Climate Dynamics
https://doi.org/10.1007/s00382-024-07296-x

Additional insights from convection-permitting scale ensembles in simulating spatiotemporal features of precipitation across the complex terrain of Peninsular India

S. M. Kirthiga[1] · B. Narasimhan[1] · C. Balaji[2,3]

Received: 12 June 2023 / Accepted: 28 May 2024

Abstract
The study assesses the predictability of rainfall patterns in India through 3-day precipitation forecasts from a regional climate model ensemble framework operating at convection-permitting (CP) scales. Initially, 149 experiments are conducted across four events representing different rainfall mechanisms. The performance of a larger set of 55 ensemble members within the multi-physics ensemble framework is evaluated using quantitative metrics such as composite scaled scores and cross-correlation analyses. This evaluation led to the development of an optimally designed smaller member ensemble framework, WRF-CP7, which reduces turnaround time while maintaining the spatial and temporal performance of simulated precipitation fields. The study further assesses the reliability of this framework over an extended period, utilizing insights from 5544 simulations (792 days × 7 ensemble members running for a 90-h lead time) conducted between September 2015 and December 2017. Comparisons between WRF-CP7 and a global climate model forecasts available at coarser resolution highlight the need of parameterization and ensemble framework at convection-permitting scale. WRF-CP7 demonstrates skill in capturing spatiotemporal variability of rainfall occurrences, evidenced by a higher spread–error correlation (0.9 vs. 0.6 in the global model) among ensemble members. The correlation remains consistent even at higher lead-times, in contrast to the reducing skill of the global model with increasing lead-time. WRF-CP7 also shows reduction in spatial and temporal errors within simulated diurnal precipitation patterns, notably during Indian Summer Monsoon, Pre-Monsoon Thunderstorm activities and North-East Monsoon. A notable 30% increase in predictability for moderate to heavy rain intensities is observed across all seasons, accompanied by a 10% decrease in false alarms compared to global model ensemble forecasts. The spatial skill of WRF-CP7 for moderate-heavy intensity events remains high (50–80%) even with a longer lead time of 72-h on an intra-seasonal timescale. With a substantial sample size, the results underscore the effectiveness of using the multi-physics ensemble framework at convective scales for operational forecasting and dynamic downscaling of climatology across the Indian subcontinent.

Keywords Multi-physics ensembles · Convective-Permitting scales · Weather Research and Forecasting (WRF) model · North-East Monsoon (NEM) · Global Ensemble Forecast System (GEFS) · IMDAA Reanalysis Data

1 Introduction

Forecasting the amount and distribution of precipitation realistically over time and space poses a significant challenge in Numerical Weather Prediction (NWP). This challenge arises from the complex interactions among various weather processes, with precipitation exhibiting more rapid and nonlinear error growth compared to other weather variables (Fritsch et al. 1998; Mullen and Buizza 2001; Ebert et al. 2003; Bei and Zhang 2007; Huang and Luo 2017). High-resolution numerical weather models, known as Convection-Permitting scale models (CPMs), have emerged

✉ S. M. Kirthiga
smkirthiga@gmail.com

[1] Hydraulics and Water Resources Division, Department of Civil Engineering, Indian Institute of Technology, Madras, Chennai 600 036, India

[2] Department of Mechanical Engineering, Indian Institute of Technology, Madras, Chennai 600 036, India

[3] Centre for Atmospheric and Climate Sciences (CACS), Indian Institute of Technology, Madras, Chennai 600 036, India

Published online: 11 June 2024

as a solution to address the limitations of traditional NWP models, particularly for predicting Quantitative Precipitation Forecasts (QPFs). At a higher resolution, CPMs are expected to more effectively capture localized weather phenomena, including mesoscale convective systems, land–ocean contrasts, and orographic lifting. These features can contribute diverse feedback to larger-scale phenomena. Consequently, CPMs have demonstrated better skill compared to the NWP models at coarser scales (Weisman et al. 2008; Clark et al. 2009; Prein et al. 2015; Clark et al. 2016; Li et al. 2018; Woodhams et al. 2018; Hanley et al. 2021; Kirthiga et al. 2021; Risanto et al. 2022).

However, uncertainties increase at finer scales (Lorenz 1969; Walser and Schär 2004; Bei and Zhang 2007; Melhauser and Zhang 2012; Zhang et al. 2019), rendering deterministic forecasts with a single realization of the given atmospheric state obsolete, especially for longer lead-times (Fritsch et al. 1998; Mullen and Buizza 2001; Bei and Zhang 2007; Prakash et al. 2016; Kirthiga et al. 2021). To tackle this issue, ensemble approaches that incorporate input and model-based errors have gained popularity (Surcel et al. 2015; Frogner et al. 2019; Risanto et al. 2022) and are operational in forecasting centers worldwide (Gebhardt et al. 2011; Bouttier et al. 2012; Tang et al. 2013; Schwartz et al. 2015, 2019; Wastl et al. 2021).

Despite their advantages, CPM ensembles tend to be under-dispersive, overconfident, and computationally demanding without significantly enhancing the quality of weather forecasts (Weisman et al. 2008; Melhauser and Zhang 2012; Clark 2019). Therefore, there is a need to develop effective ensemble designs tailored explicitly for convection-permitting scales. While early research focused on sampling uncertainties in the input space (initial and lateral boundary condition) due to the abundance of research happening in the mid-latitude (extra-tropical) regions where baroclinic perturbations primarily characterized the major atmospheric disturbances (Walser and Schär 2004). Many processes seem to be under-resolved even with finer-scale ensembles, particularly in the tropics characterized mainly by warm-season weather systems (Mullen and Buizza 2001; Bei and Zhang 2007; Woodhams et al. 2018; Hanley et al. 2021). More recent studies have emphasized the importance of incorporating model uncertainty into CPM ensembles to capture the nonlinear error growth of convective events (Berner et al. 2011, 2015; Romine et al. 2014; Wang et al. 2020). Various approaches, including multi-model (Melhauser et al. 2017), multi-physics (Berner et al. 2011; Clark et al. 2010; Gebhardt et al. 2011; Kirthiga et al. 2021), multi-parameter perturbations (Yussouf and Stensrud 2012; Wang et al. 2020), and stochastic physics schemes (Berner et al. 2011; Romine et al. 2014), have been explored. However, their applicability to convection-permitting regional climate models is not well-established (Baker et al. 2014; Wang et al. 2020).

In the context of the Indian sub-continent, CPMs have shown promise in improving the predictability of major rainfall periods, such as the South-west monsoon (Prakash et al. 2016; Hazra et al. 2020; Samanta et al. 2020), North-east monsoon (Srinivas et al. 2013; Madhulatha and Rajeevan 2018; Singh et al. 2018), and Pre-monsoon thunderstorms (Madala et al. 2014; Das et al. 2015; Halder and Mukhopadhyay 2016). The uncertainty surrounding the representation of microphysical cloud processes of major rainfall mechanisms in Indian climatology by major climate models (Madhulatha and Rajeevan 2018; Kirthiga et al. 2021; Samanta et al. 2023) has led to non-overlapping findings from earlier studies. As a result, there exists a significant gap in understanding the most suitable model configurations for convection-permitting scales. Furthermore, previous studies have focused on performance evaluations for extreme events rather than examining CPM ensembles' error growth and skill in capturing the spatiotemporal variability of all rainfall categories significant to allied sectors like agriculture and water resources management. Therefore, it remains unclear how these conclusions can be applied to an operational framework that spans both intra-seasonal and inter-annual variability. It is essential to systematically investigate and quantify the ensemble properties of the multi-physics model at the convective level (Risanto et al. 2022).

In a prior study, Kirthiga et al. (2021) emphasized the significance of a multiphysics ensemble in capturing spatiotemporal variability in QPFs from diverse rainfall mechanisms at a high 4 km spatial resolution. Building on this, the primary goal of the current research is to establish an operationally feasible convective-scale ensemble, comprising: 1) Building a larger member multi-physics ensemble framework using a Regional Climate Model (RCM)—Weather Research and Forecasting (WRF) model. Devise strategies for designing effective ensemble framework with a limited number of members – WRF-CP7 (Convection Permitting 7-member ensemble from WRF, termed as WRF-CP7 in the manuscript), while maintaining spatio-temporal quality of the precipitation forecasts. The choice of a 7-member ensemble is determined through comprehensive assessments covering deterministic and probabilistic aspects across seasons, ensuring operational feasibility. 2) Quantifying the statistical significance of the suggested smaller-member framework for predicting rainy days, no-rain days and various classes of rainfall intensities over 28 months (September 2015 to

December 2017), and 3) Investigative the added value by the multiphysics ensembles at convective-permitting scales against the coarser-scale Global Ensemble Forecast System (NCEP-GEFS20). This study aims to assess the uncertainty in the convection-permitting model and its representation of physical processes in the Indian domain during the important monsoon period (Indian Summer Monsoon and North-East Monsoon/ Retreating Monsoon). The results will contribute to improving the representation of physical processes in the Indian subcontinent, enhancing short to medium-range precipitation forecasts at finer scales. The structure of the manuscript includes Section 2 detailing the data and methods used, Section 3 presenting results and discussions on the benefits of the CPM multiphysics ensemble framework, and Section 4 summarizing key conclusions.

2 Methodology

The methodology is structured into two main sections: Section 2.1 "Formulation of Multiphysics ensemble framework" elaborates on the experimental design used to establish the optimal size of multi-physics ensembles, leading to the formulation of the WRF-CP7 ensemble framework. Section 2.2 "Added value by proposed limited member convective-permitting Multiphysics ensemble framework" provides an in-depth overview of the data and materials utilized to understand the consistency of the performance of the model and assess the added value of the proposed multi-physics ensemble across diverse temporal and spatial dimensions.

2.1 Formulation of multiphysics ensemble framework

2.1.1 Design of experiments

The study builds upon previous research conducted by Kirthiga et al. in 2021 (P1 hereafter). In the present study, we utilized the non-hydrostatic Weather Research and Forecasting (WRF) model version 4.0 developed by Skamarock et al. (2019). The model setup closely resembled that of P1, and any modifications will be elaborated further in the study. The lateral boundary conditions were obtained from the Global Forecast System (GFS) control run in forecast mode (ftp://ftp.ncep.noaa.gov/pub/data/nccf/com/gfs/prod). The Global Data Assimilation System (GDAS), employing a 4D hybrid ensemble-variational data assimilation scheme, generated initial conditions for the GFS forecast and was utilized in this study. The data was introduced into the model at 3-h intervals at a spatial resolution of 0.25 degrees (National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce 2015).

For the model configuration, two domains were used based on the conclusions drawn from P1. The parent domain – Domain 1, covered South peninsular India and encompassed 148 × 148 grid points at a resolution of 12 km (extent: 69.57E-85.63E; 0.83N-16.64N). The nested inner domain – Domain 2, used for performance assessment, was set up at a resolution of 4 km with 166 × 169 grid points (extent: 74.596E-80.604E, 5.732N-11.775N) (Fig. 1a). The model integration involved an adaptive time-step, 40 vertical

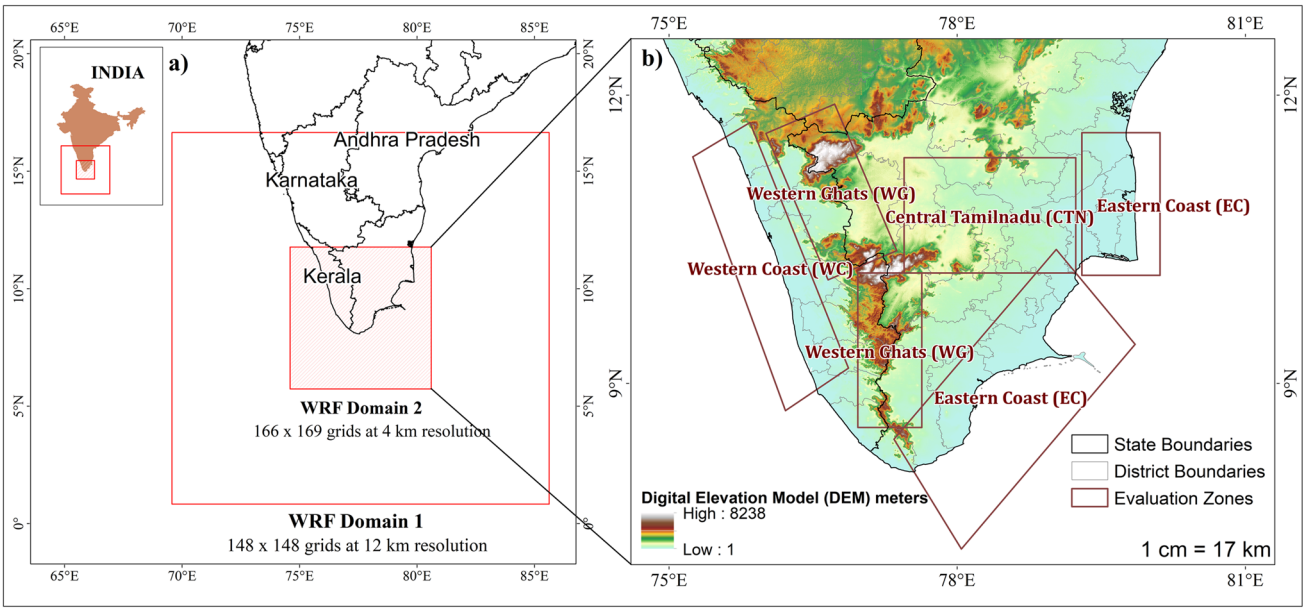

Fig. 1 (**a**) Model domains and (**b**) Analysis Domain (represented by Domain 2 red shaded in (a)), showcasing a detailed topography map with selected evaluation zones. Western Coasts (WC) regions are termed as DOM1, Western Ghats (WG) as DOM2, Eastern Coasts (EC) as DOM3 and Central TamilNadu (CTN) as DOM4. Refer to Table 1 in supplementary material for further details

levels, and a model top set at 50 hPa. The NWP models often require a user-selected constant time-step, posing a challenge as a long time-step can cause model instability for high resolution forecasts, and a shorter time-step demands excessive computing power. The optimal temporal granularity of the time-step depends on the dynamic physical processes resolved by the model at a given spatial resolution, varying daily between different rainfall and weather mechanisms. Employing an adaptive time-step dynamically adjusts to the optimal time-step supporting underlying motions, ensuring model stability and reducing total run-time compared to a static time-step. This approach was implemented in the study to improve the operational feasibility of the resulting model framework. The numerical implementation of the adaptive time-step within WRF model can be found in Hutchinson (2007).

Four specific moderate-extreme rainfall events were considered, consistent with P1, namely Ochki (OCH) event (00 UTC 28 Nov 2017 – 06 UTC 02 Dec 2017), the Gaja (GAJ) event (00 UTC 16 Nov 2018 – 06 UTC 20 Nov 2018), the Indian Summer Monsoon/South-West Monsoon (SWM) event (00 UTC 12 Aug 2018 – 06 UTC 16 Aug 2018), and a thunderstorm event from the Pre-monsoon (SUM) time period (00 UTC 13 Apr 2018 – 06 UTC 17 Apr 2018). The OCH and GAJ events correspond to episodes of tropical cyclones during the North-East Monsoon (NEM) or retreating monsoon period (October-December). These cyclonic events, intensified by synoptic conditions in the Bay of Bengal (BOB), resulted in heavy rainfall and strong winds, leading to significant damage in Tamilnadu and Kerala. On the other hand, the SWM event occurred during the significant Indian Summer Monsoon (ISM) period (July–September) and was driven by pressure differences between the Arabian Sea and BOB, resulting in extensive rainfall and subsequent flooding in Kerala. The SUM event represented a convective event characterized by evening thundercloud clusters across Tamilnadu, causing localized moderate to heavy rainfall. These events were selected to understand and measure the ability of the model to perform under different rainfall mechanisms covering different spatial scales and varying rainfall intensities. Detailed discussions and justification regarding the selected four events can be found in P1, not included here for brevity.

Apart from the physics schemes discussed in Sect. 2.1.2, the remaining physics schemes were maintained the same across all experiments. The radiation physics component included the RRTM Longwave Scheme by Mlawer et al. 1997, which accounted for longwave radiation processes, and the Dudhia shortwave radiation scheme developed by Dudhia 1989, which represented the shortwave component. The planetary boundary layer was represented using the Yonsei University Scheme (YSU) of Hong et al. 2006, while the surface layer formulation was used on the Revised MM5 Scheme of Jiménez et al. 2012. The model setup used the Noah Land Surface Model (LSM) (Tewari et al. 2016), which uses a four-layer soil representation to simulate the land surface processes.

2.1.2 Selection of physics schemes

In the previous study (P1), the impact of input data quality, model domain setup, and physics schemes on the model's performance was investigated. The spatiotemporal rainfall variability in the four selected events was better represented by employing multiphysics ensembles, including variants from representation of cumulus and microphysics schemes. However, it is worth noting that P1 only explored a limited range of physics combinations and did not fully examine the interaction effects between different cumulus and microphysics schemes. The present study is focused on evaluating the performance of a diverse set of physics suites, including both established and recently introduced cumulus and microphysics schemes in the WRF model. The assessment also considered their combined efficacy.

Seven cumulus schemes were considered viz. The Kain–Fritsch (KF) scheme, a slightly modified version of KF, the Kain–Fritsch cumulus potential scheme (MKF), Betts–Miller–Janjic (BMJ) scheme, Simplified Arakawa–Schubert scheme (SAS), Grell-Freitas (GF) ensemble cumulus scheme, New Tiedtke (NT) cumulus scheme and Multi–scale Kain–Fritsch (MsKF) cumulus scheme. In addition, two trigger functions for KF scheme that use moisture advection and relative humidity-based trigger were also tested. All the cumulus schemes are mass-flux schemes, except for BMJ which is an adjustment type static scheme. Detailed information about the specifications of each of the cumulus schemes can be found in the supplementary material, Section 5.2.

The primary finding of P1 was that simulations without cumulus schemes in the 4 km domain were unable to adequately represent the full range of precipitation features, especially within the Indian climatological context. Consequently, experiments were conducted in this study to evaluate the performance of major schemes with (referred to as the CUM cluster) and without implementation in domain 2 (referred to as the No-CUM cluster). The scale-aware mass-flux cumulus schemes, including Grell-Freitas (GF) ensemble cumulus scheme, New-Tiedtke cumulus scheme and Multi–scale Kain–Fritsch (MsKF), were also tested for CUM and No-CUM variants. However, only minimal differences in simulation output were observed with these schemes, attributed to their scale-aware nature. Consequently, the study presents results only from implementing the schemes in both domains (the CUM variant).

