https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

Post-Processing High-Resolution Ensemble Forecasts for Extreme Rainfall: Short-Term Skill Evaluation over Kyushu, Japan

Magfira Syarifuddin[1], Hamada Atsushi[1,2], Houtian He[1], Kazuaki Yasunaga[1]

[1]Faculty of Sustainable Design, University of Toyama, Toyama, 930-8555, Japan
[2]Center for Environmental Remote Sensing (CeRES), Chiba University, Chiba, 263-0022, Japan

Correspondence to: Magfira Syarifuddin (magfira@sus.u-toyama.ac.jp)

Abstract. Extreme rainfall in Japan, exemplified by the August 2021 Kyushu event with multiple linear rainbands, continues to cause severe societal impacts, underscoring the need for reliable ensemble rainfall forecasts. This study evaluates how ensemble size, horizontal resolution, and integration period influence forecast skill using the SCALE-RM model. Four ensembles were examined: a coarser-resolution set (S1; 3.2 km, 100 members) and three finer-resolution sets (S2–S4; 800 m, 50 members), all initialized from ERA5 but with different setup and time integration. Mean Bias (MB) and Quantile Mapping (QM) corrections were applied, and skill was assessed using RMSE, probability maps, ETS, and BS. Before correction, none of the ensembles reproduced the moderate-to-heavy rainfall accumulation in northern Kyushu. S1 produced the lowest RMSE but failed to capture localized maxima, decayed rainfall too early, and missed the second peak on 12 August. After correction, performance diverged. S1 shows noticeable improvement, producing moderate to higher rainfall values in the northern region, though peak intensities remain slightly underestimated. S4 shows the strongest enhancement, successfully generating the extreme rainfall intensities in the rainband core and closely matching observations, indicating that its systematic biases were effectively removed. Overall, the findings demonstrate that high resolution alone does not guarantee improved skill; ensemble size and robust post-processing are equally critical. These insights inform both operational forecasting and controlled weather-modification experiments.

1 Introduction

Numerical weather models are indispensable tools for understanding atmospheric processes and diagnosing extreme rainfall events (Hiraga and Meza, 2025; Taylor et al., 2021). They also provide the foundation for simulating and evaluating weather-modification strategies aimed at enhancing or mitigating precipitation (Hiraga et al., 2025). However, simulations of extreme convection remain challenging due to substantial uncertainties in model physics, initial and boundary conditions, and microphysical processes. These uncertainties limit the reliability of numerical simulations for both disaster preparedness and weather-modification verification. Establishing a robust baseline forecast accuracy is therefore essential before such simulations can be used with confidence.

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

Improving short-term rainfall forecast skill has been a persistent challenge in numerical modeling. Increasing model resolution and employing ensemble prediction systems (EPSs) can better represent convective dynamics and quantify forecast uncertainty (Brotze et al., 2023). Yet these benefits are often constrained by the high computational cost of generating sufficiently large ensembles. The trade-off between spatial resolution and ensemble size introduces sampling errors and systematic biases, limiting predictive skill (Gascón et al., 2019). Post-processing techniques such as Mean Bias (MB) correction and Quantile Mapping (QM) have been developed to mitigate these errors and improve probabilistic reliability, but their performance for short-duration, high-resolution convective forecasts remains insufficiently examined.

In Japan, the growing frequency of extreme rainfall events underscores the urgent need for more accurate short-term forecasting. Between 2013 and 2024, such events caused approximately 660 fatalities and over 7.3 trillion JPY in economic losses (MLIT, 2024). A notable example is the August 2021 Kyushu extreme rainfall, in which warm, humid inflow toward a stalled frontal zone triggered multiple linear rainbands (線状降水帯, read as *senjō-kōsuitai*) between 11–14 August. Particularly on 12 August and again on 14 August, successive rainbands repeatedly formed over northern Kyushu, producing continuous, torrential rainfall. Several locations recorded historically unprecedented totals, including 555.5 mm of 24-hour rainfall in Ureshino City, Saga Prefecture (JMA, 2021). The event exposed limitations in existing forecasting systems, especially in capturing narrow, repeatedly forming rainbands and their rapid evolution.

This study revisits a high-resolution ensemble simulation originally designed to evaluate potential weather-modification impacts during the August 2021 event. Here, the simulation is repurposed as a controlled testbed to examine how model resolution, ensemble size, and integration period influence forecast accuracy during a multi-band *senjō-kōsuitai* episode. Using the SCALE-RM (Scalable Computing for Advanced Library and Environment–Regional Model) (Sato et al., 2015), we apply MB and QM correction to multiple ensemble configurations to assess their ability to reduce amplitude biases and improve spatial reliability.

Previous studies have shown that post-processing can substantially improve ensemble rainfall forecasts and sometimes offset limitations in model design (Gascon et al., 2019; Staehle et al., 2024). However, most post-processing application target daily–monthly timescale (Gascón et al., 2019; Golian and Murphy, 2022; Ilaboya et al., 2024; Pierce et al., 2015; Scherrer et al., 2024), and only a few have examined its performance for sub-hourly, convective rainfall dominated by narrow, rapidly forming rainbands. Calibration of heavy precipitation further requires large training samples (Hamill et al., 2017), which are difficult to obtain for high-resolution, short-term EPSs with limited ensemble size.

To the best of our knowledge, this is the first study to apply bias correction directly to 10-minute to hourly high-resolution ensemble rainfall forecasts for a Japanese extreme rainfall event characterized by multiple linear rainbands. The findings provide practical insight into the relative importance of resolution versus ensemble size, the limits of post-processing when structural errors dominate, and the implications for both operational disaster forecasting and the design of weather-modification scenario experiments.

