Integrating WRF forecasts at different scales for pluvial flood forecasting using a rainfall threshold approach and a real-time flood model

Young, Adele; Bhattacharya, Biswa; Daniëls, Emma; Zevenbergen, Chris

DOI
10.1016/j.jhydrol.2025.132891

Publication date
2025

Document Version
Final published version

Published in
Journal of Hydrology

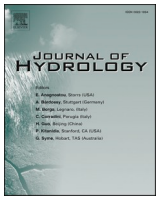
Citation (APA)
Young, A., Bhattacharya, B., Daniëls, E., & Zevenbergen, C. (2025). Integrating WRF forecasts at different scales for pluvial flood forecasting using a rainfall threshold approach and a real-time flood model. *Journal of Hydrology*, *656*, Article 132891. https://doi.org/10.1016/j.jhydrol.2025.132891

Journal of Hydrology 656 (2025) 132891

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol



Research papers

Integrating WRF forecasts at different scales for pluvial flood forecasting using a rainfall threshold approach and a real-time flood model☆

Adele Young [a,c,*], Biswa Bhattacharya [b], Emma Daniëls [d,1], Chris Zevenbergen [a,c]

[a] IHE Delft, Department of Coastal and Urban Risk Resilience, P.O.Box 3015, 2601 DA Delft, the Netherlands
[b] IHE Delft, Department of Hydroinformatics and Socio-Technical Innovation, P.O. Box 3015, 2601 DA Delft, the Netherlands
[c] Delft University of Technology, Faculty of Civil Engineering and Geosciences, 2628 CN Delft, the Netherlands
[d] Royal Netherlands Meteorological Institute KNMI, R&D Weather and Climate Modelling, P.O.Box 201, 3730 AE De Bilt, the Netherlands

ARTICLE INFO

Associate editor: A. Bardossy

Keywords:
Data scarce areas
Urban flood forecasting
Rainfall threshold approach
Early warning system
WRF modelling
Alexandria
Egypt

ABSTRACT

High-resolution rainfall forecasts are essential for forecasting urban pluvial floods and providing local decision support before an event when using rainfall or model-based approaches. Regional Numerical Weather Prediction (NWP) models increase the resolution and extend forecasts by several days when initialised with global models. These models produce forecasts at higher spatial and temporal resolutions but are computationally demanding and do not necessarily result in more accurate forecasts especially in data-scarce cities. This research evaluated how rainfall forecasting model scale dependencies satisfy high-resolution hazard forecasting requirements with two different flood forecasting approaches. Rainfall forecasts of different spatial resolutions, cumulus schemes and lead times from a high-resolution Weather Research and Forecasting (WRF) model were first evaluated and then used as input in a rainfall threshold and 1D MIKE urban drainage model for flood forecasting in the data-scarce city of Alexandria City, Egypt. Results indicate the flood forecast severity class and flood model simulation results vary with the neighbourhood size, forecast horizon, and chosen cumulus configuration but in general the smallest resolutions evaluated did not improve the hazard estimation for both flood forecasting approaches. Therefore, trade-offs must be made regarding model configurations, resolution, lead times and how the forecast output will be used. This study demonstrates the opportunities and limitations for better integrating high-resolution meteorological for the development of a rainfall threshold-based and model-based flood forecasting in cities with similar conditions. It also highlights the need to align the selected model configuration with the goals of the flood forecasting application which is critical for effective early warning systems and anticipatory flood management.

1. Introduction

The frequency and intensity of extreme rainfall events are on the rise (Prein et al., 2017), and as a result, of higher population densities, cities will be disproportionately at higher flood risk (Ermagun et al., 2024). Urban pluvial (rain-driven) or surface water flooding occurs due to high rainfall intensity in urban pluvial catchments with insufficient drainage capacity and water accumulation in low-lying areas, resulting in property damage, disruption of services, and loss of life. In urban areas with combined drainage systems, there is the additional risk of untreated effluent surcharging manholes, which mix with surface water, creating unsafe health conditions (Houston et al., 2011). With continued urbanisation and rapidly expanding cities under threat of flooding, the call to make cities more disaster resilient through preparedness for extreme weather events is increasing.

This research aims to further our knowledge of applying early warning systems supported by real-time rainfall and flood forecasting for Anticipatory Flood Management (AFMA). AFMA promotes forecast-based actions taken ahead of time to increase preparedness and reduce residual risk in data-scare areas. AFMA actions include warnings to avoid flood-prone areas, flow redistribution and pumping out water, mobilising emergency management, and cleaning the drainage system.

☆ This article is part of a special issue entitled: 'Real-time Flood' published in Journal of Hydrology.

* Corresponding author.
E-mail address: a.young@un-ihe.org (A. Young).
[1] Present address: Utrecht University, Faculty of Science, Princetonplein 5, 3584 CC Utrecht, The Netherlands.

https://doi.org/10.1016/j.jhydrol.2025.132891

Available online 21 February 2025

The effectiveness of actions depends on the information provided by rainfall forecasts and flood forecasting (FF) approaches, which include rainfall thresholds, pre-simulated scenarios and/or real-time simulation approaches. Flood forecasting and warning approaches in urban pluvial catchments require accurate predictions about peak rainfall rates and the magnitude, timing, location and impacts of flooding and require high spatial and temporal rainfall resolution forecasts suitable for small-scale urban flood modelling, which is often unavailable in data-scarce regions.

Further, delivering urban flood forecasts with sufficient lead times is challenged by uncertainties in forecasting rainfall at high spatial and temporal resolutions and running flood inundation models in real-time. Radar-based nowcasting has been used to improve the spatial and temporal resolution of rainfall for triggering warnings and FF (Schellartet al., 2011; René et al., 2013; Thorndahl et al., 2013; Iqbal, 2017; Flack et al., 2019). However, radar-based nowcasting for FF applications is provided at short lead times (< 6 h), which may be insufficient for implementing actions requiring more than 6 h lead time or unavailable, especially in data-scarce regions where nowcasting is unavailable.

When initialised with global data, the Weather Research and Forecasting Model (WRF) (Skamarock et al., 2008) forecasts rainfall at longer lead times and allows downscaling to smaller high-resolution domains. The smaller domains capture high-scale features and processes which can better simulate precipitation characteristics (Lean et al., 2008; Prein et al., 2015); Kendon et al., 2017). Increasing the simulation domain resolution alone is not enough to significantly improve the model performance. Care must also be given to the treatment of convection (Fowler et al., 2016) and the microphysics and planetary boundary layer (PBL) schemes (Clark et al., 2015; Kain et al., 2008). Many studies have investigated the impacts of different model configurations on rainfall distribution (Jankov et al., 2007; Goswami et al., 2012; Wang et al., 2016. Jankov et al.,(2005) found the most significant variability in forecasts from changes in the choice of Cumulus Parametrisation Scheme (CPS) which influences the average rain rate. (Umer et al., 2021) evaluated different CPS schemes and demonstrated the potential of WRF models as a valuable asset to capture peak rainfall intensity for flash flood modelling in a city where high quality direct and remotely sensed observations of rainfall are limited. Therefore, in high resolution WRF modelling, the model configuration has a significant impact on rainfall quantities and distributions and subsequently on flood forecasts.

There are interdependencies between the different components of the flood forecasting chain; meteorological forecasting input, hazard estimation methodologies and decision support systems that trigger a warning or action (Fig. 1). The flood forecasting approach or hazard estimation method uses empirical methods such as rainfall threshold or real-time flood simulations and data-driven models to identify the critical rainfall or flood depths which will trigger a warning or action. Physically based hydrodynamic flood forecasting models operate in real-time, integrating forecasted rainfall or observed flows to update conditions when available. The suitability of the hazard estimation approach is influenced by several factors: forecasting scale, quality of information and models available, triggering mechanism, requisite forecast variables and end-user needs (Henonin et al., 2013; Speight et al., 2020). Data driven models have the advantage of fast computational times but the lack of observed time series data in data scarce regions makes their application limited.

Using high horizontal resolution output from Numerical Weather Prediction models such as WRF with rainfall threshold approach or in real-time simulations can be valuable for forecasting urban pluvial floods at various lead times and has been done by other authors (Thorndahl et al., 2016; Brendel et al., 2020; Ming et al., 2020; Davis et al., 2022; Wang et al, 2022). The main problem this paper aims to address is whether increasing the resolution of WRF rainfall forecasts to satisfy urban flood modelling requirements will lead to improved quantification of rainfall magnitude and distribution for flood forecasting in urban pluvial cities.

This will be done by assessing the usability and performance of high-resolution WRF rainfall to produce flood forecasts for implementing anticipatory flood management actions 12–72 h before an event. Given the influence of the model configurations, WRF rainfall output is evaluated at different lead times, domain resolutions, treatments of convection and neighbourhood sizes. Further, the study investigates the influence of increasing the resolution and model configuration with two FF approaches: a rainfall threshold approach when urban models are unavailable and a real-time hydrodynamic model to forecast flood depths. Whereas previous research has assessed the influence of increased rainfall resolution with radar nowcasting or post-processing with observations, the novelty of this research is in jointly exploring the scale dependencies of both the meteorological forecasting model and different flood forecasting approaches, which has not been explored sufficiently, especially in data-scarce cities.

The research has been carried out for extreme events in the Mediterranean coastal city of Alexandria in Egypt. Every winter, Alexandria experiences extreme weather events that cause significant flooding and disruption to lives throughout the city (Zevenbergen et al, 2016). This research does not aim to determine the optimal model configuration for the WRF model given the challenges in data-scarce cities but instead explores the trade-offs in model accuracy and usefulness of rainfall forecasts to generate flood forecasts in data-scarce regions. Evaluating high-resolution models and their application for rainfall and flood

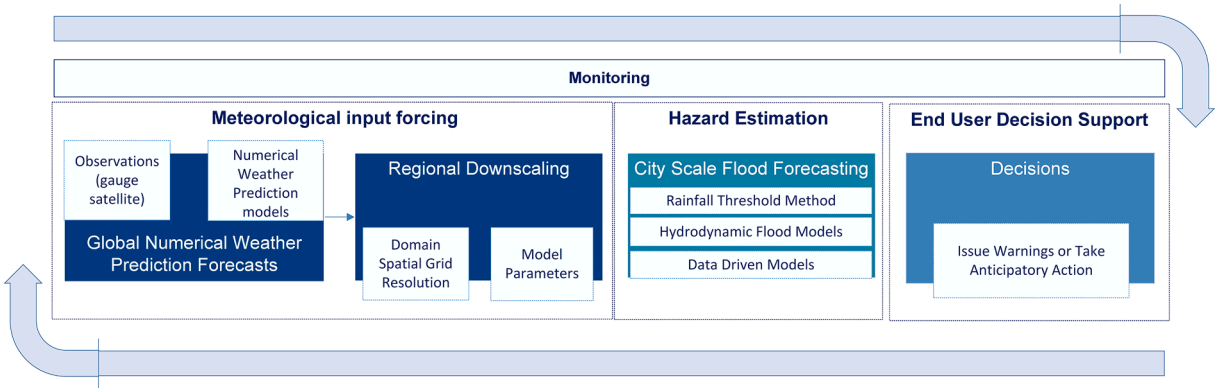


Fig. 1. Components of a pluvial flood forecasting system for early warnings and anticipatory flood management. Observations and models are used to create Global Numerical Weather Prediction Forecasts. Regional Downscaling is done with the WRF model combining domain variable spatial grid resolutions with different convection parametrisations. Hazard estimation methods using a rainfall threshold approach, real-time flood or data driven models are combined with decision support systems to trigger warnings and actions.

forecasting in data-scarce cities can highlight the interdependencies of the flood forecast chain's meteorological and hydrological aspects, which are often evaluated separately. The findings can inform the design, development and improvement of existing and future FF and early warning systems for increased preparedness by implementing anticipatory actions in cities similar to Alexandria.

2. Study area

Alexandria is located in the Alexandria Governorate, Nile Delta on the southern boundary of the Mediterranean Sea (Fig. 2). A renowned tourist destination, the city is known for its cultural heritage and landmarks. It is the second largest city in Egypt spanning over 2,300 km^2 with a population of approximately 5.5 million (UN, 2023) and comprises smaller districts. Like many Middle Eastern and North African Region cities, it has experienced urban expansion over the years, has low adpative capacity(Abdrabo et al., 2023) and informal settlements account for a third of Alexandria's total population (AASTMT and Egis BCEOM International, 2011). The coastal city is characterised by an arid Mediterranean climate with rainy winters, occasional storms (from October to April), and long warm summer months (May to September) with no rain. There is high annual rainfall variability with 360 mm(max) and 70 mm(min), averaging 195 mm/year. In the winter months, Alexandria is associated with a high temporal rainfall variability which is a distinguishing characteristic of the Mediterranean climate (Hasanean, 2004). Winter storms, locally referred to as "Al-Nawat", are a result of migratory cyclones and fronts, upper-level troughs (low pressure) accompanied by strong winds, rains and storm surges or strong winds only.

Alexandria City was selected as a case study given the increased number of flood events in recent years and the challenges faced with developing a flood forecasting system with limited data and resources. The city suffers from occasional pluvial flooding from runoff accumulation and surcharging of the city's dense sewer network and these risks are expected to increase with a changing climate (AASTMT and Egis BCEOM International, 2011). For example, the El Gomork district experiences occasional flooding and traffic disruptions along the Corniche. Storm water is collected via a combined sewer system and pumped to the West Treatment plant via lift stations and force main pipes. Alexandria does not have an operational flood forecasting system and lacks reliable sub-daily rainfall gauge data. City officials currently use information from the National Meteorological Agency and historical knowledge of floods to issue warnings and take action. With the threat of emerging risk, AFMA has been proposed as a viable solution to increase preparedness and reduce damage (Zevenbergen et al, 2016; Bhattacharya et al., 2018).

This research evaluates three extreme rainfall events, on October 25th, 2015, Dec 5th, 2018 and Nov 20th 2020 during the winter months which resulted in significant flooding and a state of emergency was declared in the governorate. Several studies using WRF have been carried out in Egypt using the National Center for Environmental Prediction's (NCEP), Global Forecasting System (GFS) and the European Centre for Medium-range Weather Forecast's ERA5 and ERA interim data sets. (Robaa and Wahab, 2019) and Ibrahim (2020) evaluated the sensitivity of WRF to convection schemes and both found there was an overestimation of rainfall but results improved with finer resolution. (Eltahan and Magooda, 2018) investigated rainfall sensitivity to microphysics schemes and found all schemes produced higher precipitation than observed. Eltahan and Magooda, (2018) found the Morrison double moment scheme achieved the closest to observed rainfall while the Thompson scheme successfully simulated the cloud pattern. El Afandi et al. (2020) found favourable results using a Single-Moment 3-class WSM3 scheme and Kain-Fritsch scheme for the development of an early warning system in the Sinai Peninsula. Ibrahim and Afandi (2014) and Cools et al. (2012) previously evaluated the use of a WRF to predict extreme rainfall over the Sinai Peninsula and Egypt. Most recently, Liu et al. (2021) used a progressive multi-metric configuration optimisation method and ERA5 reanalysis data to find an optimal configuration of the WRF model for the study area. All previous studies evaluated different resolutions. To the best of our knowledge, no studies have evaluated the use of WRF with resolutions less than 3 km which are the resolutions suggested for flood forecasting in urban scale modelling (Berne et al., 2004).