Additionally, based on the insights gained from P1, we acknowledge the role of microphysics schemes in accurately representing peak rainfall intensities by realistically representing the distribution of raindrops. The interactive behavior between each cumulus scheme and the microphysics schemes requires careful evaluation, as certain combinations have exhibited notable spatial and temporal errors (Jeworrek et al. 2019). We have considered four microphysics schemes, including two double-moment schemes (Morrison and WDM6) and two single-moment schemes (Lin and Goddard). These schemes offer a comprehensive range of options for simulating hydrometeors, encompassing water vapor, graupel, rain, cloud, cloud ice, and snow. Although they differ in their approach to simulating the number concentrations and mixing ratios of these hydrometeors, all four schemes have proven effective in the context of the Indian subcontinent. Thus, these four schemes were selected for analysis in the manuscript based on the conclusions from P1 and previous studies (Duda et al. 2014; Halder and Mukhopadhyay 2016; Pithani et al. 2019; Musaid et al. 2023).

Thus, a total of 55 combinations (mentioned as WRF-CP55 hereafters in the manuscript) were formulated and are presented in Table 1. Cases C0, C6, C8, C14, C16, C22, C24, C30, C32, C34, C36, C38, C40, C42, C44, C46, C48, C50, C52, and C54 represent explicit simulations without cumulus schemes in domain 2 which are referred to as No-CUM cluster (cases marked with * in Table 1).

2.1.3 Final set of experiments

A run-time limit of 7 h was set (based on the resource constraints elaborated in Section 5.1 of supplementary material) to enable the practical feasibility of applying the multiphysics ensemble for real-time weather forecasting. Thus, due to the dynamics of the selected events, 39 combinations practically converged for OCH, 36 combinations for GAJ, 40 combinations for SWM, and 36 combinations for SUM. The details of the individual case numbers for each event are given in Table 1. Overall, 149 experiments for the four events were evaluated, and the results are presented.

2.1.4 Methods for a comprehensive ensemble framework design

The study employed data from the Integrated Multi-satellitE Retrievals for Global Precipitation Mission (GPM-IMERG) at a spatial resolution of 10 km (referred to as GPM), as described by Huffman et al. (2014). GPM data served as a benchmark for relative comparison to evaluate the performance of the simulations. Gupta et al. (2020) highlighted the superior skill of GPM data to represent the spatial and temporal distribution of the extreme events in India. Subsequently, Kirthiga et al. (2021), in previous research, employed GPM data for the same events examined in this study. Hence, this study utilized GPM for preliminary analysis to ensure consistent evaluation across both investigations. Grid-based statistics are commonly employed for weather modeling performance assessment. While grid-to-grid comparisons offer straightforward evaluations, they can penalize model simulations that accurately represent precipitation intensity and organization but exhibit slight spatial or temporal displacement—a phenomenon known as the double penalty effect (Rossa et al. 2008). This issue is particularly pronounced in high-resolution convective-scale precipitation forecasting, necessitating the inclusion of additional statistics. Therefore, this study explores verification metrics ranging from grid-based continuous and categorical metrics to spatial performance and probabilistic skill metrics, as listed in Table 2. Rainfall intensity classification followed the India Meteorological Department (IMD) classification for hourly and daily rainfall amounts, including no rain (<1 mm/6 h, <2.5 mm/day), light rain (1– 10 mm/6 h; 2.5–15.5 mm/day), moderate rain (10.1–20 mm/6 h; 15.6–64.4 mm/day), and heavy to extremely heavy rainfall (>20.1 mm/6 h; >64.5 mm/day) (India Meteorological Department, https://mausam.imd.gov.in/imd_latest/contents/pdf/pubbrochures/Heavy%20Rainfall%20Warning%20Services.pdf).

2.2 Added value by proposed limited member convective-permitting Multiphysics ensemble framework

2.2.1 Design of experiments

The model configuration detailed in Section 2.1.1 was applied, with the exception of adjustments made to the cumulus and microphysics schemes from results from Section 3.2. The lateral boundary conditions and initial boundary conditions were same as those detailed in Section 2.1.1 (ftp://ftp.ncep.noaa.gov/pub/data/nccf/com/gfs/prod). An assessment of the proposed limited member ensemble framework, denoted as WRF-CP7 and comprising 7 ensemble members (see Section 3.2 for clarity on the selection of the members), was conducted over an extended period to assess the operational feasibility of the framework. The simulations spanned from September 2015 to December 2017, covering a total of 792 days (28 months), with a lead time of 90 h. A total of 5544 simulations (792 days $\times$ 7 ensemble members running for a 90-h lead time) were executed, and this study provides an assessment of the performance of the proposed framework in simulating precipitation. The evaluations focused on the prediction of various rainfall intensities during specific seasons, also considering performance across lead times (Refer to Table 2 for clarity).

Table 1 List of physics combinations investigated in the study

Cumulus Schemes	Microphysics schemes	Cases ID	Events	Selected ensembles and ID
Kain–Fritsch (KF) cumulus scheme (Kain 2004) in D1 termed as KF0	Lin (Chen and Sun 2002)	C0*	OCH, GAJ, SWM, SUM	
KF in D1, D2 termed as KF1	Lin	C1	OCH, GAJ, SWM, SUM	
Moisture–advection-based Trigger for KF cumulus scheme (Ma and Tan 2009) – in D1, D2 termed as KF1_1	Lin	C2		
Relative humidity-dependent Additional Perturbation for the KF cumulus scheme –in D1 and D2 termed as KF1_2	Lin	C3	OCH, GAJ, SWM, SUM	ENS5
Grell-Freitas (GF) ensemble cumulus scheme (Grell and Freitas 2014) in D1, D2 termed as GF1	Lin	C4	OCH, GAJ, SWM, SUM	
Betts–Miller–Janjic (BMJ) cumulus scheme (Janjić 1994) in D1, D2 termed as BMJ1	Lin	C5	OCH, GAJ, SWM, SUM	
BMJ in D1 termed as BMJ0	Lin	C6*	OCH, GAJ, SWM, SUM	
New Tiedtke (NT) cumulus scheme (Zhang and Wang 2017) in D1, D2 termed as NT1	Lin	C7	OCH, GAJ, SWM, SUM	ENS4
KF0 in D1	Morrison (Morrison et al. 2009)	C8*	OCH, GAJ, SWM, SUM	ENS1
KF1 in D1, D2	Morrison	C9	OCH, GAJ, SWM, SUM	ENS2
KF1_1 in D1, D2	Morrison	C10	OCH, GAJ, SWM, SUM	ENS6
KF1_2 in D1, D2	Morrison	C11	OCH, GAJ, SWM, SUM	
GF1 in D1, D2	Morrison	C12		
BMJ1 in D1, D2	Morrison	C13	OCH, GAJ, SWM, SUM	
BMJ0 in D1	Morrison	C14*	OCH, GAJ, SWM, SUM	
NT1 in D1, D2	Morrison	C15	OCH, GAJ, SWM, SUM	
KF0 in D1	WDM6 (Lim and Hong 2010)	C16*		
KF1 in D1, D2	WDM6	C17	OCH, GAJ, SWM, SUM	
KF1_1 in D1, D2	WDM6	C18		
KF1_2 in D1, D2	WDM6	C19	OCH, GAJ, SWM, SUM	
GF1 in D1, D2	WDM6	C20	OCH, GAJ, SWM, SUM	
BMJ1 in D1, D2	WDM6	C21	OCH, GAJ, SWM, SUM	
BMJ0 in D1	WDM6	C22*	OCH, GAJ, SWM, SUM	ENS3
NT1 in D1, D2	WDM6	C23	OCH, GAJ, SWM, SUM	ENS7
KF0 in D1	Goddard Scheme (Tao et al. 2016)	C24*		
KF1 in D1, D2	Goddard Scheme	C25	OCH, GAJ, SWM, SUM	
KF1_1 in D1, D2	Goddard Scheme	C26		
KF1_2 in D1, D2	Goddard Scheme	C27		
GF1 in D1, D2	Goddard Scheme	C28		
BMJ1 in D1, D2	Goddard Scheme	C29		
BMJ0 in D1	Goddard Scheme	C30*	OCH, GAJ, SWM, SUM	
NT1 in D1, D2	Goddard Scheme	C31	OCH, GAJ, SWM, SUM	
Simplified Arakawa–Schubert Scheme (SAS) (Pan and Wu 1995) in D1 termed as SAS0	Lin	C32*	OCH, GAJ, SWM, SUM	
SAS in D1, D2 termed as SAS1	Lin	C33	OCH, GAJ, SWM, SUM	
Modified Kain–Fritsch Cumulus Potential Scheme (MKF) (Berg et al. 2013) in D1 termed as MKF0	Lin	C34*	OCH, SWM, SUM	
MKF in D1, D2 termed as MKF1	Lin	C35	OCH, GAJ, SWM	
Multi–scale Kain–Fritsch (MsKF) cumulus scheme (Zheng et al. 2016) in D1 termed as MsKF0	Lin	C36*	OCH, SUM	
MsKF in D1, D2 termed as MsKF1	Lin	C37	SWM, SUM	
SAS0 in D1	Morrison	C38*	OCH, GAJ, SWM	

Table 1 (continued)

Cumulus Schemes	Microphysics schemes	Cases ID	Events	Selected ensembles and ID
SAS1 in D1, D2	Morrison	C39	OCH, SWM	
MKF0 in D1	Morrison	C40*	OCH, SWM, SUM	
MKF1 in D1, D2	Morrison	C41	OCH, GAJ, SWM, SUM	
MsKF0 in D1	Morrison	C42*	GAJ, SWM, SUM	
MsKF0 in D1, D2	Morrison	C43	OCH, SWM	
SAS0 in D1	WDM6	C44*	OCH, GAJ, SWM, SUM	
SAS1 in D1, D2	WDM6	C45	OCH, GAJ, SWM, SUM	
MKF0 in D1	WDM6	C46*	OCH, SWM, SUM	
MKF1 in D1, D2	WDM6	C47	OCH, GAJ, SWM	
MsKF0 in D1	WDM6	C48*		
MsKF1 in D1, D2	WDM6	C49		
SAS0 in D1	Goddard Scheme	C50*		
SAS1 in D1, D2	Goddard Scheme	C51		
MKF0 in D1	Goddard Scheme	C52*	OCH, GAJ, SWM, SUM	
MKF1 in D1, D2	Goddard Scheme	C53	OCH, GAJ, SWM, SUM	
MsKF0 in D1	Goddard Scheme	C54*		
MsKF1 in D1, D2	Goddard Scheme	C55		

In the study, D1 and D2 refer to Domain 1 and Domain 2, respectively. An asterisk (*) denotes cases where cumulus schemes were implemented only in Domain 1, and these cases, where convection is explicitly resolved by the microphysics schemes, termed as No-CUM cases in the manuscript. The third column, labeled "Events," specifies the combinations that converged for each event. Empty rows or missing event in any row indicate that the particular combination did not converge within the runtime limit set by resource constraints. The last column, "Selected ensembles and ID," highlights members from the larger group that were eventually chosen for the limited member ensemble framework—WRF-CP7

2.2.2 Evaluation zones

Selected evaluation zones (DOM1, DOM2, DOM3, and DOM4) in Tamil Nadu and Kerala (Fig. 1) were chosen to assess the consistency of simulated precipitation across different agro-climatic conditions (refer to Table 1 in the supplementary material). The demarcation of these zones was primarily based on major agro-ecological classes from the Food and Agriculture Organization (FAO), Global Agro Ecological Zones v4 (GAEZv4) classification (Fischer et al. 2021), considering temperature regime, soil-moisture regime, soil/terrain class, and land-cover classes. DOM1 is predominantly in the 'Tropics, lowland humid' zones (GAEZv4 class numbers 5, 6), DOM2 in 'highland, humid with dominantly steep terrain' zones (GAEZv4 class numbers 12, 49, 50), DOM3 in 'Tropics, lowland semi-arid' zones (GAEZv4 class numbers 1, 2), and DOM4 in 'Land with ample irrigated soils' zones (GAEZv4 class numbers 51). These chosen classes are widely prevalent in the tropics and the evaluation of the predictability of precipitation in these zones allows for scaling up the findings from the study to the broader Indian and Asian context. In addition to the GAEZv4 classes, the delineation of the final four evaluation zones (Fig. 1a) also considered monsoonal patterns and local administrative boundaries (refer to the supplementary material, Section 5.2 for more details).

2.2.3 Coarser-resolution convection-parameterizing ensembles

The assessment of added value from the convection-scale multiphysics ensemble framework utilized the Global Ensemble Forecast System (GEFS) version V11.0 from the National Centers for Environmental Prediction (NCEP). This system consists of 20-member ensembles at a resolution of 0.5 degrees (NCEP-GEFS20). Data was retrieved from the THORPEX Interactive Grand Global Ensemble project (TIGGE 2021). Selecting GEFS data to benchmark the performance of the proposed ensemble framework and quantify the added value by convective-scale ensembles was conducted thoughtfully, as it utilizes the same input as used by the regional WRF model. NCEP-GEFS ensembles primarily result from perturbations in initial and lateral boundary conditions from the control member (NCEP-GFS), utilizing the Ensemble Kalman Filter (EnKF) technique. Additionally, they include recently introduced model uncertainty perturbations through the Stochastic Total Tendency Perturbation (STTP) method.

Table 2 Performance metrics and verification strategies employed to evaluate the model simulations

Metrics	Limited Member Ensemble Framework	Long-term assessment and quantifying added value
Continous metrics *	• Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias	
Categorical metrics *	Category based metrics aid in assessing the performance of binary forecasts for specific rainfall thresholds. The following were some of the categorical metrics used in the study • Probability of Detection (POD), False Alarm Ratio (FAR), False Positive Rate, True Positive Rate, False Positve Rate (detailed explanation found in https://www.cawcr.gov.au/projects/verification/ and Kirthiga et al. 2021) • Probability of detection (POD) also termed a predictability or True Positive Rate (TPR) – POD = $\frac{Hits}{Hits+Misses}$, where Hits: The number of events correctly predicted by the model. Misses: The number of events observed but not predicted by the model. POD quantifies the ability of the model to correctly identify the occurrence of an event (e.g., precipitation) when it actually occurs. Higher values of POD indicate better performance in detecting the events of interest • The False Alarm Ratio (FAR) = $\frac{FalseAlarms}{FalseAlarms+Hits}$ where False Alarms: The number of events predicted by the model but not observed. Hits: The number of events correctly predicted by the model as occurring. The FAR indicates the proportion of predicted events that did not actually happen. Lower values of FAR suggest better accuracy in forecasting • The False Positive Rate (FPR) = $\frac{FalseAlarms}{FalseAlarms+TrueNegatives}$, where False Alarms: The number of events predicted by the model but not observed. True Negatives represents the number of true negative predictions (instances where the model correctly predicts the negative class). The False Positive Rate measures the proportion of negative instances that are incorrectly classified as positive by the model, providing insight into the tendency if the model to generate false alarms • The Extreme Event Score (EES), adapted from Sofokleous et al. 2021, which combines hit rate/POD and frequency bias using the contingency table/error matrix was also used. Frequency Bias (FB) = $\frac{Hits+FalseAlarms}{Hits+Misses}$. When the FB ≤ 1, then EES = H x FB, when FB > 1, then EES = H/FB. The EES penalizes underestimation or overestimation of events, providing a comprehensive score that combines the characteristics of categorical indices. EES values range from 0 (no skill) to 1 (perfect skill)	
Selecting ensemble members *	• A composite scaled score (CSS) (Sofokleous et al. 2021) was used to combine scores from categorical metrics (POD, 1-FAR, EES) and continuous metrics (RMSE, mean bias, mean absolute error). A total of six metrics were used. The Composite scale score for a member i is given by $CSS_i = \frac{1}{N_s}\sum_{s=1}^{N}\frac{x_{s,i}-x_{s,worst}}{x_{s,best}-x_{s,worst}}$, where i is the index identifying the member, s is the index of the statistical measure out of a number of N_s (here six) measures, $x_{s,i}$ is the value of measure s obtained by member i, and $x_{s,worst}$ and $x_{s,best}$ are the worst and the best values for measure s among all the ensemble members. Calculated individually for each ensemble member across various rainfall intensities and events, CSS values range from 0 to 1. Higher CSS values indicate superior performance across all six evaluation metrics (POD, 1-FAR, EES, RMSE, mean bias, and MAE) Additionally, Pearson's correlation coefficient (ranging from 0 to 1) was employed to analyze the inter-correlation among ensemble members. This metric is strategically utilized to minimize redundancy among ensemble members at an intra-seasonal timescale	
Spatial performance metrics	Fractional Skill Score (FSS) (Roberts and Lean 2008) was utilized to evaluate the performance of the ensemble framework across various spatial scales and daily rainfall categories. In general, the FSS is derived for different rainfall thresholds and neighborhood sizes. In this study, 5 × 5 pixels was selected as neighborhood size (neighborhood scale of 50 km) after testing neighborhood scales of 20 km, 30 km, 50 km and 70 km. The binary map (if the value of a pixel in the raster is above the selected rainfall threshold, the value 1 will be assigned to the pixel and vice versa) was derived for both observed and simulated average rainfall rate raster. Later a non-overlapping 5 × 5 size window (50 km neighborhood scale) was moved across the raster and the fractions within the window was calculated. The FSS for every member was computed as $FSS = 1 - \frac{\frac{1}{N}\sum_1^N (P_f - P_o)^2}{\frac{1}{N}\left[\sum_1^N P_f^2 + \sum_1^N P_o^2\right]}$, where P_f is forecast fraction, P_o is observed fraction, N is the number of spatial windows in the evaluation domain	Relying solely on grid-based statistics may not enable a seamless comparison due to variations in spatial resolutions. Thus, in addition to FSS, the following spatial statistics were also explored primarily to quantify the added value at the convective scale • Seasonal averages of simulated precipitation were computed to assess the spatial accuracy in simulating long-term climate patterns • Rain object analysis—Determination of the count and area coverage of the simulated rain objects based on specific thresholds and spatial shifts in both the north–south and east–west directions

Table 2 (continued)

Metrics	Limited Member Ensemble Framework	Long-term assessment and quantifying added value
Probabilistic methods to compare ensemble frameworks	• The rank histogram, continuous ranking probability score (CRPS) (Baker et al. 2014), outlier statistics (Kirthiga et al. 2021), ensemble spread, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC) (Measures the discrimination ability of binary forecasts) (Bouallègue and Richardson 2022), to gain insights into the probabilistic behavior of the ensembles • The Continuous Ranking Probability Score measures the discrepancy between the cumulative distribution function (CDF) of the observation dataset and the simulated CDF. Lower CRPS values indicate more skillful forecasts. CRPS = $\int_{-\infty}^{\infty}[F_{forecast}(y) - F_{observed}(y)]\,2\,dy$, where $F_{forecast}(y)$ is the forecast cumulative distribution function at value y and $F_{observed}(y)$ is the observed cumulative distribution function at value y. The CRPS integrates the squared differences between the forecast and observed CDFs over all possible values, providing a measure of the overall accuracy of the forecast distribution. Lower CRPS values indicate better forecast performance, with a perfect forecast having a CRPS of zero	• The rank histogram, CRPS, Probability Density Function and spread among the ensembles were utilized to understand the value on the probalistic nature of the forecasts • Additionally, the correlation plots between the ensemble variance and the mean squared error (MSE) of the ensemble mean precipitation was used, following the approach outlined by Clark et al. 2010

* Statistics calculated through grid-to-grid comparisons

2.2.4 Verification data and methods

IMDAA reanalysis data (Indirarani et al. 2021) was used to assess the long-term performance. The reanalysis data has a horizontal resolution of 12 km (equivalent to 0.12-degree) at 3-h intervals. A resampling procedure was followed to bring other datasets to the same resolution as IMDAA. IMDAA data was chosen for its integration of a larger network of observation data from the India Meteorological Department (IMD) into the reanalysis dataset (Indirarani et al. 2021). Notably, IMDAA data has demonstrated superior performance in capturing precipitation extremes and spatio-temporal weather patterns across the Indian subcontinent compared to other reanalysis datasets (Singh et al. 2021).

