

Optimization of WRF Physics Parameterization Schemes for the Simulation of Extreme Rainfall Events: A Multi-Scheme Ensemble and Rank-Based Framework

Dr. Hiren Lekhadiya¹, Ms. Prakruti Dave²

^{1,2}Science and Humanities Department, School of Engineering, P P Savani University

Abstract - The accurate numerical simulation of extreme rainfall events remains a persistent challenge because the precipitation produced by the Weather Research and Forecasting (WRF) model is strongly sensitive to the choice of physical parameterization schemes. This paper presents a systematic optimization framework for identifying the configuration of microphysics (MP), cumulus (CU), planetary boundary layer (PBL), and land surface model (LSM) schemes that best reproduces a high-impact rainfall event over western India. A multi-scheme ensemble is constructed by combining candidate options for each physical process, and every member is integrated on a triple-nested (27/9/3 km) domain configuration. Simulated rainfall is verified against gridded observations using both continuous measures (correlation, root-mean-square error, percentage bias) and categorical skill scores (probability of detection, false-alarm ratio, critical success index, and equitable threat score). A rank-based aggregation procedure, implemented through both a rank-sum statistic and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), collapses the multi-metric evaluation into a single ordering of ensemble members and thereby isolates the optimal configuration. The methodology is fully reproducible and is designed so that event-specific simulation output can be substituted directly into the evaluation pipeline. Results indicate that cumulus and microphysics schemes exert the dominant control on simulated extreme rainfall, and that an optimal combination selected by objective ranking outperforms any arbitrarily chosen default configuration.

Keywords

Weather Research and Forecasting (WRF) model, extreme rainfall, parameterization optimization, microphysics, cumulus convection, ensemble verification, TOPSIS, numerical weather prediction.

1. INTRODUCTION

Extreme rainfall events are among the most damaging hydrometeorological hazards, producing flash floods, urban inundation, landslides, and large socio-economic losses. Over the Indian subcontinent the frequency and intensity of widespread heavy-rain episodes have risen significantly in recent decades [1], increasing the demand for skilful, high-resolution quantitative precipitation forecasts. Mesoscale numerical weather prediction (NWP) has become the principal tool for anticipating such events, and the steady improvement in forecast skill over the past decades—driven by better models, data assimilation, and computing—has been described as a quiet revolution in the atmospheric sciences [2].

The Weather Research and Forecasting (WRF) model, in its Advanced Research dynamical core (WRF-ARW), is the most widely used community mesoscale model for research and operational regional forecasting [3, 4]. A defining feature of WRF is its modular design: the user must select, for each sub grid scale physical process, one scheme from a large menu of available parameterizations. The principal processes are cloud micro-physics (MP), cumulus convection (CU), the planetary boundary layer (PBL), radiative transfer, and the land surface (LSM). Because these schemes represent processes that are unresolved or only partially resolved on the model grid, and because precipitation is the integrated end-product of their nonlinear interactions, simulated rainfall is acutely sensitive to the chosen configuration [5, 6].

This sensitivity is both a difficulty and an opportunity. It is a difficulty because no single configuration is universally optimal: the best-performing combination varies with region, season, synoptic regime, and event type. It is an opportunity because, for a given region and class of events, a carefully optimized configuration can yield forecast skill substantially exceeding that of an arbitrarily chosen default. A large body of work has therefore been devoted to sensitivity and optimization studies for heavy-rainfall simulation over India and elsewhere, including events over

Mumbai, the Kerala floods of 2018, and several monsoon-driven episodes [5–7]. These studies consistently report that cumulus and microphysics schemes dominate the simulated rainfall response, while PBL and land-surface choices act as secondary, but non-negligible, modulators.

A further complication specific to extreme rainfall is the role of model resolution. Convection in heavy-rain events is intrinsically multi-scale, and at the kilometre-scale grid spacings now routinely affordable the model enters the convective “grey zone,” in which deep convection is neither fully parameterized nor fully resolved. In this regime the interaction between the cumulus scheme (where still active) and the explicit microphysics is delicate, and scale-aware convective closures have been developed specifically to behave sensibly across the transition [19]. The configuration that is optimal at 9 km therefore need not be optimal at 3 km, which reinforces the need for a systematic, resolution-aware evaluation rather than the wholesale adoption of a default name list.

Despite this rich literature, two methodological gaps recur. First, many studies evaluate candidate configurations using a single metric (for example, domain-average bias or correlation), which can rank a configuration highly even when it badly mis-places the rainfall maximum. Robust evaluation of extreme rain- fall requires both continuous error measures and categorical, threshold-based skill scores that reward correct placement and intensity of heavy rain. Second, when multiple metrics are re- ported, the final choice of an “optimal” scheme is frequently made subjectively, without a transparent rule for aggregating possibly conflicting rankings.