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

2 Methods

2.1 Simulation Set up

This study employed the SCALE-RM (Scalable Computing for Advanced Library and Environment–Regional Model) version 5.5.4 developed by RIKEN (Sato et al., 2015; Taylor et al., 2021), to simulate the August 2021 Kyushu heavy rainfall event, focusing on the early rainfall episode that occurred on 11–12 August 2021. Although the heaviest rainfall of the August 2021 Kyushu event occurred on 14 August, this study focuses on the earlier rainfall episode of 11–12 August. This choice is dictated by the original design of the simulation, which was developed for **weather modification assessment**, targeting conditions conducive to precipitation enhancement. By evaluating this period, we can leverage the existing high-resolution ensemble framework while assessing forecast skill and post-processing improvements in a controlled setting. The earlier event still features significant convective activity and precipitation variability, providing a representative testbed for examining ensemble performance, spatial resolution effects, and the effectiveness of bias-correction methods prior to peak rainfall conditions.

SCALE-RM was selected for its demonstrated capability to reproduce mesoscale convective systems with high spatial and temporal fidelity (Onishi et al., 2015). Its flexible domain configuration and comprehensive physics make it well-suited to represent intense, localized rainfall and complex surface–atmosphere interactions typical of Kyushu's warm-season precipitation systems. The COUPLE scheme, in particular, improves the simulation of boundary-layer moisture fluxes and local convergence that modulate convective rainfall intensity.

SCALE-RM includes advanced physical parameterizations such as the SN14 six-class two-moment bulk microphysics scheme for cloud and precipitation processes, the Kain–Fritsch (KF) convection scheme, the MSTRNX broadband radiation scheme, the Mellor–Yamada–Nakanishi–Niino (MYNN) level 2.5 planetary boundary layer scheme, and the COUPLE surface flux scheme. The COUPLE module explicitly represents the exchanges of heat, moisture, and momentum across heterogeneous land, ocean, and urban surfaces using detailed land-use data, enabling realistic treatment of surface–atmosphere feedbacks (Nishizawa et al., 2024).

The model was initialized and forced with ERA5 reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF), which provide hourly atmospheric fields at ~0.25° horizontal resolution (Hersbach et al., 2020). Different simulation periods and spatial resolutions were employed to examine the effects of model configuration on ensemble rainfall forecasts (Table 1). The simulation was initialized using boundary conditions derived from atmospheric variables at hourly intervals, including geopotential height, horizontal and vertical wind components, temperature, and specific humidity. Surface and sea-related boundary conditions consisted of sea surface temperature, 10-m horizontal wind components, 2-m air temperature, skin temperature, and surface pressure, which were updated every 6-hours. Land surface conditions were also updated every 6-hours and included soil temperature and soil moisture at four vertical levels.

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

Four ensemble sets were produced. The coarser-resolution ensemble (S1) used a grid spacing of 3.2 km over a domain covering 124°–132° E, 30°–34° N, while the three finer-resolution ensembles (S2–S4) used 0.8 km grid spacing over an extended domain (124°–136.5° E, 28.8°–36° N). Vertical levels were stretched from 20 m near the surface to 200 m aloft, with 160 total layers. All simulations were conducted on the Fugaku supercomputer, leveraging its large-scale parallel performance to manage the computational cost of high-resolution ensemble forecasts (Sato et al., 2022).

Table 1: Set of ensemble members include in this study.

Set	Integration Period (UTC)	Time step	Duration	Number of ensemble	Grid size (x×y×z)	Spatial Resolution
S1 (coarse)	10 Aug (0000) → 13 Aug (0000)	1 hour	73 hours	100	600×360×160	3.2 km
S2	11 Aug (1200) → 12 Aug (1200)	10 minutes	24 hours	50	1440×960×160	0.8 km
S3	11 Aug (0000) → 12 Aug (1200)	10 minutes	36 hours	50	1440×960×160	0.8 km
S4	10 Aug (1200) → 12 Aug (1200)	10 minutes	48 hours	50	1440×960×160	0.8 km

Following previous findings that emphasize the sensitivity of convective rainfall to moisture-induced diabatic heating (e.g.,Holloway and Neelin, 2009; Lenderink et al., 2025), initial perturbations were applied to represent uncertainty in lower-tropospheric humidity. Specifically, random perturbations of ±3% were added to the specific humidity below 600 m at the initial time (Tomita, 2008). No data assimilation was applied due to the limitations of time, cost, and data availability. This perturbation method provides a practical means of representing the uncertainty in low-level moisture fields that strongly influence convective initiation, precipitation intensity, and feedbacks to the larger-scale circulation.

2.2 Observational Data

The JMA-R (Radar/Rain Gauge–Analyzed Precipitation) dataset is a high-resolution quantitative precipitation estimation (QPE) product developed by the Japan Meteorological Agency (JMA). It provides one-hour accumulated rainfall at a 1 km grid resolution of the entirety of Japan and is updated every 10 minutes. While radar observations offer broad spatial coverage, they often deviate from ground truth due to factors such as beam height and attenuation. Conversely, rain gauges provide accurate point measurements but lack spatial continuity. JMA-R addresses these limitations by merging radar and gauge data, thereby generating spatially consistent rainfall estimates with improved quantitative accuracy.

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

JMA-R integrates data from approximately 10,000 rain gauges operated by JMA, MLIT, and local governments, together with 46 C-band radars across Japan. The product is generated through a three-step process: (1) accumulation of radar echoes accounting for advection, (2) calibration of radar-estimated rainfall using co-located rain gauge observations, and (3) composition of calibrated radar fields into a unified national grid (Makihara, 2000). The two-stage calibration, regional and local, ensures that JMA-R closely matches ground-based observations while preserving spatial rainfall patterns essential for mesoscale analysis.