An assessment was carried out across evaluation zones (Section 2.2.2) to comprehend the behavior of the model simulations across various seasons, topography, and climatology. Evaluation metrics employed for the long-term assessment are detailed in Table 2. For practical applications, determining the most useful deterministic value from the ensemble members involved utilizing both the ensemble mean (EM) and the value derived using the probability matching method (PMM) proposed by Ebert et al. (2003).

3 Results and discussion

3.1 Simulation of precipitation by multi-physics combinations across four extreme events

3.1.1 Model behavior across the different physics schemes

In this study, we classified the combinations of multiphysics simulations into two primary clusters: the CUM cluster and the No-CUM cluster. The CUM cluster incorporated cumulus schemes in domain 2, while the No-CUM cases represented explicit convection using microphysics schemes without a cumulus scheme in domain 2 (Table 1). Analysis of the domain-averaged temporal precipitation rates (mm/6 h) (Supplementary material Fig. S1) revealed a significant spread in the range of 1 mm/6 h to 5.5 mm/6 h across lead-times and events, largely evident when varying cumulus physics schemes in both domains (CUM cases). The cases within the CUM cluster (dashed lines in supplementary material Fig. S1) displayed earlier initiation of events and closely matched the intensities of peak events, aligning well with the temporal profile of GPM-IMERG data. Conversely, the No-CUM cases (C0, C6, C8, C14, C16, C22, and C24) showed a narrower spread of < 2 mm/6 h on average, suggesting lower internal variability among these cases. This indicates that the different cumulus schemes in the parent domain demonstrated minimal impact on the precipitation simulated in domain 2 for the selected events. Furthermore, No-CUM cases simulated delayed precipitation and lower peak intensities across extreme rainfall events. However, it was observed that No-CUM cases better represented weak convective activities (light rainfall of < 5 mm/6 h) with a higher success ratio exceeding 0.6. Research findings suggest that microphysics schemes are more effective in resolving stratiform clouds associated with low to moderate intensity rains (Samanta et al. 2021, 2023). The CUM clusters consistently overestimated light rainfall events, leading to a lower success ratio (< 0.5) for that category. Fractional Skill Scores (FSS) (Fig. 2) illustrate that

CUM cases outperformed No-CUM cases (first 7 rows in Fig. 2) in simulating spatio-temporal patterns of moderate to higher rain thresholds across all events. The No-CUM cases exhibited low skill for higher rain thresholds with FSS ranging between 0 – 0.28.

The KF schemes, both the default version (cases C0, C1, C8, C9, C16, C17, C24, and C25) and those with modified triggers (cases C2, C3, C10, C11, C18, C19, C26, and C27), effectively captured the variability observed across events with FSS exceeding 0.8 for rainfall occurrences. The cases in this group demonstrated the highest skill scores of FSS > 0.65 across all rainfall categories (Fig. 2), indicating higher sensitivity and specificity in the predicted events with CSS > 0.6 across all the event and rainfall thresholds (Table 3). However, the cases that incorporated relative humidity-dependent additional perturbation for the KF scheme (cases C3, C11, C19, and C27) overestimated rainfall (positive bias of 2.5 mm/6 h) for low-moderate convective events during the GAJ and SUM episodes, with the magnitude of positive bias increasing as the lead-time increased (≥ 36 h, supplementary Table 3). It is worth noting that there was an observable difference in performance

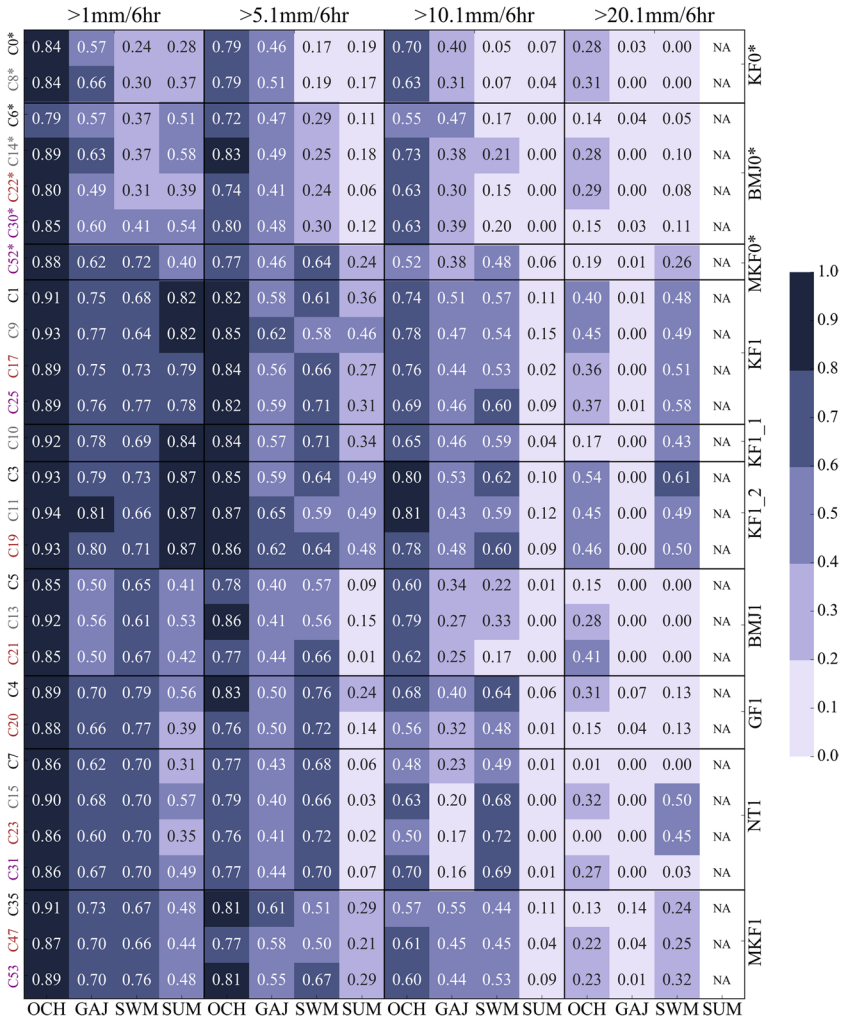

Fig. 2 Fractional Skill Score (FSS) at 50 km radius (5×5 grids) across different rain thresholds. The x-axis represents the events arranged in the sequence of OCH, GAJ, SWM, and SUM. These events are reiterated for each rain threshold, namely > 1 mm/6 h, > 5 .1 mm/6 h, > 10.1 mm/6 h, and 20.1 mm/6 h (indicated on the secondary x-axis located at the top). The y-axis represents case numbers arranged to depict the order of cumulus schemes, viz. KF0*, BMJ0*, MKJ0*, KF1, KF1_1, KF1_2, BMJ1, GF1, NT1, MKF1 (mapped in the secondary y-axis). Cases with no cumulus schemes used in the innermost domain D2 (No-CUM) are marked with *. Within each cumulus scheme class, microphysics schemes are sorted, and different colors on the case numbers indicate the microphysics schemes: **black** – Lin scheme; **grey** – Morrison scheme; brown – WDM6 scheme; purple – Goddard scheme. For a clearer understanding of the case numbers and various physics combinations, refer to Table 1. The lower x-axis denotes the events and y-axis denotes the case numbers. The black lines and corresponding secondary x-axis represent grouping based on the cumulus schemes. The event SUM corresponds to a moderate rainfall event, where values do not exceed 20.1 mm/6 h. Therefore, in the figure, it is represented as NA (Not Available)

between the CUM and No-CUM variants of KF scheme, i.e. when comparing the utilization of the KF scheme in both domains (FSS of 0.6 and bias of -2 mm/6 h) versus using KF only in the parent domain (FSS of 0.35 and bias of -5 mm/6 h). The CUM cases employing the KF scheme in the 4-km domain demonstrated improved performance during the Southwest Monsoon (SWM), achieving an FSS of > 0.5 across the rainfall intensities. In contrast, the No-CUM cases showed no skill (FSS ~ 0) for rainfall intensities exceeding 20.1 mm/6 h for the SWM event (Fig. 2). Furthermore, for the OCH and SWM event, the KF-based No-CUM cases depicted a narrower spatial distribution of rainfall compared to the KF-based CUM cluster results (not shown here). An interesting finding is that during the SUM event, the CUM variant of KF schemes consistently achieved an FSS above 0.4 for rainfall intensities exceeding 5.1 mm/6 h, while the No-CUM cases recorded lower FSS values (< 0.15).

The BMJ cases (cases C5, C6, C13, C14, C21, C22, C29, and C30) significantly under-predicted rainfall (bias of -10 mm/6 h) compared to other cumulus schemes for the four rain episodes considered, resulting in poor performance with FSS < 0.25 for higher rain thresholds (Fig. 3). The difference between the BMJ0 and BMJ1 cases were found to be very minimal, following similar spatial and temporal patterns while simulating rainfall. The MKF cumulus scheme (cases C52, C35, C47, and C53) slightly overestimated the spatial spread of rainfall during the initial periods of the SWM episode but better captured the peak event (54–90 h) during the episode (Fig. S1). Among all the events analyzed, the MKF scheme demonstrated superior performance in predicting extremes (particularly when used in domain 2), with an EES exceeding 0.45, compared to the KF schemes (EES—0.35) and other schemes (EES—0.2). The NT scheme consistently demonstrated high POD values (> 0.5) and low FAR values (< 0.3) for occurrences of low to moderate rainfall across all

Table 3 Performance of each physics combination considered in the study across the selected events

Cumulus cluster	Cases	Extreme Event Score (EES) Averaged across thresholds	Composite Scaled Score (CSS) Combining RMSE, MAE, MB, Hit Score, Success Ratio & EES
KF0	C0	0.18	0.51
	C8	0.23	0.54
BMJ0	C6	0.21	0.42
	C14	0.24	0.44
	C22	0.15	0.40
	C30	0.22	0.44
MKF0	C52	0.18	0.56
KF1	C1	0.37	0.57
	C9	0.41	0.55
	C17	0.34	0.55
	C25	0.32	0.58
KF1_1	C10	0.36	0.48
KF1_2	C3	0.40	0.64
	C11	0.39	0.63
	C19	0.37	0.60
BMJ1	C5	0.16	0.52
	C13	0.22	0.58
	C21	0.15	0.53
GF1	C4	0.22	0.49
	C20	0.17	0.53
NT1	C7	0.23	0.59
	C15	0.26	0.57
	C23	0.20	0.54
	C31	0.20	0.57
MKF1	C35	0.50	0.44
	C47	0.49	0.47
	C53	0.47	0.53

Detailed RMSE and mean bias values across the lead times can be found in the supplementary material Table 3. *The table showcases only representative cases (with results available across all four events, as outlined in Section 2.3) for each cumulus and microphysics scheme to maintain brevity*

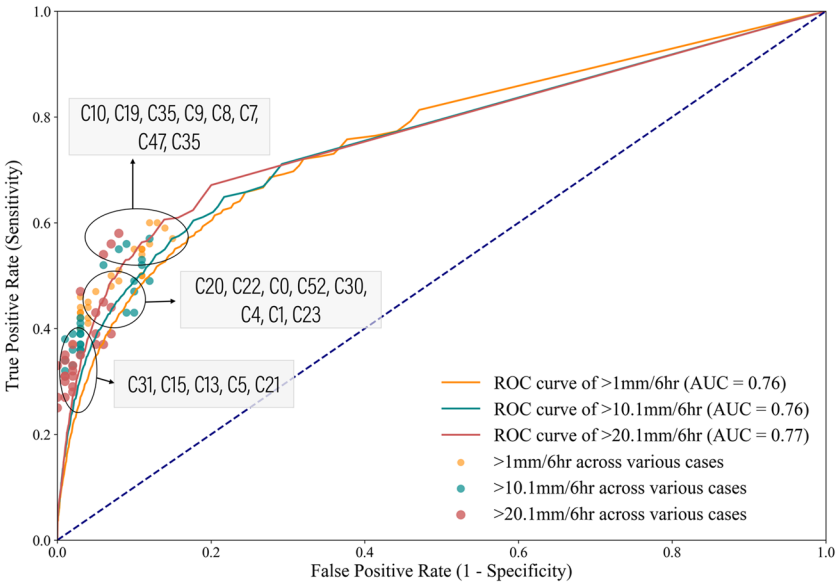

Fig. 3 Receiver Operating Characteristic (ROC) curves for various rainfall thresholds with False Positive Rate (FPR) in the x-axis and True Positive Rate (TPR) in the y-axis. The blue dashed line represents the 1:1 line. A point at the top-left corner of the ROC plot represents perfect classification (TPR = 1, FPR = 0), while a random classifier would produce a diagonal line from the bottom-left to the top-right (AUC = 0.5). The closer the ROC curve is to the top-left corner, the better the model's discrimination ability. Scatter points represent individual case skill, and black circles highlight cases with similar performance

analyzed events. Cases involving MsKF and SAS schemes did not fully converge within the given run-time constraint (Section 2.1.3) for the four events discussed here and were therefore excluded from further discussions.

The plots of domain-averaged temporal precipitation (mm/6 h) for different microphysics schemes (averaged across the cumulus schemes) (Fig. S2) revealed noticeable variations in the intensity of peak events and the spatial extent of rain clusters when the microphysics (MPS) schemes were modified. Particularly, for the SWM and SUM events, a deviation of ± 5 mm/6 h was noticed between the selected microphysics schemes, with Goddard (FSS – 0.35) and Morrison schemes (FSS – 0.3) recording superior performance. However, there was no significant difference in the initiation time and timing to peak of rainfall among the microphysics cases for the simulated events.

3.1.2 Spatio-Temporal attributes of rainfall events and influences on model performance

The OCH and GAJ events are tropical cyclones driven by large-scale dynamics, but the microphysical processes associated with tropical cyclones become highly complex during landfall and post-landfall as they move over the land. Tropical cyclones typically originate as clusters of thunderstorms over warm ocean waters. As these clusters intensify, warm, moist air converges toward the center of the disturbance at low levels. Further intensification leads to the formation of a deep layer of towering cumulonimbus clouds known as the Central Dense Overcast (CDO), characterized by intense convective activity and heavy rainfall consisting of numerous isolated convective elements. Tropical cyclone intensity is closely linked with Latent Heat Release (LHR), which is influenced by the distribution of hydrometeors above and below the melting layers (Nekkali et al. 2022). Studies have demonstrated that the choice of cumulus and microphysics schemes significantly impacts the track and rainfall intensities of tropical cyclones (Mahala et al. 2015; Kirthiga et al. 2021). The cyclonic storm Ockhi (OCH) is considered a rare event characterized by rapid intensification from a depression to a cyclone within a span of 9 h, further developing to a very severe cyclone within next 24 h (Singh et al. 2020). The pronounced increase in CAPE within the low to mid-level atmosphere during this event is largely linked to oceanic interactions under the influence of a prominent large-scale upper-level trough. Thus, KF, MKF and NT, which utilizes CAPE for closure assumptions was observed to perform better than other schemes. The rapid intensification was picked well by the simulations 48 h ahead of the event. Previous studies have indicated that the BMJ scheme was skillful at accurately representing the track and intensity of tropical cyclones. The scheme was reported to have simulated the typical warm core structure and wind pattern associated with these weather systems better than the other schemes (Kanase et al. 2020). However, in this study, the BMJ case, recorded late initiation for the OCH cyclone event forecasting a low intensity event. The BMJ scheme adjusts the current profile to a pre-determined convective sounding. However, as this event is a rare event which deviates from the average climatology of cyclone storms in India, resulted in lower skill of using a pre-defined reference profile in the scheme. The atmospheric sounding profiles revealed lower moisture levels and inadequately represented low-level vorticity, resulting in a slow and weaker development of the weather system. Cases that relied solely on microphysics schemes to resolve cloud processes in domain 2 (No-CUM cluster) failed to capture

the full spectrum of precipitation features for OCH event (Fig. 2). This limitation may be attributed to issues with available microphysics schemes in adequately resolving isolated convective elements present in the core of the cyclone (Samanta et al. 2023). The Lin and Morrisons scheme along with KF schemes (C1, C2, C3, C9, C10 and C11) performed better in capturing the spatio-temporal pattern of the rainfall features for the event (CSS—0.6 on average; Table 3).