This paper addresses both gaps by formulating WRF scheme selection as an explicit multi-criteria optimization problem. The contributions are as follows.

1. A reproducible multi-scheme ensemble design that spans the dominant physical processes (MP, CU, PBL, LSM) for an extreme rainfall event over western India.
2. A combined verification suite using continuous and categorical metrics computed against gridded reference data.
3. A transparent rank-based optimization that aggregates all metrics into a single ordering using both a rank-sum statistic and the TOPSIS multi-criteria decision method [8], al- lowing the optimal configuration to be identified objectively.

The remainder of the paper is organized as follows. Section II describes the study region, the target event, and the observational and forcing data. Section III details the WRF configuration, the physics ensemble, the verification metrics, and the rank-based optimization framework. Section IV presents and discusses the sensitivity and ranking results. Section V concludes.

2. RELATED WORK

Research on the sensitivity of WRF-simulated rainfall to physics options is extensive, and it is useful to organize it along three lines.

A. Microphysics-Focused Studies A first body of work isolates the microphysics scheme. For the catastrophic Kerala flood of 2018, a comparison of bulk microphysics schemes found that the simulated rainfall intensity and vertical hydrometeor structure depended strongly on whether the scheme was single- or double-moment and on its treatment of graupel [5]. Similar conclusions have been reached for Himalayan and peninsular events, where graupel-bearing schemes intensified convective cores and double-moment schemes improved the magnitude of the rainfall peak at the cost of additional computation. The general lesson is that microphysics governs intensity and structure more than placement.

B. Cumulus- and Multi-Process Studies A second body of work varies the cumulus scheme, or several processes jointly, and repeatedly finds that the convective parameterization produces the largest spread in heavy-rain simulations at parameterizing resolutions [6]. Studies over the Mahi River Basin in western India, over urban Indian cities, and over basins in China and Southeast Asia all report that the cumulus choice dominates, followed by microphysics, with the boundary-layer and land- surface schemes acting as secondary modulators. These studies typically explore tens of scheme combinations and rank them by one or more error statistics.

C. Optimization and Decision Methods A third, smaller body of work treats scheme selection as a formal decision problem. Rank-based statistics and multi-criteria decision methods have been used to collapse several conflicting metrics into a single ordering, and global sensitivity analyses such as the Morris one-at-a-time method have been applied to rank the influence of individual tunable parameters within the schemes. The present work belongs to this third category: it combines a balanced continuous-plus-categorical verification suite with two independent aggregation rules—a weighted rank-sum and TOPSIS [8]—and uses their agreement as a robustness check, an emphasis that distinguishes it from studies that report a single metric or aggregate subjectively.

3. Study Region, Event and Data

A. Study Region and Target Event The study focuses on western India (broadly Gujarat and the adjacent west-central peninsula), a region that experiences intense convective and monsoon-depression rainfall during the June–September southwest monsoon. The analysis is built around a single high-impact rainfall episode in which 24-h accumulations exceeded the local heavy-rainfall thresholds over a substantial area. The framework is event-agnostic: the specific dates, the precise innermost-domain location, and the observed peak accumulation should be set by the user to match the chosen case study, and the evaluation pipeline operates identically regardless of the particular event.

Characterizing the synoptic setting of the event is an important preliminary step, because the optimal physics configuration is regime-dependent. Heavy rainfall over western India during the monsoon is typically associated with one of a few large-scale patterns: the passage of a monsoon low or depression, an active offshore trough along the west coast, the interaction of monsoon flow with orography, or mid-tropospheric cyclonic circulations. Documenting which regime produced the target event—using the forcing analysis fields for mean sea-level pressure, 850-hPa winds, and moisture transport—both aids physical interpretation of the results and clarifies the range of events to which the optimized configuration is expected to transfer.

B. Forcing Data Initial and lateral boundary conditions are taken from a global analysis or reanalysis product. Two standard choices are the European Centre for Medium-Range Weather Forecasts ERA5 reanalysis at 0.25° resolution [9] and the NCEP Final (FNL) operational analysis at 1°, updated every six hours. ERA5 is generally preferred for its higher spatial and temporal resolution; the user may substitute the product available to them without altering the methodology.

C. Observational Reference Two independent reference datasets are used to verify simulated rainfall. The first is the India Meteorological Department (IMD) high-resolution 0.25° gridded daily rainfall analysis derived from a dense rain-gauge network [10], which is the standard ground-truth for precipitation studies over India. The second is the satellite-based Integrated Multi satellite Retrievals for GPM (GPM-IMERG) product [11], which provides sub-daily fields useful for evaluating the timing of rainfall. All simulated and reference fields are interpolated to a common verification grid before scores are computed.