Complementary ground-based rainfall observations were obtained from the AMeDAS (Automated Meteorological Data Acquisition System) network (Kobayashi et al., 2017), which provides hourly precipitation and meteorological measurements across Japan (Takuto et al., 2017). For this study, AMeDAS data from stations surrounding Kumamoto City or on a selected sub-region (130°–130.5°E, 32.2°–32.7°N) were included. They were 田浦 (Taura), 牛深 (Ushibuka), 一里山 (Ichiriyama), 人吉 (Hitoyoshi), 山江 (Yamae), 八代 (Yatsushiro), 本渡 (Hondo), 松島 (Matsushima), 三角 (Misumi),宇土 (Uto), and 熊本 (Kumamoto), were used to derive the mean areal rainfall representative of the high-accumulation region in the simulation domain. These multi-station datasets enable more robust temporal validation of the model and bias-corrected forecasts.

Figure 1 shows the accumulated 48-hour rainfall distribution from JMA-R during the simulation period, together with the corresponding AMeDAS station observations at Kumamoto and Ureshino Stations. We included the Ureshino Station data as it recorded the record-breaking rainfall during this extreme rainfall event on August 14, 2021. These datasets serve as the reference for evaluating the rainfall magnitude and temporal evolution reproduced by the ensemble simulations.

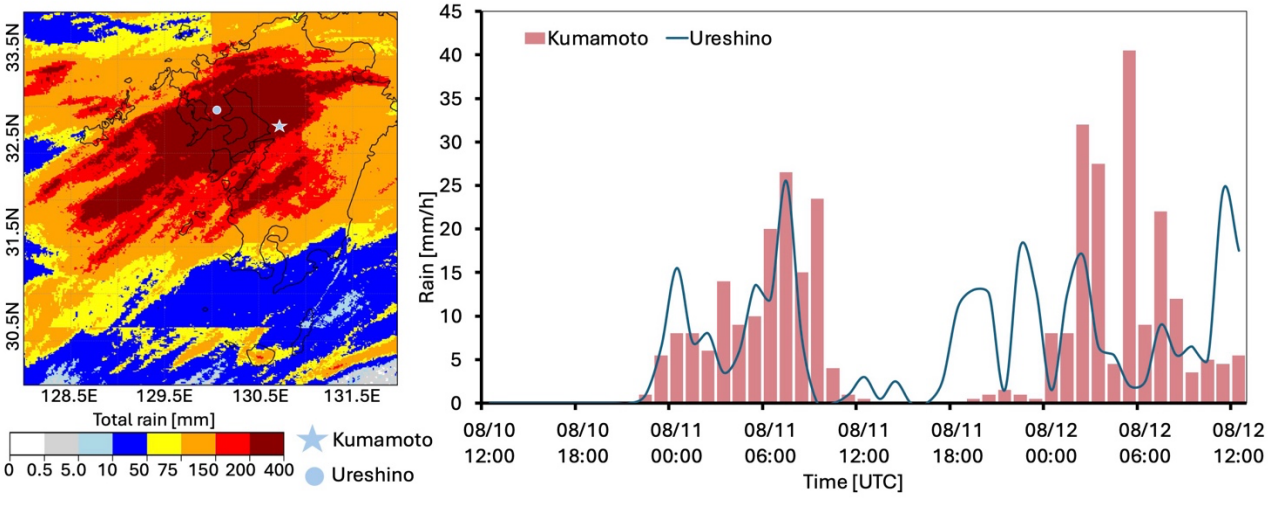


Figure 1: 48-hour of accumulated rainfall in Kyushu region at 12 UTC on 10 to 12 UTC on 12 August 2021 (left) and temporal rainfall for the same period in Kumamoto and Uershino Station.

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

2.3 Post Processing

Before post-processing, differences in spatial and temporal resolution among the ensemble sets and observational datasets were first reconciled. All observational data were clipped to match the spatial extent of each ensemble set, and the model outputs were re-grided to a 1 km spatial resolution to align with the JMA-R data. After this step the 2D grid size (*xy*) presented in Table 1 were reshaped to be 477 × 318, or equals to 151,686 data points. For the coarse-resolution ensemble (S1), the JMA-R data were aggregated to an hourly temporal resolution for consistency.

The post-processing was conducted in two sequential steps: Mean Bias (MB) correction and Quantile Mapping (QM), as outlined in Fig. 2. Two correction approaches were tested—(1) a one-step method, where only QM was applied, and (2) a two-step method, combining MB and QM corrections. The detailed procedures for each method are described in Sections 2.3.1 and 2.3.2.

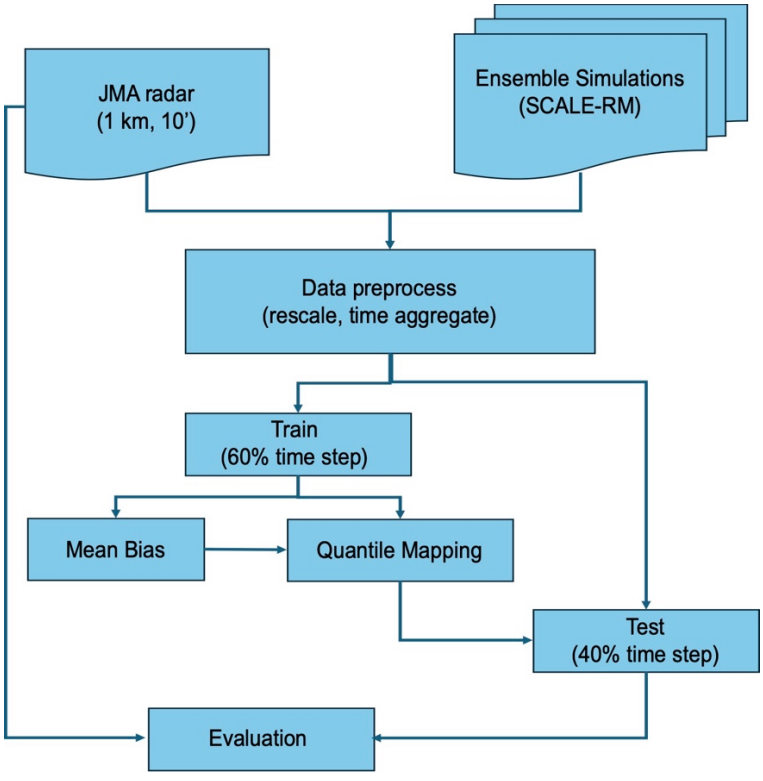


Figure 2: Procedural step of bias correction of the ensemble rainfall simulation

2.3.1 Mean bias (MB) correction

The MB is a commonly used approach assuming a constant offset between the model and observations. As an initial step, we derive the average difference of the historical model and observational percentiles. Alternatively, the difference between the mean values of both empirical cumulative distribution functions (ECDFs) can be computed. For ensemble member m, the mean bias $MB^{(m)}$ relative to the reference rainfall (JMA-R) is defined as