The conclusions drawn from the OCH event were not evidently applicable to the Gaja cyclonic event (GAJ). The GAJ event presented unique challenges, as the initial and lateral boundary conditions themselves were poorly represented from the input data system (Kirthiga et al. 2021), and the multiphysics combinations attempted here were not able to improve the simulations significantly for the major cyclonic event during the first 18 h of the simulation (Supplementary material Fig. S1). Previous studies have also highlighted the importance of near-perfect boundary forcing for accurate predictions of tropical cyclones driven mainly by synoptic-scale processes (Bucci et al. 2018). However, a significant variability of ± 7.5 mm/6 h was observed in the rainfall occurrences linked to the pull effect as the cyclone advanced westward over land during the 78–96 h lead-time.

The SWM event was part of South-West Monsoonal circulations of 2018, enhanced by presence of a low-pressure zone in Bay of Bengal. The off-shore monsoon trough was intensified by mid-tropospheric cyclonic circulation over peninsular India causing the high intensity events (Kirthiga et al. 2021). The rainfall occurred over larger spatial extent covering Kerala, Tamilnadu and parts of Karnataka, also recording highly variable rainfall intensities. Higher variability of hydrometeor distributions was noticed during the event causing huge spatio-temporal variability in the rainfall intensities (Sumesh et al. 2022). The No-CUM cases simulated high intensity events but the rainfall clusters were isolated to specific regions in the domain, thus hugely underestimating the spatial spread of the rainfall occurrence (the spatial distribution of the event simulations can be found in Kirthiga et al. 2021, not shown here for brevity). The CUM cluster was able to represent the intensification and temporal pattern of the rainfall progression, however, the areal distribution was still underestimated. The MKF1 scheme captured the accumulated rainfall for event, comparatively well when compared to GPM. However, the case simulated the event as a weak rainfall event existent for a longer duration deviating from the characteristics of the real episode.

Across the simulations, cumulus schemes influenced the spatial distribution of moderate rainfall and initiation of events, while changes to microphysics schemes introduced variability in both the intensity and spatial patterns of high-intensity rainfall occurrences within the simulated environment. Lower Convective Inhibition (CIN) and higher Convective Available Potential Energy (CAPE) values were prevalent, leading to higher skill of the CUM variant of KF schemes (implemented in two domains), with a Fractional Skill Score (FSS) exceeding 0.5 across all events considered. Cases utilizing relative humidity-based triggers with the KF scheme (C3, C11, and C19) recorded higher overall skill scores, with CSS exceeding 0.6. The Goddard, Lin, and Morrison schemes consistently demonstrated higher skill scores compared to others. For events dominated by warm rain processes like OCH and GAJ, single-moment microphysics schemes like Lin and Goddard performed well. However, for events (SWM and SUM) dominated by hydrometeors from ice and graupel categories, combinations with double-moment schemes Morrison and WDM6 schemes showed superior performance. No-CUM cases resolved precipitation intensities for events driven by large-scale dynamics and stratiform cloud processes. The ratio of simulated convective precipitation to total simulated precipitation in domain 2 consistently exceeded 0.6 across all simulated events, with contributions varying from isolated convective elements to well-structured convective processes. When microphysics schemes were used in domain 2 to explicitly resolve processes without cumulus scheme, narrow bands of precipitation, as compared to GPM, and delays in event initiation were noticed. It is noteworthy that the CUM cluster simulated a large spatial spread but weak rainfall activity, extending slightly longer than the actual duration for some events (SWM and SUM).

3.2 Reducing the number of ensemble members from an operational point of view

Figure 3 illustrates the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) for precipitation across various thresholds. This metric is commonly utilized to assess the discrimination capability of forecasts for both rare and common events (Bouallègue and Richardson 2022). The KF schemes (cases C9, C10) showed higher TPR (> 0.55) and low FPR (< 0.05), particularly when paired with the Morrison microphysics scheme (Fig. 3), thus showing a superior performance for the simulated events. Case C8, a No-CUM variant, also exhibited good performance with a Critical Success Index (CSS) of 0.54. Case C11, which utilized a relative humidity perturbation with the KF scheme in both domains along with the double-moment mixed-phase Morrison scheme, exhibited a higher False Alarm Rate (FAR) exceeding 0.4 (not shown here). Conversely, case C3, employing a relative humidity trigger with the KF scheme but with a single-moment scheme (Lin) in both domains, recorded a higher True Positive Rate (TPR) exceeding 0.6 and a False Positive Rate (FPR) of less than 0.05. The NT schemes, which were scale-aware, demonstrated consistent performance across cases C7, C15, C23, and C31 with

moderate CSS exceeding 0.55. The BMJ0 cluster, along with complex microphysics schemes such as cases C14 and C22, exhibited a moderate True Positive Rate (TPR ~ 0.4) across rainfall thresholds, accompanied by a slightly higher False Positive Rate (FPR ~ 0.1) with False Alarm Rate (FAR) exceeding 0.3. These cases also recorded an Extreme Event Score (EES) value not exceeding 0.2. Cases with the GF scheme (C12, C20) demonstrated a TPR not exceeding 0.3, along with FAR > 0.3, showing a lower skill when compared to other cases. The performance of the cases was largely found to be dependent on the specific event characteristics, as elaborated in Section 3.1.2.

Run-time constraints in the study were established due to resource limitations determined by the implemented cluster (supplementary material Section 5.1), aiming to ensure operational feasibility. A maximum allotted time of 3 h was set for each ensemble member to converge. The runtime of each case was influenced by the dynamics and complexity of the selected event, as well as the formulations of the physics schemes utilized. Instability in the grid sometimes led to longer runtimes for the CUM cluster due to frequent calls to the cumulus schemes at each time step. Events with mixed hydrometeors and complex cloud processes resulted in slightly longer runtime for the No-CUM variant compared to the CUM cluster. This aspect was not extensively investigated in previous studies, as most of them utilized a constant time step, except for a few studies (Jeworrek et al. 2019, 2021). To ensure operational feasibility, dynamic time steps were preferred to manage this complexity and decrease the overall turnaround time for various cases on a daily basis. The KF schemes with default trigger consistently exhibited the longest runtime among both CUM and No-CUM clusters. However, they consistently converged within the allocated 2-h runtime for all events. The KF schemes with perturbations (KF1_1 and KF1_2) recorded a runtime 10–20% shorter than KF0 and KF1 variants. Scale-aware NT schemes consistently demonstrated shorter turnaround times (18–32%) compared to other schemes. BMJ0 variants with different microphysics schemes demonstrated enhanced performance and reduced turnaround time (10–15%) compared to other schemes. However, when implemented in both domains as part of the CUM cluster, the BMJ scheme exhibited longer runtime, equivalent to the KF cluster. Furthermore, when combined with the Goddard microphysics scheme, it failed to converge for any of the event within the given time limit (Case C29). Certain combinations, such as MsKF with complex microphysics schemes, often failed to converge within the given time limit. Hence, selecting compatible cumulus and microphysics schemes for specific rain events is crucial, especially when explicitly resolving precipitation in finer domains (Yano et al. 2018). Combining KF cumulus schemes in domain 1 with the Morrison microphysics scheme in domain 2 resulted in the best performance (CSS and required less integration time than other combinations with KF scheme. However, the MKF cumulus scheme in domain 1 performed well when paired with the Goddard microphysics scheme (case C53) in domain 2 (CSS – 0.53). The WDM6 microphysics scheme showed improved performance (CSS – 0.6) when combined with the KF scheme utilizing the relative humidity-based trigger (case C19). While the Goddard scheme performed well during the SWM episode combined with the MKF (C53 – CSS of 0.53) and NT (C31 – CSS of 0.57) cumulus schemes, combinations involving the Goddard scheme failed to converge within the given time when combined with other cumulus scheme combinations across events. The MKF scheme required more runtime (20–30% more than other schemes) for implementation and occasionally failed to converge in cases with complex microphysics schemes and system dynamics (also observed by Berg et al. 2013).

The Pearson's correlation values presented in Fig. 4 provide insights into the inter-correlation among the different members of the multiphysics approach. The degree of correlation between these members depended on the theory behind each physics scheme formulation and the underlying mechanisms driving the event. Previous studies (Charron et al. 2010; Leutbecher et al. 2017) noted that the clustering of ensemble members is a common issue associated with the multi-model or multiphysics approach, leading to multimodality in the outputs. From the results, it was observed that increasing the number of multiphysics members did not necessarily enhance the predictability of the event. The study also underscored the importance of addressing the clustering phenomenon observed among highly correlated ensembles. The highest inter-correlation was found among the cases for the GAJ event, which is more influenced by large-scale depression. Imperfect boundary conditions led to low variability between the multiphysics combinations, resulting in the highest intercorrelation exceeding 0.8 between the different cases analyzed (Fig. 4b). Moreover, a significant occurrence of high intercorrelation (> 0.7) was observed among different microphysics schemes within a given cumulus scheme cluster, as discussed in Section 3.1.1. For other events, particularly the OCH and SWM events (Fig. 4a and Fig. 4c), the KF1 variant with default trigger showed correlation (> 0.6) with KF1, NT1, and MKF1 schemes. Conversely, the BMJ0 cluster exhibited the least correlation (< 0.3) with other schemes. It is interesting to note that for SUM event there was least correlation (< 0.35) between the multi-physics members (Fig. 4d).

Based on the performance assessment outlined above, including runtime analysis, correlation evaluation, and CSS scores (Fig. 3, Fig. 4 and Table 3), seven members have been selected for inclusion in the convection-allowing multiphysics ensemble of smaller size, denoted as WRF-CP7. These members, labeled as ENS1 to ENS7,

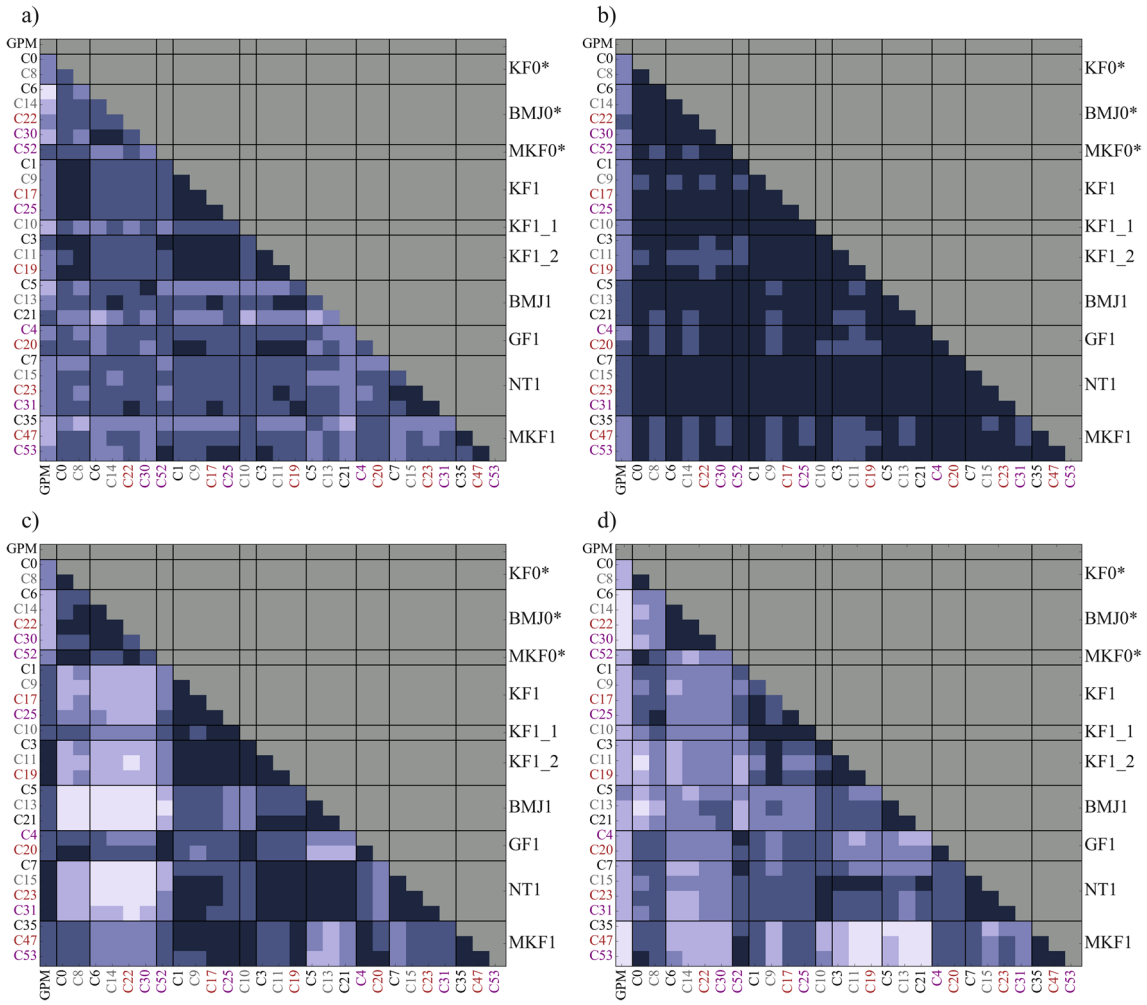

Fig. 4 Inter-correlation plots between cases for different events (a) OCH, (b) GAJ, (c) SWM, and (d) SUM, with light–dark shades of blue to indicate low inter-correlation (value of 0) to higher inter-correlation (value of 1). The first column presents the correlation with reference dataset GPM. The y-axis and x-axis depict correlation values, with GPM as first element, followed by the ensemble members (case numbers) of WRF-CP55. The alignment of case numbers on the y-axis corresponds to the order of cumulus schemes (labeled on the secondary y-axis, with '*' denoting No-CUM cases), and the color on the case numbers signifies the microphysics schemes similar to Fig. 2

correspond to case numbers C8, C9, C22, C7, C3, C10, and C23, respectively, as indicated in Table 1, 5th column (Selected ensembles and ID). The time plots presented in Fig. 5 illustrate the spread of ensembles as the lead times progress. Comparing the precipitation forecasts from WRF-CP55 with the WRF-CP7 (shaded in green), we observed that the spread was highly comparable. Furthermore, the CRPS, MAE, and RMSE values reported in Table 4 indicate that the proposed WRF-CP7 recorded similar error statistics and overall performance compared to the larger multiphysics ensembles WRF-CP55. Additionally, the outlier statistics in Table 5 demonstrate that the difference between the actual outlier base rate and the expected outlier rate for the WRF-CP7 was equivalent or even smaller when compared to the outlier statistics of WRF-CP55. Reducing the number of ensemble members did not significantly affect the distribution of the rank histogram for the four heavy rainfall events analyzed in this study (Fig. S3). However, the rank histogram, highlights skewness in both multiphysics frameworks (WRF-CP55 and WRF-CP7), indicating a tendency for the multi-physics model to slightly underestimate.

It is important to note that the maximum achievable True Positive Rate (predictability), indicative of the sensitivity of the framework, consistently remains below 0.65. This suggests that the multiphysics ensemble framework accounted for a limited amount of predictability in simulated rainfall for the selected events. Addressing input and model uncertainties is imperative for enhancing the performance of ensemble frameworks, particularly in tropical regions (Prakash et al. 2016; Huang and Luo 2017).

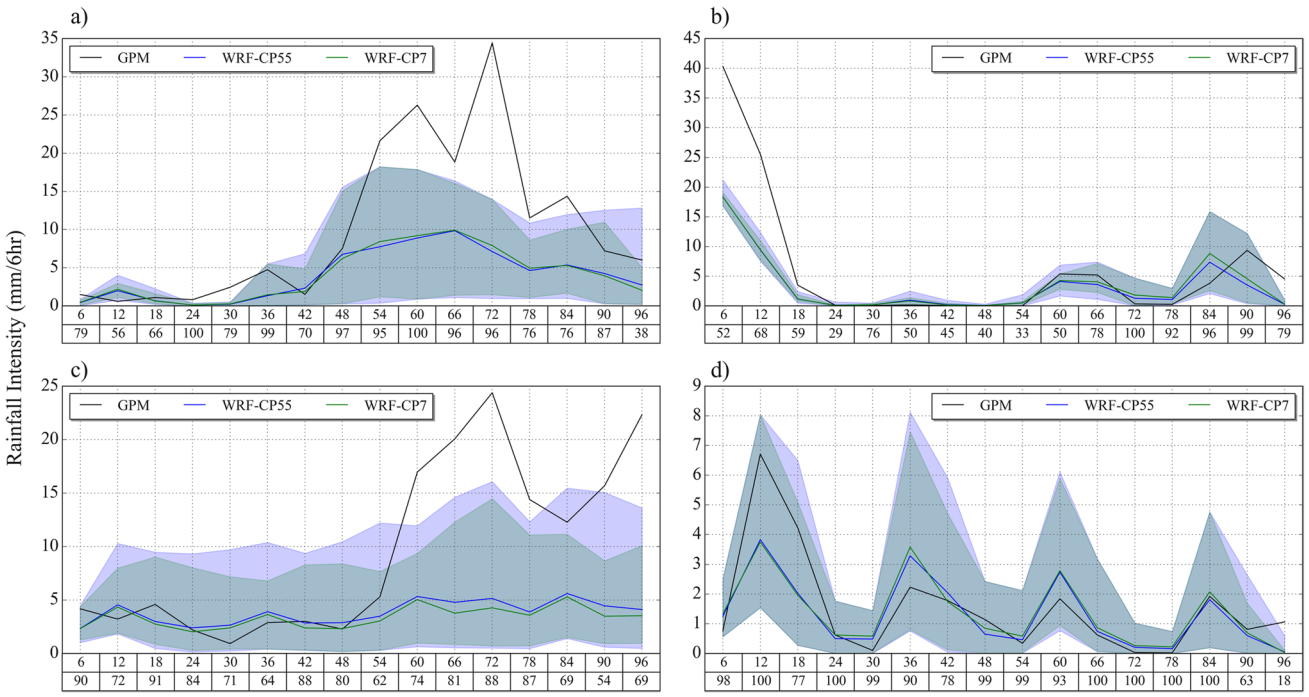

Fig. 5 Temporal plots (domain-averaged) illustrating the ensemble frameworks' spread across events, (**a**) OCH, (**b**) GAJ, (**c**) SWM, and (**d**) SUM. The x-axis of the first row represents the forecast lead time (6-hourly), while the second row displays values indicating the ratio (spread of WRF-CP7/spread of WRF-CP55) * 100. A higher ratio signifies that the spread generated by WRF-CP7 (7-member ensemble) is nearly identical to that of WRF-CP55, whereas lower values indicate a reduced spread by the WRF-CP7 model

Table 4 Performance analysis of ensemble properties across two versions of CP ensembles- WRF-CP55 and WRF-CP7 models

Statistics	CRPS (mm/6 h)		MAE (mm/6 h)		RMSE (mm/12 h)		Outlier Statistics (%)	
Lead-Time (hrs)	WRF-CP55	WRF-CP7	WRF-CP55	WRF-CP7	WRF-CP55	WRF-CP7	WRF-CP55	WRF-CP7
12	2.75	2.76	4.10	4.09	7.57	7.58	34 (7)	42 (25)
24	3.65	3.77	5.64	5.56	9.21	9.31	25 (7)	38 (25)
36	2.45	2.75	5.29	5.34	9.00	9.49	16 (7)	27 (25)
48	10.42	10.59	15.85	16.01	36.15	35.85	14 (7)	30 (25)
60	14.26	14.26	19.89	19.93	43.76	43.73	23 (7)	36 (25)
72	14.36	14.81	18.71	18.72	36.43	36.51	29 (7)	44 (25)
84	7.75	8.20	10.25	10.25	15.50	15.93	25 (7)	39 (25)
96	6.31	6.63	8.16	8.06	15.06	15.74	30 (7)	44 (25)

*The expected base rate for outlier statistics is given within the brackets, i.e. 7% ((2/(N+1))*100) for 27 member ensemble and 25% for the small 7-member ensemble*

However, much of the bias in the multiphysics ensemble may stem from over-compensation of unsampled input errors.

