Evaluation of Microphysics and Boundary Layer Schemes for Simulating Extreme Rainfall Events over Saudi Arabia using WRF-ARW v4.4

Rajesh Kumar Sahu[1], Hamza Kunhu Bangalath[1], Suleiman Mostamandi[1], Jason Evans[2], Paul A. Kucera[3], and Hylke E. Beck[1]

[1]Physical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
[2]Climate Change Research Centre, University of New South Wales, Sydney, Australia
[3]COMET Program, University Corporation for Atmospheric Research, Boulder, Colorado, USA

Correspondence: Hylke E. Beck (hylke.beck@kaust.edu.sa)

Abstract. Extreme Rainfall Events (EREs) and resulting flash floods in Saudi Arabia pose major threats, frequently causing fatalities and significant economic losses. Accurate ERE simulations are crucial for weather forecasting, climate change assessment, and disaster management. This study evaluates planetary boundary layer (PBL) and cloud microphysics (MP) schemes to simulate EREs in the Arabian Peninsula (AP) using the Advanced Research version of the Weather Research and Forecasting (WRF-ARW) model V4.4. Thirty-six combinations of four PBL and nine MP schemes were tested across 17 EREs at a convection-permitting 3-km resolution and compared with IMERG gridded satellite data for rainfall and station observations for temperature, humidity, and wind speed. The Kling-Gupta Efficiency (KGE), which incorporates correlation, variability, and bias, was used as performance metric. We found a good agreement between observed and simulated rainfall patterns, though some over- and underestimations were present. Among the PBL schemes, Yonsei University (YSU; BL1) tended to perform best in terms of rainfall, while Thompson (MP8) ranked the highest among the MP schemes. Goddard (MP7) also delivered strong results. Among all 36 combinations, the Thompson-YSU (MP8_BL1) combination consistently produced the highest mean KGE across the 17 EREs for rainfall, performing statistically significantly better than 21 other combinations. While MP8_BL1 also performed best for the other three meteorological variables, performance rankings varied across variables, likely because different physical processes govern the simulation of different variables. This study highlights the complexity of scheme evaluation and the importance of analyzing multiple EREs with high-quality reference data. The results offer practical guidance for scheme selection and lay the foundation for improving ERE forecasting and regional climate modeling over the AP.

1 Introduction

Extreme Rainfall Events (EREs) are episodes of intense rainfall over a short duration, often resulting in flash floods, landslides, severe damage to infrastructure and property, and loss of life (Easterling et al., 2000; Houze Jr, 2012; Kundzewicz et al., 2014; Srinivas et al., 2018). These events are becoming more frequent and intense as atmospheric moisture increases by about 7%

per degree of warming, following Clausius-Clapeyron scaling (e.g., Held and Soden, 2006; O'Gorman and Schneider, 2009; Muller and Takayabu, 2020; Fowler et al., 2021; Neelin et al., 2022). Although mean rainfall increases at a slower rate of 2–3% per degree, EREs can intensify by as much as 6–10% depending on their spatial and temporal scales (e.g., Allan and Soden, 2008; O'Gorman and Schneider, 2009), significantly increasing their potential for destructive impacts.

Despite its arid desert climate and low annual rainfall, Saudi Arabia regularly experiences significant EREs (Almazroui, 2011; Haggag and El-Badry, 2013; Deng et al., 2015; Yesubabu et al., 2016; Almazroui et al., 2018; Atif et al., 2020; Attada et al., 2022), particularly during the rainy season from November to April. These events are frequently associated with intrusions of an intensified subtropical jet stream, mid-latitude cyclonic disturbances, and the low-level advection of warm, moist air from nearby water bodies, including the Red Sea, Arabian Gulf, and Arabian Sea (Evans et al., 2004; Barth and Steinkohl, 2004; Evans and Smith, 2006; De Vries et al., 2013, 2016). Though infrequent, these EREs cause substantial damage (Al Saud, 2010; Youssef et al., 2016), making accurate forecasting and projection essential for disaster management, early warning systems, and climate adaptation in the region (Hijji et al., 2013; Abosuliman et al., 2014).

Advanced Research version of the Weather Research and Forecasting (WRF-ARW;Skamarock et al., 2019) is a widely used Numerical Weather Prediction (NWP) model in the Arabian Peninsula (AP) to simulate and forecast EREs (Deng et al., 2015; Almazroui et al., 2018; Taraphdar et al., 2021; Risanto et al., 2024; Luong et al., 2025; Taraphdar et al., 2025; Francis et al., 2025). These models are subject to various sources of uncertainty, particularly due to parameterizations. Two key parameterization schemes that strongly influence ERE simulations include the Planetary Boundary Layer (PBL) and cloud microphysics (MP) schemes.

The PBL scheme governs the vertical exchange of momentum, heat, and moisture between the surface and the atmosphere, playing a critical role in simulating near-surface conditions. It regulates vertical mixing and turbulence, which are essential for atmospheric instability and convective initiation — key processes that directly impact rainfall development (Kumar et al., 2008). The selection of an appropriate PBL scheme is especially important in arid and arid regions such as the AP, as intense surface heating in desert environments leads to the formation of unusually deep PBLs, sometimes extending up to 5 km during the day (Gamo, 1996; Marsham et al., 2008; Ntoumos et al., 2023). This necessitates the use of a scheme capable of accurately modeling the vertical distribution of heat, moisture, and momentum within such a deep layer. Furthermore, deserts are characterized by complex thermodynamic profiles, including sharp temperature gradients and significant humidity variations, which complicate the modeling process. Strong diurnal temperature variations also require a PBL scheme capable of effectively capturing short-term fluctuations in energy and moisture fluxes.

The MP scheme governs the evolution of cloud particles, including cloud droplets, rain, snow, and ice, which are essential to determine the intensity and duration of rainfall (Dudhia, 2014). It controls cloud formation, rainfall processes, and interactions between different water phases. It also influences radiative transfer by affecting cloud optical properties such as droplet size distribution, phase, and concentration (Stull, 1988; Garratt, 1994; Stull, 2012; Dudhia, 2014). Additionally, MP schemes govern key hydrometeor processes like condensation and coalescence, which directly impact the timing, intensity, and spatial distribution of rainfall. Both single-moment and double-moment schemes exist; the latter provide a more detailed representation by also predicting number concentrations of hydrometeors (see, e.g., Kessler, 1969; Chen and Sun, 2002; Hong et al., 2004; Rogers et al., 2001; Hong and Lim, 2006; Tao et al., 2016; Thompson et al., 2008; Morrison et al., 2009).

Although several previous studies have evaluated different WRF-ARW parameterization schemes in the AP (e.g., Deng et al., 2015; Schwitalla et al., 2020; Attada et al., 2022; Abida et al., 2022) and in other arid and semi-arid regions (e.g., Zittis et al., 2014; Tian et al., 2017; Liu et al., 2021; Messmer et al., 2021; Khansalari et al., 2021; Mekawy et al., 2022; Pegahfar et al., 2022), they have typically focused on individual EREs and conducted limited sensitivity analyses using a small number of parameterization schemes (Table 1). The case-specific nature of these studies often restricts the generalizability of their results to other EREs and varying conditions, reducing their broader applicability for predicting EREs in the complex climate dynamics of the AP.

Our study addresses this gap by conducting an extensive evaluation of WRF-ARW PBL and MP schemes for simulating EREs in the AP at convection-permitting resolution (3 km) to determine the best combination of PBL and MP schemes that consistently performs well across different EREs. We analyze 17 EREs from 2010 to 2022 across the AP, testing 36 different combinations of PBL and MP schemes to identify the optimal configuration for ERE simulation in the AP. We simulate the EREs using a two-way nested domain configuration with 53 vertical levels and horizontal resolutions of 9 and 3 km. While our primary focus is on the evaluation of rainfall, we also examine air temperature, relative humidity, and wind speed.

To guide the reader, the paper is structured according to nine key questions:

a. Which PBL scheme performs best in terms of rainfall?

b. Which MP scheme performs best in terms of rainfall?

c. Which component of the Kling-Gupta Efficiency (KGE) affects the final rainfall scores the most?

d. How statistically significant are the differences in performance between scheme combinations in terms of rainfall?

e. How consistent are the temporal and spatial performance assessments for rainfall?

f. How consistent is the performance ranking among different variables?

g. What do the spatial patterns in simulated and observed rainfall look like for the EREs?

h. How well does the model perform in terms of the other variables?

i. How do the PBL and MP schemes used in previous studies compare with those identified as optimal in our evaluation?

2 Physical geography and climatic description of the study area

Saudi Arabia, covering 80% of the AP, spans from 16°N to 33°N and 34°E to 56°E, with an area of approximately 2.1 million km^2, making it the largest country in the Middle East and the 12th largest globally. The terrain includes highlands, volcanic fields, mountain ranges, and the vast Arabian desert, featuring the Rub' al Khali, the world's largest continuous sand desert.

Table 1. Previous studies evaluating WRF-ARW physics schemes in the Middle East.

Region	Kind of schemes	Number of Events	Model layer/ Vertical levels used	Conclusion	Reference
Jeddah, Saudi Arabia	Microphysics schemes: Lin, Eta Ferrier	Three flash floods events	50	The WRF-ARW Model effectively simulates flash floods in Jeddah, with 1 km resolution improving rainfall accuracy and 5 km requiring careful parameterization due to observed spatial displacement.	Deng et al. (2015)
AP	Cumulus schemes: KF, BMJ, GF	Winter simulation from 2001 to 2016	52	Selecting subgrid convective parameterization is crucial for accurate high-resolution rainfall simulations over the AP.	Attada et al. (2020)
AP	MP schemes: Thompson 2-moment, Thompson aerosol-aware and WDM6 and PBL schemes: MYNN Level 2.5 and YSU	Case study on July 14, 2015	100	The best performance was obtained using a convection-permitting model resolution with aerosol-aware Thompson MP combined with the MYNN Level 2.5 PBL scheme, which effectively captured rainfall.	Schwitalla et al. (2020)
Middle East	PBL schemes: ACM2, QNSE, MYNN Level 2.5	Single year run for 2017	45	Gray-zone simulations enhance rainfall modeling but are highly dependent on resolution and the selection of physics schemes.	Taraphdar et al. (2021)
AP	Cumulus schemes: KF, BMJ, GF	Winter simulation from 2001 to 2016	52	ERE is best simulated using the KF scheme, highlighting the importance of cumulus parameterization in WRF-ARW for reliable modeling in the hyper-arid AP region.	Attada et al. (2022)

Despite lacking permanent rivers, it has many wadis, alluvial deposits (Vincent, 2008; WeatherOnline, 2024), and about 1,300 islands in the Arabian Gulf and the Red Sea. The central plateau stretches from the Red Sea to the Arabian Gulf, while the Asir province reaches 3,002 m above sea level at Jabal Ferwa, and the Hejaz region contains approximately 2,000 extinct volcanoes across 180,000 km^2. The climate is characterized by vast deserts, rugged mountains, and a hyper-arid climate, with extreme summer temperatures of 45–54°C and winters rarely below 0°C (De Vries et al., 2016; El Kenawy et al., 2014; Mostamandi et al., 2022; Ukhov et al., 2020). The average annual rainfall over the region is about 63 mm, except in the southwest, where monsoons bring over 300 mm of rain from October to March (Wang et al., 2025).

The primary mechanisms driving rainfall vary between the eastern and western coasts. On the western coast, the Asir mountain chains play a significant role in capturing moist northwesterly winds along the Red Sea coast, particularly during winter, extending up to the Bab el-Mandeb Strait (Pedgley, 1974; El Kenawy et al., 2014; Mostamandi et al., 2022). From East Africa through the Red Sea towards the eastern Mediterranean, the Red Sea Trough (RST) creates a geographical environment conducive to forming strong low-pressure systems over the central Red Sea. These systems can generate substantial rainfall within the region (De Vries et al., 2013; El Kenawy et al., 2014). In contrast, the eastern coast, influenced by the Hajar Mountains and its proximity to the Arabian Sea, receives convective rainfall driven by the summer monsoon and moisture-laden winds from the Indian Ocean (Babu et al., 2016).

3 Data and Methods

3.1 Selection of Historical Extreme Rainfall Cases

We selected 17 EREs across from 2010 to 2022 that caused significant damage to infrastructure and property, as well as loss of life, and received widespread media coverage. Table 2 lists the EREs analyzed in this study. We included 17 cases to increase the likelihood of obtaining statistically significant results regarding the relative performance of different schemes. We did not analyze more cases due to the significant processing, storage, and computational demands.