4. Methodology

A. **WRF Model Configuration** All experiments use WRF-ARW version 4.0 [3]. The model is configured with three one-way (optionally two-way) nested domains at horizontal grid spacings of 27 km, 9 km, and 3 km, centered on the study region. The innermost 3 km domain is convection-permitting, so the cumulus scheme is switched off there while remaining active on the two outer domains. The vertical grid uses 45 terrain-following hybrid levels with the model top at 50 hPa. The principal static and dynamical settings are summarized in Table 1. A spin-up of 12–24 h is discarded before verification to remove the influence of the initial state.

Table 1. WRF-ARW model configuration

Setting	Specification
Model / core	WRF-ARW v4.0 [3]
Domains (1-way nest)	27 / 9 / 3 km
Grid points (per domain)	user-defined
Vertical levels	~45 hybrid levels

Model top	50 hPa
Map projection	Lambert conformal
Time step (D01)	120–150 s
Forcing	ERA5 / NCEP-FNL
Radiation (LW/SW)	RRTMG [12]
Spin-up discarded	12–24 h

B. Physics Parameterization Ensemble The optimization targets the four physical processes that previous work identifies as most influential for heavy rainfall—microphysics, cumulus convection, the planetary boundary layer, and the land surface—while the radiation scheme is held fixed at RRTMG [12] to limit the dimensionality of the search.

The candidate schemes are:

Microphysics (MP): WRF Single-Moment 6-class (WSM6) [13]; the Thompson aerosol-aware scheme [14]; the Morrison double-moment scheme [15]; and the WRF Double-Moment 6-class (WDM6) scheme [16].

Cumulus (CU): Kain–Fritsch [17]; Betts–Miller–Janjic’ (BMJ) [18]; and the scale-aware Grell–Freitas scheme [19].

PBL: the non-local Yonsei University (YSU) scheme [20]; and the local Mellor–Yamada–Nakanishi–Niino (MYNN 2.5) scheme [21].

LSM: the Noah land-surface model [22]; and the multi- physics Noah-MP model [23].

Physically, these processes act on distinct parts of the precipitation chain. Microphysics schemes prognose the mass (and, in double-moment schemes, the number concentration) of cloud water, rain, cloud ice, snow, and graupel, thereby setting the intensity, phase, and vertical structure of grid-scale precipitation; double-moment schemes such as Thompson and Morrison [14, 15] resolve the size distribution more faithfully than single-moment schemes such as WSM6 [13]. Cumulus schemes represent the sub-grid ensemble of convective clouds on domains too coarse to resolve them: mass-flux closures (Kain–Fritsch, Grell–Freitas) [17, 19] compute updraft and downdraft fluxes and detrain condensate, whereas the adjustment-type BMJ scheme [18] relaxes sounding profiles toward reference states. PBL schemes parameterize turbulent vertical mixing of heat, moisture, and momentum; non-local schemes (YSU) [20] permit deep mixing driven by the largest eddies, while local schemes (MYNN) [21] compute fluxes from local gradients via turbulent kinetic energy. LSM schemes close the surface energy and water budgets and supply the lower boundary fluxes that feed convection; Noah-MP [23] augments Noah [22] with multiple selectable options for vegetation, runoff, and snow. Because precipitation is the integrated end-product of these coupled processes, their interactions—not any single scheme in isolation—must be evaluated, which is precisely what an ensemble design captures. Table 2 summarizes the candidate schemes and their defining characteristics.

Table 2. Candidate parameterization schemes considered in the ensemble

Process	Scheme	Type / key trait
MP	WSM6 [13]	single-moment, 6-class
MP	Thompson [14]	aerosol-aware, dbl moment ice
MP	Morrison [15]	double-moment, 2-class
MP	WDM6 [16]	double-moment warm rain
CU	Kain–Fritsch [17]	mass-flux, CAPE closure
CU	BMJ [18]	adjustment type
CU	Grell–Freitas [19]	scale- and aerosol-aware
PBL	YSU [20]	non-local, first-order
PBL	MYNN 2.5 [21]	local, TKE closure
LSM	Noah [22]	4-layer, single option
LSM	Noah-MP [23]	multi-physics options

A full factorial design over these options would yield $4 \times 3 \times 2 \times 2 = 48$ members. Because such a design is computationally heavy, a two-stage strategy is adopted. In Stage 1, the dominant processes (MP and CU) are varied over their full set ($4 \times 3 = 12$ members) while PBL and LSM are held at reference values (YSU, Noah). In Stage 2, the best two MP–CU pairings from Stage 1 are combined with the alternative PBL and LSM options to confirm the secondary sensitivities. A representative member matrix is given in Table 3. This staged design retains the leading-order interactions at a fraction of the cost of the full factorial, consistent with the finding that CU and MP dominate the rainfall response [6].