3. Data and methods

This research analysed WRF rainfall for different lead times, domain resolutions and treatments of convection at different neighbourhood

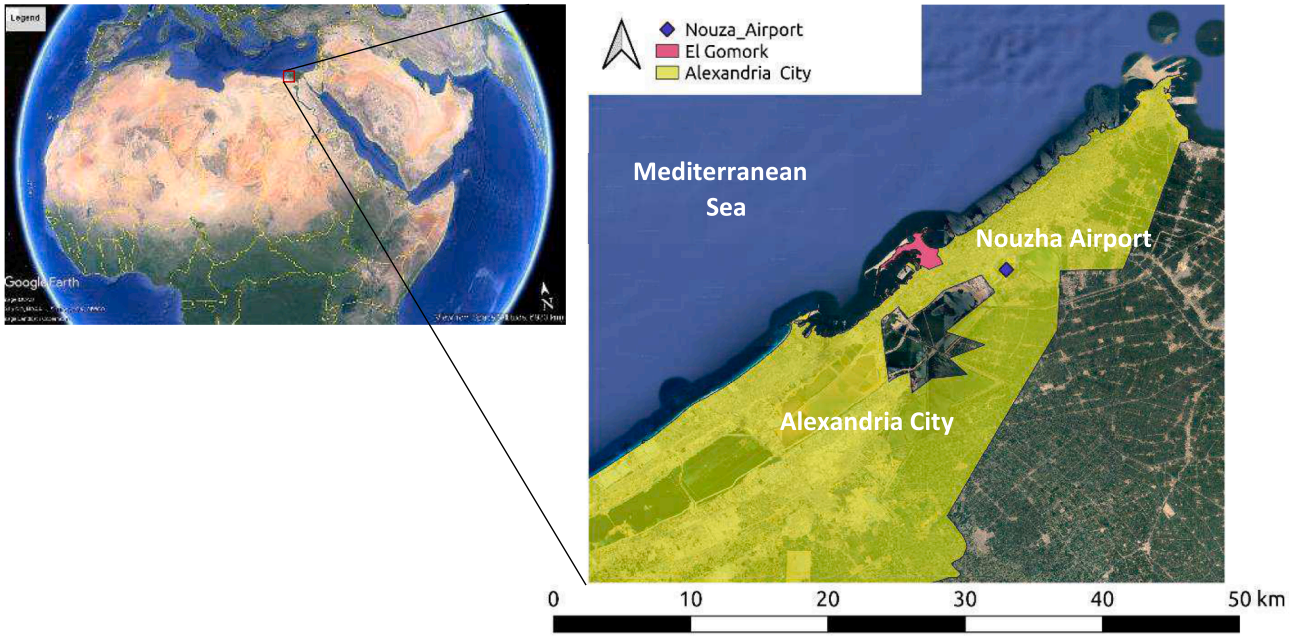


Fig. 2. Geographical location of Alexandria City. The El Gomork district is shown in pink. Refer to Fig. 5 for more details on the El Gomork drainage system. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sizes and evaluated with gridded rainfall estimates. The WRF rainfall was then used at different scales with two flood forecasting approaches: a rainfall threshold approach and an urban drainage flood model to evaluate how well they forecasted a hazard. A summary of the research framework is presented in Fig. 3. The following section briefly describes the WRF model setup and the flood forecasting approaches used.

3.1. Data and WRF setup

The Advanced Research WRF (ARW-WRF), (Skamarock et al., 2008) is a convection-permitting dynamic downscaling model configured using grid parameterisations and physics schemes, and nested domains at finer resolutions. This research used 0.25° Global Forecast Grids Historical Archive ds084.1 for lateral boundary conditions, MODIS15 land-use map and two-way nesting. Analyses were first carried out using three domains and then four domains to evaluate if forecast skill is improved by increasing the number of domains and higher spatial resolutions. The domains used boundary conditions that were updated every 3 h, a nesting ratio of 1:3:3:3 and a spatial resolution for domain 1 (D1) 10 km, domain 2 (D2) 3.3 km, domain 3 (D3) 1.1 km and domain 4 (D4) 0.37 km (Fig. 4). The outermost domain (D1) covered Egypt and the wider Mediterranean. This is consistent with the mother domain used by the Egyptian Meteorological Agency.

All spatial domains had a vertical resolution of 45 layers and the model used a timestep of 60 s with adaptive timestep. The selection of the physics schemes (Table 1) was guided by the most recent study of (Liu et al., 2021) along with the consideration for convection-permitting scales, (1–2 km) which can run without convection schemes because the grid spacing is small enough to explicitly resolve convection. However, in the convective grey zone (3–10 km) cumulus-cloud processes become partially resolved and traditional closure assumptions break down and scale-aware schemes can be used (Grell and Freitas, 2014; Zheng et al., 2016; Jeworrek et al., 2019). There is still some disagreement on how CPS should be treated in smaller domains. Paul et al. (2018) found for the Mumbai's coastal areas turning the CPS on in a 1 km simulated best results, whereas Han and Hong, (2018) found better simulation when CPS in turned off in a 3 km domain. Therefore, additional sensitivity experiments were tested for Domain 2 and 3 in which convection was parameterised and cumulus was kept off in D4 (Table 1). R1, R2 and R3 correspond to runs with 3 domains and R1_4d, R2_4d and R3_4d correspond to runs with 4 domains. The runs were initialised at the 72 h, 48 h, 24 h and 12 h lead times with a spin-up time of 12 h for three events. The objective of this research was not to determine the optimal model configuration for the WRF model but to investigate if increasing the resolution improves rainfall forecasts and to determine the usability of high resolution WRF output in flood forecasts and early warning systems. A summary of the WRF setup and methodology used is shown in Fig. 4.

Table 1
WRF model parameterisation is used, and cumulus parametrisation is used for the different runs. Parameterisation schemes are kept the same for all runs except cumulus where the Grell Freitas scheme cu = 3 is used or when no cumulus scheme is used cu = 0. R1, R2 and R3 correspond to runs with 3 domains (3d runs) and R1_4d, R2_4d and R3_4d correspond to runs with 4 domains (4d runs).

Physics schemes (Based on Lui *et al.*, 2021) same for all runs	Name	SCHEME #					
Microphysics	Single-moment 6-classWSM6	**6**					
Longwave Radiation	RRTM	**1**					
Shortwave Radiation	RRTM	**1**					
Boundary Layers	Mellor-Yamada-Janjic	**2**					
Surface Layers	Eta Similarity Scheme	**2**					
Land Surface	Unified Noah Land Surface Model	**2**					
Cumulus	Grell Freitas						
Runs (scheme configuration)		**3d runs**			**4d runs**		
	Domain	**R1**	**R2**	**R3**	**R1_4d**	**R2_4d**	**R3_4d**
	D1	3	3	3	3	3	3
	D2	0	3	3	0	3	3
	D3	0	0	3	0	0	3
	D4	–	–	–	0	0	0

Table 2
Thresholds (in mm) used to trigger warnings.

Flood Hazard Classification and Corresponding Rainfall Depth Thresholds			
0–11.99 mm No to minimal flooding	12–19.99 mm Minor flooding	20–31.99 mm Significant flooding	>=32 mm Severe flooding

3.2. Flood forecasting approaches rainfall threshold approach

Thresholds for the 24 h accumulation were selected based on previous analysis of critical rainfall thresholds (Young et al., 2021). These correspond to the severity of flooding: "No to minimal flooding" with rainfall within 0 to 11 mm, "Minor flooding" within 12 to 19 mm, "Significant flooding" within 20 to 31 mm or "Severe" with rainfall ≥ 32 mm. Young et al. (2021) derived thresholds using observed daily rainfall accumulation and rainfall intensity for a 3-year training period from

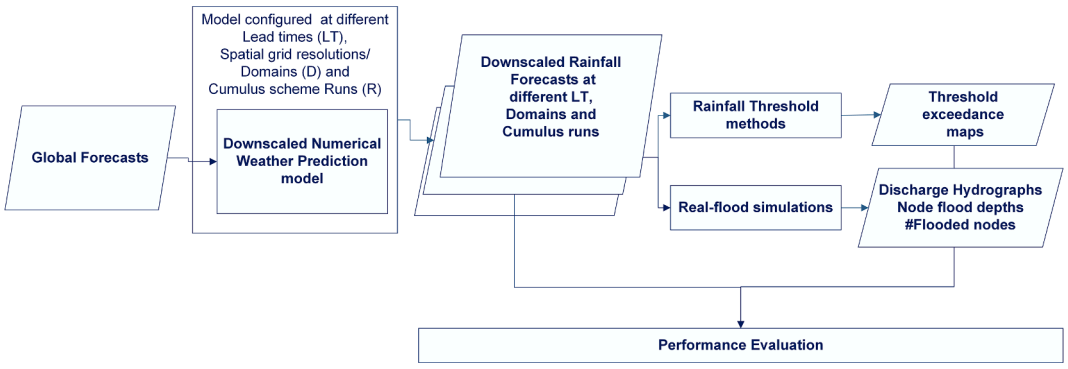


Fig. 3. Research Framework illustrating the methods for coupling WRF model to produce high resolution rainfall forecasts and hazard estimation using a rainfall threshold at the city scale and real time urban flood simulations for part of the city.

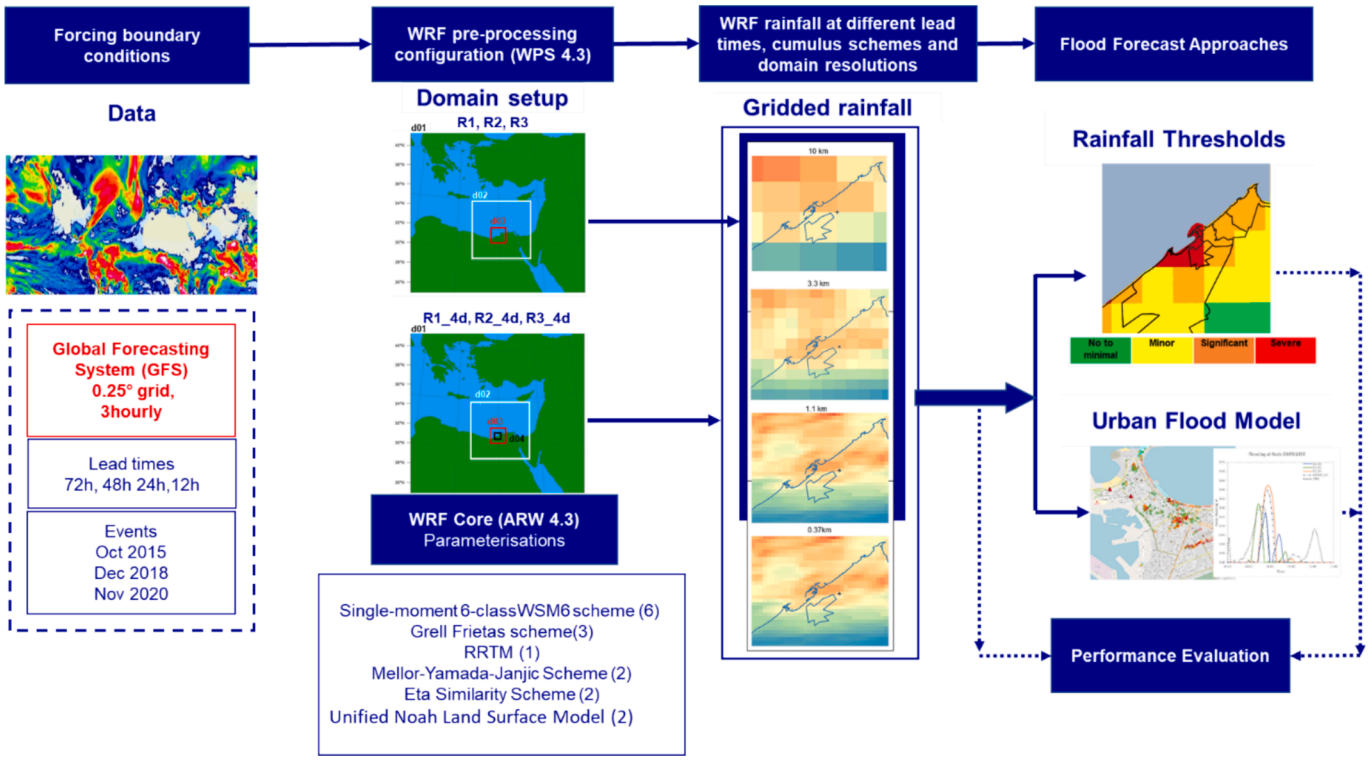


Fig. 4. Summary of data and methods used for the WRF model setup, rainfall thresholds and real-time simulation.

2012 to 2015 (using previous floods over a specific duration). Local drainage capacity (2-year return period) for Alexandria is reported as 26 mm/day for 2–3 hr events and flooding assumed to occur once exceeded (AASTMT and Egis BCEOM International, 2011). This knowledge of the system drainage capacity was combined with historical flood data derived from social media, archived newspapers, blogs, and eyewitness accounts, which have proven useful in assessing evidence of flood. Only one daily rainfall gauge was available for the study thus spatially varying thresholds were not considered. The thresholds were applied to the spatially varying rainfall forecasts to derive hazard classes over the entire governorate.

3.3. Urban flood model setup

A flood model based on the 1D MIKE Plus model (from the Danish Hydraulic Institute), capable of real-time FF, was developed to simulate flood inundation. Alexandria city is a very densely urbanized area. A land cover GIS dataset comprising buildings, roads, and green infrastructure was used to derive imperviousness for the rainfall-runoff model. The model was developed and calibrated for the Gomork

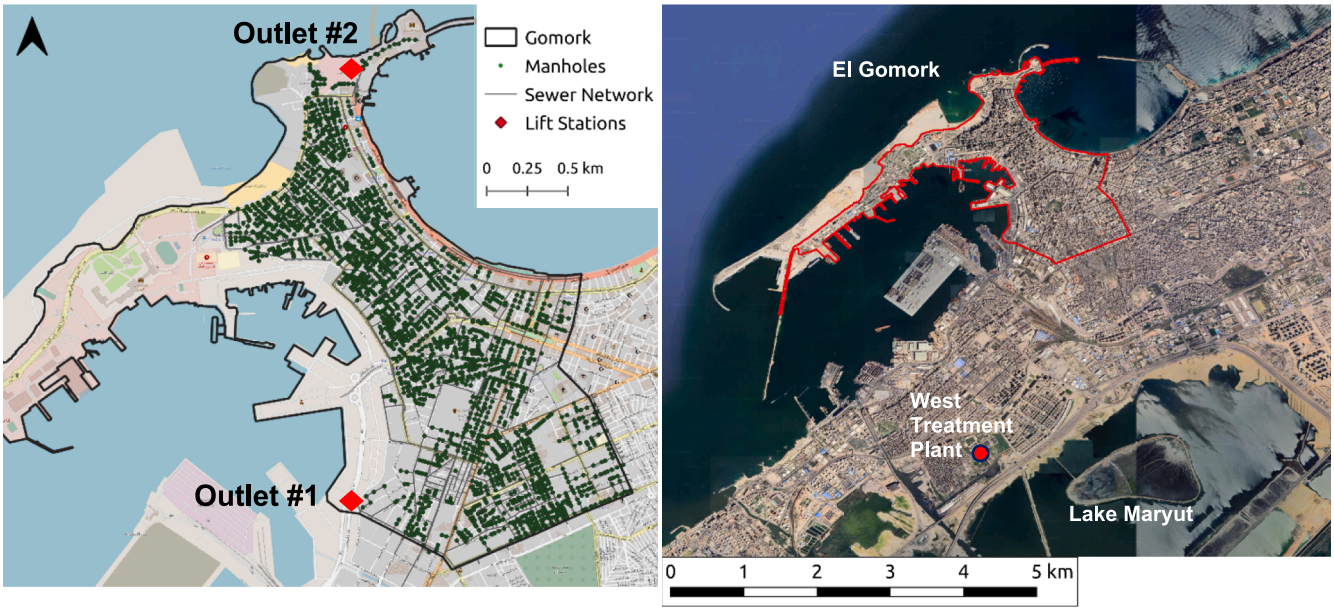


Fig. 5. Mike Urban plus drainage model set up for El Gomork district showing manholes, sewer pipe layout and location outlet lift stations (left) which pumps combined stormwater and sewage to the West Treatment Plan.

district (~5 km^2) (Mahmood, 2021) using available drainage network data.

In the real-time simulation, the hydrodynamic model was run with forecasted spatially varying gridded rainfall at 10 km, 3.3 km, 1.1 km and 0.37 km resolution to simulate flooding in Gomrok. The combined sewer and stormwater network model was used to simulate flows and water levels in urban storm drainage. It consists of a rainfall-runoff model which transforms rainfall time series to runoff hydrographs to be used as input into the hydraulic network model. The Gomork network comprises 3482 local and collector manholes connected by sewer pipes ranging from 300 to 2000 mm in diameter. Pipes flow via gravity to two sanitary lift stations; Outlet #1 (Kaytbay lift station) and Outlet #2 (El Meena). The combined sewage is then pumped from the lift stations and treated at the West Treatment plant before discharging into Lake Maryut.The model setup, location of the West Treatment Plant and Lake Maryut are shown in Fig. 5. Tides and wave action are not considered to be contributing to flooding in the model setup.