3.2 Initial and Boundary Conditions

ERA5 pressure-level data (with 37 levels, extending up to approximately 30 km altitude; 0.25° spatial resolution) was utilized to provide initial and boundary conditions for each 3-hour time step to run WRF-ARW. ERA5 is the most reliable reanalysis currently available and was therefore used for this purpose (Hersbach et al., 2020). The data were obtained from the Copernicus Climate Data Store (CDS; https://cds.climate.copernicus.eu).

3.3 Observations

As a reference for our assessment, we used rainfall estimates from the satellite-based Integrated Multi-satellite Retrievals for GPM (IMERG) Final V07 product (Huffman et al., 2023). The product covers 2000 to the present, has a 30-minute 0.1° resolution, and was aggregated to hourly for our analysis.

We also used 2-m air temperature (°C), 2-m relative humidity (%), and 10-m wind speed (m/s) observations from the IOWA Environmental Mesonet (METAR) data provided by Iowa State University (https://mesonet.agron.iastate.edu/request/download.phtml?network=SA__ASOS; for locations, see Supplement Fig. S1).

3.4 WRF-ARW Model Configuration

This study uses the WRF-ARW model version 4.4, a non-hydrostatic, fully compressible model with a terrain-following coordinate system (Skamarock et al., 2019). The model is configured with two-way nested domains with horizontal grid dimensions of 493 × 418 for the parent domain (D01) and 1012 × 889 for the nested domain (D02), and a model top pressure of 30 hPa, comprising 53 vertical hybrid sigma levels and a horizontal resolution of 3 km in the innermost domain, as shown in Fig. 1. The D01 domain covers a vast region of the AP from 21°E to 65°E in the zonal direction and from 2°N to 40°N in the meridional direction, allowing for the representation of large-scale atmospheric features and internal dynamics. We conducted 612 simulations, spanning all 36 possible PBL–MP scheme combinations, to assess their joint performance across 17 EREs. Convection is explicitly resolved in D02, while D01 uses the Kain-Fritsch parameterization (Kain and Fritsch, 1993) for sub-grid convective processes (Snook et al., 2019).

We considered 36 combinations involving nine MP and four PBL schemes. The PBL schemes tested include Mellor-Yamada Nakanishi Niino (MYNN) Level 2.5 and Level 3 (BL5, BL6; Nakanishi and Niino, 2006), Yonsei University (YSU; BL1; Hong et al., 2006), and Bougeault-Lacarrère (BouLac; BL8; Bougeault and Lacarrere, 1989), while the MP schemes include Kessler (MP1; Kessler, 1969), Purdue Lin (MP2; Chen and Sun, 2002), WRF Single-Moment 3-class and 5-class (MP3 and MP4, respectively; (Hong et al., 2004)), Eta Ferrier, (MP5; Rogers et al., 2001), WRF Single-Moment 6-class (MP6; Hong and Lim, 2006), Goddard (MP7; Tao et al., 2016), Thompson (MP8; Thompson et al., 2008), and Morrison 2-Moment (MP10; Morrison et al., 2009). These combinations were selected based on their compatibility with the surface layer physics Revised MM5 scheme (Jiménez et al., 2012), and additional schemes were not included due to the higher computational and storage demands. Previous studies focusing on the AP have also utilized these schemes, including Deng et al. (2015); Attada et al. (2022); Luong et al. (2020); Schwitalla et al. (2020).

Initial and boundary conditions were extracted from ERA5 reanalysis data at 3-hour intervals with a 0.25° resolution. All model simulations were conducted for 84 hours, including a 48-hour spin-up period to ensure model stability and reduce initialization biases. The analysis was focused on a 24-hour window corresponding to the peak rainfall period of each ERE (Table 2). Our study specifically targets short-duration, event-based simulations of ERE. In such cases, the primary drivers are typically large-scale atmospheric instabilities and moisture advection rather than slower processes like land–surface interactions. Consequently, a 48-hour spin-up period is sufficient to allow the model to dynamically and thermodynamically adjust to the initial and boundary conditions. Refer to Table 2 for the simulation start dates and Table 3 for the model configuration.

3.5 Model Assessment Approach

Each combination of MP and PBL schemes was extensively evaluated using the Kling-Gupta Efficiency (KGE; Gupta et al., 2009; Kling et al., 2012). The KGE is an aggregate performance metric that integrates correlation, bias ratio, and variability ratio into a single score, providing a holistic assessment of model performance. Several studies have successfully used KGE for spatial performance assessment of hydrometeorological models (e.g., Gupta et al., 2009; Patil and Stieglitz, 2015; Beck et al., 2019a; Nguyen et al., 2022; Tudaji et al., 2025), supporting its application in our analysis. The formula for KGE is given

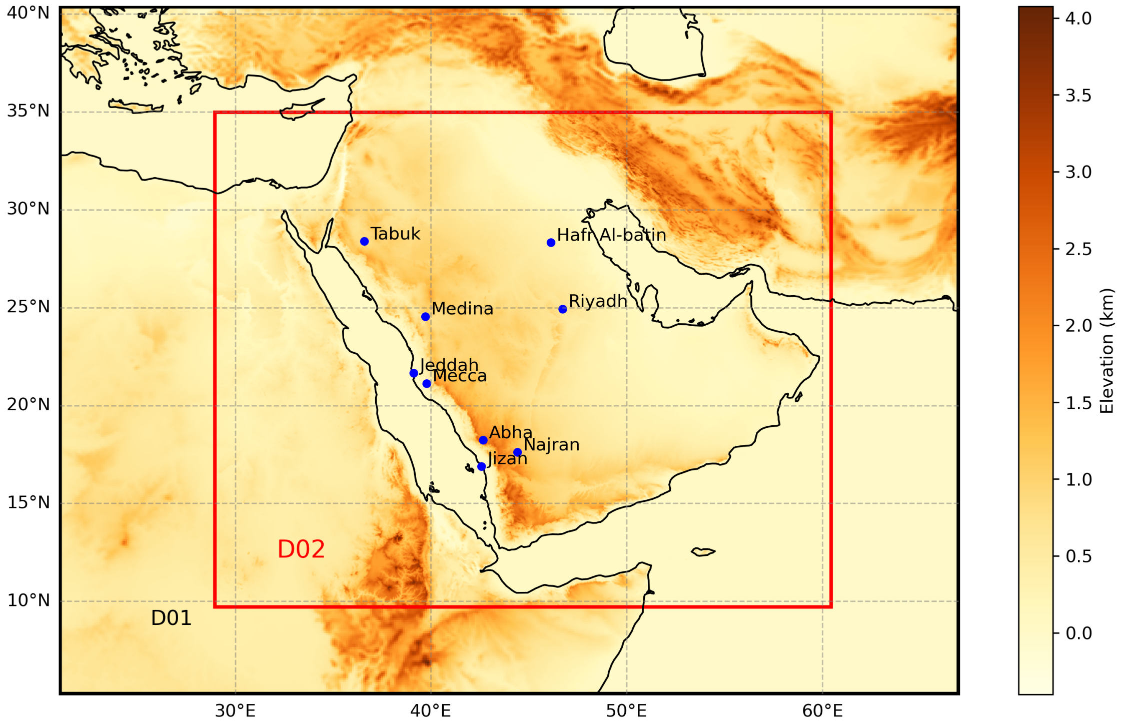

Fig. 1. WRF-ARW domain for the AP region showing the elevation in the background. METAR locations are indicated with blue markers.

by:

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}, \tag{1}$$

where r is Pearson's correlation coefficient between the observed and simulated data, β is the ratio of the mean simulated data to the mean observed data, assessing the bias, and γ is the ratio of the coefficients of variation of the simulated and observed data, evaluating the variability. A perfect but unattainable KGE score is 1, indicating complete agreement between simulated and observed data. A hypothetical simulation predicting only the observed mean would achieve a KGE of –0.41 (Knoben et al., 2019).

For rainfall, the KGE was calculated separately in space and in time. For the temporal KGE (Fig. 2), we first calculated, for each hour of the event day (Table 2), the spatial average of observed and simulated rainfall across D02 (Fig. 1). The KGE was derived from these 24 pairs of observed and simulated spatially averaged values. For the spatial KGE, for each grid cell within D02, the daily mean of observed and simulated rainfall was computed (Supplement Fig. S2). The KGE was subsequently calculated using these observed and simulated grid-cell daily means. To enable a consistent grid-cell-to-grid-cell comparison with IMERG observations, we resampled the WRF-ARW simulated rainfall data to the 0.1° IMERG grid using averaging. This resampling was performed using the xarray package in Python (Hoyer and Hamman, 2017).

Table 2. EREs across the AP selected to determine the performance of different MP and PBL scheme combinations. Simulation start times are provided in UTC. IMERG rainfall values represent 24-hour totals from simulation start for the 0.1° grid-cell with the highest amount for each ERE. Abbreviations: N=north, E=east, S=south, W=west, P=people.

Event Date	**Location**	**Simulation Start**	**Reported Rainfall**	**IMERG Rainfall**	**Fatalities / Impact**	**Source**
24-11-2022	Jeddah, Makkah, Saudi Arabia (W)	22-11-2022 00:00	179 mm	120 mm	2 P died in flooding	FloodList (www.floodlist.com)
27-04-2021	Makkah (W)	25-04-2021 00:00	Unknown	32 mm	Severe flooding reported	FloodList
04-02-2021	Tabuk (NW), Hafr Al-Batin (E)	02-02-2021 00:00	unknown	60 mm	7 P died; 1,100 P affected	General Directorate of Civil Defense (CDD)
27-10-2019	Hafr Al-Batin (E)	25-10-2019 00:00	43 mm in 30 min	30 mm	18 P died, 11 P injured	FloodList
23-05-2019	Jazan, Najran (SW)	21-05-2019 00:00	Unknown	33 mm	1 P missing in floods	FloodList
08-02-2019	Madinah (W), Tabuk (NW), Riyadh (E)	06-02-2019 00:00	36.6 mm in 24 hrs	20 mm	4 P died; many rescued	FloodList
28-01-2019	Tabuk (NW), Riyadh (C), Jeddah	26-01-2019 00:00	Unknown	41 mm	1 P died; 30 P evacuated	CDD
20-11-2017	Jeddah, Hail (W)	18-11-2017 00:00	115.5 mm/hr	73 mm	4 P died; 481 rescued	FloodList
14-02-2017	Asir (SW), Dammam (E)	12-02-2017 00:00	90 mm in 24 hrs	97 mm	1 P died; 10 P injured	CDD
28-11-2016	Asir (SW), Riyadh (C)	26-11-2016 00:00	Unknown	47 mm	8 P died; 120 evacuated	FloodList
08-04-2016	Asir, Baha, Taif (S)	06-04-2016 00:00	Unknown	36 mm	3 P died	FloodList
24-11-2015	Riyadh, Al-Qassim	22-11-2015 00:00	Unknown	35 mm	1P died	FloodList
28-10-2015	Saudi Arabia (N)	26-10-2015 00:00	Unknown	24 mm	6 P died	FloodList
23-03-2015	Riyadh (C), Al Bahah (NW)	21-03-2015 00:00	Unknown	29 mm	11 P died; 300 P rescued	FloodList
20-11-2013	Riyadh (C), Arar	18-11-2013 00:00	Unknown	42 mm	4 P died	CDD
14-01-2011	Jeddah (W)	12-01-2011 00:00	110 mm in 3 hrs	101 mm	10 P died	CDD
30-12-2010	Jeddah (W)	28-12-2010 00:00	Unknown	46 mm	No data	CDD

Table 3. WRF-ARW (Version 4.4) model configuration used in this study.