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

$$MB^m = \frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(m)} - Y_i\right), \quad (1)$$

where $X_i^{(m)}$ and Y_i denote the simulated and observed rainfall at the i-th time or grid cell, respectively. The bias-corrected forecast $\hat{X}_i^{(m)}$ is obtained by subtraction this constant offset:

$$\hat{X}_i^{(m)} = X_i^{(m)} - MB^{(m)}, \quad (2)$$

This procedure assumes a uniform bias across the distribution, thereby adjusting only the mean value while preserving the variance and higher-order distributional features of the ensemble forecast.

2.3.1 Quantile Mapping (QM)

The term *quantile mapping* (QM) summarizes a range of similar bias-correction approaches widely used in the climate research community (Lehner et al., 2023). Here, we adopt an isotonic regression (IR) formulation (Henzi et al., 2021), a non-parametric method that estimates a monotonic, piecewise-constant mapping between forecasts and observations while preserving the rank order of the input data.

Let $X_i^{(m)}$ denote the forecasted value from ensemble member m and Y_i the corresponding observation. Then, all paired samples $(X_i^{(m)}, Y_i)$ are first sorted in ascending order of $X_i^{(m)}$:

$$\left(X_{(1)}^{(m)}, Y_{(1)}\right), \left(X_{(2)}^{(m)}, Y_{(2)}\right), \cdots \left(X_{(n)}^{(m)}, Y_{(n)}\right), \quad (3)$$

Here parentheses indicate the sorted order.

Adjacent samples are then merged into non-decreasing blocks so that the block-mean observed rainfall values are monotonically increasing. Let each block B_k contain consecutive indices, and define the block means as

$$\bar{X}_{(k)}^{(m)} = \frac{1}{|B_k|}\sum_{j\in B_k} X_{(j)}^{(m)}, \quad (4)$$

$$\bar{Y}_k = \frac{1}{|B_k|}\sum_{j\in B_k} Y_{(j)}, \quad (5)$$

The resulting stepwise isotonic mapping function $f_{\mathrm{IR}}^{(m)}(X)$ is defined in eq. (6)

$$f_{\mathrm{IR}}^{m}(X) = \begin{cases} \bar{Y}_{(1)}, & X \le \bar{X}_{(1)}^{(m)} \\ \bar{Y}_{(k)}, & \bar{X}_{(k-1)}^{(m)} < X \le \bar{X}_{(k)}^{(m)}, k = 2, \cdots, K, \\ \bar{Y}_{(K)}, & X \le \bar{X}_K^m \end{cases} \quad (6)$$

The corrected rainfall for each ensemble member is finally obtained as

$$X_{\mathrm{QM},i}^{(m)} = f_{\mathrm{IR}}^{(m)}\left(X_i^{(m)}\right), \quad (7)$$

2.3.3 Mean bias corrected Quantile Mapping (MB+QM)

In this combined approach, the ensemble rainfall is first adjusted by the MB method, then passed through the same isotonic mapping function described in sub-section 2.3.2. The derived formula was combined from eq. (7) and eq. (2), as follows.

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

$$X^{(m)}_{\mathrm{Bias+QM},i} = f^{(m)}_{\mathrm{IR}}\left(X^{(m)}_i - MB^{(m)}\right), \quad (8)$$

This sequential correction first removes the constant offset between model and observation, and then corrects the residual non-linear quantile mismatch using the monotonic isotonic regression function.

2.4 Evaluation

The evaluation was conducted both spatially (on a grid basis) and temporally, focusing on the ensemble rainfall fields after correction. Four statistical metrics were used to assess the correspondence between the ensemble simulations and the reference JMA-radar data: root mean square error (RMSE), probability map, equitable threat score (ETS), and bias score (BS). These metrics collectively provide a comprehensive measure of accuracy, reliability, and event detection skill.

The evaluation was performed only on the test period, rather than the full dataset, to ensure an independent assessment of the bias correction and quantile mapping performance. This separation prevents overfitting to the training data used in the calibration phase and better reflects the generalization capability of each correction method.

As the simulation sets differ in their integration periods (Table 1), it is important to note that the common overlapping window among all ensemble sets corresponds to the S2 period, from 12:00 UTC on 11 August to 12:00 UTC on 12 August. Considering that only approximately 40% of this period was used as test data, the interval between 03:00 and 12:00 UTC on 12 August was selected for the evaluation.

For each ensemble member, RMSE was computed grid-wise against JMA-Radar rainfall (eq. 9). The probability maps were generated for the threshold θ of 5 mm h^{-1}, representing the occurrence probability of significant rainfall events, as presented in eq. 10.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2}, \quad (9)$$

$$P_i^{(S)}(\theta) = \frac{1}{M_s}\sum_{m=1}^{M_s}\mathbf{1}\{F_i^{S,m} > \theta\}, \quad (10)$$

For reference comparison, the probability map for set S4 was used as the baseline, while S1–S3 maps were presented relative to S4 to highlight inter-set differences in rainfall representation. For each set $S \in \{S1, S2, S3\}$:

$$\Delta P_i^{(S|S4\})}(\theta) = P_i^{(S)}(\theta) - P_i^{(S4)}(\theta), \quad (11)$$

ETS and BS were then calculated using binary event detection for the same threshold, quantifying the model's ability to correctly capture observed rainfall occurrences while accounting for random hits and systematic bias (Eqs. 12-13).

$$ETS = \frac{H - H_r}{H + M + F - H_r}, \quad (12)$$

$$BS = \frac{H+F}{H+M}, \quad (13)$$

In Eqs. 12-13, H is Hit, M is Miss, F is False explained in a contingency table provided as Table 2, and H_r is random hit, calculated as follows.