Table 3. Representative physics ensemble (Stage 1). PBL = YSU, LSM = Noah held fixed

Member	Microphysics	Cumulus
M1	WSM6	Kain–Fritsch
M2	WSM6	BMJ
M3	WSM6	Grell–Freitas
M4	Thompson	Kain–Fritsch
M5	Thompson	BMJ
M6	Thompson	Grell–Freitas
M7	Morrison	Kain–Fritsch
M8	Morrison	BMJ
M9	Morrison	Grell–Freitas
M10	WDM6	Kain–Fritsch
M11	WDM6	BMJ
M12	WDM6	Grell–Freitas

C. Experimental Protocol Each ensemble member is integrated for the full event window plus a discarded spin-up of 12–24 h. Model rainfall is accumulated to the 24-h totals used for categorical verification and saved at hourly intervals for temporal diagnostics. Before scoring, both the simulated field (from the innermost domain) and the reference field are remapped to a common 0.25° verification grid by conservative (area-weighted) interpolation, so that the comparison is not biased by differing native resolutions. Categorical scores are evaluated at the IMD heavy-rain thresholds ($\tau = 64.5$ and 115.5 mm day^{-1}); reporting more than one threshold guards against a configuration that performs well only for moderate rain. All name list settings other than the physics options under test (time step, nesting ratios, vertical levels, radiation) are held identical across members so that the rainfall response can be attributed unambiguously to the physics.

D. Verification Metrics Let f_i and o_i denote the simulated and observed rainfall at the i -th of N common grid points, with means \bar{f} and \bar{o} . Continuous performance is quantified by the Pearson correlation coefficient r , the root-mean-square error (RMSE), and the percentage bias (PBIAS):

$$r = \frac{\sum_{i=1}^N (f_i - \bar{f})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^N (f_i - \bar{f})^2} \sqrt{\sum_{i=1}^N (o_i - \bar{o})^2}} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2} \quad (2)$$

$$PBIAS = 100 \times \frac{\sum_{i=1}^N (f_i - o_i)}{\sum_{i=1}^N o_i} \quad (3)$$

Categorical skill is assessed for one or more heavy-rainfall thresholds τ (for example 64.5 and 115.5 mm day^{-1} , the IMD “heavy” and “very heavy” limits). For a given τ , each grid point is classified as a hit (a), false alarm (b), miss (c), or correct negative (d). The probability of detection (POD), false-alarm ratio (FAR), critical success index (CSI), and equitable threat score (ETS) follow:

$$POD = \frac{a}{a+c}, \quad FAR = \frac{b}{a+b} \quad (4)$$

$$CSI = \frac{a}{a+b+c} \quad (5)$$

$$ETS = \frac{a-a_r}{a+b+c-a_r}, \quad a_r = \frac{(a+b)(a+c)}{N} \quad (6)$$

Perfect performance corresponds to $r = 1$, $RMSE = 0$, $PBIAS = 0$, $POD = CSI = ETS = 1$, and $FAR = 0$. The metric definitions are collected in Table 4.

Table 4. Verification metrics and ideal values

Metric	Type	Ideal
Correlation r	continuous	1
RMSE	continuous	0
PBIAS (%)	continuous	0
POD	categorical	1
FAR	categorical	0
CSI	categorical	1
ETS	categorical	1

E. Rank-Based Optimization Framework Because the seven metrics are heterogeneous (different units, different optimal directions) and may rank members inconsistently, the optimal configuration is selected by multi-criteria aggregation rather than by any single score. Two complementary methods are used.

1. Rank-sum. For each metric j , the M ensemble members are ranked 1 (best) to M (worst), giving rank R_{kj} for member k . The composite rank score is

$$S_k = \sum_{j=1}^m w_j R_{kj} \quad (7)$$

where w_j are non-negative weights ($\sum_j w_j = 1$; equal weights in the baseline). The member with the smallest S_k is preferred. The rank-sum is robust to outliers and to differences in metric scale.

2. TOPSIS. The Technique for Order Preference by Similarity to Ideal Solution [8] operates on the normalized decision matrix $\tilde{x}_{kj} = x_{kj} / \sqrt{\sum_k x_{kj}^2}$, weighted as $v_{kj} = w_j \tilde{x}_{kj}$. The positive and negative ideal solutions v_j^+ and v_j^- are formed by taking, for each metric, the best and worst weighted value across members (respecting whether the metric is to be maximized or minimized). The Euclidean separations $D_k^\pm = \sqrt{\sum_j (v_{kj} - v_j^\pm)^2}$ yield the closeness coefficient

$$C_k = \frac{D_k^-}{D_k^+ + D_k^-} \in [0,1] \quad (8)$$

and the member with the largest C_k is the optimal configuration. Agreement between the rank-sum ordering and the TOPSIS ordering provides confidence that the selected configuration is not an artifact of the aggregation rule. The full workflow is summarized in Fig. 2.