Configuration Parameter	**Details**
Dynamics	Non-hydrostatics
Boundary and initial conditions	ERA5 reanalysis
Data Interval	3 hours
Grid size	D01 $(116 \times 101) \times 53$, D02: $(187 \times 181) \times 53$
Resolution	D01 9 km and d02 3 km
Map Projection	Mercator
Model top pressure	30 hPa
Land category	USGS (21)
Integration time step	30 s
Vertical coordinates	Terrain-following hydrostatic pressure vertical coordinate with 53 vertical levels
Time integration scheme	3rd-order Runge-Kutta Scheme
Spatial differencing scheme	6th-order centre differencing
Microphysics Parameterization (MP)	Kessler (MP1), Purdue Lin (MP2), WRF Single-moment 3-class (WSM3; MP3), WRF Single-moment 5-class (WSM5; MP4), Eta (Ferrier;MP5), WRF Single-moment 6-class (WSM6; MP6), Goddard (MP7), Thompson graupel (MP8), Morrison 2–moment (MP10)
Cumulus Parameterization (CU)	D01 (Kain Fritsch), D02 (no CU scheme used)
Planetary Boundary Layer (PBL) Parameterization	Yonsei University Scheme (YSU; BL1), Mellor-Yamada Nakanishi and Niino Level 2.5 (BL5), Mellor-Yamada Nakanishi and Niino Level 3 (BL6), BouLac (BL8)
Surface layer parameterization	Noah Land Surface Scheme (Chen and Dudhia, 2001)
Surface Layer Physics	Revised MM5 (Jiménez et al., 2012)
Short wave radiation (ra_sw_physics)	RRTMG scheme (Iacono et al., 2008)
Long wave radiation (ra_lw_physics)	RRTMG scheme

Additionally, to determine whether the performance is significantly different between scheme combinations for rainfall, we calculated ΔKGE scores by subtracting the mean KGE across EREs from the KGE values, thereby eliminating systematic differences in scores among EREs. We then tested whether the distributions of ΔKGE values for different scheme combinations are statistically similar or different using pairwise independent t-tests (Supplement Fig. S3).

For 2-m air temperature, 2-m relative humidity, and 10-m wind speed, KGE was calculated from hourly METAR observations from the IOWA Mesonet and corresponding simulations from the nearest model grid-cell for the day of each ERE.

4 Results and Discussion

4.1 Which PBL scheme performs best in terms of rainfall?

Fig. 2 presents temporal KGE scores for 36 PBL-MP combinations across 17 EREs. As spatial KGE scores (Supplement Fig. S2) exhibit comparable patterns, the analysis here focuses on the temporal scores. The mean temporal and spatial KGE for the PBL schemes—BL1, BL5, BL6, and BL8—are summarized in Table 4. Among these, the BL1 scheme showed superior performance among the PBL schemes (mean KGE of 0.43). Notably, BL1 is the only scheme with a non-local approach, unlike the other schemes, which are all local. This non-local mixing likely explains BL1's superior performance, enabling enhanced vertical mixing across the entire PBL. Non-local schemes like BL1 can represent large eddy structures and transport heat, moisture, and momentum over considerable vertical distances, a capability that is particularly crucial in arid environments with intense surface heating and sharp thermal gradients, such as Saudi Arabia (Hong et al., 2006; Hu et al., 2010). In contrast, local schemes like the BL5, BL6, and BL8 (mean KGE values of 0.38, 0.26, and 0.41, respectively) rely on gradients at specific vertical levels and small-scale turbulence, which restricts their ability to simulate deep convection and rapid vertical mixing.

Previous research has shown that non-local schemes, including BL1, yield a deeper and more accurately structured PBL than local schemes, especially in the presence of strong surface heating and convective activity, which are characteristic of desert climates (Xie et al., 2012; Cohen et al., 2015). Specifically, BL1's non-local treatment of PBL processes allows it to develop a deeper PBL during the daytime, a typical feature in arid regions, enhancing the scheme's ability to capture severe convective activity (Cohen et al., 2015). The performance of BL1 in representing PBL processes is especially advantageous in regions where convection is often triggered by advancing frontal systems, as is common in the AP. In a case study using the WRF-ARW model, Cohen et al. (2015) demonstrated that BL1's non-local treatment improves the PBL's response to cold fronts, triggering convection more realistically and enhancing features like the formation of double lines of intense convection. This improvement arises because BL1 minimizes the dilution of moist air by dry air entrainment, maintaining a higher moisture concentration within the PBL. This "fuel" is crucial for sustaining severe convection when fronts initiate it, particularly in desert regions, where dry air entrainment can otherwise weaken or inhibit intense convective activity and thus reduce the accuracy of ERE simulations.

In contrast, local schemes like BL5 and BL6 and BL8 are optimized for stable or stratified PBLs, typically performing well by simulating small-scale turbulence. However, these schemes often struggle in unstable, highly convective environments like those in Saudi Arabia, where larger eddy structures dominate and require extensive vertical mixing to capture intense updrafts and rainfall (Hu et al., 2013; Cohen et al., 2015). Performance is particularly low for the BL6 scheme (mean KGE of 0.26; Table 4) scheme, sometimes showing negative KGE scores across different MP schemes (Fig. 2). The scheme's higher-order local closure approach can lead to over-diffusion, dampening essential vertical motions and limiting its ability to capture coherent eddies and large-scale vertical transport—critical for effective moisture and heat distribution needed for convective rainfall (Nakanishi and Niino, 2006; Shin and Hong, 2011).

Nevertheless, Schwitalla et al. (2020) reported the best performance with the MP8-BL5 scheme combination in their convection-permitting simulation over the AP for a single ERE on 14 July 2015 (Table 1), which contrasts with our findings. This contrast may be due to differences in the characteristics of that particular ERE, model setup, or surface properties. In particular, their use of a higher vertical resolution (100 levels) may have favored the performance of BL5, a local scheme that strongly depends on accurately resolved vertical gradients. Similarly, the relatively weaker performance of the BL6 and BL8 schemes in our simulations may be partly attributed to the coarser vertical resolution. However, unlike their single-event study, the present research evaluates 17 EREs across the AP spanning multiple seasons and years. This multi-ERE approach is particularly important for identifying parameterization schemes that are consistently reliable under a range of conditions. Since future climate projections cannot be directly validated against observations, selecting robust configurations based on a diverse set of past EREs is essential for improving model confidence in future applications.

4.2 Which MP scheme performs best in terms of rainfall?

Fig. 2 presents temporal KGE scores for 36 PBL–MP combinations across 17 EREs. Since spatial KGE scores (Supplement Fig. S2) demonstrate similar values, the discussion is limited to temporal scores. The mean temporal and spatial KGE for various MP schemes, including MP1, MP2, MP3, MP4, MP5, MP6, MP7, MP8, and MP10, are presented in Table 4. The MP7 and MP8 schemes achieved the highest mean KGE scores. This is likely due to their sophisticated handling of cloud microphysics, especially in representing mixed-phase and ice-phase processes essential for simulating EREs in arid regions like Saudi Arabia. Though MP7 is a single-moment scheme, it includes detailed processes for ice, snow, and graupel, making it effective for capturing intense convective storms driven by complex thermodynamics and rapid cloud development (Tao, 2003). Its optimized treatment of rain formation and melting allows it to handle the rapid updrafts and temperature variations characteristic of desert climates, where efficient particle formation and fallout are crucial for high-intensity EREs.

As a double-moment approach, the MP8 scheme further enhances these capabilities by dynamically adjusting particle size distributions, including cloud droplets, rain, ice and snow. This adaptability allows it to respond effectively to environmental changes typical of desert frontal systems, where intense updrafts can quickly alter particle sizes (Thompson et al., 2008). The double-moment structure offers flexibility in tracking a broad range of particle sizes, enabling MP8 to simulate light and heavy rainfall effectively. This capability is crucial in arid regions, where rapid shifts between intense rainfall and dry conditions are common, and tracking both mass and concentration enhances the accuracy of these transitions.

The superior performance of these schemes over simpler single-moment models, like MP1, MP2, or MP3, underscores the importance of advanced microphysical processes —including graupel and hail processes, multiple ice-phase species, prognostic treatment of various hydrometeors, and more complex interactions between cloud and rainfall particles — for capturing ERE variability and intensity. Simpler schemes lack adaptability to evolving particle size distributions, limiting their effectiveness in highly convective environments with rapid shifts. Notably, despite its advanced double-moment structure, Morrison underperformed, possibly due to interactions with other model components that may hinder accuracy in arid, convective conditions—a point warranting further research beyond this study. Overall, our results highlight the importance of selecting MP schemes with detailed ice and mixed-phase processes when modeling EREs in desert regions.

Table 4. Mean KGE values for temporal and spatial assessments of MP and PBL schemes.

Scheme	**Temporal KGE**	**Spatial KGE**
MP Schemes		
Kessler (MP1)	0.26	0.05
Purdue Lin (MP2)	0.35	0.27
WRF Single-Moment 3-class (WSM3; MP3)	0.41	0.30
WRF Single-Moment 5-class (WSM5; MP4)	0.39	0.25
Eta Ferrier (MP5)	0.39	0.28
WRF Single-Moment 6-class (WSM6; MP6)	0.36	0.28
Goddard (MP7)	0.42	0.33
Thompson (MP8)	0.42	0.31
Morrison (MP10)	0.30	0.29
PBL Schemes		
YSU (BL1)	0.43	0.29
Mellor-Yamada Nakanishi Niino Level 2.5 (MYNN Level 2.5; BL5)	0.38	0.27
Mellor-Yamada Nakanishi Niino Level 3 (MYNN Level 3; BL6)	0.26	0.21
Boulac (BL8)	0.41	0.27

4.3 Which component of the Kling-Gupta Efficiency (KGE) affects the final rainfall scores the most?

Fig. 3a presents the values of KGE and its components — correlation, bias, and variability (r, γ, and β, respectively; Eq. 1) — for all 17 EREs for the best performing MP8_BL1 scheme combination for rainfall (Fig. 2 and Table 4). In the interest of conciseness, we focus only on the temporal KGE results here, as the spatial KGE results are quite consistent (see Sections 4.1, 4.2, and 4.5, and Table 4).

Correlation is sensitive to the timing of EREs, variability ratio is sensitive to the distribution, and bias reflects the mean. For the best combination (MP8_BL1), the mean temporal KGE score for rainfall across 17 EREs is 0.48. Decomposing this score into the three components, expressed as $|r-1|$, $|\beta-1|$ and $|\gamma-1|$ to make them comparable, yields mean absolute values of 0.33, 0.23, 0.25, respectively, where values closer to 0 indicate better performance. Among the three KGE components, the scheme thus performed worst in terms of correlation, and this subcomponent thus exerted the dominant influence on the final KGE scores. This suggests that in order to get an improved KGE score, the most important component score to improve is the correlation, which, in the temporal assessment, is related to the timing of EREs. The mean KGE value across all other schemes and EREs is 0.36, and the mean values for $|r-1|$, $|\beta-1|$ and $|\gamma-1|$ are 0.34, 0.29, and 0.24, respectively. This suggests that the correlation also tends to exert the dominant influence for the other scheme combinations, while bias also plays a role. The mean KGE score for the worst-performing scheme combination MP10_BL6 is 0.13, while the mean values of the three KGE

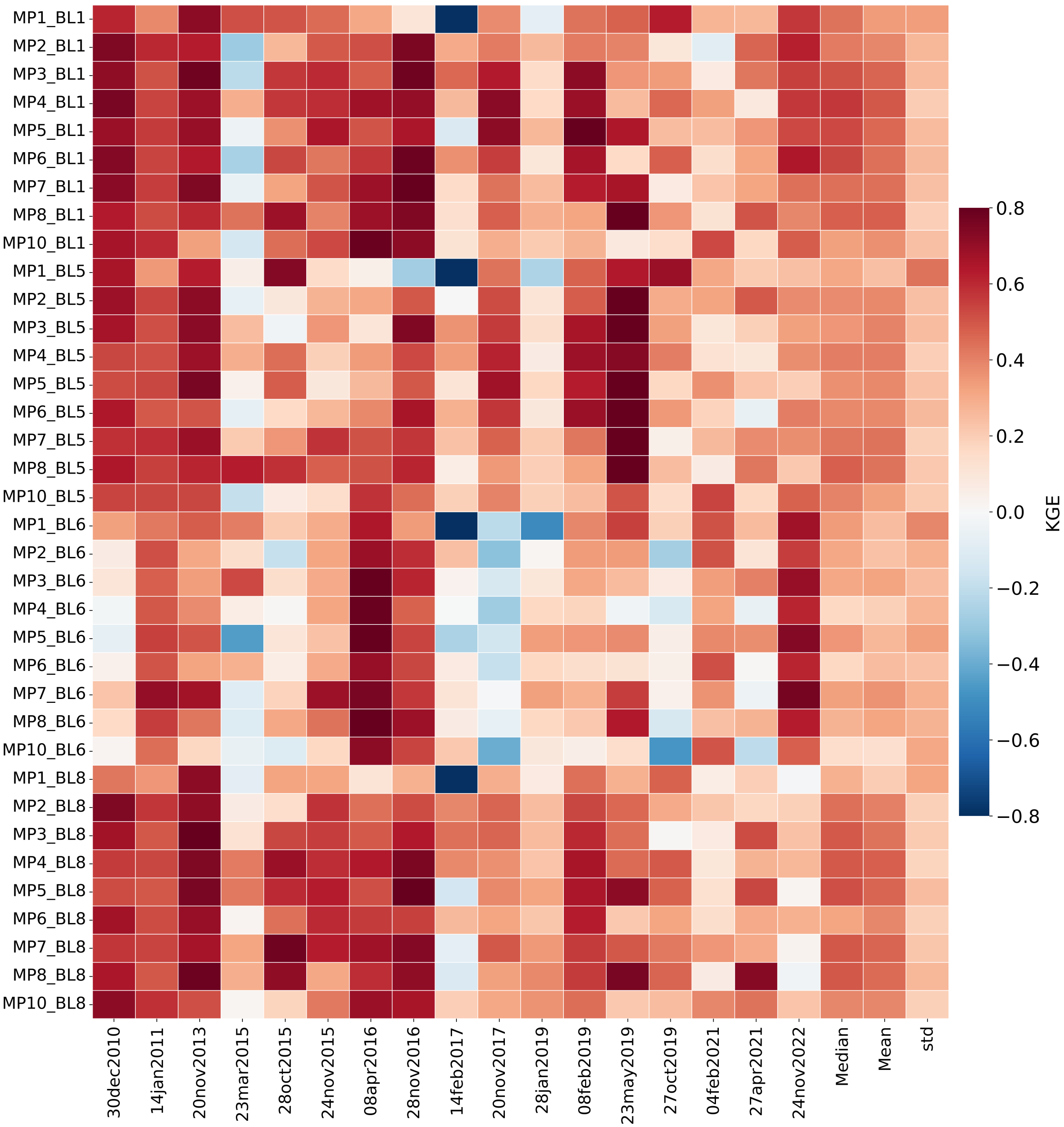

Fig. 2. Temporal KGE scores for rainfall derived from 36 WRF-ARW scheme combinations across 17 EREs. The scores were calculated by comparing hourly WRF-ARW simulated rainfall against IMERG V7 satellite rainfall data over the 24-hour event day.