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

$$H_r = \frac{(H+M)(H+F)}{H+M+F+C}, \quad (14)$$

Table 2: Contingency Table of Model Evaluation

		Ensemble Forecast	
		Yes	**No**
Observed (JMA-R)	**Yes**	Hit	Miss
	No	False	NULL

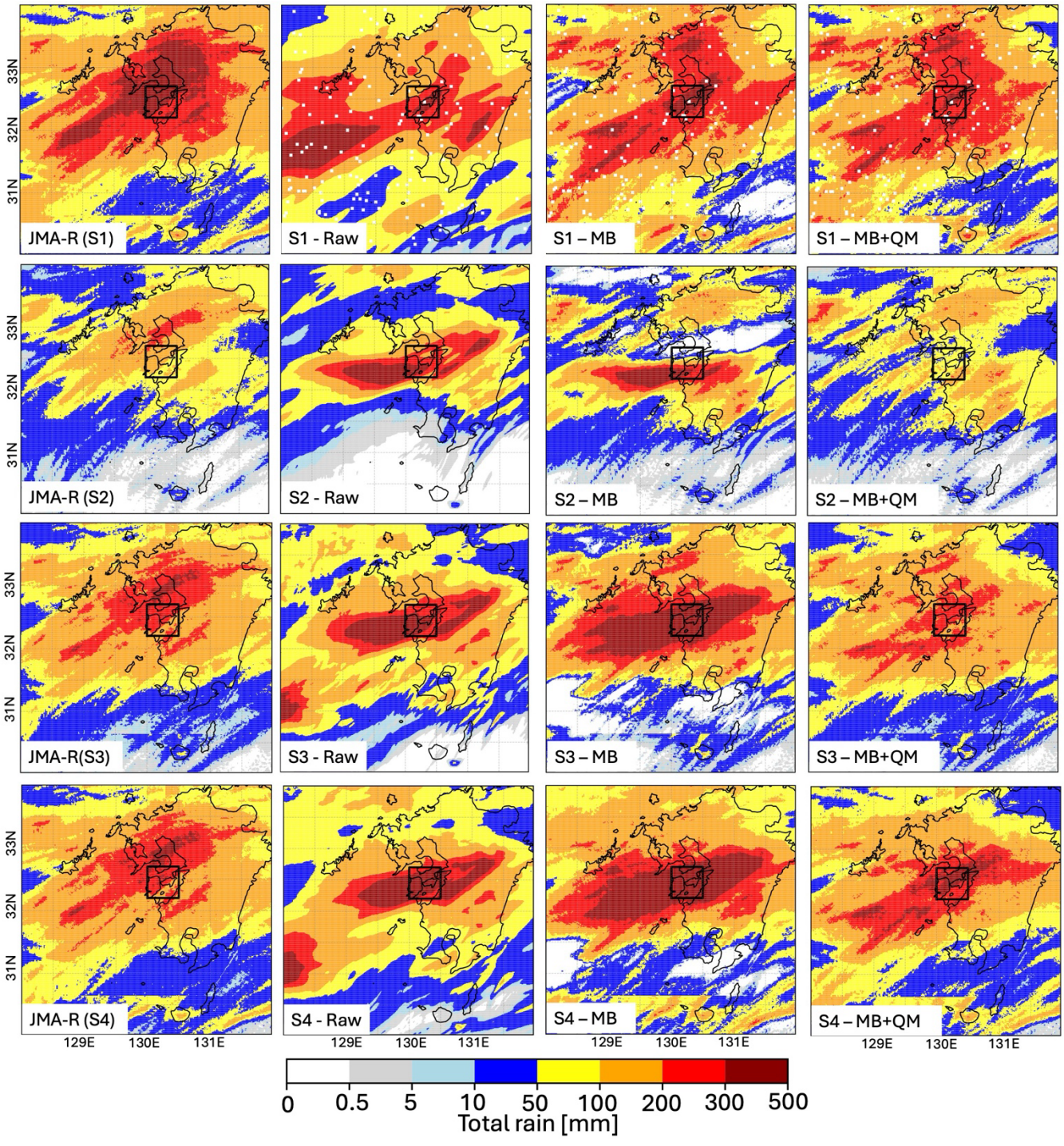
3 Results

3.1 Impact of post-processing

Total rain based on each set of ensembles and the post-processing method is presented in Fig. 3, while temporal variation in the target area, indicated by black box in Fig. 3, is presented in Fig. 4. The target area is selected due to its proximity to Kumamoto Station, which experience high intensity of rainfall on 11-12 August 2021 (Fig. 1).

Figures 3 and 4 show that across the ensemble sets S1–S4, the influence of spatial resolution and training length is evident in both the rainfall accumulation and the temporal evolution. The raw simulation failed to produce heavy rainfall occurred in the northern part of Kyushu. The finer-resolution sets (S2–S4) consistently produce much stronger rainfall intensity, particularly during the first peak of the event, than both S1 and the JMA reference. This overamplification is most pronounced in S2 to S4, while S2 specifically exhibits an extreme overestimation near 12 August 1200–1500 UTC that persists even after MB or MB+QM correction. Although S1 has the coarsest grid spacing, its longer effective training window (≈44 hours, compared with ≈14 hours for S2) allows the correction to better capture the temporal structure of the *senjō-kōsuitai*, resulting in more realistic spatial rainfall patterns. The largest improvements after applying MB+QM appear in S4, which benefit from the largest number of training samples from its time integration and temporal resolution, allowing the quantile adjustments to more effectively reshape the extreme rainfall tail and redistribute excessive intensities into physically consistent moderate-rainfall regions, especially for the heavy rainfall in the central of Kyushu, near Kumamoto station.

