., 1997. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res. Atmos.* 102 (D14), 16663–16682.

- Naabil, E., Lamptey, B.L., Arnault, J., Olufayo, A., Kunstmann, H., 2017. Water resources management using the WRF-Hydro modelling system: Case-study of the Tono dam in West Africa. *J. Hydrol.: Reg. Stud.* 12, 196–209.
- Navarro, A., García-Ortega, E., Merino, A., Sánchez, J.L., Kummerow, C., Tapiador, F.J., 2019. Assessment of IMERG precipitation estimates over Europe. *Remote Sens. (Basel)* 11 (21), 2470.
- Palomino-Ángel, S., Anaya-Acevedo, J.A., Botero, B.A., 2019. Evaluation of 3B42V7 and IMERG daily-precipitation products for a very high-precipitation region in northwestern South America. *Atmos. Res.* 217, 37–48.
- Papavasileiou, G., Kotroni, V., Lagouvardos, K., Giannaros, T.M., 2022. Observational and numerical study of a giant hailstorm in Attica, Greece, on 4 October 2019. *Atmos. Res.* 278, 106341.
- Prakash, S., Mitra, A.K., AghaKouchak, A., Liu, Z., Norouzi, H., Pai, D.S., 2018. A preliminary assessment of GPM-based multi-satellite precipitation estimates over a monsoon dominated region. *J. Hydrol.* 556, 865–876.
- Rago, V., Chiaravalloti, F., Chiodo, G., Gabriele, S., Lupiano, V., Nicastro, R., Pellegrino, A.D., Procopio, A., Siviglia, S., Terranova, O.G., Iovine, G.G., 2017. Geomorphic effects caused by heavy rainfall in southern Calabria (Italy) on 30 October–1 November 2015. *J. Maps* 13 (2), 836–843.
- Rummukainen, M., 2010. State-of-the-art with regional climate models. *Wiley Interdiscip. Rev. Clim. Chang.* 1 (1), 82–96.
- Schewe, J., Gosling, S.N., Reyer, C., Zhao, F., Ciaia, P., Elliott, J., Francois, L., Huber, V., Lotze, H.K., Seneviratne, S.I., Van Vliet, M.T., 2019. State-of-the-art global models underestimate impacts from climate extremes. *Nat. Commun.* 10 (1), 1005.
- Seck, A., Welty, C., Maxwell, R.M., 2015. Spin-up behavior and effects of initial conditions for an integrated hydrologic model. *Water Resour. Res.* 51 (4), 2188–2210.
- Skamarock, W.C., 2004. Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Weather Rev.* 132 (12), 3019–3032.
- Skamarock, W.C., Klemp, J.B., 2008. A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.* 227 (7), 3465–3485.
- Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Liu, Z., Berner, J., Wang, W., Powers, J.G., Duda, M.G., Barker, D.M., Huang, X.Y., 2019. A Description Of The Advanced Research WRF Model Version 4. National Center for Atmospheric Research, Boulder, CO, USA, p. 145.
- Talbot, C., Bou-Zeid, E., Smith, J., 2012. Nested mesoscale large-eddy simulations with WRF: Performance in real test cases. *J. Hydrometeorol.* 13 (5), 1421–1441.
- Tang, J., Wang, S., Niu, X., Hui, P., Zong, P., Wang, X., 2017. Impact of spectral nudging on regional climate simulation over CORDEX East Asia using WRF. *Clim. Dyn.* 48 (7), 2339–2357.
- Tang, G., Zeng, Z., Long, D., Guo, X., Yong, B., Zhang, W., Hong, Y., 2016. Statistical and hydrological comparisons between TRMM and GPM level-3 products over a midlatitude basin: Is day-1 IMERG a good successor for TMPA 3B42V7? *J. Hydrometeorol.* 17 (1), 121–137.
- Turco, M., Sanna, A., Herrera, S., Llasat, M.C., Gutiérrez, J.M., 2013. Large biases and inconsistent climate change signals in ENSEMBLES regional projections. *Clim. Change* 120, 859–869.
- Wang, Z., Zhong, R., Lai, C., Chen, J., 2017. Evaluation of the GPM IMERG satellite-based precipitation products and the hydrological utility. *Atmos. Res.* 196, 151–163.
- Wang, J., Zhuo, L., Han, D., Liu, Y., Rico-Ramirez, M.A., 2023. Hydrological model adaptability to rainfall inputs of varied quality. *Water Resour. Res.* 59 (2).
- Warner, T.T., 2011. Quality assurance in atmospheric modeling. *Bull. Am. Meteorol. Soc.* 92 (12), 1601–1610.
- Yang, Z.L., Dickinson, R.E., Henderson-Sellers, A., Pitman, A.J., 1995. Preliminary study of spin-up processes in land surface models with the first stage data of Project for Intercomparison of Land Surface Parameterization Schemes Phase 1 (a). *J. Geophys. Res. Atmos.* 100 (D8), 16553–16578.
- Zevenbergen, C., Bhattacharya, B., Wahaab, R.A., Elbarki, W.A.I., Busker, T., Rodriguez, C.S., 2017. In the aftermath of the October 2015 Alexandria Flood Challenges of an Arab city to deal with extreme precipitation storms. *Nat. Hazards* 86 (2), 901–917.
- Zhuo, L., Dai, Q., Han, D., Chen, N., Zhao, B., 2019. Assessment of simulated soil moisture from WRF Noah, Noah-MP, and CLM land surface schemes for landslide hazard application. *Hydrol. Earth Syst. Sci.* 23 (10), 4199–4218.