components $|r-1|$, $|\beta-1|$, and $|\gamma-1|$ are 0.33, 0.57, and 0.36, respectively. This scheme thus performs particularly poorly in terms of bias.

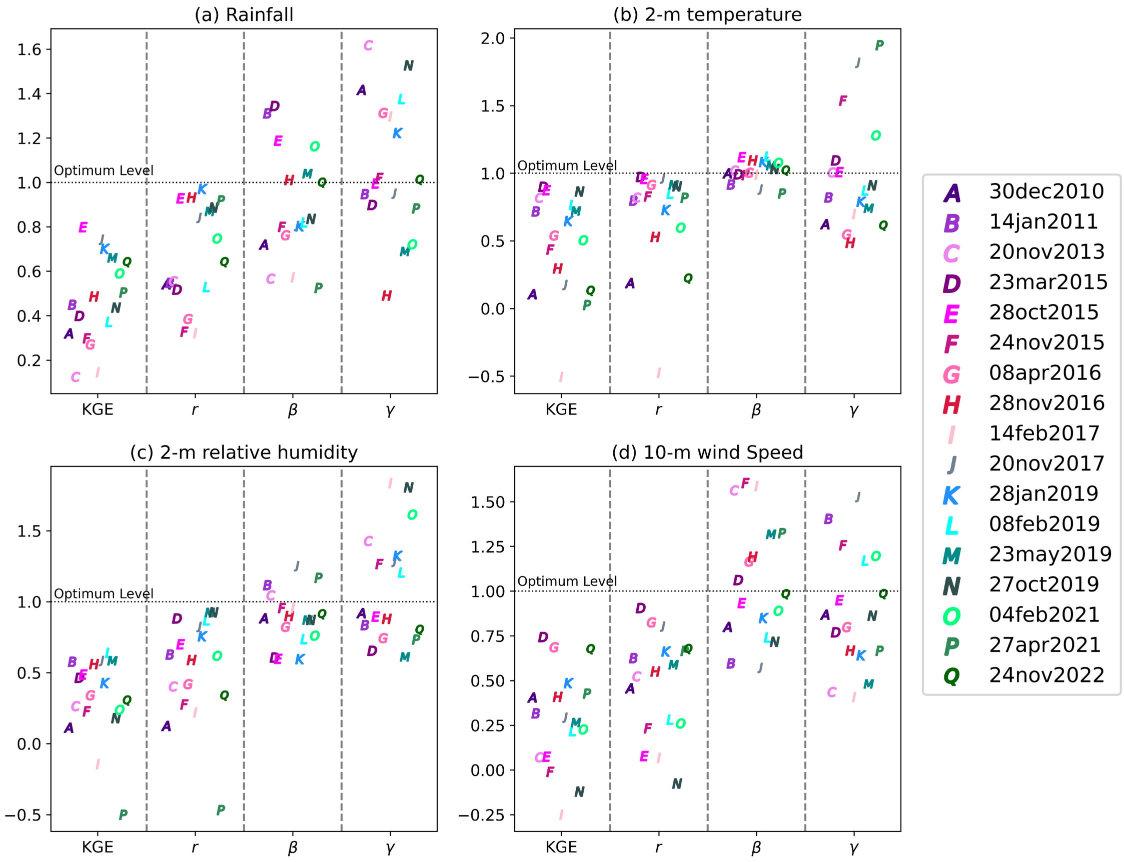

Fig. 3. Correlation coefficient (r), bias (β), and variability ratio (γ) values used to calculate the KGE values for the best-performing combination across 17 EREs for (a) Rainfall, (b) 2-m air temperature, (c) 2-m relative humidity, and (d) 10-m wind speed. Panel (a) uses IMERG V7 as reference and panels (b–d) METAR observations over each 24-hour event day. The letters (A, B, . . . ,Q) indicate the 17 different EREs (Table 2).

4.4 How statistically significant are the differences in performance among scheme combinations in terms of rainfall?

The differences in KGE between different scheme combinations for rainfall are generally relatively small. For example, the best-performing scheme combination (MP8_BL1) achieved a mean KGE of 0.48, while the second-best-performing scheme combination (MP7_BL1) achieved a mean KGE of 0.44 (Fig. 2). Furthermore, the corresponding standard deviations across EREs are 0.20 and 0.24, respectively, indicating substantial variability in scores among EREs. Additionally, the consistency in performance ranking among EREs is fairly low (Fig. 5). This raises the question whether the observed differences in performance between scheme combinations are statistically significant and, hence, whether our evaluation approach is adequate for determining the relative performance of different scheme combinations, which is the primary objective of this study.

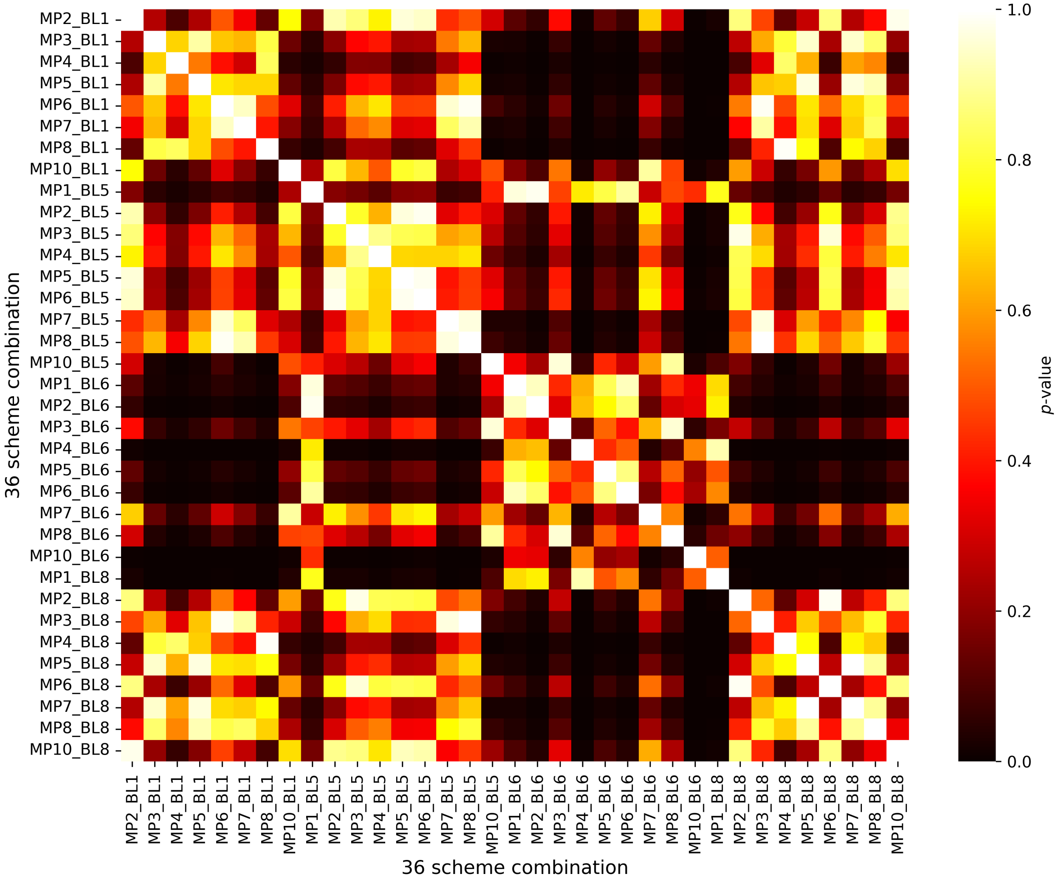

Fig. 4. Pairwise p-values from independent t-tests comparing the ΔKGE distributions of 36 scheme combinations for rainfall. ΔKGE values were calculated by subtracting the mean KGE across EREs from the KGE values presented in Fig. 2. A p threshold of 0.1 was used to identify statistically significant differences between scheme combinations.

Fig. 4 presents a 36x36 matrix of pairwise p-values from independent t-tests comparing ΔKGE distributions of 36 scheme combinations for rainfall. ΔKGE values were calculated by subtracting the mean KGE across EREs from the KGE values presented in Fig. 2, thereby eliminating systematic differences in scores among EREs. The results reveal that the best-performing scheme combination (MP8_BL1) significantly outperforms 21 other scheme combinations (at a p-level of 0.1), whereas the worst-performing scheme combination (MP10_BL6) performed significantly worse than 28 other scheme combinations (also at a p-level of 0.1). These results confirm that our assessment provides meaningful and statistically significant insights into the relative performance of different scheme combinations. However, our assessment does not definitively identify a single best-performing scheme but instead highlights groups of better- and worse-performing schemes.

The ability of an assessment such as this to detect significant differences in performance between schemes depends on the mean and standard deviation of the ΔKGE distribution. Assuming a standard deviation of 0.15 (equivalent to that of

MP8_BL1), the current sample size of 17 EREs requires a mean ΔKGE difference greater than 0.06 between schemes to yield a statistically significant difference at a p-level of 0.1. Analyzing a larger sample of EREs would reduce the required mean difference, making it easier to detect significant differences in performance between schemes. For example, if we were to analyze 50 EREs, the required difference in mean ΔKGE would be just 0.03 (assuming again a standard deviation of 0.15). However, analyzing a larger number of EREs is computationally more expensive.

The standard deviation (i.e., the variability in ΔKGE among EREs) and hence the number of EREs required to detect significant performance differences between schemes are partly influenced by the quality of the reference data. In this study, we used a microwave satellite-based rainfall product (IMERG-Final V07), which is associated with greater uncertainty than gauge-radar-based datasets (Beck et al., 2019b). This increased uncertainty may have contributed to higher variability in KGE scores (Evans and Imran, 2024). Unfortunately, radar data are only commercially available in Saudi Arabia. Due to the strong correlation between different microwave satellite-based rainfall datasets — such as IMERG, GSMaP (Kubota et al., 2024), and CMORPH-CDR (Xie et al., 2019) — and the fact that IMERG-Final V7 significantly outperforms other satellite datasets (Wang et al., 2025), we were unable to quantify the uncertainty arising from the choice of reference data as done by Evans and Imran (2024).

4.5 How consistent are the temporal and spatial performance assessments for rainfall?

We calculated KGE scores both temporally and spatially to assess the performance of the 36 PBL-MP scheme combinations across the 17 EREs. The temporal KGE results for rainfall are presented in Fig. 2, while the spatial KGE results for rainfall are provided in Supplement Fig. S2. The mean KGE values categorized by MP and PBL schemes, for both temporal and spatial assessments, are summarized in Table 4. The overall mean temporal KGE across all schemes and EREs for rainfall is 0.37, whereas the overall mean spatial KGE is 0.26. This indicates that the simulations are more effective at capturing temporal variations in rainfall than spatial variations. This is expected, as accurately simulating the location of localized convective systems remains a major challenge. Overall, we found a strong consistency in the overall ranking of schemes between the temporal and spatial assessments, with a Spearman rank correlation of 0.65 ($p < 0.001$) between the mean temporal and spatial KGE values for the scheme combinations. The MP7 and MP8 schemes, when combined with BL1, consistently ranked highest across both temporal and spatial KGE assessments (Fig. 2; Supplement Fig. S2; Table 4). Conversely, the MP1 scheme with BL6 scheme performed worst in both assessments.