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026



230 **Figure 3: Accumulated rainfall from all set simulation based on corrected method as indicated in each panel compared to its observed JMA-R, row wise from top to bottom is S1 to S3**

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

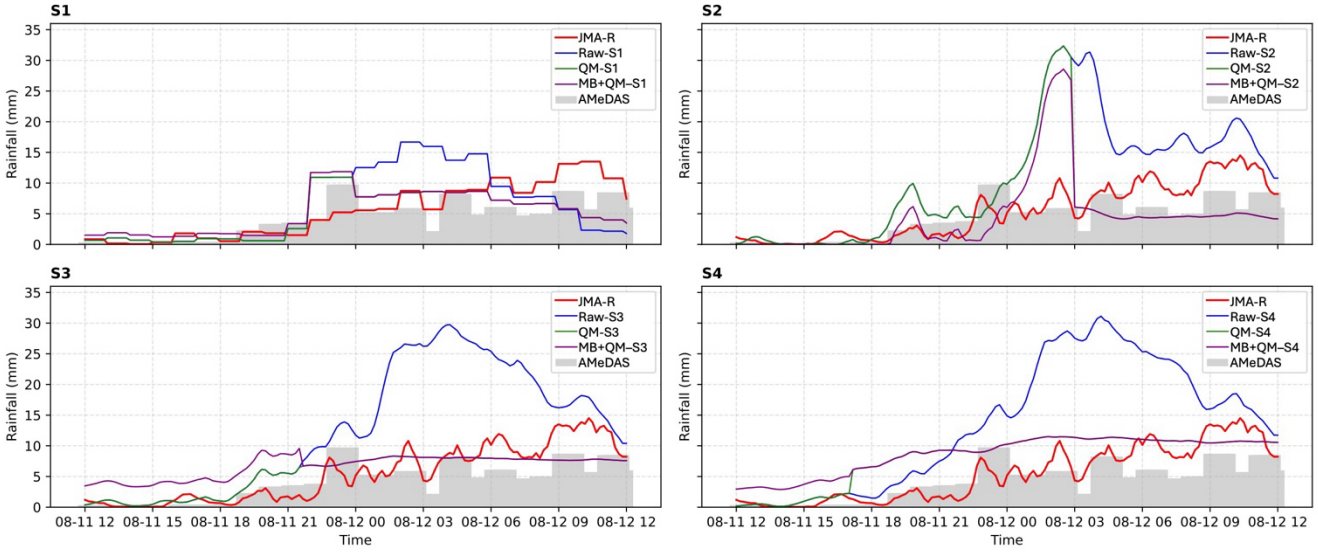


Figure 4: Temporal rainfall fluctuation compared to observed rainfall within the overlapping period on Aug 11, 12 JST to Aug 12, 12 JST.

3.2 Evaluation

The probability map of rain rates exceeding 5 mm h^{-1} for S4 (MB+QM) relative to JMA-R is shown in Fig. 5. Figure 6 presents the probability differences for the same threshold—S1–S4, S2–S4, and S3–S4—using S4 as the baseline. In S4, the combined MB+QM correction produces an extensive region with probabilities above 0.9, aligning with the observed concentrated rainfall area exceeding 50 mm during the 9-hour test period (03:00–12:00 UTC on 12 August).

In the probability-difference maps, S1 shows negative values around −0.5 in areas where S4 exhibits high probabilities. Positive differences of about 0.25–0.5 appear in northern Kyushu, and these regions correspond to areas of moderate to heavy observed rainfall, indicating that S1 captures part of the northern rainfall signal that S4 still slightly underestimates (by roughly 0.2 in probability). The negative differences are more pronounced in S2, with several grid points reaching −1.

For S3, most areas with high probabilities in S4 correspond to near-zero differences, although regions where S4 has probabilities around 0.6 show negative differences. Positive differences are present over southern Kyushu in S3; however, comparison with JMA-R indicates that these locations correspond to accumulated rainfall below 5 mm, meaning S3 produced higher probabilities than observed in that region.