Outlet discharge, total volume and node flooding simulated using the WRF rainfall were evaluated against simulation results generated using MSWEP and IMERG and flood depths were compared with photos taken on the day of the flood. Rainstorms in Alexandria range from a duration of 1 to 9 h (AASTMT and Egis BCEOM International, 2011; Mahmood, 2021) whereas the MSWEP data has a temporal resolution of 3hs. To simulate extreme cases, the 3 h MSWEP rainfall was assumed to occur in a 1 h interval. IMERG rainfall, which has a half-hourly resolution was aggregated to hourly temporal resolution.

3.4. WRF performance evaluation

This section discusses the methods used to evaluate the performance of the WRF model. Given the large-scale nature of the simulated events, results were evaluated at different neighbourhood sizes to understand the forecasting skill in relation to flood forecasting applications. Neighbourhood locations were determined based on past knowledge of weather systems. In general, rainfall quickly reduces as it moves inland. The larger neighbourhoods 1000 km & 500 km size cover the phenomenon scale, whereas the smaller sizes give coverage along the coast (200 km), the governorate (100 km), the city (50 km) and specific districts within the city (20 km and 10 km) Fig. 6.

In the absence of reliable sub-daily gauge data, results were compared with the gridded precipitation dataset: Multi-Source Weighted-Ensemble Precipitation MSWEP V2, which merges several gauge, satellite, and reanalysis-based data (Beck et al., 2019) and GPM IMERG Final Precipitation Products V7(Huffman et al., 2019) henceforth referred to as IMERG or observed rainfall estimates. There is uncertainty associated with the use of gridded rainfall products, thus both products were compared with the observed rainfall gauge at the Nouzha airport (Table 3) and the forecasts were verified against the product which matched the observed rainfall best.

Both products have a spatial resolution of 0.1 degrees and highest temporal resolution of three hours and half-hour, respectively. Both datasets have been found suitable for detecting rainfall events in North Africa (Nashwan et al., 2019; Mekonnen et al., 2023).

The performance of the WRF model was evaluated on measures of comparison such as scatter plots, Coefficient of Variation and Pearson Correlation Coefficient (CC), Standard Deviation (SD) and presented in a Taylor diagram and the derived Centred Root Mean Square Error (CRSME) (Taylor, 2001). Enhancing the small-scale detail can decrease the precipitation forecast skill due to the double penalty of erroneous rainfall placement and timing, even though the representation of rainfall accumulation and intensity may look more realistic. To overcome this, the neighbourhood spatial verification Fraction skill score (FSS) (Roberts, 2008) was used. This method directly compares the fractional coverage of forecast and observed rainfall, which exceeds a specified threshold for a neighbourhood size. FSS is calculated for each neighbourhood size (Fig. 4) using Eq. (1) where F_i and O_i are the forecast rainfall fraction and observed rainfall fraction exceeding specified thresholds, respectively and N is the number of spatial grids in a neighbourhood size. An FSS of 1 is perfect, an FSS above 0.5 is considered useful, and an FSS of 0 is not a skill.

Table 3
Comparison of the 72hr accumulated rainfall at Nouzha Airport gauge rainfall with MSWEP and IMERG precipitation products.

	Nouzha Airport	MSWEP	GPM
	m m/3d ays		
Oct 25th, 2015	32.1	38.13	21.97
Dec 5th, 2018	56.13	33.13	59.33
Nov 20th, 2020	34.03	15.31	32.13

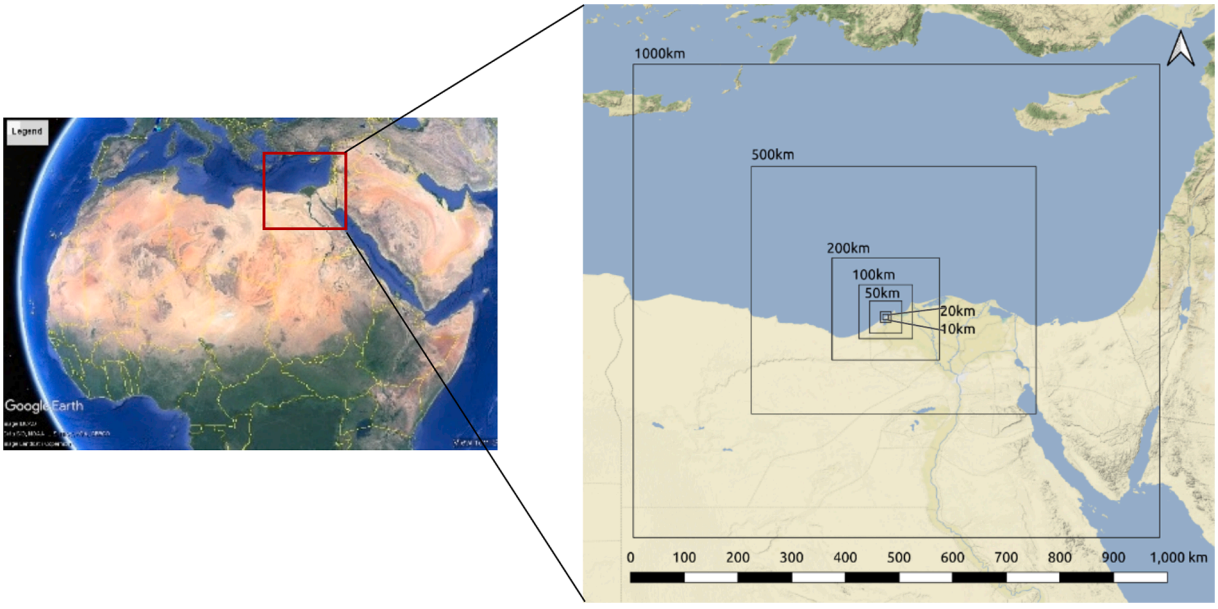


Fig. 6. Visual representation of the neighbourhood spatial sizes used for FSS calculations.

$$\boldsymbol{FSS} = 1 - \frac{\frac{1}{N}\sum_{i=i}^{N}(F_i - O_i)^2}{\frac{1}{N}\left(\sum_{i=1}^{N}(F_i)^2 + \sum_{i=1}^{N}(O_i)^2\right)} \quad (1)$$

In Eq (1), FSS was computed and evaluated for 24 h accumulations for Domain1 (10 km), Domain2 (3.3 km) and Domain3 (1.11 km) at the, 72 h, 24 h and 12 h lead time for both runs with a convective scheme and no convective scheme in respective domains (Table1). All forecast resolutions were aggregated to a 0.1° grid to match observed rainfall estimates.

Rainfall Threshold Performance Indices.

For the rainfall thresholds, categorical scoring was used to evaluate if the warning class was consistent with the observed event category. Categorical descriptions of a hit, miss and or false alarm are presented below. The Probability of Detection (POD), False Alarm Ratio (FAR) and Critical Success Index (CSI) or threat score takes both false alarms and missed events into consideration and is sensitive to hits. The metrics were evaluated for the 10 km, 1,1 km and 0,37 km spatial resolutions for all runs.

$$\text{Hit Rate (Probability of Detection)} = \frac{Hit}{Hit + Miss} \quad (2)$$

$$\text{False Alarm Ratio (FAR)} = \frac{FalseAlarm}{Hit + FalseAlarm} \quad (3)$$

$$\text{Threat Score/ Critical Success Index(CSI)} = \frac{Hit}{Hit + FalseAlarm + Miss} \quad (4)$$

Urban Flood model Metrics.

The Normalised Root Mean Square Error (NRMSE) and the Kling-Gupta Efficiency score (KGE) were used to evaluate the performance of the 1D flood model. NRMSE measures the average difference between a statistical model's predicted values and the observed values where a RMSE of 0 is considered a good score. The equation NRMSE is shown as equation (5), where P_i is the model-predicted value, O_i is the observed value, $\bar{O}$ is the mean of the observed value, and N is the number of observations. KGE is a single combined expression of correlation, variability and bias into one score and represents a goodness of fit score. A KGE of 1 is considered a perfect score. The metrics were evaluated for the 10 km, 1,1k m and 0,37 km spatial resolutions for all runs for the outlet discharge and node flooding at a known flood location.

$$NRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{n}(P_i - O_i)^2}}{\overline{O}} \quad (5)$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (6)$$

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2} \quad (7)$$

where r is the Pearson correlation between observations and simulations, α is the variability ratio and β is the bias ratio. This is further expressed in equation (7) where σ*obs* is the standard deviation in observations, σ*sim* the standard deviation in simulations, μ*sim* the simulation mean, and μ*obs* the observation mean.

4. Results

Results are presented and discussed for WRF rainfall at different spatial resolutions 10 km (D1), 1.1 km (D3) and 0.37 km (D4) for runs with different treatments of convection. Results are presented for the three events Oct 25th 2015, Dec 5th 2018 and Nov 20th, 2020. The analysis first examined how results vary with lead time and neighbourhood size, followed by the influence of the horizontal grid resolution and cumulus configuration in each domain. The WRF rainfall is then evaluated using rainfall thresholds for forecasting hazard classes and for use in a real-time model simulation for forecasting discharge and node flood depths.

4.1. WRF rainfall forecast analysis

4.1.1. Leadtime and neighbourhood size analysis

Fig. 7 shows how the rainfall event on October 25th 2015 is forecasted at the 1000 km and 50 km neighbourhood sizes for the 10 km grid resolution for Run2 (R2). The 50 km neighbourhood gives a closer snapshot of the city scale. While the event's occurrence is captured at all lead times, it shows higher forecasted values at 72 h and 24 h and lower accumulation and extent at 12 h. Generally, the forecasts show similar extent and areas of high accumulation compared to MSWEP whereas the IMERG product showed significantly higher rainfall values offshore but rainfall reduced over the land, less than MSWEP.

Fig. 8 shows scatterplots of forecasted and MSWEP area average rainfall for the event on Oct 25th, 2015, and Dec 5th 2018 for the 72 h, 24 h and 12 h lead time, 10 km domain resolution (D1) and different neighbourhood sizes for R1, R2 and R3. The 12 h lead times for all runs performed poorly, heavily under-forecasting rainfall for neighbourhood sizes smaller than 200 km. Similar trends are presented in Fig. 7. Given the similarity in the trend for lead times in all runs, it can be assumed that the poorer performance was caused by the initial and lateral boundary conditions used to initialise WRF.

Compared to the 12 h lead time, forecast results showed high skill at the 72 h and best overall performance at the 24 h lead time at all sizes, particularly at the smaller neighbourhood sizes. Noticeably, the rainfall was under-forecasted at the 10 km scale for the 72 h lead time for R1 and R3, but it performed well for R2. R3 showed more similarity for the 24 h lead time at the 10 km neighbourhood scale. Results for the event can be found in the **Appendix**.

For the spatial verification metric FSS, there were minor differences between the domain spatial resolutions and runs, but results varied with lead time, neighbourhood size and thresholds. Consistent with the previous analysis, the forecasts performed best at the 24 h and 72 h lead time for all runs and all thresholds except for the 32 mm threshold which shows no skill at the 10 km neighbourhood for the 12 h lead time. This suggests the WRF model is over-forecasting in some grids of these neighbourhood sizes. Results presented for R3 in Fig. 9 also show that despite performing poorly for the 20 mm and 32 mm thresholds for smaller neighbourhood sizes, the FSS for 12 h lead time still showed skill at the 500 km and 1000 km. The WRF model was able to show a threat was imminent and could confirm the persistency of a threat, which most often may be good enough information for early warning and anticipatory actions for reducing disaster risk. However, the forecasts were not useful (<0.5) at the 10 km neighbourhood size at the 72 h lead time, which would be more relevant to urban scale modelling. Given the computational requirements of running models at higher resolutions, running the models at a coarser resolution and at longer lead times is preferential. This would allow protocols to be put in place and the model can run at higher resolutions for shorter lead times.

4.1.2. Influence of cumulus configuration and horizontal grid resolution

Given the favourable results at the 24 h lead time, the results are presented only for this lead time for all model runs, including downscaling to a 4th domain (0.37 km) which aims to meet the requirement for higher-resolution rainfall for urban flood forecasting. For R1_D4, R2_D4 and R3_D4 run (see Table 1), the runs use the same configuration as the corresponding R1, R2 and R3 but the cumulus scheme was turned off in the 4th domain. The results are presented for the 50 km neighbourhood size which showed useful for the 24 h lead time will be most pertinent to city administration. Fig. 10 shows the spatial rainfall variability for each run across the 50 km neighbourhood size for the Oct 25th event. In general, the 24 h accumulated rainfall across grid resolutions varies by about 5–20 mm. Notably, there is a tendency for the finer grids to show areas of more intense rainfall and higher rainfall

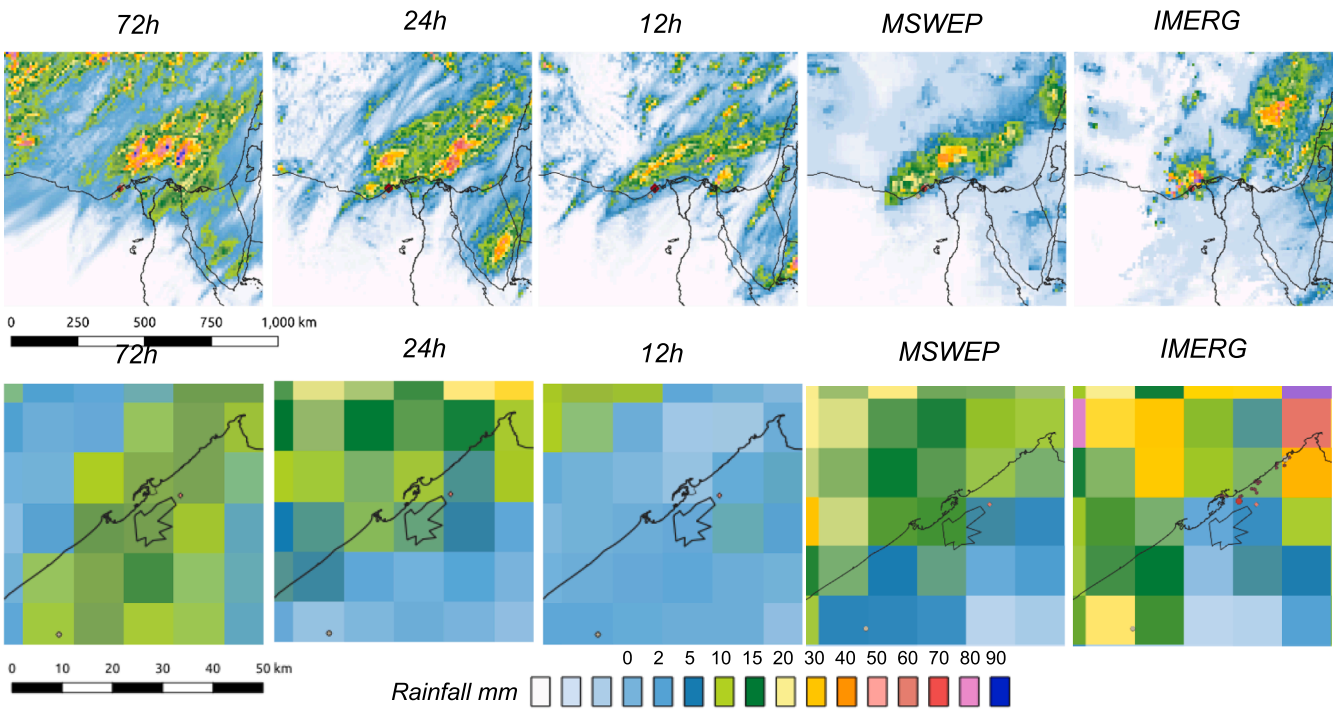


Fig. 7. Images show how the weather phenomenon is predicted for 24 h accumulated rainfall from 21:00 Oct 24th to 21:00 Oct 25th at 72 h, 24 h and 12 h lead time for the 1000 km (top) and 50 km (bottom) neighbourhood at the 10 km resolution and R2. MSWEP and IMERG are shown for the same period.