4.6 How consistent is the performance ranking among different variables?

Ideally, if our conclusions about the performance of various MP and PBL scheme combinations regarding rainfall are valid, and if this superior performance truly reflects a model that better represents reality (i.e., we are getting the right results for the right reasons; Kirchner, 2006a), then the performance ranking for rainfall should align with those of the other variables (2-m air temperature, 2-m relative humidity, and 10-m wind speed). Indeed, for all variables, MP8_BL1 provided the highest mean temporal KGE (Fig. 2 and Table 4), tentatively suggesting that this particular scheme combination does indeed yield a more robust model in all respects.

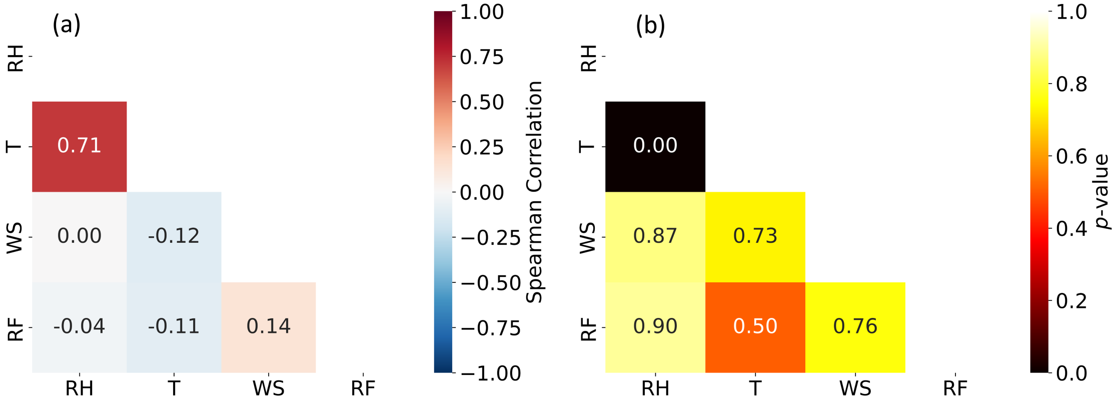

Fig. 5. (a) Mean Spearman correlation coefficients and (b) corresponding median p-values calculated among mean KGE scores for different meteorological variables, indicating the degree of consistency in performance rankings among variables. Variable definitions: 2-m relative humidity = RH; 2-m air temperature = T; 10-m wind speed = WS; and rainfall = RF.

Additionally, we calculated Spearman rank correlations and corresponding p-values between the temporal mean KGE scores for the different variables (Fig. 5), to examine the degree of consistency in performance rankings among the variables. Most variable pairs exhibited insignificant correlations except for temperature and relative humidity, which are intrinsically linked through the Clausius-Clapeyron relationship as temperature controls saturation vapor pressure and, thus, relative humidity. The lack of significant correlations might have three potential explanations. First, uncertainties in the reference data may cause discrepancies in model performance; the significant uncertainty in IMERG for rainfall (Wang et al., 2025), along with the difficulty of comparing point-based IOWA Environmental Mesonet data to WRF-ARW grid cells for other variables, makes this explanation plausible. Second, although MP and PBL schemes strongly influence rainfall simulation, other model components like land surface schemes, which affect soil moisture and heat fluxes, and radiation schemes, which affect surface and atmospheric energy balances, may have a more pronounced impact on variables such as temperature and wind speed. Third, there might be compensatory behavior within the model, where improvements in simulating one variable do not necessarily result in a more realistic simulation and may yield reduced performance in others.

This phenomenon, where models achieve the right results for the wrong reasons, is not uncommon in geosciences and poses significant challenges in model evaluation and improvement (Kirchner, 2006b; Parker, 2006; Knutti, 2010; Hourdin et al., 2017; Broecker, 2017; Krantz et al., 2021). Resolving this requires examining model structure and variable interactions more closely to determine if improvements reflect real accuracy or trade-offs, which is beyond the scope of the current study.

4.7 What do the spatial patterns in simulated and observed rainfall look like for the EREs?

Figs. 6 and 7, respectively, present observed (IMERG-Final V07) and simulated (WRF-ARW) 24-hour rainfall accumulations for the 17 selected EREs. The WRF-ARW model simulations were generated using the best-performing scheme (MP8_BL1).

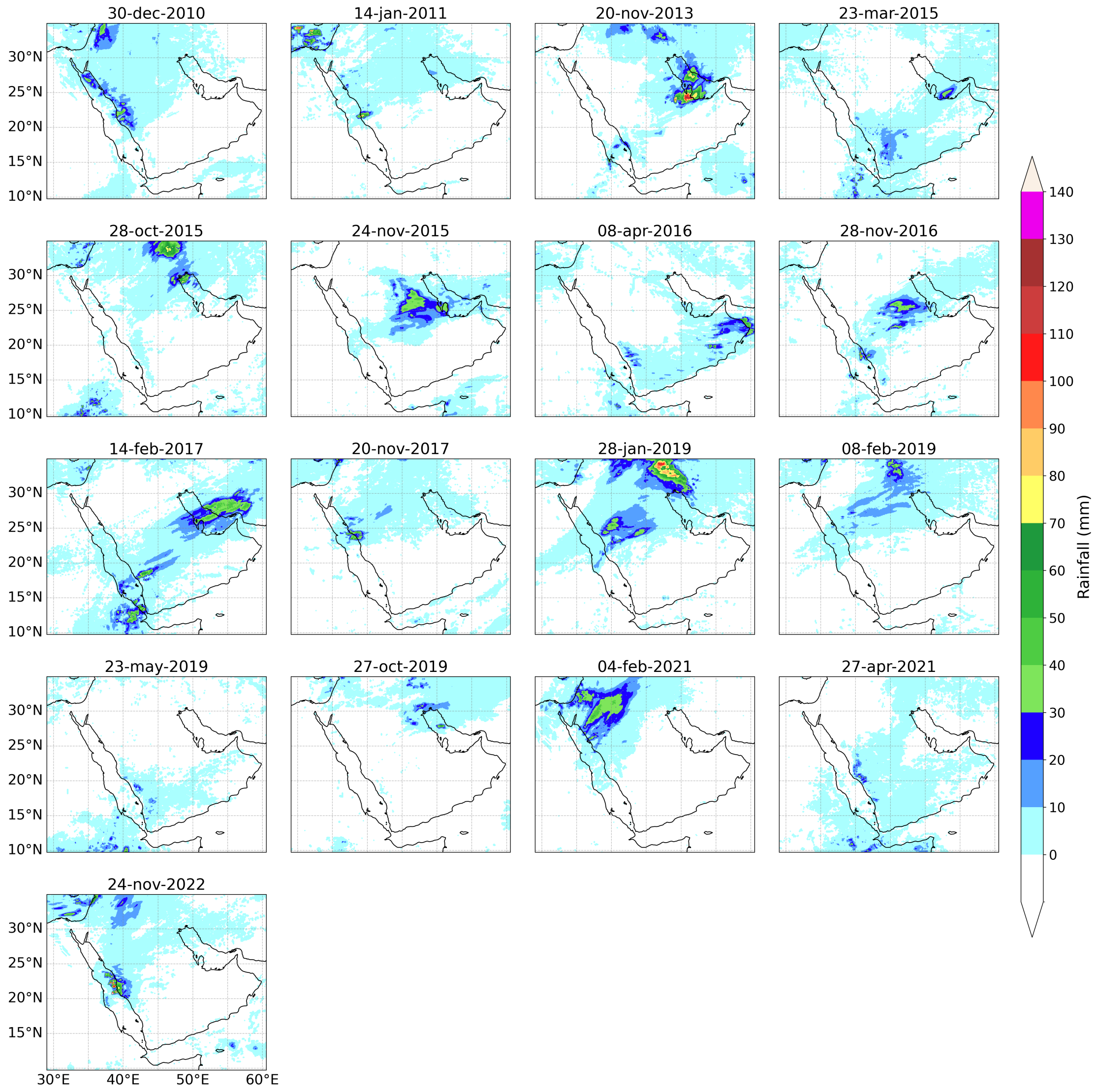

Fig. 6. Daily accumulated rainfall from our observation-based data source (IMERG-Final V07) for the 17 EREs.

Overall, WRF-ARW generally seems to capture reasonably well the location, extent, and amounts indicated by IMERG. For 325 example, the strong convective systems with high-intensity localized rainfall exceeding 120 mm on EREs like 20-Nov-2013 and 28-Jan-2019 are captured well. However, the model overestimates rainfall for several EREs (e.g., 08-Feb-2019) and underestimates rainfall for others (e.g., 28-Oct-2015). While WRF-ARW generally captures the broad patterns, the lack of a better match is attributable to several reasons. First, potential deficiencies in the MP, BL, and convection schemes, along with other modeling limitations, can lead to inaccuracies in moisture convergence and convective updrafts (Taraphdar et al., 2021; Attada

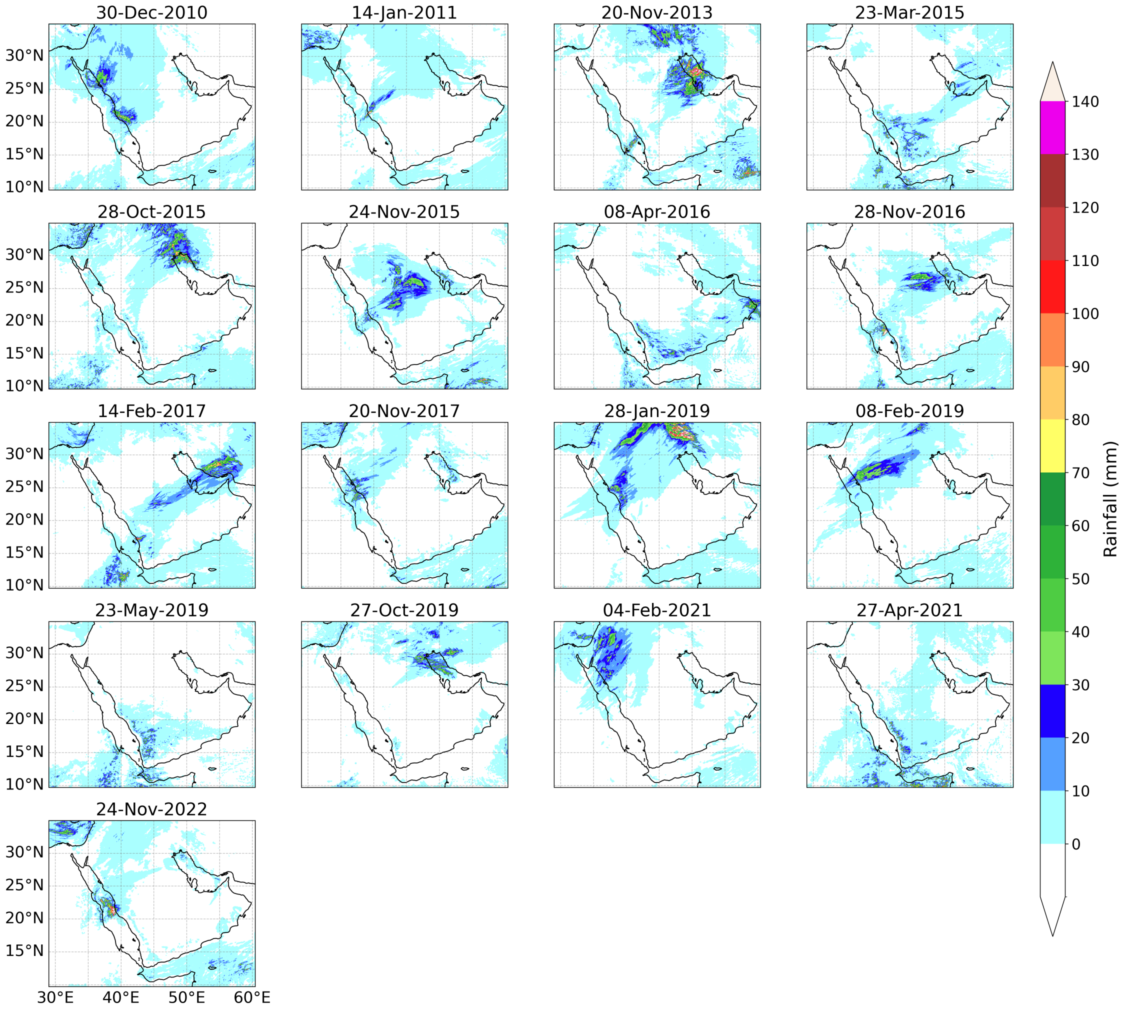

Fig. 7. Daily accumulated rainfall from WRF-ARW using the best performing scheme combination (MP8_BL1) for the 17 EREs

.

et al., 2022). These limitations include simplified representations of land–atmosphere interactions, unresolved sub-grid processes, and the use of prescribed lateral boundary conditions updated every 6 hours, which may not fully capture fast-evolving or small-scale features entering the domain. Second, we used ERA5 as boundary conditions to force the model, and while ERA5 is the best reanalysis currently available, it nonetheless is subject to random errors and bias (Hersbach et al., 2020; Soci et al., 2024). Third, we did not include data assimilation or nudging (Lei and Hacker, 2015; Feng et al., 2021), two important techniques to improve the simulations. Fourth and finally, the IMERG data, though found to perform relatively well in precipitation product evaluations (Abbas et al., 2025; Wang et al., 2025), nonetheless carries significant uncertainty in the region.