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

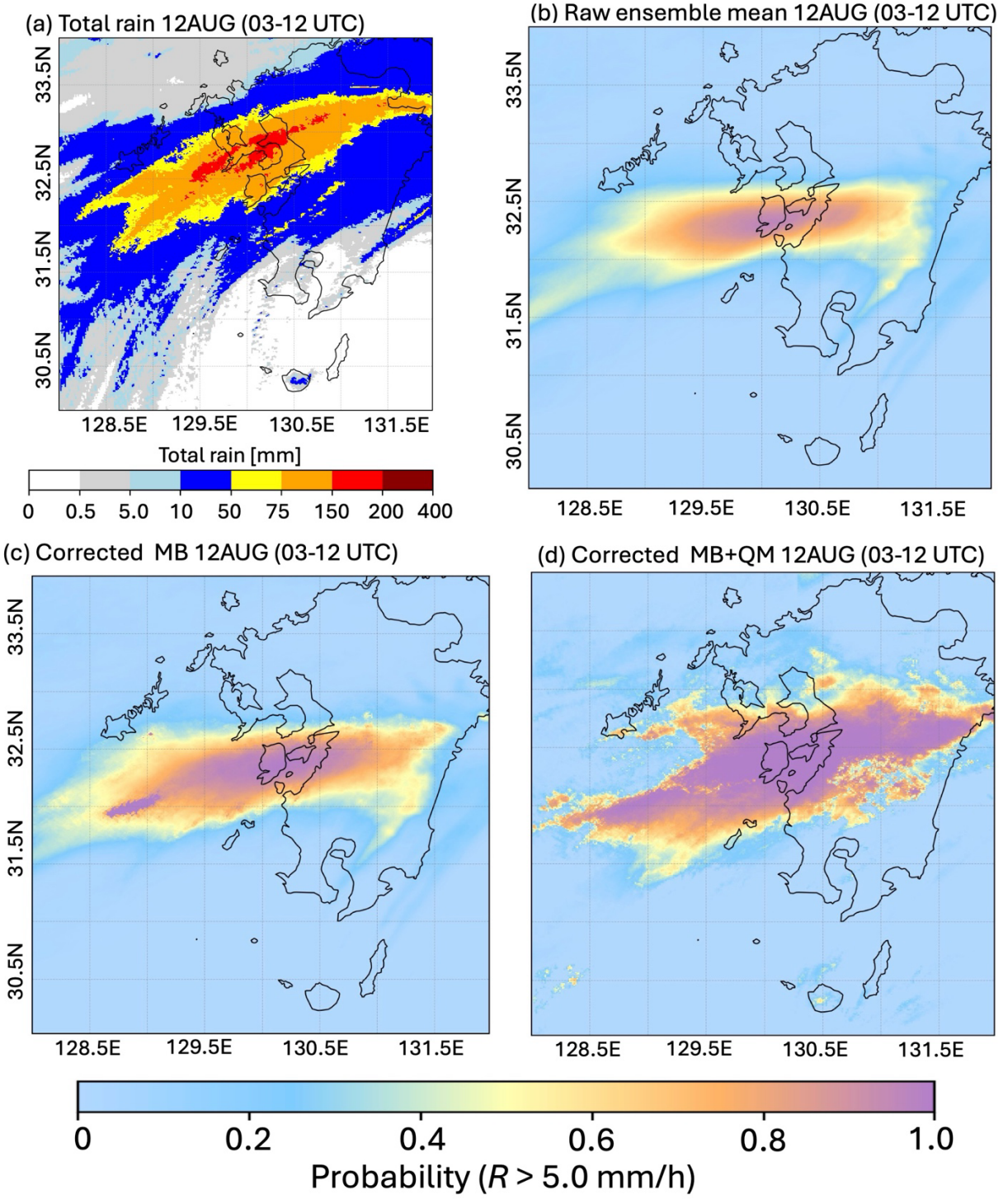


Figure 5: Probability map of rain rate higher than 5 mm/h for S4 compared to JMA-R

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

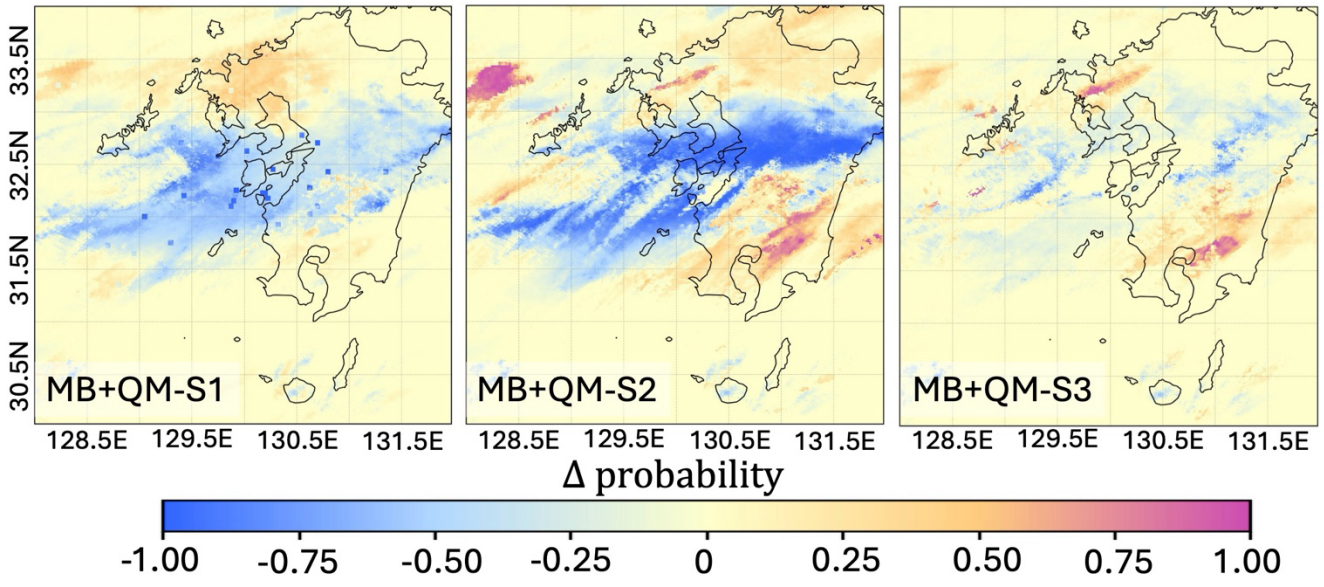


Figure 6: Difference in probability map compared to S4 (MB+QM) as the base line for S1, S2, and S3

The grid-wise RMSE distribution is shown in Fig. 7. Overall, S1 exhibits lower RMSE than S2–S4 in the raw simulations, with most values falling below 10 mm h^{-1}. Applying MB+QM does not substantially change the spatial RMSE structure for S1, although several locations with low JMA-R rainfall, which previously missed in the raw fields, become more apparent. In the finer-resolution runs, S2 displays the broadest area of RMSE exceeding 50 mm h^{-1}, concentrated along a narrow southwest–northeast corridor over central Kyushu. This high-RMSE band remains visible in S3 and S4 but with reduced spatial extent. After MB+QM, the finer-resolution RMSE patterns become more comparable to S1; however, a persistent southwest–northeast RMSE band above 50 mm h^{-1} remains in all ensembles, including S1, indicating a systematic structural error that is not removed by intensity-based correction.