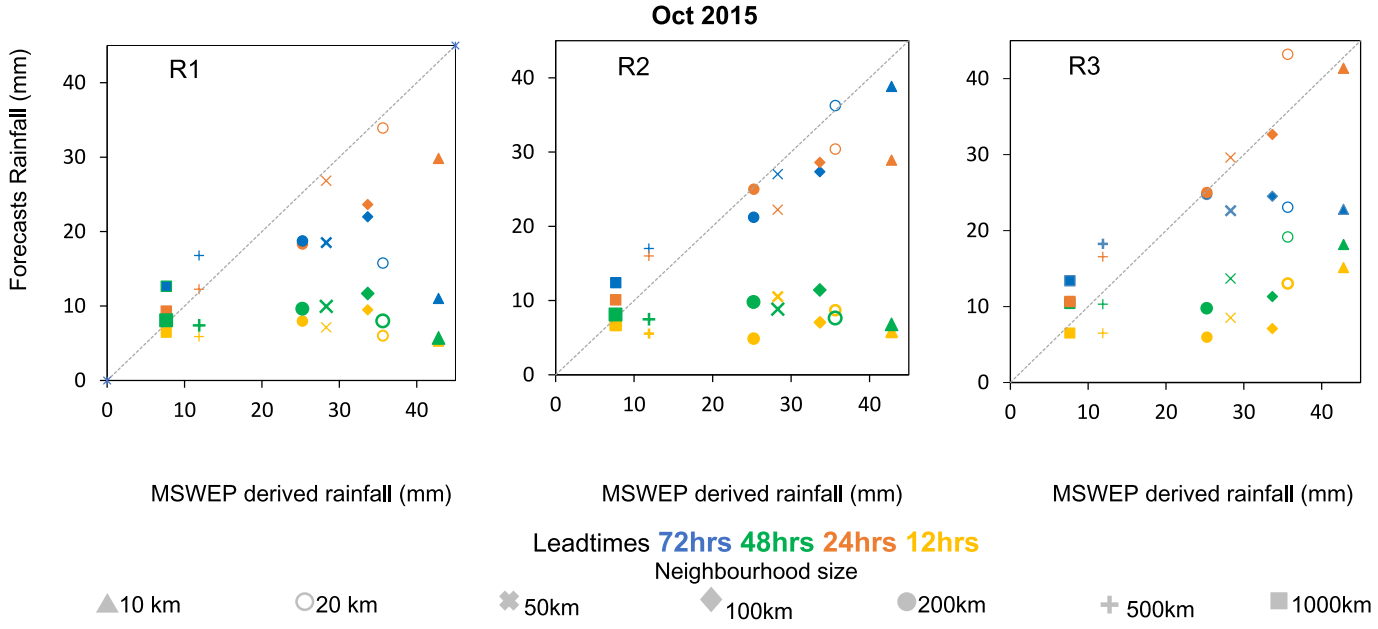


Fig. 8. Comparison of 24 h rainfall accumulation from Oct 24th 21:00 to Oct 25th 21:00 showing the accumulated area averages for the 10 km resolution (D1) for different neighbourhoods at lead times of 12, 24,48 and 72 h for R1, R2 & R3.

accumulation than D1 of the same run.

It was found that all runs correctly predicted rainfall occurrence, but there was some variability in rainfall locations, magnitudes, and intensities across runs and domains. For the 50 km neighbourhood size, R1 had the lowest coefficient of variation, 30 % and R3 had the closest coefficient of variation 40 % compared to MSWEP 43 %. This suggests grid points are more spread out across the rainfall mean than observations with smaller variability across grid points. Moreover, in Fig. 10, rainfall accumulation tends to increase when the cumulus scheme is turned on in the smaller domains for 3 domain runs. Compared to R1 and R2, R3 shows areas of higher rainfall accumulation in the magnitude of 70 mm off the coast and parts of the city, particularly on the eastern side of the city. There are also noticeable differences between the horizontal resolutions. For both R1 & R3, D1(10 km) averages the rainfall over larger areas detecting lower accumulated rainfall in the city compared to D3 which shows greater similarity. However, D3 captures greater rainfall variability in detail. In general, there appear to be fewer areas of extreme rainfall over the governorate in R1_D4 and R3_D4,

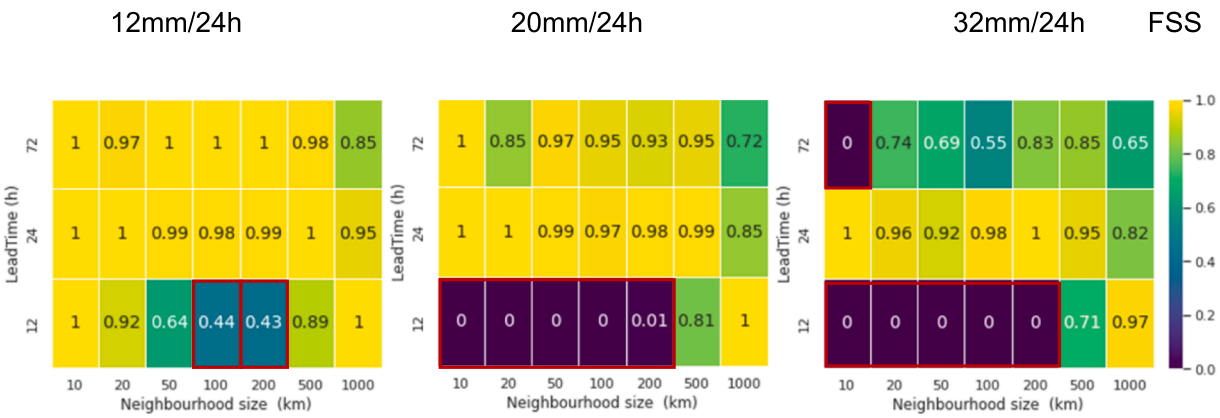


Fig. 9. Fraction Skill Score (FSS) for D1 (10 km) using different thresholds and neighbourhood sizes for run R3. Neighbourhood sizes and lead times without a useful skill are highlighted in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

whereas R2_D4 shows areas of increased rainfall compared to R2.

The differences in the runs are further displayed using a Taylor Diagram (Taylor, 2001) (Fig. 11) which compares and summarises the standard deviation (SD), and correlation coefficient (CC) of all runs and domains with the SD and CC of MSWEP or IMERG, which is referred to here as the reference data. Forecasts were re-gridded to match the observed resolution. In Fig. 11, overall the runs vary with events. R2 D3 agree best with observed rainfall values since these runs are closest to the observed ref (dashed) line and correlation coefficients 0.65–0.95 and CRMSE ranging from 8-11 mm/24 h. the general pattern well with high correlation values (>0.8) but the spread of forecast values is larger than the observed data (above the dashed line). This is primarily because these runs produce areas of extreme rainfall offshore that skew the variability in the data. There is less agreement on the other runs have correlation values between 0 and 0.8. This indicates that while it produces a range of values similar to those observed, the forecast for these runs does not match the precise pattern or distribution of the observed data, leading to a low correlation and suggesting a bias in the forecast or the observation. Domain 4 (D4) (0.37 km) for cumulus configurations runs for the three events showed lower correlation coefficients and higher CRMSEs for all runs for all events. The results further highlight that D4 runs performed poorly compared to the D3 and D1 runs for all cumulus configurations.

4.1.3. Rainfall temporal distribution over the Gomork area

The previous analysis focused on how rainfall varies spatially and with different neighbourhood scales. Further analysis examined how the temporal distribution varied within 10 km^2, corresponding to the area used in the MIKE + model and immediate surrounding areas. This gives insight into the way node flooding might vary with intensity. Results were compared with 3-hourly MSWEP (MSWEP_3h), hourly MSWEP (MSWEP_h) and hourly IMERG (IMERG). Fig. 12 presents the rainfall time series for all runs. MSWEP_h was constructed as a worst-case scenario by assuming the 3 h accumulated rainfall fell. For the Oct 25th event, the worst-case scenario assumed the 3 h accumulation fell in the last hour of a 3 h accumulation and the first hour of the subsequent 3 h accumulation.

Rainfall is represented as either a single peak or a bi-modal distribution with two peaks. R3 and R2 show primarily single peaks similar to MSWEP_h which assumes a single peak distribution but R3 and R2 both have the highest peak intensities (above 20 mm/h) for D3 (1.1 km). When downscaled using a 4th domain to 0.37 km with no cumulus scheme (Fig. 12), the rainfall distribution is bimodal with rainfall peaks less than 16 mm. Similar results were analysed for the Dec 2018 event, where the 4th domain runs all showed lower rainfall for all domains, whereas for the Nov 2020 event, the D3 and D4 domains both showed high rainfall peak timing. Although the results show similar peak intensities in Dec 2018 and Nov 2020 notably, there is a mismatch in the peak of about 3-4hr when compared with IMERG where the WRF simulated rainfall matching the timing of the MSWEP but not the magnitude (See **Appendix**).

Overall, the three domain runs all matched the accumulated observed rainfall, whereas the D4 runs either underestimated or overestimated in the case of Nov 2020. These results suggest that while the inner domains exhibit higher rainfall accumulations than the parent domain, the rainfall peak does not always increase with the higher resolution, as shown with D4. This could be explained since having no cumulus scheme in the smaller domains tends to decrease rainfall in those domains. Given two-way nesting was used, feedback between domains reduces the rainfall in the outer domains and perhaps influences the distribution in the outer domain. Although the 4th domain (except for R1_4d), better matches the peak intensities of the satellite-derived rainfall forecasts, there is a lower total rainfall forecast compared to MSWEP and IMERG.

4.2. Rainfall threshold analysis

Hazards are classified as “No to minimal flooding”, “Minor Flooding”, “Significant flooding” and “Severe flooding” based on 24 h accumulated rainfall (Table 2). Analysis of thresholds showed that results varied with the treatment of convection for the different runs. Classifications are shown for different districts in Alexandria for MSWEP and IMERG (Fig. 13). For the event on 25th October 2015, the forecast hazard classes, derived based on Table 2, varied from a significant to severe flood event along the coastal areas (Fig. 12). While MSWEP and IMERG indicated a hazard class “Significant flooding” in Montazah, Sharaq and Wasat districts, Gomork and parts of Almeria were predicted to be more extreme. There were minimal differences when comparing the 3.3 km, 1.1 km and 0.37 km grid sizes. Compared to the 10 km grid the smaller resolutions detected more spatial variability in the hazard classification and areas of higher rainfall. The 10 km grid size tends to average out the rainfall over the grid. Fig. 14 shows how the hazard warning class varied for the 10 km and 1.1 km for the 3-domain run and 0.37 km for the 4-domain run for the different runs and cumulus configurations for a 24 h accumulation for different districts in the city. Most of the runs were able to predict this extreme class except R1_4d and R3_4d, which predicted a less severe but significant class for Gomork and neighbouring districts. R1, R3 and R2_4d over-predicted a more severe event in Montazah and neighbouring areas. Some runs showed variable classes compared to MSWEP and IMERG, but all runs correctly predicted a reduction in the total rainfall inland and different runs were able to distinguish different hazard classes at districts and subdistrict levels, which is useful when location-specific measures need to be taken. This will aid decision-makers in prioritising, coordinating and allocating

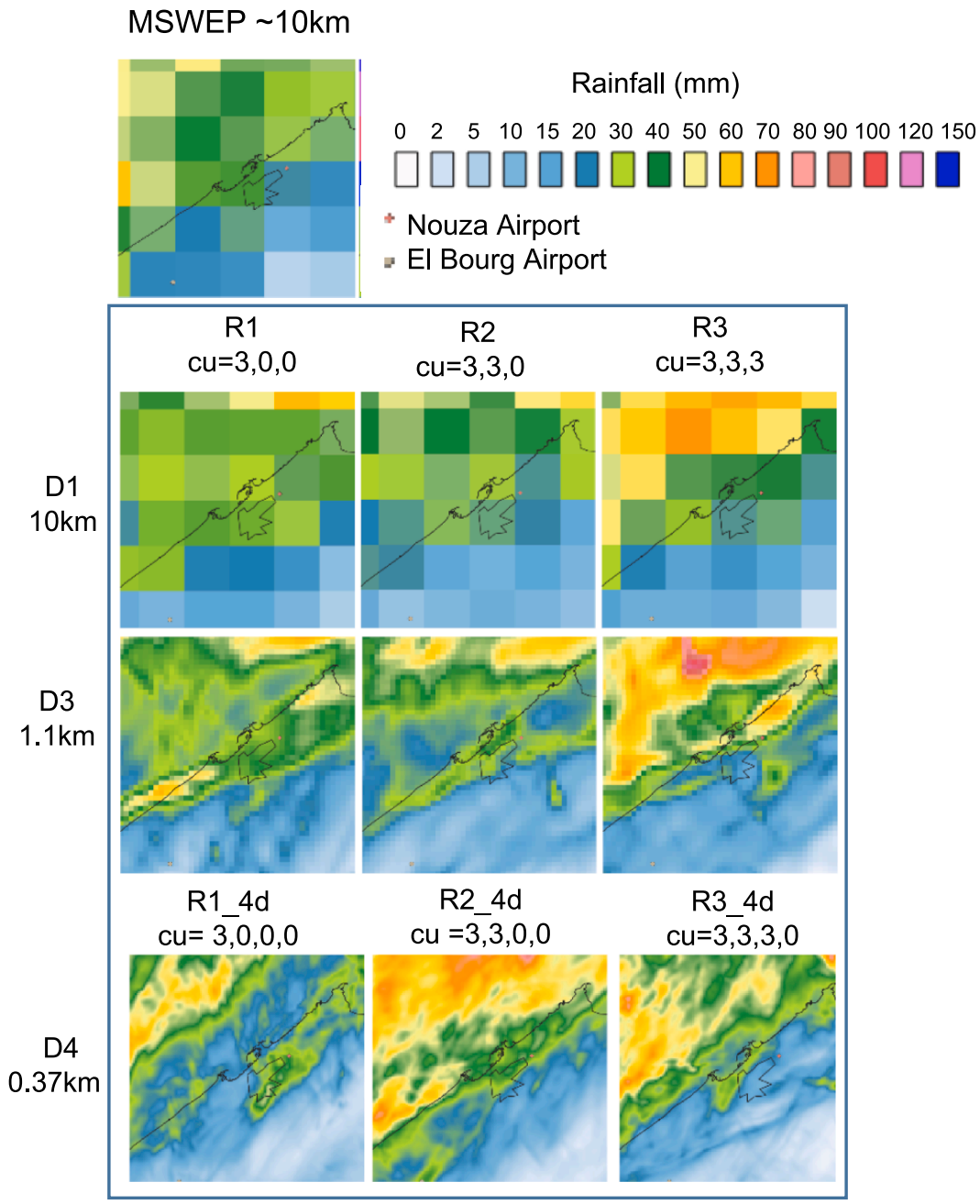


Fig. 10. Rainfall accumulation from 21:00 UTC on Oct 24th to 21:00 UTC on Oct 25th 2015 for MSWEP, WRF resolutions and different cumulus treatments cu = 3 represents the Grell Freitas Cu scheme, 0 indicates no cumulus scheme in the respective domains.

resources to areas at most risk, combined with prior knowledge of hot spot areas. This could be achieved even without building a stormwater model of the city.

Categorical scoring was used to evaluate if the warning classes in Table 2 were consistent with the observed event category. A POD and CSI of 1 and a FAR of 0 are considered a perfect forecast whereas a POD and CSI above 0.75 considered useful. The results (Table 4) show that the D3 (1.1 km) 3 domain runs consistently had higher POD and CSI scores, meaning these runs were better at detecting an event and lower false alarms. Both runs with the 4th domain R1_D4 (10 km) and D4_4d (0.37 km) achieved lower scores with all the runs with all the thresholds. Within the 3domain runs, results vary with thresholds and cumulus configuration. In general, higher FAR scores were found for rainfall accumulations over 20 and 32 mm. However, either Run2 or R3 performed better when the cumulus is turned off and on in the smallest domain. All runs had good performance above the 50 mm threshold rainfall which indicates all forecasts had agreement in not predicting "very extreme" rainfall accumulations.