4.8 How well does the model perform in terms of the other variables?

While the previous subsections focused primarily on rainfall, it is worthwhile to investigate how the model performs in terms of other meteorological variables. To this end we analyzed the KGE components for 2-m air temperature, 2-m relative humidity, and 10-m wind speed as presented in Figs. 3b, 3c, and 3d, respectively. Fig. 3b presents the KGE and its components (r, γ, and β) for all 17 EREs for temperature using the best-performing combination (MP8_BL1). For this scheme, the mean temporal KGE score across the 17 EREs is 0.47, which is similar to that obtained for rainfall (0.48). This is somewhat unexpected, as temperature is constrained by surface energy balance processes, resulting in smoother variations and less extreme variability compared to rainfall. The mean values for $|r-1|$, $|\beta-1|$ and $|\gamma-1|$ for temperature are 0.32, 0.06, and 0.33, respectively. Among the three KGE components, the scheme thus performed worst in terms of correlation and variability, which therefore exert the dominant influence on the final KGE scores.

Fig. 3c presents the KGE and its components for the 17 EREs for relative humidity using the best performing scheme (MP8_BL1). For this scheme, the mean temporal KGE score across 17 EREs is 0.31, which is lower than that obtained for rainfall and temperature. This may reflect relative humidity's nonlinear dependence on both temperature and moisture in addition to the high spatio-temporal variability. The mean values for $|r-1|$, $|\beta-1|$ and $|\gamma-1|$ are 0.47, 0.18, 0.33, respectively. Among the three KGE components, the scheme thus performed worst in terms of correlation, followed by variability, which therefore exert the dominant influence on the final KGE scores.

Fig. 3d presents the KGE and its components for the 17 EREs for wind speed using the best performing scheme (MP8_BL1). For this scheme, the mean temporal KGE score across the 17 EREs is 0.29, the lowest among the four variables, likely due to the influence of fine-scale topography and surface roughness variability on wind speed. The mean values for $|r-1|$, $|\beta-1|$ and $|\gamma-1|$ are 0.52, 0.28, 0.30, respectively. Among the three KGE components, the scheme thus performed worst in terms of correlation, which therefore exerts the dominant influence on the final KGE scores.

4.9 How do the PBL and MP schemes used in previous studies compare with those identified as optimal in our evaluation?

Although our findings are subject to uncertainty and must be interpreted with caution, as highlighted in the previous subsections, they provide a useful basis for evaluating schemes used in previous WRF-ARW studies in the region. Our review of these studies (Table 5) reveals varying choices of PBL and MP schemes, with mixed alignment to the results of this study. Several studies, such as those by Abida et al. (2022), Almazroui et al. (2018), and Patlakas et al. (2023), used the BL1 scheme, which our results confirm as the best-performing scheme for capturing the unique convective dynamics in arid climates. These studies highlighted BL1's robust vertical mixing capabilities and adaptability to desert environments. On the other hand, studies like Attada et al. (2020) and Taraphdar et al. (2021), which employed BL6 and QNSE (BL4), respectively, used local turbulence schemes that our findings show may be less suited for unstable, highly convective conditions typical in the region. Similarly, while MP schemes like MP8 and MP7, identified in our study as well-performing, were used in some cases (Taraphdar et al., 2021; Attada et al., 2020), other studies, such as Deng et al. (2015), relied on simpler MP schemes like MP2 and MP5, which

Table 5. Studies simulating EREs in the Middle East using WRF-ARW.

Study	MP Scheme	PBL Scheme	Key Findings
Luong et al. (2020)	Morrison (MP10)	Mellor-Yamada-Janjic (MYJ; BL2)	Evaluated urbanization impacts on ERE over Jeddah; high-resolution models essential for urban storm simulation.
Francis et al. (2024)	Thompson aerosol-aware (MP28)	Quasi-Normal Scale Elimination (QNSE; BL4)	Enhanced performance in capturing rainfall patterns for EREs involving atmospheric rivers in the Middle East.
Deng et al. (2015)	Lin (MP2), Eta Ferrier (MP5)	Mellor-Yamada-Janjic (MYJ; BL2)	Demonstrated role of different MP schemes in capturing Jeddah flash-flood events.
Attada et al. (2020)	Thompson (MP8)	MYNN Level 3 (BL6)	Demonstrates consistent skill in simulating rainfall associated with EREs over arid regions of the AP
Taraphdar et al. (2021)	Thompson (MP8)	Quasi-Normal Scale Elimination (QNSE; BL4)	Optimal pairing for rainfall simulation under 9-km resolution, balancing accuracy and efficiency in UAE simulations.
Abida et al. (2022)	WSM 3-class (MP3)	YSU (BL1)	Best performance in hyper-arid coastal regions, enhancing temperature, humidity, and wind accuracy at BNPP site.
Almazroui et al. (2018)	Eta Ferrier (MP5)	YSU (BL1)	Highlighted YSU's reliability for PBL dynamics in ERE conditions (e.g., Jeddah 2009 event).
Patlakas et al. (2023)	Single-moment six-class (MP6)	YSU (BL1)	YSU's adoption in operational forecasting at the Saudi National Center for Meteorology for its robustness in arid climates.

may lack the sophistication needed to capture mixed-phase processes in intense convective systems fully. Thus, while several studies employed schemes previously shown to perform well in similar regional contexts, others might have improved simulation accuracy by incorporating the BL1 scheme and advanced MP schemes identified as effective in our study. However, we would like to reiterate that our findings are subject to uncertainty, and these conclusions should therefore be interpreted with caution.

5 Conclusion

This study evaluated the performance of PBL and MP parameterizations for simulating EREs in the AP using the WRF-ARW model at a convection-permitting resolution, serving as a verification study for hydrometeorology in the region. The results show that the model captures temporal rainfall variations (mean KGE = 0.37) more effectively than spatial patterns (mean KGE = 0.26), reflecting the localized nature of rainfall in the region. Nonetheless, a strong correlation (Spearman rank correlation of 0.65, p-value = 0.00) between temporal and spatial KGE rankings highlights consistency in scheme performance. This verification is crucial for improving confidence in hydrometeorological modeling and forecasting, particularly for regions prone to flash floods and extreme rainfall. Thus, the findings guide model selection and a vital validation benchmark for future hydrometeorological research and operational forecasting in desert climates.The answers to the questions posed in the introduction, each addressed in detail in the Results and Discussion, are as follows:

a. Which PBL scheme performs best in terms of rainfall?

   The BL1 scheme outperformed the other PBL schemes, achieving a mean temporal KGE of 0.43 and a mean spatial KGE of 0.29. This superior performance is attributed to non-local mixing, which enhances vertical transport and convective processes and makes it particularly effective for simulating ERE in arid regions like the AP. In contrast, local schemes such as BL5, BL6, and BL8 performed worse because they rely on small-scale turbulence, which limits the representation of deep convection.

b. Which MP scheme performs best in terms of rainfall?

   The MP7 and MP8 schemes performed best, achieving a mean temporal KGE of 0.42, with mean spatial KGEs of 0.33 and 0.31, respectively. Their strong performance is attributed to their advanced mixed-phase and ice-phase microphysics. MP8's double-moment structure enhances adaptability, while MP7's optimized ice and graupel processes improve convective simulations. These results highlight the benefit of advanced MP schemes for accurately modeling EREs in arid regions.

c. Which component of the Kling-Gupta Efficiency (KGE) affects the final rainfall scores the most?

   Among the three KGE components (correlation, bias ratio, and variability), correlation and variability exerted the strongest influence on the temporal rainfall KGE scores. Enhancing these components should be prioritized to further improve the accuracy of ERE simulations.

d. How statistically significant are the differences in performance between scheme combinations in terms of rainfall?

   Pairwise statistical tests between distributions of temporal KGE scores obtained by the scheme combinations revealed that the BL1_MP8 combination significantly outperformed 21 other scheme combinations, while the poorest-performing combination, MP10_BL6, was statistically inferior to 28 other combinations. Thus, we could not statistically identify a single best- or worst-performing combination, despite the large sample of 17 EREs.

e. How consistent are the temporal and spatial performance assessments for rainfall?

   The assessment reveals that BL1_MP7 and BL1_MP8 performed best in both the temporal and spatial assessment for rainfall. The higher mean temporal KGE (0.37) compared to the mean spatial KGE (0.26) for all 36 combinations indicates that the model captures rainfall variability more effectively over time than across space. Although spatial KGE values were lower, the ranking of combination performance remained consistent (Spearman rank correlation of 0.65).

f. How consistent is the performance ranking among different variables?

The MP8_BL1 combination provided the best performance for all variables (rainfall, 2-m air temperature, 2-m relative humidity, and 10-m wind speed). However, we obtained weak correlations between performance rankings across the variables, indicating poor consistency. This is likely because different physical processes govern the simulations of different variables. That is, while MP and PBL schemes influence rainfall, other components, such as land surface and radiation schemes, affect temperature and wind. This underlines the complexity of model parameterization, particularly as cloud evolution is influenced not only by PBL and MP schemes but also by radiative processes, emphasizing the need for further integrated research.

g. What do the spatial patterns in simulated and observed rainfall look like for the EREs?

For the best-performing physics combination (MP8_BL1), the spatial patterns of simulated and observed rainfall were generally well captured, although occasional overestimations and underestimations were noted. These discrepancies are likely attributable to limitations in the boundary conditions (the ERA5 reanalysis) and uncertainties in the observations (the IMERG satellite-based rainfall product).

h. How well does the model perform in terms of the other variables?

Using the best-performing scheme combination (MP8_BL1), air temperature showed a mean temporal KGE score of 0.47, similar to that of rainfall (0.48), with performance limited mainly by correlation and variability. Relative humidity had a lower mean temporal KGE score (0.31), like due to its nonlinear dependence on temperature and moisture, with correlation as the dominant error source. Wind speed had the poorest performance (mean temporal KGE of 0.29), likely due to unresolved fine-scale topographic and surface roughness effects.

i. How do the PBL and MP schemes used in previous studies compare with those identified as optimal in our evaluation?

Our findings align with several previous studies in the Middle East that employed the BL1 scheme, reinforcing its effectiveness for simulating regional atmospheric dynamics. At the same time, our results suggest that studies using simpler MP schemes — such as MP2 or MP5 — may achieve improved simulation accuracy by adopting more advanced schemes like MP8.

By identifying the optimal PBL and MP combination from 36 tested configurations across 17 EREs, we established a strong foundation for improving the accuracy of ERE simulations across the AP, a region that remains understudied despite frequent flash floods and significant casualties. As the most comprehensive evaluation of PBL and MP schemes in the AP to date, our study emphasizes the importance of parameterization choices on ERE simulation performance, serving as a key reference for future modeling efforts. Our results may guide researchers and forecasters in selecting the most effective parameterization schemes, ultimately contributing to more reliable forecasting and enhanced disaster preparedness in arid environments. To further advance ERE simulation fidelity, future work should extend beyond PBL and MP schemes to systematically evaluate the impact of land surface schemes, radiation parameterizations, and data assimilation techniques.

Code availability. The code used to generate the results of this study is available from the corresponding author upon request.