3.3 Performance of the WRF-CP7 Framework – more extended period analysis

To assess the reliability across different seasons, rain mechanisms, and rainfall intensities, we validated 72-h ahead precipitation forecasts from WRF-CP7 for an extended period (September 2015 to December 2017). Two ensemble members from the No-CUM cluster were considered: ENS1 representing the KF0-Morrison configuration (case C8), and ENS3 representing the BMJ0-WDM6 configuration (case C22). Three members from the CUM cluster of KF schemes with different triggers were included: ENS2 representing KF1-Morrison (case C9), ENS5 representing KF1_2 (Relative humidity-dependent Additional Perturbation)-Lin (case

Table 5 Categorical error statistics across individual ensemble members during the long-term (Sep 2015 – Dec 2017) forecast simulations

Statistics	Rain intensities	ENS1 (KF0-Morrison)	ENS2 (KF1-Morrison)	ENS3 (BMJ0-WDM6)	ENS4 (NT1_Lin)	ENS5 (KF1_2-Lin)	ENS6 (KF1_1-Morrison)	ENS7 (NT1-WDM6)
POD	> 1 mm/6 h	0.31	0.69	0.36	0.63	0.73	0.70	0.65
	> 10.1 mm/6 h	0.25	0.53	0.24	0.30	0.60	0.44	0.35
	> 20.1 mm/6 h	0.26	0.43	0.18	0.28	0.52	0.35	0.32
Success Ratio (1-FAR)	> 1 mm/6 h	0.56	0.55	0.53	0.66	0.57	0.56	0.59
	> 10.1 mm/6 h	0.21	0.22	0.18	0.38	0.24	0.24	0.27
	> 20.1 mm/6 h	0.12	0.09	0.07	0.34	0.10	0.13	0.15
EES	> 1 mm/6 h	0.14	0.60	0.22	0.53	0.65	0.62	0.62
	> 10.1 mm/6 h	0.21	0.22	0.18	0.25	0.24	0.24	0.21
	> 20.1 mm/6 h	0.12	0.09	0.07	0.23	0.10	0.13	0.15

ENS1 to ENS7, correspond to case numbers C8, C9, C22, C7, C3, C10, and C23, respectively, as indicated in Table 1. *Continuous error metrics* viz. *RMSE, mean bias, and Pearson's correlation coefficient values across the lead times can be found in the supplementary material* Table 3

C3), and ENS6 representing KF1_1 (Moisture–advection-based Trigger)-Morrison (case C10). Two ensemble members from the CUM cluster of NT scheme were also selected: ENS4 representing NT1-Lin (case C7) and ENS7 representing NT1-WDM6 (case C23).

As discussed in earlier sections, there was a clear distinction in the performance of individual ensemble members based on the major cumulus and microphysics schemes considered. The various categorical statistics of the performance of individual members are listed in Table 5 as a function of rainfall thresholds, and continuous statistics as a function of lead-time are available in supplementary material Table 3. The members with the NT scheme (ENS4, ENS7) recorded the lowest RMSE (3–4 mm/6 h) across the lead-times. Notably, the error progression as lead-time increases was not very prominent (< 4% increasing trend) with the NT scheme (refer to Table 3 in supplementary material). These members exhibited a slight negative bias (underestimation) with an average of -0.25 mm/6 h. However, the NT scheme with the double-moment microphysics scheme WDM6 reduced the negative bias to -0.15 mm/6 h. Highest Success Ratio (0.45) was noticed for the ENS4 across the rainfall thresholds (Table 5). The NT scheme simulations demonstrated a closer match to the observed cumulative distribution function of forecasted rainfall compared to IMDAA data across various lead-times (Fig. S4 in supplementary material). This scheme, being scale-aware exhibited faster convergence, taking 25–40% less time than the longest runtime taken by the ensemble framework. Consequently, it consistently provided reliable forecasts throughout the year. The incorporation of mid-level cumulus parameterizations, cumulus downdrafts, and cumulus momentum transports, proved highly relevant for tropical setups by earlier studies (Zhang and Wang 2017; Wang 2022; Zhou et al. 2024) and thus suggested in 'NCAR tropical suite'(https://www2.mmm.ucar.edu/wrf/users/physics/ncar_tropical_suite.php). However, the NT scheme faces challenges in accurately simulating low clouds and shallow convection (also reported by Zhang and Wang 2018), particularly evident during the pre-monsoon (SUM) and winter (WIN) seasons. The underestimation of moderate to heavy rainfall during the onset of SWM, SUM and WIN season was widely noticed. Additionally, coastal regions, prone to onshore flow and the formation of low clouds that contribute to localized rainfall, displayed lower skill among members using the NT1 scheme.

For the No-CUM members, the KF0 and BMJ0 configurations (ENS1 and ENS3), a very similar RMSE of 5.1 mm/6 h and 5.7 mm/6 h, respectively, was recorded. The diurnal variability was adequately represented by the No-CUM members, although they failed to capture peak intensities across seasons. Research indicates that the convective precipitation to total precipitation ratio exceeds 0.5, particularly in peninsular India during the major monsoon seasons (Romatschke and Houze 2011; Sreenath et al. 2022). The explicit resolution of convective elements posed challenges within the current microphysics schemes (Samanta et al. 2023). Previous studies have suggested that utilizing cumulus parameterization at a 4-km scale performed better for certain events (Kirthiga et al. 2021; Wang et al. 2021). However, ENS1 displayed higher skill during the North-East Monsoon (NEM) season (POD – 0.7) and Winter (WIN) season (POD – 0.4), particularly for DOM3, the eastern coasts. The NEM season is mainly driven by large-scale dynamics. Initially, as easterly winds move inland from the ocean, there is minimal contribution from convective precipitation elements when precipitation happens in the coasts. However, increased convective interactions are observed as they interact with the land and penetrate further inland, resulting in a drastic reduction in the skill of ENS1 for DOM4 (POD < 0.45). The increase in RMSE values as lead-time increases was notably higher with the KF scheme, with an increasing error trend rate reaching about 30% in

ENS1. It is important to highlight that that despite the BMJ0 scheme not exhibiting a linear trend (< 4% increase) in RMSE values as lead-time increases, it recorded the highest error (RMSE—6.5 mm/6 h) during the late night-early morning time. However, the BMJ scheme exhibited lower bias in simulated temperatures (< 0.5 ℃) and relative humidity, and it was noted to effectively capture rainfall from weak convective systems (also reported by Kanase et al. 2020). Specifically, the performance of ENS3 was higher (POD – 0.45) during the North-East Monsoon (NEM) season, particularly for events influenced by easterly winds and low-level convergence.

The CUM cluster with KF schemes (ENS2, ENS5, ENS6) increased rain occurrence predictability (> 1 mm/6 h) from the No-CUM variant of the KF scheme (ENS1) by about 129%, with the POD improving from 0.31 to 0.71 (Table 5). Similarly, a 35% increase in the POD of > 10.1 mm/6 h rainfall and 29% increase in the POD of > 20.1 mm/6 h rainfall was recorded. It is evident that this increase in predictability occurred without a rise in false alarms, as the false alarm ratio remained below 0.45 across the members, similar to ENS1 (Table 5). The relative humidity dependent additional perturbation (KF1_2) demonstrated higher POD, particularly for higher rainfall thresholds (POD—0.65, across seasons). However, a significant false alarm ratio (0.6) and wide overestimation (mean bias – 0.5 mm/6 h) for low-moderate rains was also noticed with this member, ENS5.

The probability distribution function (wet rain intensities > 1 mm/6 h) of the forecasts followed a similar pattern to that of IMDAA (Fig. S4). The differences in the cumulative distribution function (CDF) between the ensemble members increased with lead time. However, most ensemble members overestimated extreme events (right tail), particularly as the lead time increased. Notably, ENS3 (utilizing the BMJ cumulus scheme and WDM6 microphysics scheme) and ENS4 (employing the NT cumulus scheme and Lin microphysics scheme) outperformed in simulating low to moderate rain intensities (< 10.1 mm/6 h), even with a lead time of 72 h.

In summary, the analysis of error metrics and distribution graphs revealed that the ensemble framework slightly overestimated moderate to heavy events (> 10.1 mm/6 h) and exhibited higher uncertainty in those simulations as lead time increased. Interestingly, the findings suggest that the ensemble members exhibited similar performance during the initial forecast hours (6–24 h), but afterward, their error growth rates diverged.

3.4 Investigating the added value by the multiphysics ensembles at an intra-seasonal scale across Southern India

The added value of the proposed convection-allowing resolution multiphysics framework (WRF-CP7) was assessed by comparing the ensemble mean and spatiotemporal spread to those from NCEP-GEFS20 (refer Table 2 for more details). Figure 6 presents domain-wise averaged 6-hourly precipitation plots across different seasons. During the Southwest Monsoon/Indian Summer Monsoon (SWM) season, which exhibits higher rainfall occurrences between 6–18 UTC (11.30 AM—11.30 PM IST) in DOM1 and DOM2 (core SWM regions), the WRF-CP7 ensembles accurately represented the diurnal variability (correlation coefficient – 0.75). As mentioned earlier, the spread becomes more pronounced with increased lead time. However, the NCEP-GEFS20 performed poorly in capturing diurnal variability (correlation coefficient – 0.56), rainfall intensities, and spread among ensemble members during the SWM season. In the Northeast Monsoon (NEM) season, major rainfall occurs between 0–6 UTC (5.30 AM—12 PM Indian Standard Time (IST)) in the core monsoon domain, specifically the DOM3-Eastern Coast region. DOM3 is the first inland convergence zone, as the easterlies and associated depressions bring moisture from the Bay of Bengal. The system then moves with less intensity to Central TN (DOM4), where significant rainfall occurs between 12–18 UTC (5.30 PM—11.30 PM IST). Regarding rainfall initiation, the NCEP-GEFS20 and WRF-CP7 ensemble mean showed a slight difference, with the WRF-CP7 capturing the initiation time accurately in DOM3 but not accurately representing the peak rainfall occurrences in DOM4. Interestingly, although the NCEP-GEFS20 exhibited less spread (± 1 mm/6 h) among ensemble members, it closely followed the diurnal pattern and rain intensities observed by IMDAA for the core NEM monsoon regions (DOM3 and DOM4). NCEP-GEFS20 also captured the patterns in DOM1 and DOM2 well during the NEM period, while the WRF-CP7 slightly overestimated the peak rainfall occurrences.

Both models, WRF-CP7 and NCEP-GEFS20, capture the initiation of pre-monsoon summer (SUM) convective-type rainfall (12 UTC, 2.30 PM—5.30 PM IST) well across the domains, with slight over-prediction by the WRF-CP7 mean and slight underestimation by the NCEP-GEFS20 mean. In the DOM4 region (Central Tamilnadu—Cauvery river basin), the inland heating and local moisture fluxes enhance SUM activity, and the WRF-CP7 ensemble framework well captured this pattern. The POD for SUM events showed a 3% increase in the rain/no rain category, with the WRF simulated SUM events demonstrating a POD 30% higher than that of GEFS. However, on average, the POD of WRF ensemble mean for rain/no rain occurrences stands at 0.6. This indicates a significant opportunity exists for improving predictions in this season due to its substantial uncertainty, which arises from its shorter time-scales and greater flux transfer at low-levels. Studies have suggested that improving parameterizations of land surface processes can enhance skills at convective scales (Osuri et al. 2017).

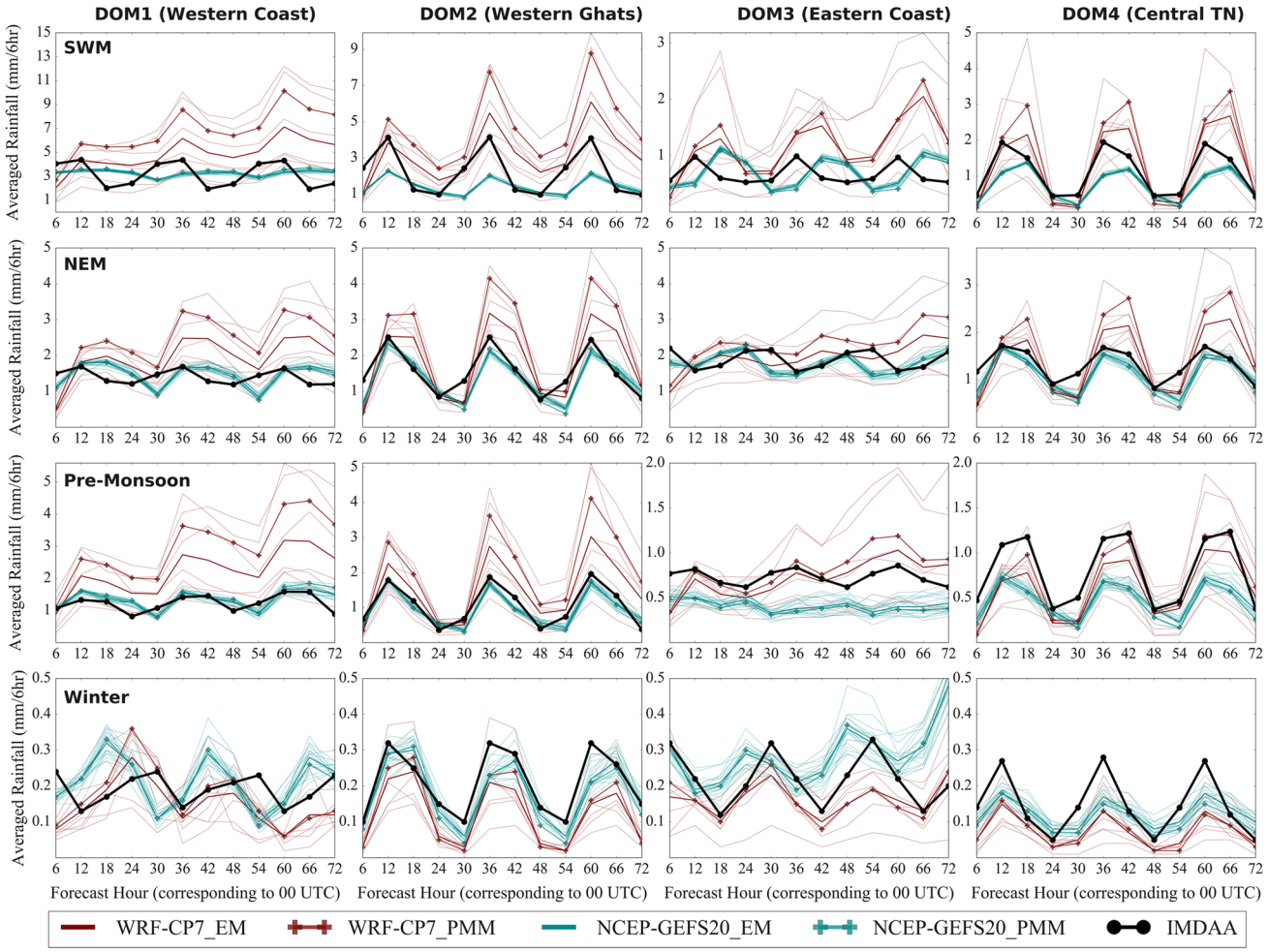

Fig. 6 Diurnal variations in 6-hourly rain intensities across lead times. The observed rainfall intensities from IMDAA are represented by black lines and circle markers. Model-simulated precipitation is displayed for the ensemble mean of WRF-CP7_EM (maroon) and NCEP-GEFS20_EM (darkcyan) ensemble framework. Additionally, the results of the Probability Matching technique Mean (PMM) for WRF-CP7_PMM (maroon with + markers) and NCEP-GEFS20_PMM (darkcyan with + markers) are presented. The individual members from WRF-CP7 and NCEP-GEFS20 are depicted in shades lighter than their respective models. The rows correspond to different seasons (SWM, NEM, Pre-Monsoon, and Winter), and the columns represent various domains (DOM1: Western Coast, DOM2: Western Ghats, DOM3: Eastern Ghats, and DOM4: Central Tamilnadu). The x-axis reflects lead times at 6-h intervals, aligning with 00 UTC (5:30 AM IST), while the y-axis indicates the averaged rain intensities (mm/6 h)

Both models perform well in capturing the characteristics of winter precipitation (WIN). Interestingly, the NCEP-GEFS20 ensemble mean tends to be higher than the WRF-CP7 ensemble mean and more comparable to IMDAA rain rates for the WIN season. The multiphysics WRF-CP7 framework showed good performance for warm-rain type processes, while the cold-rain processes and associated mesoscale forcing during winter are slightly better represented by the NCEP-GEFS20 ensembles, including perturbed initial conditions.