4.3. Coupled WRF and urban flood forecasting

The MIKE plus model was used to evaluate peak discharge and volume at the outlet of the El Gomork district. The model was run with gridded forecasts rainfall at different resolutions and runs corresponding to different cumulus configurations: D1(10 km), D2(3.3 km),D3 1.1 km and D4(0.37 km), with a 24-hour lead time. In the absence of measured flow data, results were verified against model simulation runs with satellite-derived IMERG and MSWEP 1 h derived rainfall intervals and historical knowledge of floods. The lift station pump at outlet #1 has a design capacity of 0.9 m^3/s therefore, it is assumed that flooding will

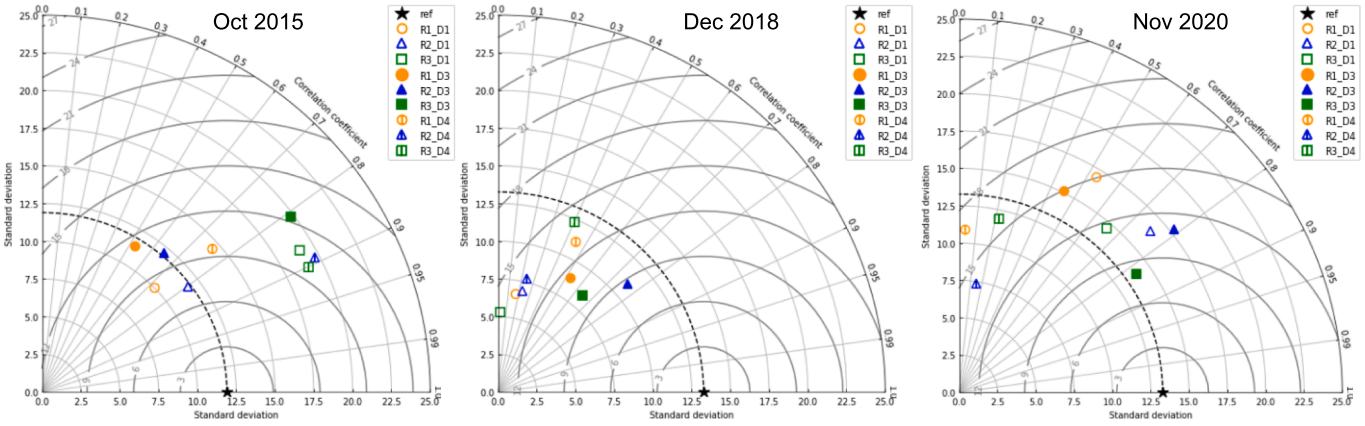


Fig. 11. Taylor diagram for all runs and three domains/resolutions (D1 10 km, D3-1.1 km and D4 0.37 km) at the 50 km neighbourhood for all three events. R1 (3,0,0), R3 (3,3,3), R1_D4(3,0,0,0), R2_D4(3,3,0,0) and R3_D4(3,3,3,0) for three events.

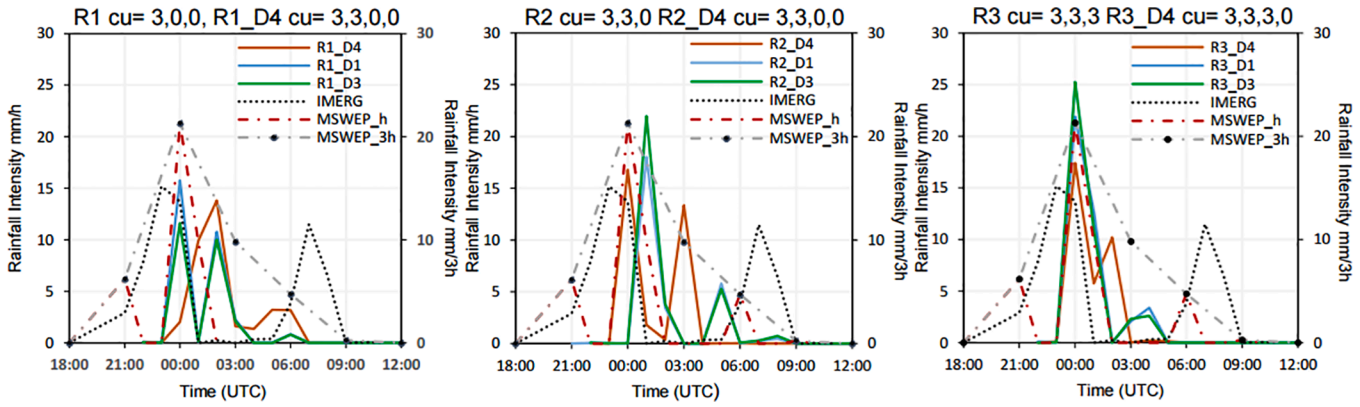


Fig. 12. Rainfall time series for all runs D1(10 km), D3(1 km) and D4 (0.37 km) from 18:00 Oct 24th to 12:00 Oct 25th 2015 at a known flood location. MSWEP is shown for 3 h accumulations and an extreme case 1 h accumulations. R1, R2& R3 correspond to runs with 3 domains and R1_D4, R2_D4 and R3_D4 correspond to 4 domains runs. Cu = 3 represents the Grell Freitas Cu scheme, and 0 indicates no cumulus scheme in the respective domains. The left axis mm/h and the right axis shows mm/3 h corresponding to MSWEP_3h.

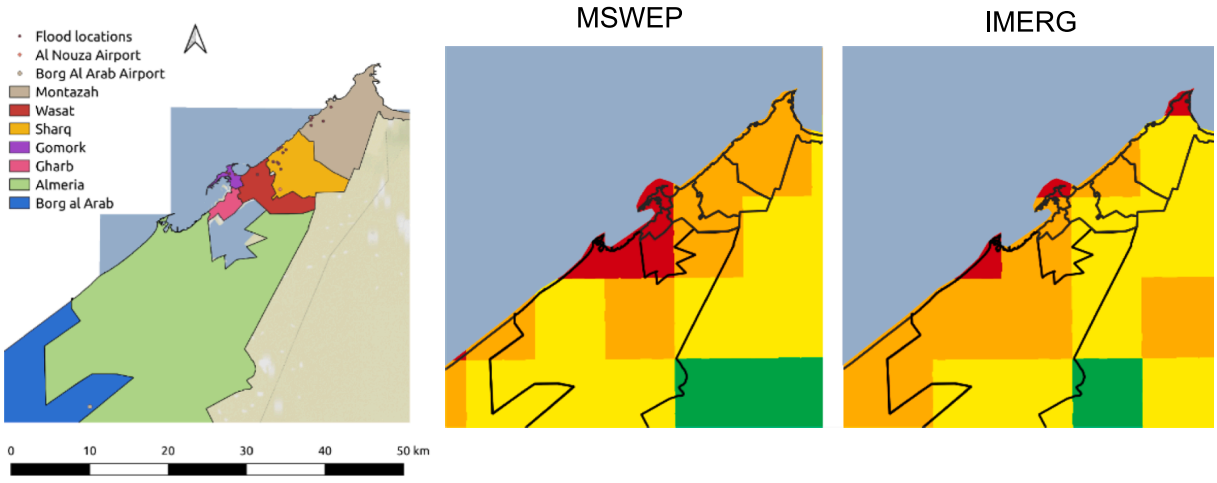


Fig. 13. Left: Alexandria Districts; the colours have been arbitrarily chosen to display districts. Right: Hazard classification for Alexandria districts for the 25th October 2015 flood event for 24 h accumulation for the Oct 25th event based on 3-hourly MSWEP and hourly IMERG data. The colour codes signifying hazard classes are also shown in Table 2.

occur once this flow is exceeded. Peak discharge reached 30mins after peak rainfall.

4.3.1. *Discharge and volume at outlet*

The results for Oct 2015 in Fig. 15 show that forecast rainfall corresponding to R3 led to the highest simulated discharge for all

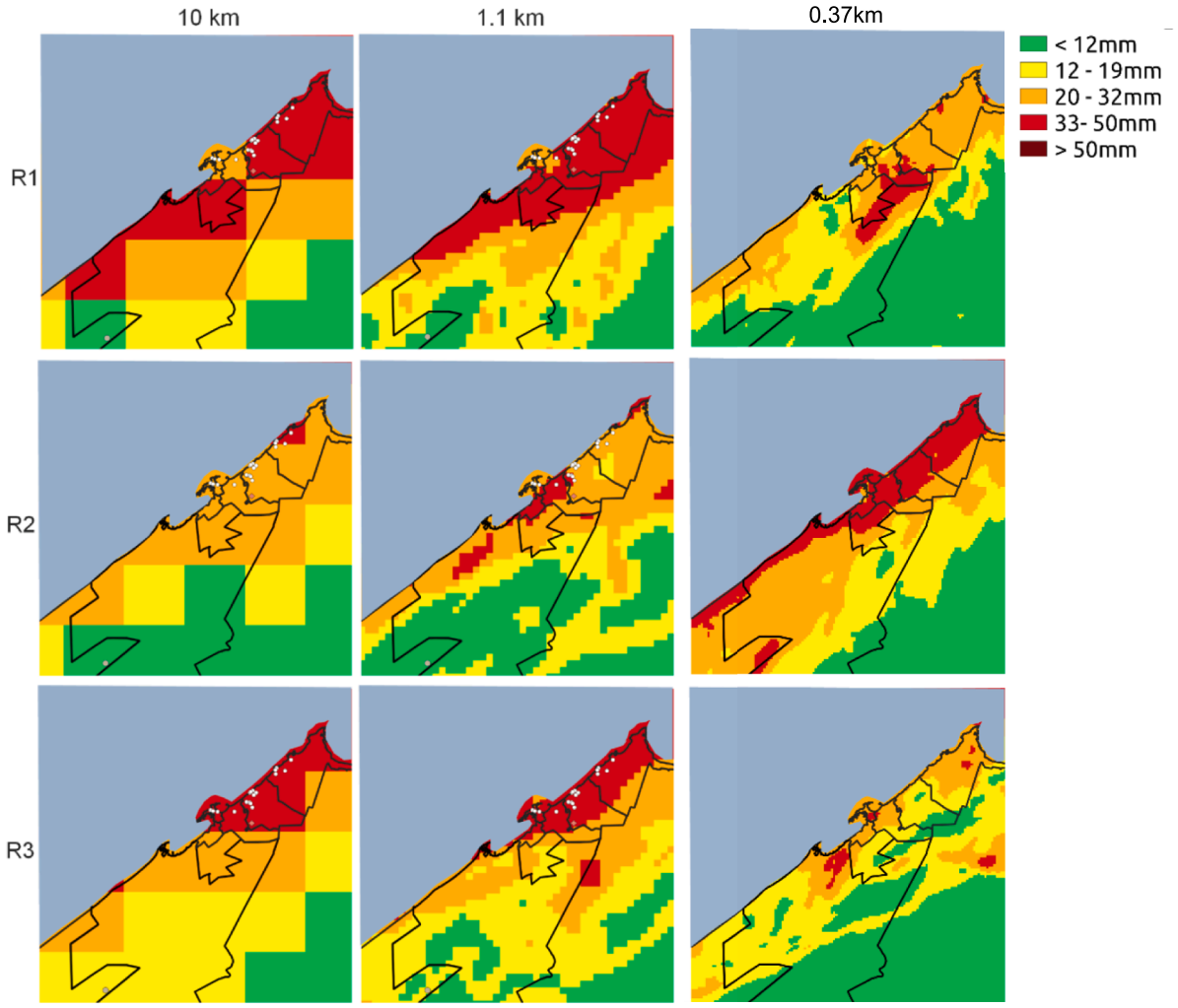


Fig. 14. Forecast of hazard classes with a lead time of 24 h for the Oct 25th, 2015 flood event for different runs based on the 24 h rainfall accumulations. Left:(10 km) & Middle 1.1 km grids for run downscaled to 3 domains. Right 0.37 km grids downscaled to 4 domains.

Table 4
Average performance metrics; Probability of detection (POD). False Alarm Ratio (FAR) and Critical Score Index (CSI) for) all runs for all events for the 50 km neighbourhood scale. Results are presented for the inner most domains 1.1 km and 0.37 km respectively.

Avg for all events	**0.1mm**			**12mm**			**20mm**			**32mm**			**50mm**		
	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3
								POD							
D1	1.00	1.00	0.95	0.73	0.68	0.72	0.70	0.57	0.66	0.63	0.64	0.72	0.91	0.93	0.81
D3	1.00	1.00	0.96	0.76	0.71	0.76	0.68	0.55	0.73	0.63	0.67	0.73	0.87	0.93	0.91
D4	0.96	0.96	0.97	0.65	0.69	0.68	0.49	0.51	0.49	0.61	0.56	0.57	0.92	0.88	0.88
								FAR							
D1	0.00	0.00	0.05	0.29	0.35	0.26	0.28	0.45	0.30	0.33	0.36	0.31	0.11	0.07	0.08
D3	0.00	0.00	0.04	0.25	0.33	0.27	0.34	0.43	0.31	0.41	0.33	0.24	0.10	0.07	0.20
D4	0.04	0.04	0.03	0.33	0.28	0.31	0.51	0.48	0.49	0.43	0.41	0.45	0.08	0.12	0.16
								CSI							
D1	1.00	1.00	0.91	0.53	0.61	0.56	0.41	0.49	0.38	0.49	0.49	0.53	0.87	0.83	0.87
D3	1.00	1.00	0.93	0.64	0.56	0.62	0.53	0.61	0.55	0.43	0.51	0.61	0.82	0.88	0.68
D4	0.93	0.93	0.95	0.54	0.61	0.57	0.35	0.37	0.36	0.41	0.43	0.39	0.86	0.79	0.73

resolutions and showed higher peaks compared to MSWEP and IMERG. The discharge for each run followed a similar pattern of the rainfall distribution. The discharge at the outlet simulated with the rainfall forecast using R2 and R3 for the 10 km (D1) & 1.1 km(D3) showed peak discharges similar to that with MSWEP_h and IMERG (~1.5 m^3/s). The simulated discharge for R1 forecasts showed a different pattern with two lower peaks. Simulations using rainfall forecasts based on R3 & R2 for D3 and D1_gave peak discharges similar to MSWEP. The runs also gave the highest KGE scores (0.4–0.78) and the lowest NRMSE of 0.43 when compared with the other runs and D4 (Table 5). The exception being Dec

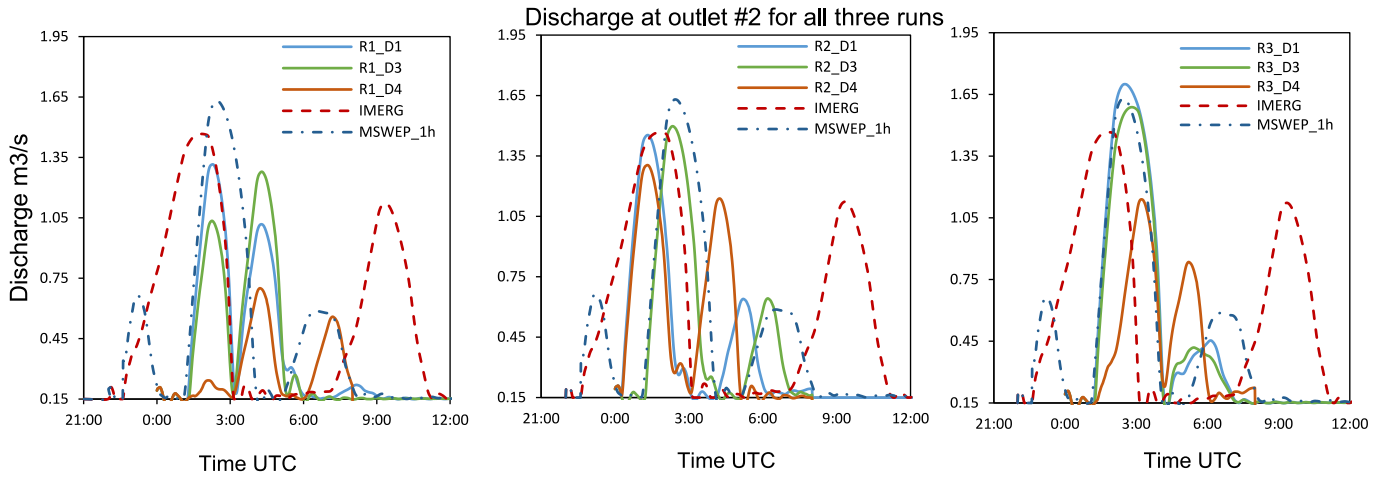


Fig. 15. Simulated Discharge at outlet #2 on Oct 25th, 2015 for R1, R2 and R3 for D1 (10 km), D3 (1.1 km) and (D4) 0,37 km which corresponds to runs with 4 domain and the cumulus scheme is turned off in the 4th domain.