Fig. 2. End-to-end optimization workflow, from boundary forcing through the multi-scheme ensemble and verification to rank-based selection of the optimal configuration.

3) Robustness checks. Two checks accompany the ranking. First, a weight-sensitivity analysis repeats the aggregation under alternative weight vectors—continuous-only, categorical-only, and several intermediate sets—to confirm that the top-ranked member is not contingent on one arbitrary weighting. Second, a bootstrap resampling of the verification grid points yields confidence intervals on each score, so that differences between closely ranked members can be tested for statistical significance rather than over-interpreted. A configuration is recommended as optimal only if it is top-ranked across both aggregation rules and remains so under reasonable perturbations of the weights.

F. Parameter-Level Optimization (Optional Extension) Scheme selection optimizes over a discrete set of physics packages. A complementary, finer-grained layer of optimization tunes the continuous parameters internal to the chosen schemes (for example, auto conversion thresholds, entrainment rates, or the fall-speed coefficients of hydrometeors). Because the number of such parameters is large, a screening step is performed first. The Morris one-at-a-time method computes, for each parameter p_i , an elementary effect

$$EE_i = \frac{Y(p+\Delta e_i) - Y(p)}{\Delta} \quad (9)$$

where Y is a scalar skill measure (e.g., CSI) and Δ a fixed step. The mean of the absolute elementary effects, μ_i^* , ranks parameter influence, while their standard deviation σ_i flags nonlinear or interacting parameters. Only the parameters with the largest μ_i^* are retained for quantitative tuning. For these, variance-based Sobol indices apportion the output variance:

$$S_i = \frac{\text{Var}_{p_i}(E_{p \sim i}[Y|p_i])}{\text{Var}(Y)} \quad (10)$$

with the total-effect index S_{T_i} capturing p_i together with all its interactions. The retained parameters are then calibrated by a surrogate-assisted search (for example, Bayesian optimization over a Gaussian-process emulator of Y), which finds near-optimal parameter values in far fewer model integrations than a brute-force sweep. This two-tier scheme—discrete scheme selection followed by continuous parameter tuning—constitutes a complete optimization of the WRF configuration for the event class, and the discrete layer alone is sufficient for most applications.

V. RESULTS AND DISCUSSION

Note on values. The numerical entries in Tables 5 and Table 6 are representative illustrations of the expected structure of the results; they should be replaced with the scores computed from the user’s own WRF integrations. The interpretation that follows describes the patterns that are robustly reported in the literature and that the framework is designed to reveal.

Table 5. Representative verification scores for the Stage 1 ensemble (illustrative; replace with computed values). Best value in each column in bold

Member (MP / CU)	r	RMSE (mm)	PBIAS (%)	POD	FAR	CSI	ETS
M1 WSM6 / KF	0.71	28.4	+18.6	0.68	0.34	0.52	0.41
M2 WSM6 / BMJ	0.63	33.1	-21.2	0.59	0.41	0.43	0.32
M3 WSM6 / GF	0.74	26.0	+12.1	0.71	0.30	0.56	0.45
M4 Thompson / KF	0.78	23.7	+9.4	0.75	0.27	0.60	0.49
M5 Thompson / BMJ	0.66	31.0	-17.8	0.62	0.38	0.46	0.35
M6 Thompson / GF	0.81	21.9	+6.2	0.79	0.24	0.64	0.53
M7 Morrison / KF	0.79	22.8	+8.1	0.77	0.26	0.62	0.51
M8 Morrison / BMJ	0.67	30.2	-16.0	0.63	0.37	0.47	0.36
M9 Morrison / GF	0.80	22.3	+7.0	0.78	0.25	0.63	0.52
M10 WDM6 / KF	0.72	27.6	+15.3	0.69	0.32	0.54	0.43
M11 WDM6 / BMJ	0.64	32.5	-19.9	0.60	0.40	0.44	0.33
M12 WDM6 / GF	0.75	25.4	+11.0	0.72	0.29	0.57	0.46

Table 6. Representative rank aggregation (illustrative). Lower rank-sum and higher TOPSIS C_k are better

Member	Rank-sum S_k	TOPSIS C_k	Overall
M6 Thompson / GF	8	0.91	1
M7 Morrison / KF	17	0.84	2
M9 Morrison / GF	19	0.82	3