The spatial maps displaying averaged daily precipitation in Fig. 7 reveal that the multiphysics ensemble mean generated by WRF-CP7 provides more detailed spatial information regarding daily precipitation distributions across seasons and domains. The NCEP-GEFS20 ensemble mean exhibits a significant dry bias for DOM2 and DOM3 in all seasons (SWM, SUM, and WIN), except for NEM. The NCEP-GEFS20 simulations failed to adequately represent the inland rainfall zone in DOM1 during the SWM season, shifting the critical rainfall zone towards the coast and ocean. Studies have also shown the lower skill of NCEP-GEFS over core monsoon regions (DOM1 and DOM2) of South-West Monsoon season with a larger dry bias (Dube et al. 2017; Saminathan et al. 2021). This dry bias was reduced in WRF-CP7 simulations, showing a 25–30% reduction in average RMSE to 6.6 mm/day across evaluation zones. The improved representation of Western Ghats terrain characteristics and better simulation of diurnal patterns

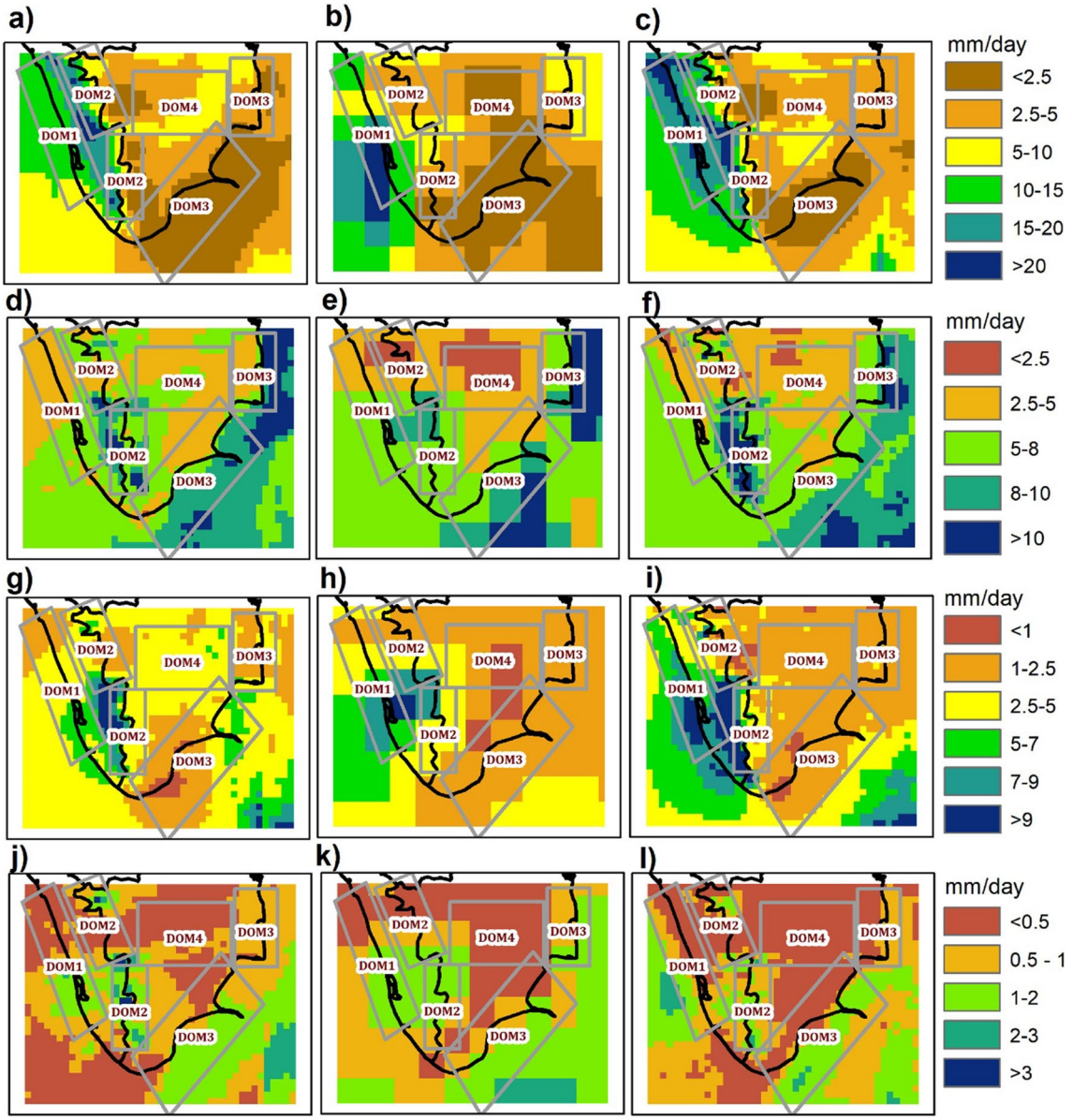

Fig. 7 Spatial averaged precipitation across seasons (unit – mm/day). The rows denote the seasons viz. SWM (**a-c**), NEM (**d-f**), SUM (**g-i**), and WIN (**j-l**). a, d, g, j denotes the IMDAA averaged precipitation climatology, b, e, h, k denotes the NCEP-GEFS20 simulated ensemble mean precipitation, and c, f, I, l denotes the WRF-CP7 multi-physics framework ensemble mean precipitation

of temperature and humidity profiles by the WRF-CP7 led to the improvement over NCEP-GEFS20 (not shown here). Studies have also recorded the superior performance of regional scale models in resolving the complex orographic related rainfall mechanisms (Kirthiga and Patel 2018). Interestingly, while NCEP-GEFS20 accurately simulated the spatial variability of NEM in eastern coastal zones (DOM3), the WRF-CP7 ensemble mean slightly underpredicted coastal rainfall. During the North-East monsoon season, which is characterized by tropical cyclones driven by large-scale dynamics and easterly trough-related activities, the synoptic to mesoscale processes were well represented by GEFS ensembles. Despite being driven by the control member of this system, WRF-CP7 showed minimal influence on RMSE (10–15% reduction compared to GEFS) in precipitation simulations during the NEM season for the DOM3. However, as NEM systems moved over land, microphysical processes became highly complex, with more localized convective elements, which was not well represented by coarser resolution NCEP-simulations. This resulted in a dry bias in the inland region of central Tamil Nadu (DOM4), which was better resolved by the WRF-CP7 ensemble mean at the convective scale, reducing spatial shift errors by 20–30%. During pre-monsoon convective thunderstorms, WRF-CP7

outperformed NCEP-GEFS, reducing RMSE by 20–25% in the Western Ghats regions, with an average improvement of 10–15% across other domains. However, WRF-CP7 simulations showed limitations in capturing the complexity of fluxes, with a 60% probability of detection and a 60% False Alarm Ratio. It is to be noted that the pre-summer monsoon rainfall is characterized by thunderstorm activity with huge spatial variability of mixed-phase hydrometeors. Addressing this, increasing spatial resolution to 1 km or less and improving parameterization of local fluxes through better representation of surface processes can help reduce uncertainty for this season (Kirthiga and Patel 2018; Sati and Mohan 2021; Prasad et al. 2024).

The Fractional Skill Score (FSS) plots (Fig. 8) displayed a clear superiority of the WRF-CP7 forecasts over the NCEP-GEFS20 forecasts in representing the intra-seasonal spatial variability. When a target FSS score of 0.5 was considered satisfactory, as Woodhams et al. 2018 mentioned, the FSS values for the NCEP-GEFS20 simulations only exceeded 0.5 for the threshold > 2.5 mm/day (rain/no rain event). On the other hand, the FSS values for the WRF-CP7 ensembles exceeded 0.5 even for higher thresholds (> 10 mm/day). Analyzing the count of rain objects in Figure S5, the NCEP-GEFS20 ensembles showed significant under-prediction of the count and exhibited low spatial spread. In contrast, the WRF-CP7 members realistically simulated the count of rain objects, matching the estimated objects by IMDAA. Interestingly, the most significant deviation among the WRF-CP7 ensembles occurs in DOM3, the eastern coastal region, regarding the count of simulated rain objects. Figure S6 showcases displacement plots showing the direction and number of grids displaced for the simulated rain objects > 10.1 mm/day (moderate-very heavy rainfall category). The negative values in Figure S6a represent a northward shift in the simulated rain objects, while the positive values denote a southward shift in the objects. An ideal value of zero represents the simulated rain object centroid is near the ones from IMDAA centroids. Similarly, figure S6b represents eastward (negative) and westward (positive) direction shifts. The WRF-CP7 reduced the shifts in rain object centroids in the overall shift in the simulated rain objects, and in addition, the uncertainty with the misplaced centroid was also limited. In the SWM season, the NCEP-GEFS20 mean displayed a westward (+ ve) shift in DOM1, while the WRF-CP7 multiphysics mean reduced the spatial bias.

Table 6 lists the RMSE, POD, and FAR values for different years, seasons, and forecast lead times (until 3 days). The WRF-CP7 ensemble mean showed a higher error growth rate than the NCEP-GEFS20 mean across lead times. Consequently, the POD was higher for the WRF-CP7 simulations, especially for higher thresholds. Additionally, the FAR was smaller for the WRF-CP7 ensemble mean throughout the seasons and years. Notably, the POD and FAR for rain occurrences (> 2.5 mm/day threshold) were higher for the NCEP-GEFS20 ensembles, particularly for NEM 2015. However, the POD and FAR for > 10 mm/day were noticeably enhanced by the WRF multi-physics ensemble simulations. Therefore, while the NCEP-GEFS20 ensembles were able to capture rain events, they significantly under-predicted moderate or higher intensity events throughout all years and seasons. A 30% increase in POD by WRF-CP7 was observed across all seasons for moderate to heavy rain intensities, accompanied by a 10% decrease in FAR against GEFS forecasts. The higher lead times of WRF-CP7 forecasts recorded higher spread and error growth rate for precipitation simulations in the coastal domains, particularly DOM1 and DOM3, during SWM and NEM seasons, respectively. The bias could be attributed either to the error in the input data (as reflected by GEFS performance) or lesser distance from the outflow boundary (Lavin-Gullon et al. 2021), or inherent issues with WRF in simulating coastline interactions (Hock et al. 2022).

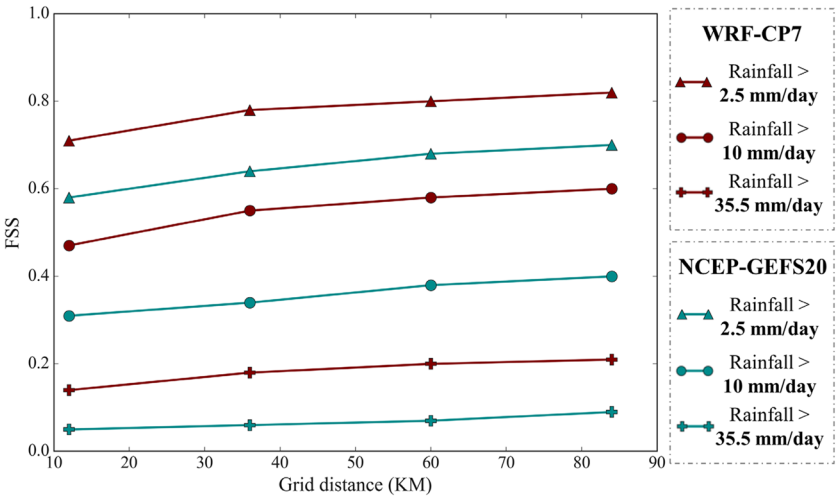

Fig. 8 Fractional Skill Scores (FSS) plots at different spatial scales and rain intensity thresholds for ensemble mean from WRF-CP7 and NCEP-GEFS20

Table 6 Performance metrics of the simulated daily precipitation from the proposed CP ensemble framework – WRFCP7 against NCEP-GEFS20 at intra-seasonal timescale and lead times

		RMSE (mm/day)		POD—Threshold > 2.5 mm/day		POD—Threshold > 10 mm/day		FAR—Threshold > 2.5 mm/day		FAR—Threshold > 10 mm/day	
	Lead Time	NCEP-GEFS20	WRF-CP7	NCEP-GEFS20	WRF-CP7	NCEP-GEFS20	WRF-CP7	NCEP-GEFS20	WRF-CP7	NCEP-GEFS20	WRF-CP7
SWM (2016)—Mean Rainfall (4.6 mm/day)	Day1	8.01	6.61	0.65	0.66	0.49	0.61	0.36	0.29	0.55	0.5
	Day2	7.83	9.75	0.63	0.69	0.41	0.58	0.37	0.35	0.49	0.46
	Day3	7.59	11.06	0.67	0.68	0.48	0.53	0.39	0.39	0.53	0.38
SWM (2017)—Mean Rainfall (6.6 mm/day)	Day1	10.01	7.81	0.66	0.67	0.3	0.43	0.33	0.26	0.68	0.49
	Day2	9.75	9.4	0.62	0.68	0.51	0.57	0.31	0.29	0.53	0.4
	Day3	9.73	11.63	0.65	0.69	0.45	0.64	0.31	0.32	0.48	0.45
NEM (2015)—Mean Rainfall (9.2 mm/day)	Day1	11.25	10.3	0.85	0.7	0.32	0.46	0.22	0.13	0.69	0.63
	Day2	10.66	11.7	0.87	0.73	0	0.13	0.23	0.14	0	0.88
	Day3	10.56	13.21	0.89	0.73	0.3	0.34	0.24	0.16	0.72	0.52
NEM (2016)—Mean Rainfall (3 mm/day)	Day1	5.61	4.61	0.53	0.55	0.51	0.63	0.53	0.3	0.54	0.6
	Day2	5.53	5.19	0.48	0.52	0.38	0.59	0.53	0.3	0.51	0.51
	Day3	5.14	5.48	0.52	0.49	0.48	0.61	0.52	0.33	0.5	0.4
NEM (2017)—Mean Rainfall (8.3 mm/day)	Day1	12.01	9.86	0.74	0.66	0.24	0.45	0.31	0.18	0.69	0.5
	Day2	11.2	11.1	0.74	0.66	0.52	0.57	0.29	0.2	0.5	0.43
	Day3	11	13.26	0.77	0.66	0.45	0.65	0.31	0.23	0.44	0.5
SUM (2016)—Mean Rainfall (4.4 mm/day)	Day1	7.02	6.04	0.6	0.63	0.25	0.46	0.42	0.34	0.62	0.6
	Day2	7.02	7.55	0.59	0.62	0	0.04	0.4	0.37	1	0.95
	Day3	6.83	10.57	0.59	0.59	0.3	0.3	0.38	0.38	0.74	0.54
SUM (2017)—Mean Rainfall (2.8 mm/day)	Day1	4.71	4.6	0.54	0.56	0.56	0.62	0.51	0.39	0.53	0.61
	Day2	4.91	5.28	0.49	0.53	0.42	0.61	0.49	0.4	0.53	0.55
	Day3	5	5.96	0.52	0.5	0.53	0.6	0.46	0.41	0.48	0.44
WIN (2016)—Mean Rainfall (0.5 mm/day)	Day1	1.29	1.27	0.19	0.29	0.29	0.41	0.76	0.66	0.65	0.55
	Day2	1.51	1.38	0.15	0.18	0.5	0.55	0.86	0.7	0.51	0.49
	Day3	1.84	1.27	0.28	0.2	0.53	0.64	0.88	0.74	0.4	0.51
WIN (2017)—Mean Rainfall (1.1 mm/day)	Day1	2.77	2.09	0.54	0.5	0.28	0.46	0.55	0.41	0.48	0.59
	Day2	2.76	2.02	0.55	0.48	0.08	0.09	0.58	0.4	0.96	0.91
	Day3	3.08	2.17	0.58	0.45	0.3	0.31	0.62	0.4	0.77	0.56

Figure 9 depicted the variance and MSE for day 1 forecasts of WRF-CP7, closely aligned with the 1:1 line and falling within the lower quadrant for most days across seasons. The NCEP-GEFS20 forecasts did not account for the error in simulated precipitation. The correlation plots for day 2 and day 3 multiphysics WRF forecasts also aligned with the 1:1 line. The correlation plots of NCEP-GEFS20 for longer lead times degraded compared to day 1 forecasts, deviating from the 1:1 line (not shown in the manuscript for brevity). The rank histogram (inset in correlation plots of Fig. 9) for NCEP-GEFS20 ensembles values displayed a U-shaped distribution, indicating a highly uncertain ensemble framework

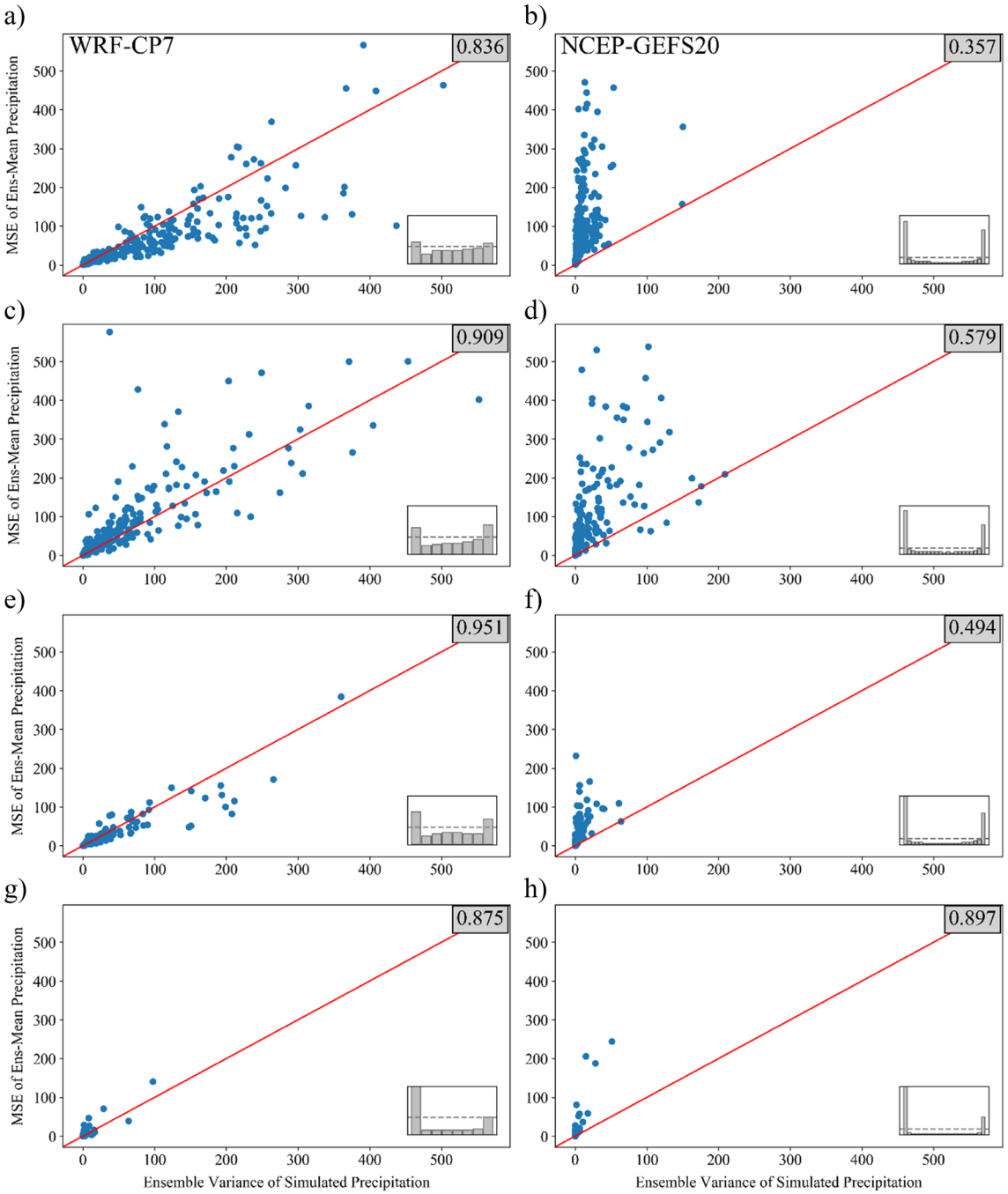

Fig. 9 Correlation plots for ensemble variance (X-axis) versus the mean squared error of ensemble mean (Y-axis). The rows represent the seasons viz. SWM (a-b), NEM (c-d), SUM (e–f), and WIN (g-h). The column represents the WRF-CP7 multi-physics ensemble framework (a, c, e, g) and NCEP-GEFS20 ensemble framework (b, d, f, h). The inset shows corresponding rank histograms with reference value in horizontal dashed line. The values in the gray box denotes the Pearson's correlation coefficient

with both under-prediction and over-prediction. The rank histograms for WRF ensembles showed a slightly over-predicting model, with an increasing tendency to overestimate as the lead time increased. Though not discussed in detail (for brevity), it is worth noting that the Probability Matching technique Mean (PMM) did not demonstrate superior performance for either of the ensemble framework (Table 6).