Table 5
NRMSE and KGE values for simulated discharge at outlet #2 for the three events. D1_R3 (10 km) performed the best which uses the adaptive cumulus scheme in the smallest domain whereas the highest error was shown the 0.37 km domain when the cumulus scheme is turned off. The best score for each run per event is highlighted.

					NRMSE				
	R1_D1	R1_D3	R1_D4	R2_D1	R2_D3	R2_D4	R3_D1	R3_D3	R3_D4
2015	0.80	1.10	1.09	1.12	0.49	1.15	0.42	0.43	0.83
2018	0.97	0.97	1.04	1.01	1.01	0.99	0.98	0.98	1.05
2020	0.83	0.95	1.06	0.96	0.96	0.97	0.93	0.64	1.07
					KGE				
	R1_D1	R1_D3	R1_D4	R2_D1	R2_D3	R2_D4	R3_D1	R3_D3	R3_D4
2015	0.43	−0.37	−0.35	−0.02	0.68	−0.08	0.77	0.78	0.31
2018	0.18	0.11	0.06	0.23	0.40	−0.09	0.22	0.02	0.09
2020	0.22	0.29	−0.04	0.21	0.24	0.27	0.20	0.66	−0.26

2018 R3_D3 which gave had a KGE of 0.02 and NRMSE of 0.98 Fig. 15 also shows discharge simulations using rainfall forecasts based on 10 km, 1.1 km (D3) and 0.37 km (D4) resolutions, the latter which are the innermost domains of WRF runs. All runs underestimated the discharge with the rainfall forecasts using the 0.37 km resolution with NRMSE ranging from 0.83 to1.09. Similar results were found for the Oct 2018 and Nov 2020 events which gave lower NRMSE values and higher KGE for 1 km domain and higher NRMSE values and lower KGE scores for the 0.37 km runs. However, except for Oct 2015, all the runs and domains resulted in high NRMSE values, which means the models still have challenges predicting the timing of the rainfall but the higher KGE indicates some agreement in the correlations and the spread. The NRMSE and KGE values for the three events in shown in Table 5.

The simulation using rainfall forecasts based on the R3_D3 (1.1 km) grid produced the highest total volumes at the outlet. Simulated discharge based on R3 forecasts gave values closest to MSWEP_h, however, simulated discharge with R1_D4 and R3_D4 forecasts gave considerably lower volumes compared to MSWEP_h and IMERG-based simulations (Table 6). Although IMERG resulted in a peak discharge similar to MSWEP_h, the total runoff volumes were higher compared to that with MSWEP_h simulated at the outlet because of the higher total accumulated rainfall.

Table 6
Total volume of water discharged at the outlet for the Oct 2015 event for the different WRF runs and resolutions.

Input rainfall /resolutions	IMERG	MSWEP_h	WRF R1	WRF R2	WRF R3
	Total Volume at Outlet m^3				
10 km resolution	21,544	14,234			
10 km			10635	10103	14171
1.1 km resolution (D3)			10,476	10741	12374
0.37 km resolution (D4)			4955	10929	9206

4.3.2. *Node flooding*

The flood results above focused on the simulation of the entire system for El Gomork and found that the R2& R3 produced good simulations of the peak of flooding for the 10 km and 1.1 km. This is also observed when analyzing node flooding at a known flood location in the city. During the October 25, 2015 event, flood depths of 0.15 – 0.3 m (Fig. 16) were reported at this location and in the vicinity. Fig. 17 shows the simulated node flooding at a selected node with the rainfall forecasts for the 10 km, 1.1 km and 0.37 km with the hourly MSWEP and IMERG rainfall. The rainfall intensity and spatial distribution of grid rainfall used for the 10 km, 1.1 km and 0.37 km simulations are shown in Fig. 18. For both the R2 and R3, the 1.1 km grid shows a flood depth of ~ 0.4 m, which is similar to the depth simulated using MSWEP but approximately 0.1 m above the flood depth simulated using IMERG derived rainfall and the reported 0.3 m flood depth. However, compared to the discharge for the system which showed similar peak values and distributions, for the 10 km and 1.1 km, higher node flooding was simulated at the manhole. This is because of the 1.1 km runs showed higher rainfall values and greater variability in the rainfall which is not detected in the larger 10 km grid as shown in Fig. 18. Although the timing of the flood was well predicted for the Oct 2015 event, there was a delay in the timing of the flood was also observed in the Oct 2018 and Nov 2020 event.

Fig. 19 shows a comparison of node flooding greater than or equal to 0.3 m for R3 and R3_4d runs for the 1.1 km grid. R3_D3 shows a higher number of nodes (101 nodes) compared to MSWEP 83 flooded nodes, R3_D1 118 nodes, R3_D4 shows 32 nodes flooded above 0.3 m,

Fig. 16. Known flooding locations in El Gomork along the Corniche Road in Alexandria. On October 25, 2015, this location experienced flooding depths ~ 0.15—0.3 m.

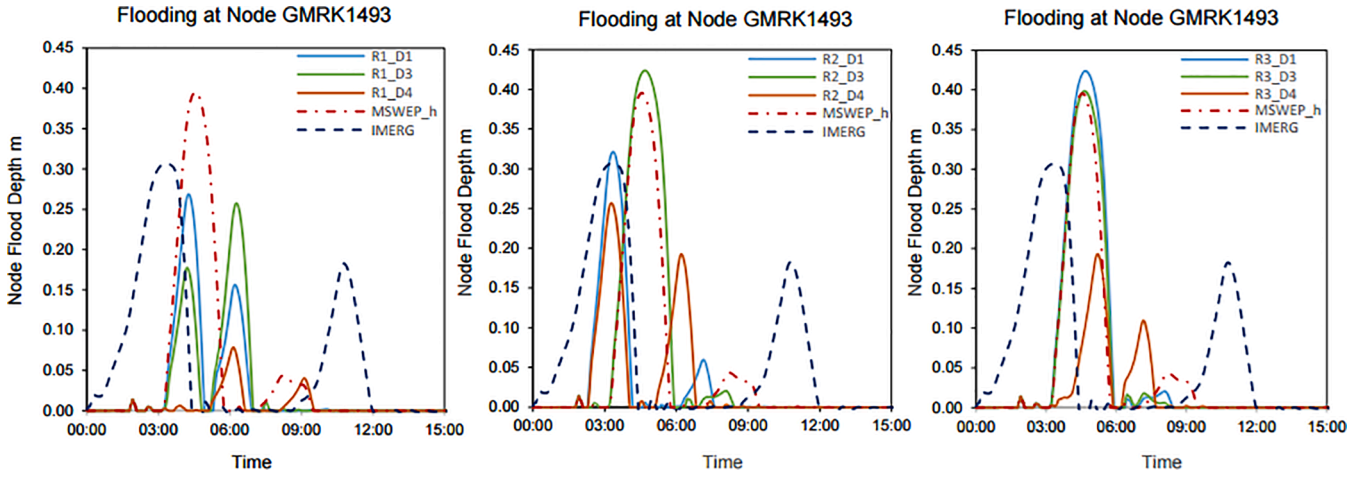


Fig. 17. Node flooding depths (m) at known flood location in Gomork on Oct 25th, 2015. For all runs at D1(10 km), D3(1.1.k m). and 0.37 km- 4 domains (orange). The horizontal dotted line indicates the observed flood depth at 0.3 *m.* (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

approximately 1/3 of nodes flooded for R3. Table 7 summarises the difference in the number of nodes flooded between the simulations with observed rainfall and the R3 forecast for the Dec 2018 and Nov 2020 events. The findings show that D4 either under forecasted or over forecasted the number of flooded nodes compared to the other grid sizes. Knowledge of which nodes are forecasted to flood can help decision-makers to implement protective measures before an event and coordinate response during an event However, false alarms on the number and location of flooded nodes can lead to the misallocation of resources, flooding and further traffic disruptions.1D-2D models could be run to give a more accurate characterisation of flood inundation however there are still challenges in reducing computational time for real-time applications and high-resolution data requirements.

5. Discussion

5.1. Trade-offs in model lead time, domain resolution, cumulus configuration and flood forecasting methodology

Our results reinforce different factors, can influence rainfall structure, magnitude, location and timing in high-resolution WRF simulations, such as spatial resolution, number of domains, lead time and model parametrisation.

This research evaluated the forecast skill at different neighbourhood sizes related to the scale of the phenomena and city level. Different neighbourhoods were evaluated to highlight how forecasts vary with the neighbourhood scales being considered. The 1000 km neighbourhood scale detected the presence of a weather system however, this was not very useful for the coastline. However, the metrics vary considerably at scales below 200 km. Overall the 24 LT performed best and the forecasts showed skill up to the 72hr LT but less at the 48 h and the 12hr LT. Although the 72 h rainfall forecast is less accurate than the 24 h rainfall forecast, the FSS scores showed there is still some skill associated with this forecast. More variability is observed at neighbourhood scales below 200 km and highlights the difficulty in evaluating forecasts at smaller neighbourhood scales and longer and shorter lead times which maybe more important for flood forecasting and emergency managers.

While many urban flood forecasters may not consider the treatment of convection, it becomes an important factor in rainfall and flood forecasting given the scale dependence of cumulus schemes at high

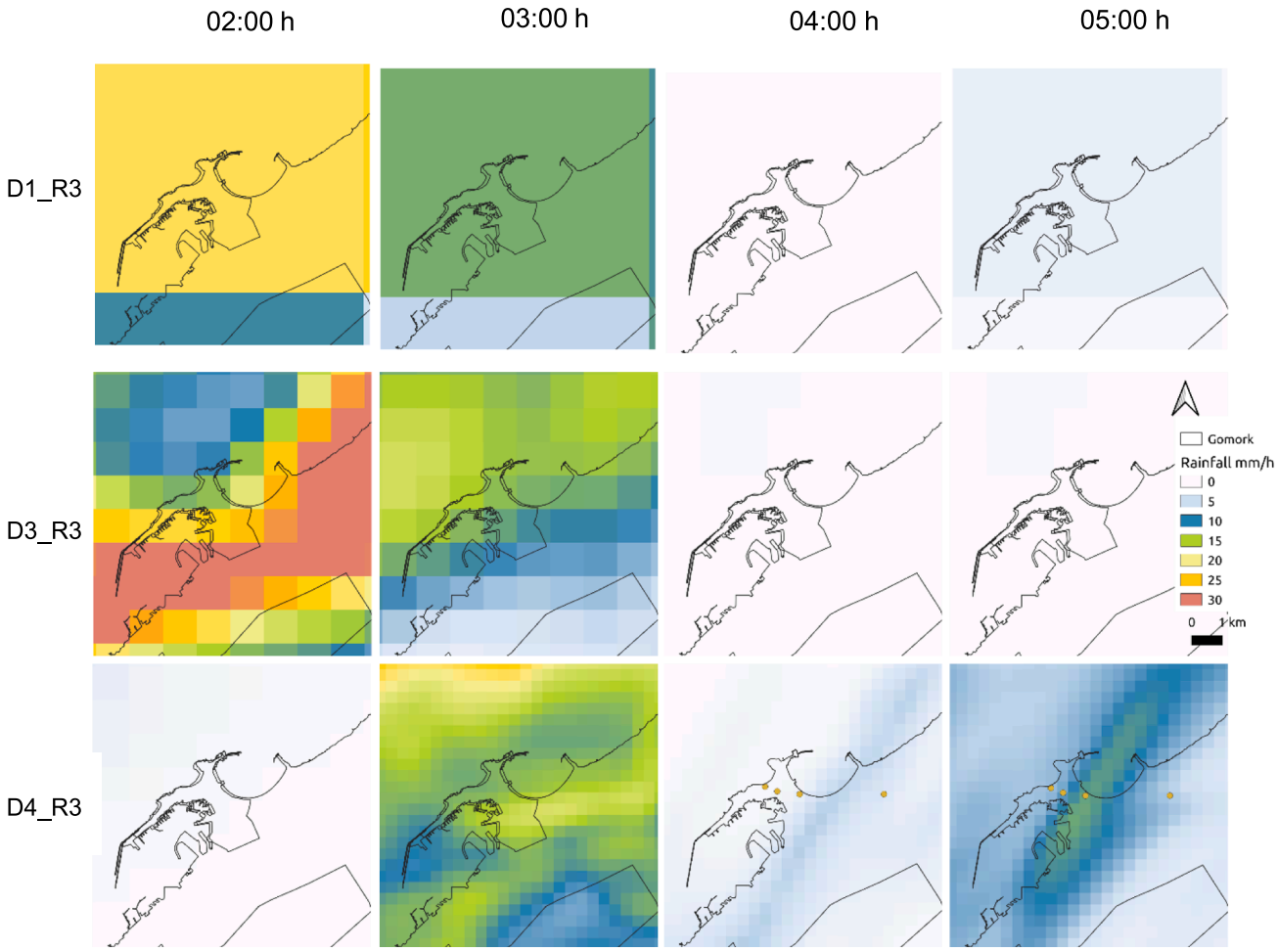


Fig. 18. The rainfall intensity and spatial distribution of grid rainfall used in the MIKE model for 10Km, 1.1 km grid resolution for run R3 and 0.37 km R3_D4.

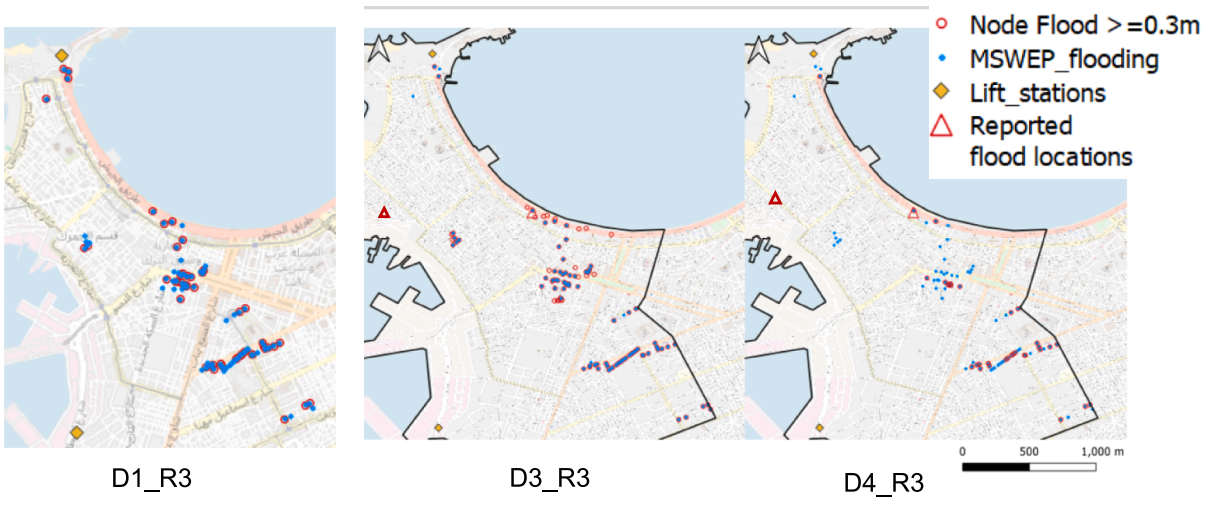


Fig. 19. For the Oct 2015 event. Maximum node flooding >= 0.3 m for R3. The 10 km grid (left), 1.1 km grid for runs (middle) and R3_4d(right). MSWEP simulated flood locations are shown in blue. More locations were detected for R3_D3 (101 flooded nodes) compared to R3_D1 (118 nodes) and R3_D4 (32 flooded nodes) and MSWEP (83 flooded nodes). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

resolutions needed for grid-based hydrodynamic models. This was specially observed in modelling the flood depths. The Grell- Freitas convection scheme was selected given it is a scale-aware scheme suitable for resolutions < 10 km –3 km. Our results indicate that turning the CPS on and off influenced simulated rainfall spatial and temporal patterns, which is in line with previous studies (Jeworrek et al., 2019). All runs and events forecasted the occurrence of an extreme rainfall event but with variability in magnitude and placement especially at

Table 7
Difference in the number of flooded noted between simulated 1D runs with MSWEP or IMERG and R3_D1 (10 km), R3_D3(1.1 km) and R3_D4(0.37 km).