Data availability. The IMERG rainfall data are available via NASA GES DISC(https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGHH_07/summary?keywords=GPM_IMERG), while the ERA5 data is available at the Copernicus Climate Data Store (CDS; https://cds.climate.copernicus.eu). The radiosonde data available on the University of Wyoming's upper-air sounding archive (https://weather.uwyo.edu/upperair/sounding.html) while surface meteorological data is accessible via the Iowa State University Environmental Mesonet website (https://mesonet.agron.iastate.edu/request/download.phtml?network=SA__ASOS).

Author contributions. RKS: modeling, analysis, visualization, and writing. HB: initial idea, conceptualization, writing, oversight of analysis, and project administration. All coauthors contributed to writing, revising, and refining the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. For part of our analysis, we used resources from the Shaheen supercomputer at King Abdullah University of Science and Technology (KAUST) in Thuwal, Saudi Arabia. The authors acknowledge the European Center for Medium-Range Weather Forecasts (ECMWF), National Aeronautical and Space Administration (NASA), University of Wyoming, and Iowa state university for the ERA5, GPM, radiosonde, and METAR datasets, respectively.

References

Abbas, A., Yang, Y., Pan, M., Tramblay, Y., Shen, C., Ji, H., Gebrechorkos, S. H., Pappenberger, F., Pyo, J. C., Feng, D., et al.: Comprehensive global assessment of 23 gridded precipitation datasets across 16,295 catchments using hydrological modeling, EGUsphere, 2025, 1–31, 2025.

Abida, R., Addad, Y., Francis, D., Temimi, M., Nelli, N., Fonseca, R., Nesterov, O., and Bosc, E.: Evaluation of the performance of the WRF model in a hyper-arid environment: A sensitivity study, Atmosphere, 13, 985, 2022.

Abosuliman, S. S., Kumar, A., and Alam, F.: Flood disaster planning and management in Jeddah, Saudi Arabia — A Survey, in: Proceedings of the 2014 International Conference on Industrial Engineering and Operations Management Bali, Indonesia, January 7–9, 2014, 2014.

Al Saud, M.: Assessment of flood hazard of Jeddah area 2009, Saudi Arabia, 2010.

Allan, R. P. and Soden, B. J.: Atmospheric warming and the amplification of precipitation extremes, Science, 321, 1481–1484, 2008.

Almazroui, M.: Calibration of TRMM rainfall climatology over Saudi Arabia during 1998–2009, Atmospheric Research, 99, 400–414, 2011.

Almazroui, M., Raju, P., Yusef, A., Hussein, M., and Omar, M.: Simulation of extreme rainfall event of November 2009 over Jeddah, Saudi Arabia: the explicit role of topography and surface heating, Theoretical and applied climatology, 132, 89–101, 2018.

Atif, R. M., Almazroui, M., Saeed, S., Abid, M. A., Islam, M. N., and Ismail, M.: Extreme precipitation events over Saudi Arabia during the wet season and their associated teleconnections, Atmospheric Research, 231, 104 655, 2020.

Attada, R., Dasari, H. P., Kunchala, R. K., Langodan, S., Kumar, K. N., Knio, O., and Hoteit, I.: Evaluating cumulus parameterization schemes for the simulation of Arabian Peninsula winter rainfall, Journal of Hydrometeorology, 21, 1089–1114, 2020.

Attada, R., Dasari, H. P., Ghostine, R., Kondapalli, N. K., Kunchala, R. K., Luong, T. M., and Hoteit, I.: Diagnostic evaluation of extreme winter rainfall events over the Arabian Peninsula using high-resolution weather research and forecasting simulations, Meteorological Applications, 29, e2095, 2022.

Babu, C., Jayakrishnan, P., and Varikoden, H.: Characteristics of precipitation pattern in the Arabian Peninsula and its variability associated with ENSO, Arabian Journal of Geosciences, 9, 1–12, 2016.

Barth, H.-J. and Steinkohl, F.: Origin of winter precipitation in the central coastal lowlands of Saudi Arabia, Journal of arid environments, 57, 101–115, 2004.

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., Van Dijk, A. I., Huffman, G. J., Adler, R. F., and Wood, E. F.: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS, Hydrology and Earth System Sciences, 23, 207–224, 2019a.

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., van Dijk, A. I. J. M., Huffman, G. J., Adler, R. F., and Wood, E. F.: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS, Hydrology and Earth System Sciences, 23, 207–224, 2019b.

Bougeault, P. and Lacarrere, P.: Parameterization of orography-induced turbulence in a mesobeta–scale model, Monthly weather review, 117, 1872–1890, 1989.

Broecker, W.: When climate change predictions are right for the wrong reasons, Climatic Change, 142, 1–6, 2017.

Chen, F. and Dudhia, J.: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity, Monthly weather review, 129, 569–585, 2001.

Chen, S.-H. and Sun, W.-Y.: A one-dimensional time dependent cloud model, Journal of the Meteorological Society of Japan. Ser. II, 80, 99–118, 2002.

Cohen, A. E., Cavallo, S. M., Coniglio, M. C., and Brooks, H. E.: A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern US cold season severe weather environments, Weather and forecasting, 30, 591–612, 2015.

De Vries, A., Feldstein, S. B., Riemer, M., Tyrlis, E., Sprenger, M., Baumgart, M., Fnais, M., and Lelieveld, J.: Dynamics of tropical–extratropical interactions and extreme precipitation events in Saudi Arabia in autumn, winter and spring, Quarterly Journal of the Royal Meteorological Society, 142, 1862–1880, 2016.

De Vries, A. J., Tyrlis, E., Edry, D., Krichak, S., Steil, B., and Lelieveld, J.: Extreme precipitation events in the Middle East: dynamics of the Active Red Sea Trough, Journal of Geophysical Research: Atmospheres, 118, 7087–7108, 2013.

Deng, L., McCabe, M. F., Stenchikov, G., Evans, J. P., and Kucera, P. A.: Simulation of flash-flood-producing storm events in Saudi Arabia using the weather research and forecasting model, Journal of Hydrometeorology, 16, 615–630, 2015.

Dudhia, J.: A history of mesoscale model development, Asia-Pacific Journal of Atmospheric Sciences, 50, 121–131, 2014.

Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R., and Mearns, L. O.: Climate extremes: observations, modeling, and impacts, science, 289, 2068–2074, 2000.

El Kenawy, A. M., McCabe, M. F., Stenchikov, G. L., and Raj, J.: Multi-decadal classification of synoptic weather types, observed trends and links to rainfall characteristics over Saudi Arabia, Frontiers in Environmental Science, 2, 37, 2014.

Evans, J. and Imran, H.: The observation range adjusted method: a novel approach to accounting for observation uncertainty in model evaluation, Environmental Research Communications, 6, 071 001, 2024.

Evans, J. and Smith, R.: Water vapor transport and the production of precipitation in the eastern Fertile Crescent, Journal of Hydrometeorology, 7, 1295–1307, 2006.

Evans, J. P., Smith, R. B., and Oglesby, R. J.: Middle East climate simulation and dominant precipitation processes, International Journal of Climatology, 24, 1671–1694, https://doi.org/https://doi.org/10.1002/joc.1084, 2004.

Feng, T., Hu, Z., Tang, S., and Huang, J.: Improvement of an extreme heavy rainfall simulation using nudging assimilation, Journal of Meteorological Research, 35, 313–328, 2021.

Fowler, H. J., Lenderink, G., Prein, A. F., Westra, S., Allan, R. P., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Do, H. X., et al.: Anthropogenic intensification of short-duration rainfall extremes, Nature Reviews Earth & Environment, 2, 107–122, 2021.

Francis, D., Fonseca, R., Bozkurt, D., Nelli, N., and Guan, B.: Atmospheric river rapids and their role in the extreme rainfall event of April 2023 in the Middle East, Geophysical Research Letters, 51, e2024GL109 446, 2024.

Francis, D., Fonseca, R., Nelli, N., Cherif, C., Yarragunta, Y., Zittis, G., and Jan de Vries, A.: From cause to consequence: examining the historic April 2024 rainstorm in the United Arab Emirates through the lens of climate change, npj Climate and Atmospheric Science, 8, 1–14, 2025.

Gamo, M.: Thickness of the dry convection and large-scale subsidence above deserts, Boundary-Layer Meteorology, 79, 265–278, 1996.

Garratt, J. R.: The atmospheric boundary layer, Earth-Science Reviews, 37, 89–134, 1994.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of hydrology, 377, 80–91, 2009.

Haggag, M. and El-Badry, H.: Mesoscale numerical study of quasi-stationary convective system over Jeddah in November 2009, Atmospheric and Climate Sciences, 3, 2013.

Held, I. M. and Soden, B. J.: Robust responses of the hydrological cycle to global warming, Journal of climate, 19, 5686–5699, 2006.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., noz Sabater, J. M., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G. D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Holm, E., Janiskova, M., Keeley, S., Laloyaux, P., Lopez, P., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, https://doi.org/10.1002/qj.3803, 2020.

Hijji, M., Amin, S., Iqbal, R., and Harrop, W.: A critical evaluation of the rational need for an IT management system for flash flood events in Jeddah, Saudi Arabia, in: 2013 Sixth International Conference on Developments in eSystems Engineering, pp. 209–214, https://doi.org/10.1109/DeSE.2013.45, 2013.

Hong, S.-Y. and Lim, J.-O. J.: The WRF single-moment 6-class microphysics scheme (WSM6), Asia-Pacific Journal of Atmospheric Sciences, 42, 129–151, 2006.

Hong, S.-Y., Dudhia, J., and Chen, S.-H.: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation, Monthly weather review, 132, 103–120, 2004.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Monthly weather review, 134, 2318–2341, 2006.

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., et al.: The art and science of climate model tuning, Bulletin of the American Meteorological Society, 98, 589–602, 2017.

Houze Jr, R. A.: Orographic effects on precipitating clouds, Reviews of Geophysics, 50, 2012.

Hoyer, S. and Hamman, J.: xarray: ND labeled arrays and datasets in Python, Journal of Open Research Software, 5, 10–10, 2017.

Hu, X.-M., Nielsen-Gammon, J. W., and Zhang, F.: Evaluation of three planetary boundary layer schemes in the WRF model, Journal of Applied Meteorology and Climatology, 49, 1831–1844, 2010.

Hu, X.-M., Klein, P. M., and Xue, M.: Evaluation of the updated YSU planetary boundary layer scheme within WRF for wind resource and air quality assessments, Journal of Geophysical Research: Atmospheres, 118, 10–490, 2013.

Huffman, G. J., Bolvin, D. T., Joyce, R., Kelley, O. A., Nelkin, E. J., Portier, A., Stocker, E. F., Tan, J., Watters, D. C., and West, B. J.: IMERG V07 Release Notes, accessed 16 May 2024, 2023.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, Journal of Geophysical Research: Atmospheres, 113, 2008.

Jiménez, P. A., Dudhia, J., González-Rouco, J. F., Navarro, J., Montávez, J. P., and García-Bustamante, E.: A revised scheme for the WRF surface layer formulation, Monthly weather review, 140, 898–918, 2012.

Kain, J. S. and Fritsch, J. M.: Convective parameterization for mesoscale models: The Kain-Fritsch scheme, in: The representation of cumulus convection in numerical models, pp. 165–170, Springer, 1993.

Kessler, E.: On the distribution and continuity of water substance in atmospheric circulations, in: On the distribution and continuity of water substance in atmospheric circulations, pp. 1–84, Springer, 1969.

Khansalari, S., Ranjbar-Saadatabadi, A., Fazel-Rastgar, F., and Raziei, T.: Synoptic and dynamic analysis of a flash flood-inducing heavy rainfall event in arid and semi-arid central-northern Iran and its simulation using the WRF model, Dynamics of Atmospheres and Oceans, 93, 101 198, 2021.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resources Research, 42, W03S04, https://doi.org/10.1029/2005WR004362, 2006a.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water resources research, 42, 2006b.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, Journal of hydrology, 424, 264–277, 2012.

Knoben, W. J., Freer, J. E., and Woods, R. A.: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, Hydrology and Earth System Sciences, 23, 4323–4331, 2019.

Knutti, R.: The end of model democracy? An editorial comment, Climatic change, 102, 395–404, 2010.

Krantz, W., Pierce, D., Goldenson, N., and Cayan, D.: Memorandum on evaluating global climate models for studying regional climate change in California, Tech. Rep.), 2021.

Kubota, T., Yamamoto, M. K., Ito, M., Tashima, T., Hirose, H., Ushio, T., Aonashi, K., Shige, S., Hamada, A., Yamaji, M., Yoshida, N., and Kachi, M.: Construction of a longer-term and more homogeneous GSMaP precipitation dataset, pp. 355–373, Springer, https://doi.org/10.1007/978-3-030-24568-9_20, 2024.