4 Summary and conclusions

The primary focus of the study was to quantify the usefulness of convection-permitting resolution multiphysics ensemble for the simulation of year-long medium-range precipitation forecasts. The study involved three major objectives.

(i) Quantifying the predictability of precipitation across peninsular India at convection-permitting (CP) scales. The initial multi-physics members (WRF-CP55) were generated using a larger spectrum of the cumulus and microphysics parameterization combinations available with a Regional Climate Model (RCM) – Weather Research and Forecasting (WRF) model. Even though the 4 km is generally considered a convection-permitting resolution, the results from the analysis approve of the fuzziness in the usage of the cumulus scheme in the innermost domain (4 km resolution). In the present study, certain events with strong large-scale forcing, complex microphysics schemes resolved the precipitation explicitly at 4-km resolution (in agreement with studies Mukhopadhyay et al. 2010; Guo et al. 2022; Ou et al. 2020). The ratio of simulated convective precipitation to total simulated precipitation in domain 2 consistently exceeded 0.6 across all simulated events, with contributions varying from isolated convective elements to well-structured convective processes. When microphysics schemes were employed in domain 2 to explicitly resolve processes without a cumulus scheme, narrow bands of precipitation were simulated, in contrast to GPM observations, and delays in event initiation were observed. Consequently, it was observed that the simulation of intensely local or mesoscale convective events improved significantly when appropriate combinations of cumulus parameterization schemes and microphysics schemes were utilized within the 4 km domain. Studies have highlighted the challenges that current versions of microphysics schemes face in accurately capturing isolated convective elements, a common feature of significant rainfall events in subtropical climatology (Srinivas et al. 2013; Madala et al. 2014; Hazra et al. 2020; Samanta et al. 2023).

(ii) Designing a computationally efficient multiphysics ensemble framework, practically feasible from an operational point of view. Investigating the physics combinations, the resulting simulations indicated that selecting compatible schemes were essential to designing a time-efficient ensemble framework. When the convection was explicitly resolved, complex microphysics schemes were more relevant (as reported by Kirthiga et al. 2021). Some of the physics combinations were found to be highly inter-correlated. Thus, we proposed a smaller 7-member ensemble framework, WRF-CP7, based on the performances across rainfall mechanisms. A composite scaled score (CSS) that combines multiple evaluation metrics and inter-correlation analysis was used to arrive at the final seven-member ensemble. The suggested WRF-CP7 framework with smaller ensembles reduces the turnaround time without compromising the spread of the simulated precipitation fields. The study demonstrated the higher skill of No-CUM cases in simulating low-moderate rain categories (stratiform clouds and precipitation from weak convective storms). These results are comparable to previous studies with similar configurations and study regions (Mukhopadhyay et al. 2010; Srinivas et al. 2013; Das et al. 2015).

(iii) Evaluating the intra-seasonal predictability and reliability of WRF-CP7, 3-day forecasts for each day from Sep 2015 to Dec 2017. A total of 5544 simulations (792 days × 7 ensembles) were made with approximately 130–180 min of run-time for each simulation. The quantitative analysis suggested that the WRF-CP7 members represented the spatiotemporal variability of rainfall occurrences with varying thresholds and were dispersive. Higher confidence was recorded in the occurrence of a rain event with the WRF-CP7, and reduced false alarm ratios were reported. The diurnal variability was adequately represented by the No-CUM members, although they failed to capture peak intensities across seasons. Research indicates that the convective precipitation to total precipitation ratio exceeds 0.5, particularly in peninsular India during the major monsoon seasons (Romatschke and Houze 2011; Sreenath et al. 2022). The explicit resolution of convective elements posed challenges within the current microphysics schemes (Samanta et al. 2023). Previous studies have suggested that utilizing cumulus parameterization at a 4-km scale performed better for certain events (Kirthiga et al. 2021; Wang et al. 2021). The CUM cluster with KF schemes (ENS2, ENS5, ENS6) increased rain occurrence predictability (> 1 mm/6 h) from the No-CUM variant of the KF scheme (ENS1) by about 129%, with the POD improving from 0.31 to 0.71. Similarly, a 35% increase in the POD of 10.1 mm/6 h rainfall and 29% increase in the POD of 20.1 mm/6 h rainfall was recorded. It is evident that this increase in predictability occurred without a rise in false alarms, as the false alarm ratio remained below 0.45 across the members. The NT scheme recorded higher success ratio (0.45) across the rainfall intensities and lead-times. Notably, the error progression as lead-time increases was not very prominent (< 4% increasing trend) with the NT scheme. This scheme, being scale-aware exhibited faster convergence, taking 25–40% less time than the

longest runtime taken by the ensemble framework. The error growth (4–44% increasing trend) of the overall ensemble framework helped in sampling the uncertainty of future atmospheric states and, thus, increasing the predictability of the extreme events. These results align with previous studies investigating precipitation simulations with multiphysics ensembles at convection-permitting resolutions (Clark et al. 2010; Duda et al. 2014; Berner et al. 2015; Francis et al. 2020). However, the overall performance of the ensemble mean suggested that the multiphysics ensemble resulted in an overpredicting model with overestimation for low-moderate rain intensities while slightly under-predicting in the heavy-very heavy rain category.

(iv) The added value of the proposed framework was assessed through the comparisons against the coarser-scale publicly-available Global Ensemble Forecast System forecasts (NCEP-GEFS20). A 30% increase in POD by WRF-CP7 was observed across all seasons for moderate to heavy rain intensities, accompanied by a 10% decrease in FAR when compared to GEFS forecasts. The NCEP-GEFS20 ensembles were less dispersive (spatially and temporally) and thus resulted in an over-confident deterministic mean which did not capture the higher intensity rains across all the seasons and years for the study domain. The spread of the simulated precipitation between the ensemble members was also not sensitive to the lead times. The mean from the WRF-CP7 simulation reduced the biases, particularly in the land regions and for heavy rain thresholds, compared to the NCEP-GEFS20 ensembles. The spatial shift from the observed rain objects was reduced with a high-resolution WRF-CP7 ensemble mean, improving the spatial accuracy of the simulated precipitation. The suggested ensemble framework gave a superior performance for SWM and SUM seasons. However, for some of the NEM events and WIN season, NCEP-GEFS20 simulation with perturbed initial conditions showed significant performance.

Physics schemes in regional/global climate models face challenges to accurately represent major rain mechanisms in the Indian subcontinent (Samanta et al. 2021, 2023), necessitating a better understanding of rain system simulations at convection-allowing scales. Intensive campaigns like CAIPEEX (Prabha et al., 2011) and INCOMPASS (Hazra et al. 2020) aid in incorporating observed features into numerical model formulations. The present study addresses model uncertainty in forecasting precipitation for a longer lead time, particularly in the convective permitting scales across India. The study uniquely evaluates the ensemble members for simulating different precipitation thresholds. Therefore, the methods employed in this study are particularly relevant for supporting vulnerability impact assessment studies (VIA), especially in agriculture and water resources. However, in the present ensemble configuration, the deterministic ensemble mean overpredicted the low-moderate rainfall, particularly the bias becoming dominant with an increase in lead time. Some studies like Thornes and Stephenson 2001 have argued that missing a heavy rainfall event was more damaging than anticipated. Since the resulting model has a systematic wet bias, post-processing techniques like bias correction can effectively be explored to improve the precipitation forecasts for a longer lead time (Clark et al. 2016). Increasing the spatial resolution to 1 km or less can also help reduce the uncertainty induced by parameterizing convection (Schwartz et al. 2017; Frogner et al. 2019; Sofokleous et al. 2021). Techniques to address the input and model uncertainties are necessary for the ensemble framework, particularly in the tropics (Prakash et al. 2016; Huang and Luo 2017). The results from the study strongly advocate for extensive sampling within the model uncertainty space, coupled with input uncertainty, to capture better the entire spectrum of rainfall mechanisms occurring in the tropics.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00382-024-07296-x.

Acknowledgements We acknowledge the funding support from the Ministry of Water Resources, RD&GR, Govt. of India under Indian National Committee on Climate Change (INCCC) Grant no: 16/22/2016-R&D. We would like to thank the Centre of Development of Advanced Computing (C-DAC), Pune, (PARAM Utkarsh cluster) for hosting the 7-member ensemble model from October 2022 to March 2024. Their support greatly facilitated our understanding of many practical aspects of the run-time, as discussed in this study.

Author contributions Author 1: S M Kirthiga

Conception and design of the study, data download and processing, analysis and interpretation of data, drafting of the manuscript and approval of the version of the manuscript to be published and corresponding author.

Author 2: B Narasimhan

Conception of the study, providing resources, revising the manuscript critically for important intellectual content and approval of the version of the manuscript to be published.

Author 3: C Balaji

Conception of the study, providing resources, revising the manuscript critically for important intellectual content, approval of the version of the manuscript to be published.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We understand that the Corresponding Author is the sole contact for the Editorial process. He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability The datasets generated during or analyzed during the current study are not publicly available as the data size is huge but are available from the corresponding author upon reasonable request.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

Baker LH, Rudd AC, Migliorini S, Bannister RN (2014) Representation of model error in a convective-scale ensemble prediction system. Nonlinear Process Geophys 21:19–39. https://doi.org/10.5194/NPG-21-19-2014

Bei N, Zhang F (2007) Impacts of initial condition errors on mesoscale predictability of heavy precipitation along the Mei-Yu front of China. Q J R Meteorol Soc 133:83–99. https://doi.org/10.1002/QJ.20

Bouallègue ZB, Richardson DS (2022) On the ROC area of ensemble forecasts for rare events. Weather Forecast 37:787–796. https://doi.org/10.1175/WAF-D-21-0195.1

Berg LK, Gustafson WI, Kassianov EI, Deng L (2013) Evaluation of a Modified Scheme for Shallow Convection: Implementation of CuP and Case Studies. Mon Weather Rev 141:134–147. https://doi.org/10.1175/MWR-D-12-00136.1

Berner J, Ha SY, Hacker JP et al (2011) Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. Mon Weather Rev 139:1972–1995. https://doi.org/10.1175/2010MWR3595.1

Berner J, Fossell KR, Ha SY et al (2015) Increasing the Skill of Probabilistic Forecasts: Understanding Performance Improvements from Model-Error Representations. Mon Weather Rev 143:1295–1320. https://doi.org/10.1175/MWR-D-14-00091.1

Bouttier F, Vié B, Nuissier O, Raynaud L (2012) Impact of stochastic physics in a convection-permitting ensemble. Mon Weather Rev 140:3706–3721. https://doi.org/10.1175/MWR-D-12-00031.1

Bucci LR, O'Handley C, Emmitt GD et al (2018) Validation of an airborne doppler wind lidar in tropical cyclones. Sensors 18. https://doi.org/10.3390/s18124288

Charron M, Pellerin G, Spacek L et al (2010) Toward random sampling of model error in the Canadian ensemble prediction system. Mon Weather Rev 138:1877–1901. https://doi.org/10.1175/2009MWR3187.1

Chen SH, Sun WY (2002) A One-dimensional Time Dependent Cloud Model. J Meteorol Soc Japan Ser II 80:99–118. https://doi.org/10.2151/JMSJ.80.99

Clark AJ (2019) Comparisons of QPFs derived from single-and multicore convection-allowing ensembles. Weather Forecast 34:1955–1964. https://doi.org/10.1175/WAF-D-19-0128.1

Clark AJ, Gallus WA, Xue M, Kong F (2009) A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. Weather Forecast 24:1121–1140. https://doi.org/10.1175/2009WAF2222222.1

Clark AJ, Gallus WA, Xue M, Kong F (2010) Growth of spread in convection-allowing and convection-parameterizing ensembles. Weather Forecast 25:594–612. https://doi.org/10.1175/2009WAF2222318.1

Clark P, Roberts N, Lean H et al (2016) Convection-permitting models: a step-change in rainfall forecasting. Meteorol Appl 23:165–181. https://doi.org/10.1002/MET.1538

Das MK, Chowdhury AM, Das S (2015) Sensitivity Study with Physical Parameterization Schemes for Simulation of Mesoscale Convective Systems Associated with Squall Events. Int J Earth Atmos Sci 2:20–36

Dube A, Ashrit R, Singh H et al (2017) Evaluating the performance of two global ensemble forecasting systems in predicting rainfall over India during the southwest monsoons. Meteorol Appl 24:230–238. https://doi.org/10.1002/met.1621

Duda JD, Wang X, Kong F, Xue M (2014) Using varied microphysics to account for uncertainty in warm-season QPF in a convection-allowing ensemble. Mon Weather Rev 142:2198–2219. https://doi.org/10.1175/MWR-D-13-00297.1

Dudhia J (1989) Numerical Study of Convection Observed during the Winter Monsoon Experiment Using a Mesoscale Two-Dimensional Model. J Atmos Sci 46:3077–3107. https://doi.org/10.1175/1520-0469(1989)046%3c3077:NSOCOD%3e2.0.CO;2

Ebert EE, Damrath U, Wergen W, Baldwin ME (2003) Supplement to The WGNE Assessment of Short-term Quantitative Precipitation Forecasts. Bull Am Meteorol Soc 84:492–492. https://doi.org/10.1175/bams-84-4-ebert

Fischer G, Nachtergaele FO, van Velthuizen HT, Chiozza F, Franceschini G, Henry M, Muchoney D, Tramberend S (2021) Global agro-ecological zones v4 – model documentation. Rome, FAO. https://doi.org/10.4060/cb4744en

Francis T, Jayakumar A, Mohandas S et al (2020) Simulation of a mesoscale convective system over Northern India: Sensitivity to convection partitioning in a regional NWP model. Dyn Atmos Ocean 92:101162. https://doi.org/10.1016/j.dynatmoce.2020.101162

Fritsch JM, Houze RA, Adler R et al (1998) Quantitative Precipitation Forecasting: Report of the Eighth Prospectus Development Team, U.S. Weather Research Program. Bull Am Meteorol Soc 79:285–299. https://doi.org/10.1175/1520-0477(1998)079

Frogner IL, Singleton AT, Køltzow M, Andrae U (2019) Convection-permitting ensembles: Challenges related to their design and use. Q J R Meteorol Soc 145:90–106. https://doi.org/10.1002/QJ.3525

Gebhardt C, Theis SE, Paulat M, Ben Bouallègue Z (2011) Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. Atmos Res 100:168–177. https://doi.org/10.1016/j.atmosres.2010.12.008

Grell GA, Freitas SR (2014) A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. Atmos Chem Phys 14:5233–5250. https://doi.org/10.5194/ACP-14-5233-2014

Guo Z, Fang J, Shao M et al (2022) Improved summer daily and sub-daily precipitation over Eastern China in convection-permitting simulations. Atmos Res 265:105929. https://doi.org/10.1016/j.atmosres.2021.105929

Gupta V, Jain MK, Singh PK, Singh V (2020) An assessment of global satellite-based precipitation datasets in capturing precipitation extremes: A comparison with observed precipitation dataset in India. Int J Climatol 40:3667–3688. https://doi.org/10.1002/joc.6419

Halder M, Mukhopadhyay P (2016) Microphysical processes and hydrometeor distributions associated with thunderstorms over India: WRF (cloud-resolving) simulations and validations using TRMM. Nat Hazards 83:1125–1155. https://doi.org/10.1007/s11069-016-2365-2

Hanley KE, Pirret JSR, Bain CL et al (2021) Assessment of convection-permitting versions of the Unified Model over the Lake Victoria basin region. Q J R Meteorol Soc 147:1642–1660. https://doi.org/10.1002/qj.3988

Hazra V, Pattnaik S, Sisodiya A et al (2020) Assessing the performance of cloud microphysical parameterization over the Indian region: Simulation of monsoon depressions and validation with INCOMPASS observations. Atmos Res 239:104925. https://doi.org/10.1016/j.atmosres.2020.104925