	R3_D1(10 km)	R3_D3(1.1 km)	R3_D4(0.37 km)
2015	30	13	−56
2018	1	34	−8
2020	−10	−8	108

neighbourhood scales less than 200 km. The effect of the parameterisation varied with events, which is expected as effects vary with the size and characteristics of the event and depend on the role of convection in the event formation. One notable difference is the mismatch of the timing of the peak, which results in a delay in peak flood time. When comparing the 10 km and 1.1 km resolution it was shown that even though the inner domains are exhibiting higher rainfall accumulations than the parent domain, rainfall peaks do increase when the CPS in turned on in the smallest domains (R3). This is similar to the findings of Robaa and Wahab (2019) and Ibrahim (2020) who evaluated the sensitivity of WRF to convection schemes and found there was an overestimation of rainfall when cumulus schemes are turned on in the smaller domains.

When a 4th domain was used to capture a smaller resolution, it appears the runs either under-forecasts or over forecasted. In general, downscaling to 0.37 km using a 4th domain did not improve the rainfall forecasts but instead resulted in lower or higher rainfall totals and flood depths when compared to MSWEP and IMERG. While the findings indicate using a 4th domains to achieve finer resolution did not provide any improvement. It was not possible to draw a definitive conclusion about whether R2 or R3 gave better results as the results varied according to the events and their characteristics. However, when the cumulus schemes are turned off in R1, forecasts intended to under forecasts which suggest that CPS scheme should be used for grid sizes less than 10 km.

5.2. *Varying scales with flood forecasting approaches*

Generally, in urban flood simulation and forecasting, accuracy is a key factor (Wang et al, 2022). The results show challenges in using WRF for flood forecasting, especially when trying to predict flooding at manholes. Although there is mismatch in the timing, there is better agreement on the magnitudes of the peak. Despite these differences in the results, the models were able to predict an extreme event which resulted in flooding. For both the rainfall threshold method and flood model, the runs were successful in detecting an extreme event but performed better with the 10 km and 1.1 km (D3) resolutions for the higher thresholds 20 mm and 32 km while the 0.3 km gave the lowest POD and CSI scores and higher scores for the FAR. All runs performed well above the 50 mm rainfall, indicating that all forecasts agreed with not predicting “very extreme” rainfall accumulations. For the 1D real time simulated discharge, the 10 km and 1.1 km(D3) grid also performed best with NRMSEs of 0.42 and 0.43 and KGE above 0.4 for R2 and R3 but in general, the 0.37 km performed worst Our study indicated no gain in increasing resolution below 1 km when evaluating the flood model results irrespective of the cumulus configuration for Alexandria, Egypt Therefore, while it is assumed that higher resolutions will improve the spatial extent and durations of rainfall outputs by capturing more realistic sub-grid model dynamics, the finer spatial resolution can also introduce errors which does not justify the increased computational resources required for running such high-resolution models (Roberts, 2008; Kain et al., 2008; Schwartz et al., 2009).

There are some challenges to overcome and improvements to gain in achieving the 1 km resolution. The big advantage of this resolution is that it is suitable for running small-scale flood simulation models and can capture the variability in rainfall, thereby improving the quality of the flood forecasts and the opportunity to increase the effectiveness of flood risk management actions at specific locations. However, when more city scale actions are to be taken, such as issuing warnings, the coarser 10 km is suitable. Previous studies (Liu et al., 2012; Woodhams et al., 2018; Goodarzi et al., 2019; Liu et al.,2021) also found good rainfall simulation results at 3–4 km horizontal grid spacing, but it should be noted these studies did not consider a hazard estimation method. Thorndahl et al. (2016) highlighted the limitations and trade-offs of using a 3 km grid NWP model vs 0.5 km radar rainfall for urban pluvial flood forecasting to capture the rainfall variability and intensity needed for accurate flood predictions in urban environments. While the study found the 3 km grid NWP rainfall was limited, our study found the 1 km grid resolved NWP rainfall had some skill in detecting the correct flood class, flood depths and number of flooded manholes, especially at lead times up to 24 hrs. Therefore, when developing a flood forecasting system, lead time, cumulus configuration, and domain resolutions should also be aligned with the flood forecasting approach, the capability of local decision-makers, and the requirements for specific actions.

5.3. *Making imperfect forecasts useful*

The rainfall threshold method is quick, simple, and convenient, giving forecasted flood classes and does not require complex models. The 1D model, forecasted discharge, flood depths, and the number of flooded manholes. The analysis results indicate challenges in using WRF forecasts for flood forecasting at the city scale for both methods. Therefore, even though the results do not always match the flood classes per pixel or flood depths of the observed runs exactly, the forecasts have sufficient skill to raise alert and preparedness levels. Lead time is essential for forecast-based actions considered in the analysis and depends on the specific action. Hazard-reducing actions such as cleaning drains and increasing storage via pumping can range from a few days to a few hours, depending on the event’s severity, whereas exposure-reducing measures such as traffic measures and evacuations require hours.

Results found that although the 72 h rainfall forecast is less accurate than the 24 h rainfall forecast, there is still enough skill to use these forecasts to detect the phenomena and the city’s scale. However, below the 50 km scale, these other lead times are shown to be less skilful. Therefore, instead of using one method, the rainfall threshold and flood simulation results can be combined and used by decision-makers to make a decision especially when incorporated with predefined emergency preparedness protocols and triggers. Once a threat is detected in these forecasts, the rainfall threshold methods can be used to identify at-risk areas at the governorate scale. Given the skill of the 24 h lead time, the WRF can then be run at a higher resolution at this lead time so that rainfall forecasts can be used in the hydraulic model to forecast areas/ districts/ manholes at risk of flooding.

While our research only examined three events, further studies are essential to evaluate more events and the robustness of thresholds that capture the non-linearity of flooding. The real-time flood model requires more data, computational resources and model calibration and uncertainty in model inputs especially when physically-based modelled are used to represent hydrological processes. The advantage of real time approaches lies in their ability to directly model spatial variability and its impacts at city scale which is beneficial to areas that do not have high resolution observations (Hofmann and Schüttrumpf, 2019; Speight et al., 2020). Given the reliance of the peak intensity, it may be more useful to use peak intensity vs accumulation or timing as an indication for potential flooding (Umer et al., 2021). The practical value of these results depends on how a decision maker values forecast accuracy and which variables are most important to decision making: peak intensity, peak discharge, flood depth, total number of flooded nodes or the identification of a threat.

This allows decisions to be reviewed and updated in tandem which increases preparedness at longer lead times. This can be especially

valuable for complex data-scarce urban areas that lack information to refine models or where long lead times are needed to increase preparedness (Speight et al., 2020). Corral et al., 2018) compared a rainfall-based forecasting system using hazard level and a rainfall-runoff model to compute the stream flows at pixel scale and found similarities in the use for early warnings for flash flooding.

5.4. Limitations and future research

The unavailability of sub-daily data is one of the biggest challenges in performance evaluation in data scarce regions. Gauge data is only available at a daily resolution and the satellite-derived rainfall estimates and reanalysis data are at coarser resolutions than the forecasts and have their own bias. This was highlighted by the differences in MSWEP and IMERG data. Still, our results showed that MSWEP and IMERG data were able to corroborate the occurrence of extreme rainfall and node flooding for respective events. However, there was some variability in the severity. This further supports the use of satellite and reanalysis data in the absence of gauge data (Nashwan et al., 2019; Mekonnen et al., 2023) for the rainfall threshold approach sensitive to thresholds and in the absence of measured discharge data to evaluate the forecast data.

This research evaluated three events. It did not aim to make a conclusive statement about best model configurations and setup but rather to assess how hazard classes vary when high resolution precipitation at different scales are used with different flood forecasting approaches in data-scarce regions. Post-processing and data assimilation of forecast precipitation are methods used to improve the accuracy, sharpness and reliability of the forecast by removing systematic errors (Verkade et al., 2013; Crochemore et al., 2016), but it is still reliant on the availability robust observational data at suitable resolutions which were not available for this study. Data assimilation and bias correction methods such as Kalman Filter or quantile mapping (Bárdossy et al., 2021; Mapiam et al., 2022) should be incorporated into the flood forecasting chain to update forecasts when real-time data is available.

Future research should consider how performance varies with rainfall and storm characteristics. Sources of uncertainty have not been elaborated in this research. In particular, there is uncertainty in initial conditions and urban model uncertainty, which are important components. Probabilistic or ensemble approaches can be used to explore uncertainty in the placement and distribution of rainfall (Rico-Ramirez et al., 2015; Ravazzani et al., 2016; Yang et al., 2016; Böing et al., 2020; Brendel et al., 2020; Young et al., 2021). The ultimate goal will be more accurate, higher resolution spatial and temporal rainfall estimates through weather radar or an increased number of gauges and collection of discharge data. However, a database of historical flood hazards and impacts will remain invaluable and should be prioritised for verifying and determining expected impacts.

6. Conclusions

This research employed a limited area Weather Research and Forecasting (WRF) model in the context of urban-scale flood forecasting for Alexandria, Egypt, to evaluate if increasing the spatial resolution along with varying treatments of convection improves flood forecasting using a rainfall threshold and real-time urban model. It was found that the WRF model effectively identified extreme weather phenomena, and the rainfall threshold method correctly distinguished hazard classes at the district level. However, rainfall magnitudes, distributions and classes depended on the model configuration, and cumulus scheme used, particularly at convection-resolving resolutions (>3km). A timing delay of peak flows and flood depths was observed, however, the 1.1 km grid resolution was found to provide the best performance, especially when comparing flooding at specific manholes. The rainfall threshold gave POV and CSI values above 0.5 for all events for the 20 mm threshold, and the outflow discharge KGE values ranged from 0.4 to 0.8 for the 1.1 km. For both the rainfall threshold and urban flood modelling approaches, there was no improvement in using a fourth domain at (0.37 km) over using the 1.1 km or 10 km grid resolution regardless of the cumulus configuration. POV and CSI values were below 0.5 for all events for all runs, and KGE values below 0.27. Both the 1.1 km and the 10 km resolution require fewer computing resources than the 0.37 km. Therefore, while higher rainfall resolutions are required for urban scale modelling, this study highlighted that smallest resolution did not lead to improved hazard estimation. Additionally, the runs exhibited good skill in detecting extreme events at the 72LT but less skill at the 12hLT lead times and at various useful neighbourhood sizes, with the rainfall threshold method for all runs at coarser resolutions.

Despite challenges in using high resolution rainfall for flood forecasting, the good performance of both the rainfall threshold and real-time simulation methods suggests the potential for their combined use rather than using one over the other. Cities can use real-time simulating methods as more data becomes available to build models. Combining WRF forecasts with rainfall threshold methods allows early warning and flood forecasting at longer lead times without having to run complex flood models. When flood models are available, they can be coupled with WRF forecasts to provide more detailed information on the specific flood locations and depths at shorter lead times, but these results are also sensitive to the WRF model configurations as smaller resolutions do not necessarily predict more accurate flood forecasts. Combining a rainfall threshold and real-time forecasting model with a suitable cumulus configuration capitalises on the strengths of each method, which is invaluable for decision-makers in coordinating resources and implementing location-specific measures, particularly in conjunction with prior knowledge of high-risk areas.

Limitations for high-resolution modelling have been focused on the limits of computational resources but this research highlights there are also limits on quality. It is best to understand how methods can complement each other in data-scarce regions while being mindful of lead time, model configurations, cumulus schemes and resolutions. Forecasters and end-users must determine the useful resolutions and neighbourhood sizes that align with the specific goals of flood forecasting applications. As technology continues to advance and more real-time data becomes available, the use of high-resolution rainfall models hold promise for improving local forecasts and enhancing our ability to mitigate flood-related risks. Overall, this study advances our understanding of urban-scale flood forecasting in data-scarce regions: it stresses the importance of considering lead time, neighbourhood size, resolution, cumulus configuration, and integration with drainage models. The findings offer a foundation for the development of effective flood forecasting methodologies for anticipatory flood management and decision support tools despite the challenges in data-scarce cities. Finally, this research reiterates the need to understand the independencies and limitations of the aspects of the flood forecast chain; the meteorological output, the flood forecasting method and the data and methods used for verification with the overall goal for improved decision making and preparedness.

CRediT authorship contribution statement

Adele Young: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Biswa Bhattacharya:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization. **Emma Daniels:** Writing – review & editing, Methodology, Conceptualization. **Chris Zevenbergen:** Writing – review & editing, Supervision.

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.

Acknowledgements

This publication has been supported by a project which is (partly) financed by the Dutch Research Council (NWO). This work used the Dutch national e-infrastructure with the support of the SURF Cooperative using grant no. EINF-3489.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2025.132891.

Data availability

Data will be made available on request.

References

AASTMT and Egis BCEOM International (2011) Phase 1 Risk Assessment for the Present Situation and Horizon 2030-Climate Change Adaptation and Natural Disasters Preparedness in the Coastal Cities of North Africa-. Available at: http://www.abhatoo.net.ma/content/download/61166/1315121/version/1/file/Climate+Change+Adaptation+and+Natural+Disasters+Preparedness+in+the+Coastal+Cities+of+North+Africa.+Phase+1%2C+Risk+Assessment+for+the+Present+Situation+and+Horizon+2030+–+Alexandria+Area++Draft+Final+Version.pdf.

Abdrabo, K.I., Kantoush, S.A., Esmaiel, A., Saber, M., Sumi, T., Almamari, M., Elboshy, B., Ghoniem, S., 2023. An integrated indicator-based approach for constructing an urban flood vulnerability index as an urban decision-making tool using the PCA and AHP techniques: A case study of Alexandria, Egypt. Urban Climate 48, 101426. https://doi.org/10.1016/j.uclim.2023.101426.

Bárdossy, A., Seidel, J., El Hachem, A., 2021. The use of personal weather station observations to improve precipitation estimation and interpolation. Hydrol. Earth Syst. Sci. 25 (2), 583–601. https://doi.org/10.5194/hess-25-583-2021.

Beck, H.E., Wood, E.F., Pan, M., Fisher, C.K., Miralles, D.G., Van Dijk, A.I.J.M., McVicar, T.R., Adler, R.F., 2019. MSWep v2 Global 3-hourly 0.1° precipitation: methodology and quantitative assessment. Bull. Am. Meteorol. Soc. 100 (3), 473–500. https://doi.org/10.1175/BAMS-D-17-0138.1.

Berne, A., Delrieu, G., Creutin, J.D., Obled, C., 2004. Temporal and spatial resolution of rainfall measurements required for urban hydrology. J. Hydrol. 299 (3–4), 166–179. https://doi.org/10.1016/S0022-1694(04)00363-4.

Bhattacharya, B., Zevenbergen, C., Young, A. and Radhakrishnan, M. (2018) 'Extreme Flooding in Alexandria: Can Anticipatory Flood Management be a Solution?'. In: Loggia, G. La, Freni, G., Puleo, V., and Marchis, M. De Eds): *HIC 2018. 13th International Conference on Hydroinformatics*. EasyChair (EPiC Series in Engineering), pp. 252–257. doi: 10.29007/wvth.