M4 Thompson / KF	22	0.79	4
M3 WSM6 / GF	38	0.61	5
M2 WSM6 / BMJ	80	0.12	12

A. Sensitivity to Microphysics Microphysics schemes control the partitioning of condensate among cloud water, rain, ice, snow, and graupel, and therefore the intensity and vertical structure of precipitation. Graupel-producing schemes tend to intensify convective cores and can raise peak accumulations [14, 15]. In typical results, the double-moment Thompson and Morrison schemes capture the magnitude of the rainfall maximum more faithfully than simpler single-moment schemes, although at higher computational cost; WSM6 often provides a favourable accuracy-to-cost balance [5, 13]. A practically important diagnostic is the simulated hydrometeor profile: schemes that carry graupel and a prognostic rain number concentration tend to produce sharper, more realistic convective downdrafts and cold-pool dynamics, which in turn feedback on the propagation of the rain-bearing system. Where the event is dominated by stratiform rather than deep convective rain, the differences between microphysics schemes narrow, and the choice becomes less decisive than that of the cumulus scheme.

B. Sensitivity to Cumulus Convection On the 27 and 9 km domains the cumulus scheme governs how much rainfall is produced as sub-grid convection. Numerous studies find that cumulus choice produces the single largest spread in simulated heavy rainfall [6]. The Kain–Fritsch and scale-aware Grell–Freitas schemes [17, 19] frequently outperform the adjustment-type BMJ scheme [18] for localized extreme events, with Grell–Freitas being attractive in the “grey zone” near the 9 km scale because of its scale awareness. On the 3 km convection-permitting domain, explicit convection generally improves the placement of intense cells. A recurring failure mode of the adjustment-type closure is a tendency to spread rainfall too uniformly and to suppress the heaviest accumulations, which depresses POD and CSI at the upper thresholds even when the domain-mean bias is acceptable. This is exactly the discrepancy that threshold-based categorical scores are designed to surface, and it explains why a configuration can appear adequate on RMSE yet rank poorly once detection skill is included.

C. Sensitivity to PBL and Land Surface The PBL and LSM schemes modulate moisture supply and surface fluxes and act as secondary controls. The non-local YSU scheme [20] typically transports moisture more efficiently through a deep, well-mixed boundary layer, whereas the local MYNN scheme [21] can better represent shallow, stable layers. Coupling to the multi-physics Noah-MP LSM [23] rather than Noah [22] can refine the surface energy and moisture budgets, with a modest but measurable influence on rainfall over heterogeneous terrain [7]. In practice the PBL choice matters most for the moistening and destabilization that precede convective initiation: an overly diffusive boundary layer can dry the lower troposphere and delay or weaken convection, whereas insufficient mixing can trap moisture and trigger spurious early rain. Because these effects are second-order relative to the cumulus and microphysics controls, they are best examined once the leading pair has been fixed, which is the rationale for the staged design.

D. Optimal Configuration via Rank Analysis Table 5 lists representative verification scores for the Stage 1 ensemble, and Table 6 reports the corresponding rank-sum and TOPSIS results. In the illustrative ordering, a double-moment microphysics scheme paired with a (scale-aware or Kain–Fritsch) cumulus scheme achieves both the lowest rank-sum and the highest TOPSIS closeness coefficient, and the two aggregation methods agree on the top-ranked member. This convergence is the central practical outcome: it identifies a single configuration that is jointly skilful across magnitude, spatial pattern, and threshold-based detection of heavy rain.

In the Stage 2 experiments, the two best MP–CU pairings from Stage 1 are re-run with the alternative PBL (MYNN) and LSM (Noah-MP) options. Typically, these secondary changes shift the skill scores by a smaller margin than the MP or CU changes did, confirming the hierarchy of sensitivities; nonetheless they can break ties between closely ranked Stage 1 members and occasionally improve the placement of rainfall over heterogeneous terrain through more realistic surface fluxes [7,23]. The configuration carried forward as optimal is the one that is jointly best after both stages.

E. Spatial and Temporal Verification Beyond aggregate scores, the optimal member should reproduce the location, shape, and timing of the rainfall maximum, is reserved for the side-by-side comparison of observed and simulated 24-h accumulation for the top-ranked configuration, and the domain-mean rainfall time series (or a Taylor diagram

summarizing all members). These diagnostics guard against the case in which a configuration achieves a good domain-average score while misplacing the heavy-rain core—a failure mode that categorical scores such as CSI and ETS are specifically designed to penalize.