Kumar, A., Sarin, M., and Sudheer, A.: Mineral and anthropogenic aerosols in Arabian Sea–atmospheric boundary layer: Sources and spatial variability, Atmospheric Environment, 42, 5169–5181, 2008.

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., Mechler, R., Bouwer, L. M., Arnell, N., Mach, K., et al.: Flood risk and climate change: global and regional perspectives, Hydrological Sciences Journal, 59, 1–28, 2014.

Lei, L. and Hacker, J. P.: Nudging, ensemble, and nudging ensembles for data assimilation in the presence of model error, Monthly Weather Review, 143, 2600–2610, 2015.

Liu, Y., Chen, Y., Chen, O., Wang, J., Zhuo, L., Rico-Ramirez, M. A., and Han, D.: To develop a progressive multimetric configuration optimisation method for WRF simulations of extreme rainfall events over Egypt, Journal of Hydrology, 598, 126 237, 2021.

Luong, T. M., Dasari, H. P., and Hoteit, I.: Impact of urbanization on the simulation of extreme rainfall in the city of Jeddah, Saudi Arabia, Journal of Applied Meteorology and Climatology, 59, 953–971, 2020.

Luong, T. M., Dasari, H. P., Attada, R., Chang, H.-I., Risanto, C. B., Castro, C. L., Zampieri, M., Vitart, F., and Hoteit, I.: Rainfall climatology and predictability over the Kingdom of Saudi Arabia at subseasonal scale, Quarterly Journal of the Royal Meteorological Society, p. e5015, 2025.

Marsham, J. H., Parker, D. J., Grams, C. M., Johnson, B. T., Grey, W. M., and Ross, A. N.: Observations of mesoscale and boundary-layer scale circulations affecting dust transport and uplift over the Sahara, Atmospheric Chemistry and Physics, 8, 6979–6993, 2008.

Mekawy, M., Saber, M., Mekhaimar, S. A., Zakey, A. S., Robaa, S. M., and Abdel Wahab, M.: Evaluation of WRF microphysics schemes performance forced by reanalysis and satellite-based precipitation datasets for early warning system of extreme storms in hyper arid environment, Climate, 11, 8, 2022.

Messmer, M., González-Rojí, S. J., Raible, C. C., and Stocker, T. F.: Sensitivity of precipitation and temperature over the Mount Kenya area to physics parameterization options in a high-resolution model simulation performed with WRFV3. 8.1, Geoscientific Model Development, 14, 2691–2711, 2021.

Morrison, H., Thompson, G., and Tatarskii, V.: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one-and two-moment schemes, Monthly weather review, 137, 991–1007, 2009.

Mostamandi, S., Predybaylo, E., Osipov, S., Zolina, O., Gulev, S., Parajuli, S., and Stenchikov, G.: Sea breeze geoengineering to increase rainfall over the Arabian Red Sea coastal plains, Journal of Hydrometeorology, 23, 3–24, 2022.

Muller, C. and Takayabu, Y.: Response of precipitation extremes to warming: what have we learned from theory and idealized cloud-resolving simulations, and what remains to be learned?, Environmental Research Letters, 15, 035 001, 2020.

Nakanishi, M. and Niino, H.: An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog, Boundary-Layer Meteorology, 119, 397–407, 2006.

Neelin, J. D., Martinez-Villalobos, C., Stechmann, S. N., Ahmed, F., Chen, G., Norris, J. M., Kuo, Y.-H., and Lenderink, G.: Precipitation extremes and water vapor: Relationships in current climate and implications for climate change, Current Climate Change Reports, 8, 17–33, 2022.

Nguyen, T. V., Uniyal, B., Tran, D. A., and Pham, T. B. T.: On the evaluation of both spatial and temporal performance of distributed hydrological models using remote sensing products, Remote Sensing, 14, 1959, 2022.

Ntoumos, A., Hadjinicolaou, P., Zittis, G., Constantinidou, K., Tzyrkalli, A., and Lelieveld, J.: Evaluation of WRF Model Boundary Layer Schemes in Simulating Temperature and Heat Extremes over the Middle East–North Africa (MENA) Region, Journal of Applied Meteorology and Climatology, 62, 1315–1332, 2023.

O'Gorman, P. A. and Schneider, T.: The physical basis for increases in precipitation extremes in simulations of 21st-century climate change, Proceedings of the National Academy of Sciences, 106, 14 773–14 777, 2009.

Parker, W. S.: Understanding pluralism in climate modeling, Foundations of Science, 11, 349–368, 2006.

Patil, S. D. and Stieglitz, M.: Comparing spatial and temporal transferability of hydrological model parameters, Journal of Hydrology, 525, 409–417, 2015.

Patlakas, P., Stathopoulos, C., Kalogeri, C., Vervatis, V., Karagiorgos, J., Chaniotis, I., Kallos, A., Ghulam, A. S., Al-omary, M. A., Papageorgiou, I., et al.: The Development and Operational Use of an Integrated Numerical Weather Prediction System in the National Center for Meteorology of the Kingdom of Saudi Arabia, Weather and Forecasting, 38, 2289–2319, 2023.

Pedgley, D.: An outline of the weather and climate of the Red Sea, L'oceanographie physique de la Mer Rouge, pp. 9–27, 1974.

Pegahfar, N., Gharaylou, M., and Shoushtari, M. H.: Assessing the performance of the WRF model cumulus parameterization schemes for the simulation of five heavy rainfall events over the Pol-Dokhtar, Iran during 1999–2019, Natural Hazards, 112, 253–279, 2022.

Risanto, C. B., Chang, H.-I., Luong, T. M., Dasari, H. P., Attada, R., Castro, C. L., and Hoteit, I.: Retrospective sub-seasonal forecasts of extreme precipitation events in the Arabian Peninsula using convective-permitting modeling, Climate Dynamics, 62, 2877–2906, 2024.

Rogers, E., Black, T., Ferrier, B., Lin, Y., Parrish, D., and DiMego, G.: Changes to the NCEP Meso Eta Analysis and Forecast System: Increase in resolution, new cloud microphysics, modified precipitation assimilation, modified 3DVAR analysis, NWS Technical Procedures Bulletin, 488, 15, 2001.

Schwitalla, T., Branch, O., and Wulfmeyer, V.: Sensitivity study of the planetary boundary layer and microphysical schemes to the initialization of convection over the Arabian Peninsula, Quarterly Journal of the Royal Meteorological Society, 146, 846–869, 2020.

Shin, H. H. and Hong, S.-Y.: Intercomparison of planetary boundary-layer parametrizations in the WRF model for a single day from CASES-99, Boundary-Layer Meteorology, 139, 261–281, 2011.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J. G., Duda, M. G., Barker, D. M., et al.: A description of the advanced research WRF version 4, NCAR tech. note ncar/tn-556+ str, 145, 2019.

Snook, N., Kong, F., Brewster, K. A., Xue, M., Thomas, K. W., Supinie, T. A., Perfater, S., and Albright, B.: Evaluation of convection-permitting precipitation forecast products using WRF, NMMB, and FV3 for the 2016–17 NOAA hydrometeorology testbed flash flood and intense rainfall experiments, Weather and Forecasting, 34, 781–804, 2019.

Soci, C., Hersbach, H., Simmons, A., Poli, P., Bell, B., Berrisford, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., et al.: The ERA5 global reanalysis from 1940 to 2022, Quarterly Journal of the Royal Meteorological Society, 2024.

Srinivas, C., Yesubabu, V., Prasad, D. H., Prasad, K. H., Greeshma, M., Baskaran, R., and Venkatraman, B.: Simulation of an extreme heavy rainfall event over Chennai, India using WRF: Sensitivity to grid resolution and boundary layer physics, Atmospheric Research, 210, 66–82, 2018.

Stull, R. B.: Mean boundary layer characteristics, in: An introduction to boundary layer meteorology, pp. 1–27, Springer, 1988.

Stull, R. B.: An introduction to boundary layer meteorology, vol. 13, Springer Science & Business Media, 2012.

Tao, W.-K.: Goddard Cumulus Ensemble (GCE) model: Application for understanding precipitation processes, Meteorological Monographs, 29, 107–138, 2003.

Tao, W.-K., Wu, D., Lang, S., Chern, J.-D., Peters-Lidard, C., Fridlind, A., and Matsui, T.: High-resolution NU-WRF simulations of a deep convective-precipitation system during MC3E: Further improvements and comparisons between Goddard microphysics schemes and observations, Journal of Geophysical Research: Atmospheres, 121, 1278–1305, 2016.

Taraphdar, S., Pauluis, O. M., Xue, L., Liu, C., Rasmussen, R., Ajayamohan, R., Tessendorf, S., Jing, X., Chen, S., and Grabowski, W. W.: WRF gray-zone simulations of precipitation over the Middle-East and the UAE: Impacts of physical parameterizations and resolution, Journal of Geophysical Research: Atmospheres, 126, e2021JD034 648, 2021.

Taraphdar, S., Gopalakrishnan, D., Liu, C., Pauluis, O. M., Xue, L., Ajayamohan, R., Leung, L. R., Hagos, S., Grabowski, W. W., Chen, S., et al.: Subtropical jet regulates Arabian winter precipitation: A viable mechanism, Journal of the Atmospheric Sciences, 82, 713–732, 2025.

Thompson, G., Field, P. R., Rasmussen, R. M., and Hall, W. D.: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization, Monthly weather review, 136, 5095–5115, 2008.

Tian, J., Liu, J., Wang, J., Li, C., Yu, F., and Chu, Z.: A spatio-temporal evaluation of the WRF physical parameterisations for numerical rainfall simulation in semi-humid and semi-arid catchments of Northern China, Atmospheric Research, 191, 141–155, 2017.

Tudaji, M., Nan, Y., and Tian, F.: Assessing the value of high-resolution rainfall and streamflow data for hydrological modeling: an analysis based on 63 catchments in southeast China, Hydrology and Earth System Sciences, 29, 1919–1937, 2025.

Ukhov, A., Mostamandi, S., Da Silva, A., Flemming, J., Alshehri, Y., Shevchenko, I., and Stenchikov, G.: Assessment of natural and anthropogenic aerosol air pollution in the Middle East using MERRA-2, CAMS data assimilation products, and high-resolution WRF-Chem model simulations, Atmospheric Chemistry and Physics Discussions, 2020, 1–42, 2020.

Vincent, P.: Saudi Arabia: an environmental overview, CRC Press, 2008.

Wang, X., Alharbi, R. S., Baez-Villanueva, O. M., Green, A., McCabe, M. F., Wada, Y., Van Dijk, A. I., Abid, M. A., and Beck, H.: Saudi Rainfall (SaRa): Hourly 0.1 Gridded Rainfall (1979–Present) for Saudi Arabia via Machine Learning Fusion of Satellite and Model Data, EGUsphere, 2025, 1–32, 2025.

WeatherOnline: Saudi Arabia Weather, https://www.weatheronline.co.uk/reports/climate/Saudi-Arabia.htm, accessed 16 May 2024, 2024.

Xie, B., Fung, J. C., Chan, A., and Lau, A.: Evaluation of nonlocal and local planetary boundary layer schemes in the WRF model, Journal of Geophysical Research: Atmospheres, 117, 2012.

Xie, P., Joyce, R., Wu, S., Yoo, S.-H., Yarosh, Y., Sun, F., Lin, R., and Program, N. C.: NOAA Climate Data Record (CDR) of CPC Morphing Technique (CMORPH) High Resolution Global Precipitation Estimates, Version 1 [indicate subset], https://doi.org/10.25921/w9va-q159, nOAA National Centers for Environmental Information. Accessed: 2025-07-01, 2019.

Yesubabu, V., Srinivas, C. V., Langodan, S., and Hoteit, I.: Predicting extreme rainfall events over Jeddah, Saudi Arabia: Impact of data assimilation with conventional and satellite observations, Quarterly Journal of the Royal Meteorological Society, 142, 327–348, 2016.

Youssef, A. M., Sefry, S. A., Pradhan, B., and Alfadail, E. A.: Analysis on causes of flash flood in Jeddah city (Kingdom of Saudi Arabia) of 2009 and 2011 using multi-sensor remote sensing data and GIS, Geomatics, Natural Hazards and Risk, 7, 1018–1042, 2016.

Zittis, G., Hadjinicolaou, P., and Lelieveld, J.: Comparison of WRF model physics parameterizations over the MENA-CORDEX domain, American Journal of Climate Change, 3, 490–511, 2014.