Böing, S.J., Birch, C.E., Rabb, B.L., Kay, Shelton, L., 2020. A percentile-based approach to rainfall scenario construction for surface-water flood forecasts meteorological applications science and technology for weather and climate. Meteorol. Appl. 27. https://doi.org/10.1002/met.1963.

Brendel, C.E., Dymond, R.L., Aguilar, M.F., 2020. Integration of quantitative precipitation forecasts with real-time hydrology and hydraulics modeling towards probabilistic forecasting of urban flooding. Environ. Model. Softw.. Elsevier Ltd 134, 104864. https://doi.org/10.1016/j.envsoft.2020.104864.

Clark, A.J., Coniglio, M.C., Coffer, B.E., Thompson, G., Xue, M., Kong, F., 2015. Sensitivity of 24-h forecast dryline position and structure to boundary layer parameterizations in convection-allowing WRF model simulations. Weather Forecast. 30, 613–638. https://doi.org/10.1175/WAF-D-14-00078.1.

Cools, J., Vanderkimpen, P., El Afandi, G., Abdelkhalek, A., Fockedey, S., El Sammany, M., Abdallah, G., El Bihery, M., Bauwens, W., Huygens, M., 2012. An early warning system for flash floods in hyper-arid Egypt. Natural Hazards Earth Syst. Sci. 12 (2), 443–457. https://doi.org/10.5194/nhess-12-443-2012.

Corral, C., Berenguer, M., Sempere-Torres, D., Poletti, L., Silvestro, F., Rebora, N., 2018. (2019) 'Comparison of two early warning systems for regional flash flood hazard forecasting'. J. Hydrol. Elsevier 572, 603–619. https://doi.org/10.1016/j.jhydrol.2019.03.026.

Crochemore, L., Ramos, M.-H., Pappenberger, F., 2016. Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts. Hydrol. Earth Syst. Sci. 20, 3601–3618. https://doi.org/10.5194/hess-20-3601-2016.

Davis, S., Pentakota, L., Saptarishy, N., Mujumdar, P.P., 2022. A flood forecasting framework coupling a high resolution WRF ensemble with an urban hydrologic model. Front. Earth Sci. 10. https://doi.org/10.3389/feart.2022.883842.

El Afandi, Morsy, M., 2020. Developing an Early Warning System for Flash Flood in Egypt: Case Study Sinai Peninsula. In: Negm, A.M. (Ed.), Flash Floods in Egypt. Springer International Publishing, Cham, pp. 45–60. https://doi.org/10.1007/978-3-030-29635-3_4.

Eltahan, M., Magooda, M., 2018. Sensitivity of WRF microphysics schemes: case study of simulating a severe rainfall over Egypt. J. Phys. Conf. Ser. 1039 (1). https://doi.org/10.1088/1742-6596/1039/1/012024.

Ermagun, A., Smith, V., Janatabadi, F., 2024. High urban flood risk and no shelter access disproportionally impacts vulnerable communities in the USA. Commun. Earth Environ. 5 (1), 2. https://doi.org/10.1038/s43247-023-01165-x.

Flack, D., Skinner, C., Hawkness-Smith, L., O'Donnell, G., Thompson, R., Waller, J., Chen, A., Speight, L., 2019. Recommendations for improving integration in national end-to-end flood forecasting systems: an overview of the FFIR (Flooding From Intense Rainfall) programme. Water 11 (4), 725. https://doi.org/10.3390/w11040725.

Fowler, L.D., Skamarock, W.C., Grell, G.A., Freitas, S.R., Duda, M.G., 2016. Analyzing the grell–freitas convection scheme from hydrostatic to nonhydrostatic scales within a global model. Mon. Weather Rev. 144 (6), 2285–2306. https://doi.org/10.1175/MWR-D-15-0311.1.

Goodarzi, L., Banihabib, M.E., Roozbahani, A., 2019. A decision-making model for flood warning system based on ensemble forecasts. Journal of Hydrology. Elsevier 573, 207–219. https://doi.org/10.1016/j.jhydrol.2019.03.040.

Goswami, P., Shivappa, H., Goud, S., 2012. Comparative analysis of the role of domain size, horizontal resolution and initial conditions in the simulation of tropical heavy rainfall events. Meteorol. Appl. 19 (2), 170–178. https://doi.org/10.1002/met.253.

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. https://doi.org/10.5194/acp-14-5233-2014.

Han, J.-Y., Hong, S.-Y., 2018. Precipitation forecast experiments using the Weather Research and Forecasting (WRF) model at gray-zone resolutions. Weather Forecast. 33 (6), 1605–1616. https://doi.org/10.1175/WAF-D-18-0026.1.

Hasanean, H.M., 2004. Precipitation variability over the mediterranean and its linkage with El Nino Southern Oscillation (ENSO). J. Meteorol. 29 (289), 151–160.

Henonin, J., Russo, B., Mark, O., Gourbesville, P., 2013. Real-time urban flood forecasting and modelling – a state of the art. Journal of Hydroinformatics 15 (3), 717. https://doi.org/10.2166/hydro.2013.132.

Hofmann, J., Schüttrumpf, H., 2019. Risk-Based early warning system for pluvial flash floods: approaches and foundations. Geosciences 9 (3), 127. https://doi.org/10.3390/geosciences9030127.

Houston, D., Werrity, A., Bassett, D., Geddes, A., Hoolachan, A. and McMillan, M. (2011) 'Pluvial (rain-related) flooding in urban areas: the invisible hazard', *Joseph Rowntree Foundation*, pp. 1–92. Available at: https://eprints.gla.ac.uk/162145/ (Accessed: 1 June 2021).

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) Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) Algorithm Theoretical Basis Document (ATBD) Version 06.

Ibrahim, A., 2020. A Comparative Study of Two Extreme Cases Hit Egypt in January 2008 and 2009 Using WRF Different Convective Schemes. Biomed J Sci & Tech Res 26(2)-2020. BJSTR. MS.ID.004326.

Iqbal, A. (2017) 'Radar rainfall forecasting for sewer flood modelling to support decision-making in sewer network operations'. Available at: https://ore.exeter.ac.uk/repository/handle/10871/32565.

Ibrahim, S., Afandi, G., 2014. Short-range Rainfall Prediction over Egypt using the Weather Research and Forecasting Model. Open J. Renew. Energy Sustain. Dev. 2014 (2), 56–70. https://doi.org/10.15764/RESD.2014.02006.

Jankov, I., Gallus, 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. https://doi.org/10.1175/WAF998.1.

Jankov, I., Gallus, W.A., Segal, M., Shaw, B., Koch, S.E., 2005. The impact of different WRF model physical parameterizations and their interactions on warm season MCS rainfall. Weather Forecast. 20 (6), 1048–1060. https://doi.org/10.1175/WAF888.1.

Jeworrek, J., West, G., Stull, R., 2019. Evaluation of cumulus and microphysics parameterizations in WRF across the convective gray zone. Weather Forecast. 34 (4), 1097–1115. https://doi.org/10.1175/WAF-D-18-0178.1.

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., Thomas, K.W., 2008. Some Practical Considerations Regarding Horizontal Resolution in the First Generation of Operational Convection-Allowing NWP. Weather Forecast. 23 (5), 931–952. https://doi.org/10.1175/WAF2007106.1.

Kendon, E.J., Ban, N., Roberts, N.M., Fowler, H.J., Roberts, M.J., Chan, S.C., Evans, J.P., Fosser, G., Wilkinson, J.M., 2017. Do Convection-Permitting Regional Climate Models Improve Projections of Future Precipitation Change? Bull. Am. Meteorol. Soc. 98 (1), 79–93. https://doi.org/10.1175/BAMS-D-15-0004.1.

Lean, H.W., Clark, P.A., Dixon, M., Roberts, N.M., Fitch, A., Forbes, R., Halliwell, C., 2008. Characteristics of high- resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. Mon. Wea. Rev. 136, 3408–3424. https://doi.org/10.1175/2008MWR2332.1.

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. https://doi.org/10.1002/hyp.8247.

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 rainfall events over Egypt. J. Hydrol. 598, 126237. https://doi.org/10.1016/j.jhydrol.2021.126237.

Mahmood, F. (2021) Rainfall Characterization and Urban Flood Modelling in Data Scarce Regions: A Case Study of Alexandria, Egypt. IHE Delft Institute for Water Education.

Mapiam, P.P., Methaprayun, M., Bogaard, T., Schoups, G., Ten Veldhuis, M.-C., 2022. Citizen rain gauges improve hourly radar rainfall bias correction using a two-step Kalman filter. Hydrol. Earth Syst. Sci. 26 (3), 775–794. https://doi.org/10.5194/hess-26-775-2022.

Mekonnen, K., Velpuri, N.M., Leh, M., Akpoti, K., Owusu, A., Tinonetsana, P., Hamouda, T., Ghansah, B., Paranamana, T.P., Munzimi, Y., 2023. Accuracy of satellite and reanalysis rainfall estimates over Africa: a multi-scale assessment of eight products for continental applications. J. Hydrol.: Reg. Stud. 49. https://doi.org/10.1016/j.ejrh.2023.101514.

Ming, X., Liang, Q., Xia, X., Li, D., Fowler, H.J., 2020. Real-Time flood forecasting based on a high-performance 2-D hydrodynamic model and numerical weather predictions. Water Resour. Res. 56 (7). https://doi.org/10.1029/2019WR025583.

Nashwan, M.S., Shahid, S., Wang, X., 2019. Assessment of satellite-based precipitation measurement products over the hot desert climate of Egypt. Remote Sens. (Basel) 11 (5). https://doi.org/10.3390/rs11050555.

Paul, S., Ghosh, S., Mathew, M., Devanand, A., Karmakar, S., Niyogi, D., 2018. 'Increased spatial variability and intensification of extreme monsoon rainfall due to urbanization'. Sci. Rep. Springer, US 8 (1), 1–10. https://doi.org/10.1038/s41598-018-22322-9.

Prein, A.F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tölle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., van Lipzig, Leung, R., 2015. A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. Rev. Geophys. 53, 323–361. https://doi.org/10.1002/2014RG000475.

Prein, A.F., Rasmussen, R.M., Ikeda, K., Liu, C., Clark, M.P., Holland, G.J., 2017. The future intensification of hourly precipitation extremes. Nat. Clim. Chang. 7 (1), 48–52. https://doi.org/10.1038/nclimate3168.

Ravazzani, G., Amengual, A., Ceppi, A., Homar, V., Romero, R., Lombardi, G., Mancini, M., 2016. Potentialities of ensemble strategies for flood forecasting over the Milano urban area. J. Hydrology. Elsevier B.V. 539, 237–253. https://doi.org/10.1016/j.jhydrol.2016.05.023.

René, J.-R., Madsen, H., Mark, O., 2013. A methodology for probabilistic real-time forecasting – an urban case study. Journal of Hydroinformatics 15 (3), 751–762. https://doi.org/10.2166/hydro.2012.031.

Rico-Ramirez, M.A., Liguori, S., Schellart, A.N.A., 2015. Quantifying radar-rainfall uncertainties in urban drainage flow modelling. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2015.05.057.

Robaa, S. M. and Wahab, M. M. A. (2019) 'Sensitivity of WRF Model to convection schemes for rainfall forecast over Egypt', pp. 12–17.

Roberts, N., 2008. Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. Meteorol. Appl. 15 (1), 163–169. https://doi.org/10.1002/met.57.

Schellart, A., Ochoa, S., Simões, N., Wang, L., Rico-ramirez, M., Duncan, A., Chen, A.S., Maksimovi, Č., 2011. Urban pluvial flood modelling with real time rainfall information – UK case studies. 12nd International Conference on Urban Drainage, Porto Alegre/brazil 2, 10–15.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D. O., Barker, D., Duda, M. G., Huang, X.-Y., Wang, W., & Powers, J. G. (2008). *A Description of the Advanced Research WRF Version 3*. University Corporation for Atmospheric Research. https://doi.org/10.5065/D68S4MVH (Original work published 2008).

Schwartz, C.S., Kain, J.S., Weiss, S.J., Xue, M., Bright, D.R., Kong, F., Thomas, K.W., Levit, J.J., 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), 3351–3372. https://doi.org/10.1175/2009MWR2924.1.

Speight, L.J., Cranston, M.D., White, C.J., Kelly, L., 2020. (2021) 'Operational and emerging capabilities for surface water flood forecasting'. WIREs Water 1–24. https://doi.org/10.1002/wat2.1517.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. Atmos. 106 (D7), 7183–7192. https://doi.org/10.1029/2000JD900719.

Thorndahl, S., Nielsen, J.E., Jensen, D.G., 2016. Urban pluvial flood prediction: a case study evaluating radar rainfall nowcasts and numerical weather prediction models as model inputs. Water Sci. Technol. 74 (11), 2599–2610. https://doi.org/10.2166/wst.2016.474.

Thorndahl, S., Poulsen, T.S., Bøvith, T., Borup, M., Ahm, M., Nielsen, J.E., Grum, M., Rasmussen, M.R., Gill, R., Mikkelsen, P.S., 2013. Comparison of short-term rainfall forecasts for modelbased flow prediction in urban drainage systems. Water Sci. Technol. 68 (2), 472–478. https://doi.org/10.2166/wst.2013.274.

Umer, Y., Ettema, J., Jetten, V., Steeneveld, G.-J., Ronda, R., 2021. Evaluation of the WRF Model to Simulate a High-Intensity Rainfall Event over Kampala, Uganda. Water 13 (6), 873. https://doi.org/10.3390/w13060873.

Verkade, J.S., Brown, J.D., Reggiani, P., Weerts, A.H., 2013. Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. J. Hydrol. 501, 73–91. https://doi.org/10.1016/j.jhydrol.2013.07.039.

Wang, H., Hu, Y., Guo, Y., Wu, Z. and Yan, D. (2022) 'Urban flood forecasting based on the coupling of numerical weather model and stormwater model: A case study of Zhengzhou city', Journal of Hydrology: Regional Studies. Elsevier B.V., 39 (December 2021), p. 100985. doi: 10.1016/j.ejrh.2021.100985.

Wang, X., Steinle, P., Seed, A., Xiao, Y., 2016. The sensitivity of heavy precipitation to horizontal resolution, domain size, and rain rate assimilation: case studies with a convection-permitting model. Adv. Meteorol. https://doi.org/10.1155/2016/7943845.

Woodhams, B.J., Birch, C.E., Marsham, J.H., Bain, C.L., Roberts, N.M., Boyd, D.F.A., 2018. What is the added value of a convection-permitting model for forecasting extreme rainfall over tropical East Africa? Mon. Weather Rev. 146 (9), 2757–2780. https://doi.org/10.1175/MWR-D-17-0396.1.

Yang, T.H., Hwang, G Do, Tsai, C.C., Ho, J.Y., 2016. Using rainfall thresholds and ensemble precipitation forecasts to issue and improve urban inundation alerts. Hydrol. Earth Syst. Sci. 20 (12), 4731–4745. https://doi.org/10.5194/hess-20-4731-2016.

Young, A., Bhattacharya, B., Zevenbergen, C., 2021. A rainfall threshold-based approach to early warnings in urban data-scarce regions: a case study of pluvial flooding in Alexandria, Egypt. J. Flood Risk Manage. 14 (2), 1–16. https://doi.org/10.1111/jfr3.12702.

Zevenbergen, C., Bhattacharya, B., Wahaab, R. A., Elbarki, W. A. I., Busker, T. and Salinas Rodriguez, C. N. A. (2016) 'In the aftermath of the October 2015 Alexandria Flood Challenges of an Arab city to deal with extreme rainfall storms', *Natural Hazards*. Springer Netherlands, (October 2015). doi: 10.1007/s11069-016-2724-z.

Zheng, Y., Alapaty, K., Herwehe, J.A., Del Genio, A.D., Niyogi, D., 2016. Improving high-resolution weather forecasts using the Weather Research and Forecasting (WRF) model with an updated kain-fritsch scheme. Mon. Weather Rev. 144 (3), 833–860. https://doi.org/10.1175/MWR-D-15-0005.1.