Hock N, Zhang F, Pu Z (2022) Numerical Simulations of a Florida Sea Breeze and Its Interactions with Associated Convection: Effects of Geophysical Representation and Model Resolution. Adv Atmos Sci 39:697–713. https://doi.org/10.1007/s00376-021-1216-6

Hong SY, Noh Y, Dudhia J (2006) A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes. Mon Weather Rev 134:2318–2341. https://doi.org/10.1175/MWR3199.1

Huang L, Luo Y (2017) Evaluation of quantitative precipitation forecasts by TIGGE ensembles for south China during the presummer rainy season. J Geophys Res Atmos 122:8494–8516. https://doi.org/10.1002/2017JD026512

Hutchinson TA (2007) An adaptive time-step for increased model efficiency. In: Extended abstracts, eighth WRF users' workshop. p 4. https://ams.confex.com/ams/pdfpapers/153842.pdf

Indirarani S, Arulalan T, George JP et al (2021) IMDAA: High-resolution satellite-era reanalysis for the indian monsoon region. J Clim 34:5109–5133. https://doi.org/10.1175/JCLI-D-20-0412.1

Janjić ZI (1994) The Step-Mountain Eta Coordinate Model: Further Developments of the Convection, Viscous Sublayer, and Turbulence Closure Schemes. Mon Weather Rev 122:927–945. https://doi.org/10.1175/1520-0493(1994)122%3c0927:TSMECM%3e2.0.CO;2

Jeworrek J, West G, Stull R (2019) Evaluation of Cumulus and Microphysics Parameterizations in WRF across the Convective Gray Zone. Weather Forecast 34:1097–1115. https://doi.org/10.1175/WAF-D-18-0178.1

Jeworrek J, West G, Stull R (2021) WRF Precipitation Performance and Predictability for Systematically Varied Parameterizations over Complex Terrain. Weather Forecast 36:893–913. https://doi.org/10.1175/WAF-D-20-0195.1

Jiménez PA, Dudhia J, González-Rouco JF et al (2012) A Revised Scheme for the WRF Surface Layer Formulation. Mon Weather Rev 140:898–918. https://doi.org/10.1175/MWR-D-11-00056.1

Kain JS (2004) The Kain-Fritsch Convective Parameterization: An Update. J Appl Meteorol Climatol 43:170–181. https://doi.org/10.1175/1520-0450(2004)043

Kanase RD, Deshpande MS, Krishna RPM, Mukhopadhyay P (2020) Evaluation of convective parameterization schemes in simulation of tropical cyclones by Climate Forecast System model: Version 2. J Earth Syst Sci 129:168. https://doi.org/10.1007/s12040-020-01433-w

Kirthiga SM, Narasimhan B, Balaji C (2021) A multi-physics ensemble approach for short-term precipitation forecasts at convective permitting scales based on sensitivity experiments over southern parts of peninsular India. J Earth Syst Sci 130. https://doi.org/10.1007/s12040-021-01556-8

Kirthiga SM, Patel NR (2018) Impact of updating land surface data on micrometeorological weather simulations from the WRF model. Atmósfera 31:165–183

Lavin-Gullon A, Fernandez J, Bastin S et al (2021) Internal variability versus multi-physics uncertainty in a regional climate model. Int J Climatol 41:E656–E671. https://doi.org/10.1002/joc.6717

Leutbecher M, Lock SJ, Ollinaho P et al (2017) Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. Q J R Meteorol Soc 143:2315–2339. https://doi.org/10.1002/qj.3094

Li P, Furtado K, Zhou T et al (2018) (2018) The diurnal cycle of East Asian summer monsoon precipitation simulated by the Met Office Unified Model at convection-permitting scales. Clim Dyn 551(55):131–151. https://doi.org/10.1007/S00382-018-4368-Z

Lim KSS, Hong SY (2010) Development of an Effective Double-Moment Cloud Microphysics Scheme with Prognostic Cloud Condensation Nuclei (CCN) for Weather and Climate Models. Mon Weather Rev 138:1587–1612. https://doi.org/10.1175/2009MWR2968.1

Lorenz EN (1969) The predictability of a flow which possesses many scales of motion. Tellus 21:289–307. https://doi.org/10.3402/tellusa.v21i3.10086

Ma LM, Tan ZM (2009) Improving the behavior of the cumulus parameterization for tropical cyclone prediction: Convection trigger. Atmos Res 92:190–211. https://doi.org/10.1016/J.ATMOSRES.2008.09.022

Madala S, Satyanarayana ANV, Rao TN (2014) Performance evaluation of PBL and cumulus parameterization schemes of WRF ARW model in simulating severe thunderstorm events over Gadanki MST radar facility - Case study. Atmos Res 139:1–17. https://doi.org/10.1016/j.atmosres.2013.12.017

Madhulatha A, Rajeevan M (2018) Impact of different parameterization schemes on simulation of mesoscale convective system over south-east India. Meteorol Atmos Phys 130:49–65. https://doi.org/10.1007/s00703-017-0502-4

Mahala BK, Mohanty PK, Nayak BK (2015) Impact of Microphysics Schemes in the Simulation of Cyclone Phailinusing WRF Model. Procedia Eng 116:655–662. https://doi.org/10.1016/j.proeng.2015.08.342

Melhauser C, Zhang F (2012) Practical and intrinsic predictability of severe and convective weather at the mesoscales. J Atmos Sci 69:3350–3371. https://doi.org/10.1175/JAS-D-11-0315.1

Melhauser C, Zhang F, Weng Y et al (2017) A Multiple-Model Convection-Permitting Ensemble Examination of the Probabilistic Prediction of Tropical Cyclones: Hurricanes Sandy (2012) and Edouard (2014). Weather Forecast 32:665–688. https://doi.org/10.1175/WAF-D-16-0082.1

Mlawer EJ, Taubman SJ, Brown PD et al (1997) Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J Geophys Res Atmos 102:16663–16682. https://doi.org/10.1029/97JD00237

Morrison H, Thompson G, Tatarskii V (2009) Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes. Mon Weather Rev 137:991–1007. https://doi.org/10.1175/2008MWR2556.1

Mukhopadhyay P, Taraphdar S, Goswami BN, Krishnakumar K (2010) Indian summer monsoon precipitation climatology in a high-resolution regional climate model: Impacts of convective parameterization on systematic biases. Weather Forecast 25:369–387. https://doi.org/10.1175/2009WAF2222320.1

Mullen SL, Buizza R (2001) Quantitative precipitation forecasts over the United States by the ECMWF ensemble prediction system. Mon Weather Rev 129:638–663. https://doi.org/10.1175/1520-0493(2001)129%3c0638:QPFOTU%3e2.0.CO;2

Musaid PP, Manoj MG, Panda SK et al (2023) Dynamical influence of West Pacific Typhoons on the 2018 historic flood of Kerala as revealed by the weather research and forecasting (WRF) model. Clim Dyn 1:1–19. https://doi.org/10.1007/S00382-022-06648-9/FIGURES/19

Nekkali YS, Osuri KK, Das AK, Niyogi D (2022) Understanding the characteristics of microphysical processes in the rapid intensity changes of tropical cyclones over the Bay of Bengal. Q J R Meteorol Soc 148:3715–3729. https://doi.org/10.1002/qj.4384

Osuri KK, Nadimpalli R, Mohanty UC et al (2017) Improved prediction of severe thunderstorms over the Indian Monsoon region using high-resolution soil moisture and temperature initialization. Sci Rep 7:41377. https://doi.org/10.1038/srep41377

Ou T, Chen D, Chen X et al (2020) Simulation of summer precipitation diurnal cycles over the Tibetan Plateau at the gray-zone grid spacing for cumulus parameterization. Clim Dyn 54:3525–3539. https://doi.org/10.1007/s00382-020-05181-x

Pan H-L, Wu W-S (1995) Implementing a mass flux convection parameterization package for the NMC medium-range forecast model. https://repository.library.noaa.gov/view/noaa/11429

Pithani P, Ghude SD, Prabhakaran T et al (2019) WRF model sensitivity to choice of PBL and microphysics parameterization for an advection fog event at Barkachha, rural site in the Indo-Gangetic

basin, India. Theor Appl Climatol 136:1099–1113. https://doi.org/10.1007/S00704-018-2530-5/FIGURES/10

Prakash S, Mitra AK, Momin IM et al (2016) Skill of short- to medium-range monsoon rainfall forecasts from two global models over India for hydro-meteorological applications. Meteorol Appl 23:574–586. https://doi.org/10.1002/MET.1579

Prasad SK, Saha K, Shanker G et al (2024) Evaluating lightning forecasts of a convective scale ensemble prediction system over India. Theor Appl Climatol. https://doi.org/10.1007/s00704-024-04880-3

Prein AF, Langhans W, Fosser G et al (2015) A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. Rev Geophys 53:323–361. https://doi.org/10.1002/2014RG000475

Risanto CB, Chang HI, Luong TM et al (2022) Retrospective subseasonal forecasts of extreme precipitation events in the Arabian Peninsula using convective-permitting modeling. Clim Dyn 1:1–30. https://doi.org/10.1007/S00382-022-06336-8/FIGURES/2

Roberts NM, Lean HW (2008) Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. Mon Weather Rev 136:78–97. https://doi.org/10.1175/2007MWR2123.1

Romatschke U, Houze RA (2011) Characteristics of Precipitating Convective Systems in the South Asian Monsoon. J Hydrometeorol 12:3–26. https://doi.org/10.1175/2010JHM1289.1

Romine GS, Schwartz CS, Berner J et al (2014) Representing forecast error in a convection-permitting ensemble system. Mon Weather Rev 142:4519–4541. https://doi.org/10.1175/MWR-D-14-00100.1

Rossa A, Nurmi P, Ebert E (2008) Overview of methods for the verification of quantitative precipitation forecasts. In: Michaelides S (ed) Precipitation: Advances in Measurement, Estimation and Prediction. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 419–452

Samanta S, Gayatri K, Murugavel P et al (2020) Case Study of a Convective Cluster Over the Rain Shadow Region of Western Ghats Using Multi-platform Observations and WRF Model. Pure Appl Geophys 177:2931–2957. https://doi.org/10.1007/s00024-019-02360-8

Samanta S, Prabha TV, Murugavel P, Suneetha P (2021) Rainfall types in the lifecycle of a stationary cloud cluster during the Indian Summer Monsoon: An investigation with numerical simulations and radar observation. Atmos Res 263:105794. https://doi.org/10.1016/j.atmosres.2021.105794

Samanta S, Prabha TV, Murugavel P, Suneetha P (2023) Morphological and microphysical characteristics associated with the lifecycle of a stationary cloud cluster during the Indian Summer Monsoon: A comparative study with numerical simulations and radar observation. Atmos Res 281:106464. https://doi.org/10.1016/j.atmosres.2022.106464

Saminathan S, Medina H, Mitra S, Tian D (2021) Improving short to medium range GEFS precipitation forecast in India. J Hydrol 598:126431. https://doi.org/10.1016/j.jhydrol.2021.126431

Sati AP, Mohan M (2021) Impact of urban sprawls on thunderstorm episodes: Assessment using WRF model over central-national capital region of India. Urban Clim 37:100869. https://doi.org/10.1016/j.uclim.2021.100869

Schwartz CS, Romine GS, Sobash RA et al (2015) Ncar's experimental real-time convection-allowing ensemble prediction system. Weather Forecast 30:1645–1654. https://doi.org/10.1175/WAF-D-15-0103.1

Schwartz CS, Romine GS, Fossell KR et al (2017) Toward 1-km ensemble forecasts over large domains. Mon Weather Rev 145:2943–2969. https://doi.org/10.1175/MWR-D-16-0410.1

Schwartz CS, Romine GS, Sobash RA et al (2019) NCAR's real-time convection-allowing ensemble project. Bull Am Meteorol Soc 100:321–343. https://doi.org/10.1175/BAMS-D-17-0297.1

Singh KS, Bonthu S, Purvaja R et al (2018) Prediction of heavy rainfall over Chennai Metropolitan City, Tamil Nadu, India: Impact of microphysical parameterization schemes. Atmos Res 202:219–234. https://doi.org/10.1016/j.atmosres.2017.11.028

Singh V, Koll R, Deshpande M (2020) The Unusual Long Track and Rapid Intensification of Very Severe Cyclone Ockhi. Current Science 119:771–779. https://doi.org/10.18520/cs/v119/i5/771-779

Singh T, Saha U, Prasad VS, Das GM (2021) Assessment of newly-developed high resolution reanalyses (IMDAA, NGFS and ERA5) against rainfall observations for Indian region. Atmos Res 259:105679. https://doi.org/10.1016/j.atmosres.2021.105679

Skamarock C, Klemp B, Dudhia J, et al (2019) A Description of the Advanced Research WRF Model Version 4. https://doi.org/10.5065/1DFH-6P97

Sofokleous I, Bruggeman A, Michaelides S et al (2021) Comprehensive methodology for the evaluation of high-resolution wrf multiphysics precipitation simulations for small, topographically complex domains. J Hydrometeorol 22:1169–1186. https://doi.org/10.1175/JHM-D-20-0110.1

Sreenath AV, Abhilash S, Vijaykumar P, Mapes BE (2022) West coast India's rainfall is becoming more convective. npj Climate pheric Science 5:36. https://doi.org/10.1038/s41612-022-00258-2

Srinivas CV, Bhaskar Rao DV, Yesubabu V et al (2013) Tropical cyclone predictions over the bay of bengal using the high-resolution advanced research weather research and forecasting (ARW) model. Q J R Meteorol Soc 139:1810–1825. https://doi.org/10.1002/qj.2064

Sumesh RK, Resmi EA, Unnikrishnan CK et al (2022) The extreme precipitation events of August 2018 and 2019 over southern Western Ghats, India: A microphysical analysis using in-situ measurements. Atmos Res 277:106322. https://doi.org/10.1016/j.atmosres.2022.106322

Surcel M, Zawadzki I, Yau MK (2015) A study on the scale dependence of the predictability of precipitation patterns. J Atmos Sci 72:216–235. https://doi.org/10.1175/JAS-D-14-0071.1

Tang Y, Lean HW, Bornemann J (2013) The benefits of the Met Office variable resolution NWP model for forecasting convection. Meteorol Appl 20:417–426. https://doi.org/10.1002/MET.1300

Tao WK, Wu D, Lang S et al (2016) High-resolution NU-WRF simulations of a deep convective-precipitation system during MC3E: Further improvements and comparisons between Goddard microphysics schemes and observations. J Geophys Res Atmos 121:1278–1305. https://doi.org/10.1002/2015JD023986

Tewari M, Wang W, Dudhia J et al (2016) Implementation and verification of the united NOAH land surface model in the WRF model. In: 20th Conference on Weather analysis and forecasting/16th conference on numerical weather prediction. pp 11–15. http://n2t.net/ark:/85065/d7fb523p

Thornes JE, Stephenson DB (2001) How to judge the quality and value of weather forecast products. Meteorol Appl 8:307–314. https://doi.org/10.1017/S1350482701003061

Walser A, Schär C (2004) Convection-resolving precipitation forecasting and its predictability in Alpine river catchments. J Hydrol 288:57–73. https://doi.org/10.1016/j.jhydrol.2003.11.035

Wang W (2022) Forecasting Convection with a "Scale-Aware" Tiedtke Cumulus Parameterization Scheme at Kilometer Scales. Weather Forecast 37:1491–1507. https://doi.org/10.1175/WAF-D-21-0179.1

Wang L, Shen X, Liu J, Wang B (2020) Model Uncertainty Representation for a Convection-Allowing Ensemble Prediction System Based on CNOP-P. Adv Atmos Sci 37:817–831. https://doi.org/10.1007/s00376-020-9262-z

Wang R, Qiao F, Liang X-Z et al (2021) Role of convection representation across the gray zone in forecasting warm season extreme precipitation over Shanghai from two typical cases. Atmos Res 253:105370. https://doi.org/10.1016/j.atmosres.2020.105370

Wastl C, Wang Y, Atencia A et al (2021) C-LAEF: Convection-permitting Limited-Area Ensemble Forecasting system. Q J R Meteorol Soc 147:1431–1451. https://doi.org/10.1002/QJ.3986

Weisman ML, Davis C, Wang W et al (2008) Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model. Weather Forecast 23:407–437. https://doi.org/10.1175/2007WAF2007005.1

Woodhams BJ, Birch CE, Marsham JH et al (2018) What is the added value of a convection-permitting model for forecasting extreme rainfall over tropical East Africa? Mon Weather Rev 146:2757–2780. https://doi.org/10.1175/MWR-D-17-0396.1

Yano JII, Ziemian'skiMi Z, Mi C et al (2018) Scientific challenges of convective-scale numerical weather prediction. Bulletin of the American Meteorological Society 99:699–710. https://doi.org/10.1175/BAMS-D-17-0125.1

Yussouf N, Stensrud DDJ (2012) Comparison of single-parameter and multiparameter ensembles for assimilation of radar observations using the ensemble kalman filter. Mon Weather Rev 140:562–586. https://doi.org/10.1175/MWR-D-10-05074.1

Zhang C, Wang Y (2017) Projected Future Changes of Tropical Cyclone Activity over the Western North and South Pacific in a 20-km-Mesh Regional Climate Model. J Clim 30:5923–5941. https://doi.org/10.1175/JCLI-D-16-0597.1

Zhang C, Wang Y (2018) Why is the simulated climatology of tropical cyclones so sensitive to the choice of cumulus parameterization scheme in the WRF model? Clim Dyn 51:3613–3633. https://doi.org/10.1007/s00382-018-4099-1

Zhang F, Qiang Sun Y, Magnusson L et al (2019) What is the predictability limit of midlatitude weather? Journal of the Atmospheric Sciences 76:1077–1091. https://doi.org/10.1175/JAS-D-18-0269.1

Zheng Y, Alapaty K, Herwehe JA et al (2016) Improving High-Resolution Weather Forecasts Using the Weather Research and Forecasting (WRF) Model with an Updated Kain-Fritsch Scheme. Mon Weather Review 144:833–860. https://doi.org/10.1175/MWR-D-15-0005.1

Zhou P, Ma M, Shao M, Tang J (2024) Sensitivity of summer precipitation simulation to the physical parameterizations in WRF over the Tibetan Plateau: A case study of 2018. Atmospheric Research 299:107174. https://doi.org/10.1016/j.atmosres.2023.107174