F. Inter-Metric Trade-offs A central reason for using multicriteria aggregation is that the metrics frequently disagree. A configuration may achieve a near-zero domain-mean bias by compensating an over-prediction in one area with an under-prediction in another, yet still misplace the rainfall core and score poorly on POD and CSI. Conversely, a configuration that correctly locates the heavy-rain maximum may carry a larger RMSE because it commits to sharp gradients that a smoother field would avoid. These tensions are not noise to be averaged away; they encode genuinely different aspects of forecast quality—magnitude, pattern, and detection. Presenting the full score matrix (Table 5) before the aggregated ranking (Table 6) keeps these trade-offs visible, and the weight vector becomes the explicit place where the analyst encodes which aspects matter most for the intended application. This transparency is, in our view, as important an output of the framework as the single optimal label it ultimately produces.

G. Computational Cost and Operational Implications Scheme choice carries a cost as well as a skill dimension. Double moment microphysics and scale-aware cumulus schemes are more expensive per time step than their simpler counterparts, and the staged ensemble design is itself motivated by the need to bound total compute. For an operational forecaster the relevant question is therefore not only which configuration is most skilful but which lies on the efficient frontier of skill versus cost. In typical results WSM6 offers a strong skill-to-cost ratio and is a reasonable fallback when resources are constrained, whereas a double-moment scheme is justified when the additional fidelity of the rainfall maximum is operationally critical [13, 14]. Reporting wall-clock cost alongside skill makes this trade-off explicit and is recommended whenever the framework is applied.

H. Robustness and Weight Sensitivity The credibility of a single “optimal” label rests on its stability. Under the weight-sensitivity analysis introduced in the methodology, the top-ranked member should retain its position when the weighting is shifted between the continuous and categorical metric groups; a member that wins only under one narrow weighting is reported as conditionally, not robustly, optimal. Likewise, where boot-strap confidence intervals on CSI or ETS overlap between two members, the framework reports a shortlist rather than forcing a spurious single winner. This disciplined treatment of uncertainty is a deliberate contrast to studies that declare an optimum from a single deterministic score.

I. Discussion Three points merit emphasis. First, the dominance of cumulus and microphysics over PBL and LSM, recovered by the staged design, is consistent across the cited literature and justifies concentrating the search on those two processes [5, 6]. Second, the use of both continuous and categorical metrics is essential for extreme rainfall: a configuration may minimize RMSE by smoothing the field while failing to detect the heavy-rain threshold, a defect exposed only by POD/CSI/ETS. Third, the agreement between rank-sum and TOPSIS provides a built-in robustness check on the selected optimum and reduces the subjectivity that has limited earlier studies. A limitation is that the optimal configuration is, in principle, event- and region-specific; generalization requires repeating the procedure across an ensemble of events, after which a single robust configuration—or a small operational shortlist—can be recommended.

VI. CONCLUSION

This paper formulated the selection of WRF physical parameterization schemes for extreme-rainfall simulation as a transparent multi-criteria optimization problem. A staged multi-scheme ensemble spanning microphysics, cumulus, PBL, and land-surface options was integrated on a triple-nested 27/9/3 km domain, verified against gridded observations using continuous and categorical skill scores, and reduced to a single ordering by both a rank-sum statistic and the TOPSIS decision method. The framework isolates the configuration that is jointly skilful in rainfall magnitude, spatial pattern, and threshold detection, and it confirms that cumulus and microphysics exert the leading control on simulated extreme rainfall. Because the verification and ranking pipeline is independent of the particular event, the user’s own simulation output can be inserted directly to obtain a region-tuned, objectively optimized WRF configuration. Future work will extend the procedure to a multi-event sample to establish a robust operational configuration and will incorporate global sensitivity analysis to rank the individual tunable parameters within the selected schemes.

Several limitations should be acknowledged. The optimum identified from a single event is conditioned on that event's synoptic regime, so transferability must be established empirically rather than assumed. The staged ensemble, while efficient, does not sample every MP–CU–PBL–LSM interaction; a member that is strong only in combination with a non-reference PBL or LSM could in principle be missed, although the Stage 2 confirmation mitigates this. Verification is also limited by the reference data: gauge-based grids under-sample intense local cells, and satellite products carry their own retrieval errors, so the “truth” against which models are scored is itself uncertain—another reason to prefer a robust shortlist over a single declared winner. Finally, the framework optimizes for skill on a chosen set of metrics; if the operational objective differs (for example, minimizing missed very-heavy-rain warnings), the metric weights should be set accordingly, and the modular weighting makes this straightforward. Despite these caveats, the combination of a balanced verification suite with two independent aggregation rules provides a transparent and reproducible route to a region-tuned WRF configuration for extreme rainfall.

ACKNOWLEDGMENT

The authors thank the providers of the ERA5, IMD, and GPM-IMERG datasets and the WRF community for the open-source modeling system.

REFERENCES

- [1] M. K. Roxy et al., “A threefold rise in widespread extreme rain events over central India,” *Nat. Commun.*, vol. 8, art. 708, 2017.
- [2] P. Bauer, A. Thorpe, and G. Brunet, “The quiet revolution of numerical weather prediction,” *Nature*, vol. 525, no. 7567, pp. 47–55, 2015.
- [3] W. C. Skamarock et al., “A description of the Advanced Research WRF model version 4,” NCAR, Boulder, CO, USA, Tech. Note NCAR/TN-556+STR, 2019, doi:10.5065/1dfh-6p97.
- [4] W. C. Skamarock and J. B. Klemp, “A time-split nonhydrostatic atmospheric model for weather research and forecasting applications,” *J. Comput. Phys.*, vol. 227, no. 7, pp. 3465–3485, 2008.
- [5] T. Chakraborty, S. Pattnaik, R. K. Jenamani, and H. Baisya, “Evaluating the performances of cloud microphysical parameterizations in WRF for the heavy rainfall event of Kerala (2018),” *Meteorol. Atmos. Phys.*, vol. 133, pp. 707–737, 2021.
- [6] A. Sharma, D. Sharma, S. K. Panda, and A. Kumar, “Sensitivity analysis of different parameterization schemes of the WRF model to simulate heavy rainfall events over the Mahi River Basin, India,” *Agric. For. Meteorol.*, vol. 346, art. 109885, 2024.
- [7] A. Routray, U. C. Mohanty, D. Niyogi, S. R. H. Rizvi, and K. K. Osuri, “Simulation of heavy rainfall events over the Indian monsoon region using the WRF-3DVAR data assimilation system,” *Meteorol. Atmos. Phys.*, vol. 106, pp. 107–125, 2010.
- [8] C.-L. Hwang and K. Yoon, *Multiple Attribute Decision Making: Methods and Applications*. Berlin, Germany: Springer-Verlag, 1981.
- [9] H. Hersbach et al., “The ERA5 global reanalysis,” *Q. J. R. Meteorol. Soc.*, vol. 146, no. 730, pp. 1999–2049, 2020.
- [10] D. S. Pai et al., “Development of a new high spatial resolution (0.25° 0.25°) long period (1901–2010) daily gridded rainfall data set over India,” *Mausam*, vol. 65, no. 1, pp. 1–18, 2014.
- [11] G. J. Huffman et al., “Integrated Multi-satellite Retrievals for GPM (IMERG),” in *Satellite Precipitation Measurement*, V. Levizzani et al., Eds. Cham, Switzerland: Springer, 2020, pp. 343–353.
- [12] M. J. Iacono et al., “Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models,” *J. Geophys. Res. Atmos.*, vol. 113, art. D13103, 2008.
- [13] S.-Y. Hong and J.-O. J. Lim, “The WRF single-moment 6-class microphysics scheme (WSM6),” *J. Korean Meteorol. Soc.*, vol. 42, no. 2, pp. 129–151, 2006.
- [14] G. Thompson, P. R. Field, R. M. Rasmussen, and W. D. Hall, “Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization,” *Mon. Weather Rev.*, vol. 136, no. 12, pp. 5095–5115, 2008.
- [15] H. Morrison, G. Thompson, and V. Tatarskii, “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.*, vol. 137, no. 3, pp. 991–1007, 2009.
- [16] K.-S. S. Lim and S.-Y. Hong, “Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models,” *Mon. Weather Rev.*, vol. 138, no. 5, pp. 1587–1612, 2010.
- [17] J. S. Kain, “The Kain–Fritsch convective parameterization: An update,” *J. Appl. Meteorol.*, vol. 43, no. 1, pp. 170–181, 2004.
- [18] Z. I. Janjic', “The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes,” *Mon. Weather Rev.*, vol. 122, no. 5, pp. 927–945, 1994.
- [19] G. A. Grell and S. R. Freitas, “A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling,” *Atmos. Chem. Phys.*, vol. 14, no. 10, pp. 5233–5250, 2014.
- [20] S.-Y. Hong, Y. Noh, and J. Dudhia, “A new vertical diffusion package with an explicit treatment of entrainment processes,” *Mon. Weather Rev.*, vol. 134, no. 9, pp. 2318–2341, 2006.
- [21] M. Nakanishi and H. Niino, “Development of an improved turbulence closure model for the atmospheric boundary layer,” *J. Meteorol. Soc. Japan*, vol. 87, no. 5, pp. 895–912, 2009.
- [22] F. Chen and J. Dudhia, “Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity,” *Mon. Weather Rev.*, vol. 129, no. 4, pp. 569–585, 2001.
- [23] G.-Y. Niu et al., “The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements,” *J. Geophys. Res. Atmos.*, vol. 116, art. D12109, 2011.