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

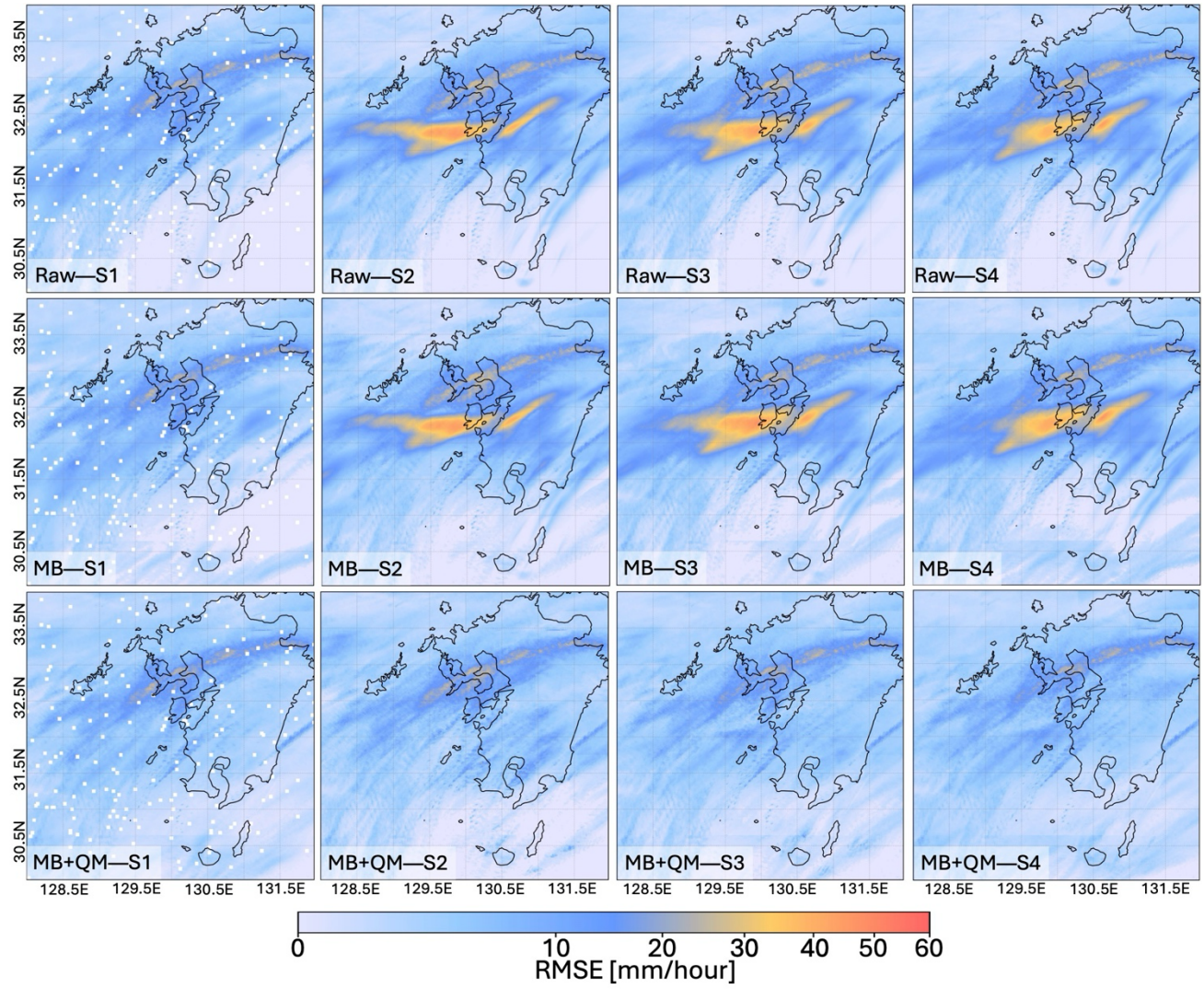


Figure 7: Temporal average of grid-wise RMSE for each ensemble. Presented is for Raw (top row), Mean Bias (second row), and combination of Mean Bias and Quantile Mapping (third row).

The BS of the ensemble set are presented in Fig. 8., where we presented average grid wise BS for full domain and target area (Fig. 3). The BS values for both the full domain and the target box generally exhibit the same trend, where the postprocessing method increases the BS.

For the full-domain evaluation in Fig. 8, the raw BS for S1 at the lowest threshold is around 0.75, slightly higher than S2–S4, which fall between 0.6–0.7. All ensembles show a similar decreasing trend with increasing thresholds, indicating that underforecasting becomes more severe at higher rain-rate categories. Applying QM alone yields only a marginal increase in BS. In contrast, MB+QM produces a clearer improvement for S3 and S4 at the lower thresholds (<1.5 mm h^{-1}), where their

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

BS values rise to approximately 0.8, moving closer to the ideal value of 1. At higher thresholds (>4 mm h^{-1}), the BS for all ensembles converges toward the raw-data values, reflecting persistent underestimation at these intensities.

Within the target-area evaluation—where heavier rainfall is expected near the Kumamoto segment of the rainband—S3 and S4 maintain BS values close to 1 up to the 3 mm h^{-1} threshold, indicating good agreement between forecasted and observed rainfall coverage. At higher thresholds (4–5 mm h^{-1}), BS remains above 0.75, showing moderate underforecast but still better performance than in the full-domain case. S1 follows a gradually decreasing pattern across thresholds, dropping below 0.75 at high intensities, consistent with its tendency to underestimate localized extremes.

S2 displays a distinct behavior: within the target area, the corrected BS values (QM and MB+QM) are lower than the raw BS at both low and high thresholds, with only the midrange threshold near 2 mm h^{-1} showing an increase. This pattern suggests that S2's raw fields overestimate the spatial extent of rainfall in the narrow *senjō-kōsuitai* corridor, and QM substantially reduces this overforecasting. This behavior is not evident in the full-domain results for S2 but becomes apparent in the target-area analysis, where the overintensification and spatial overextension of S2's convective core are most pronounced.

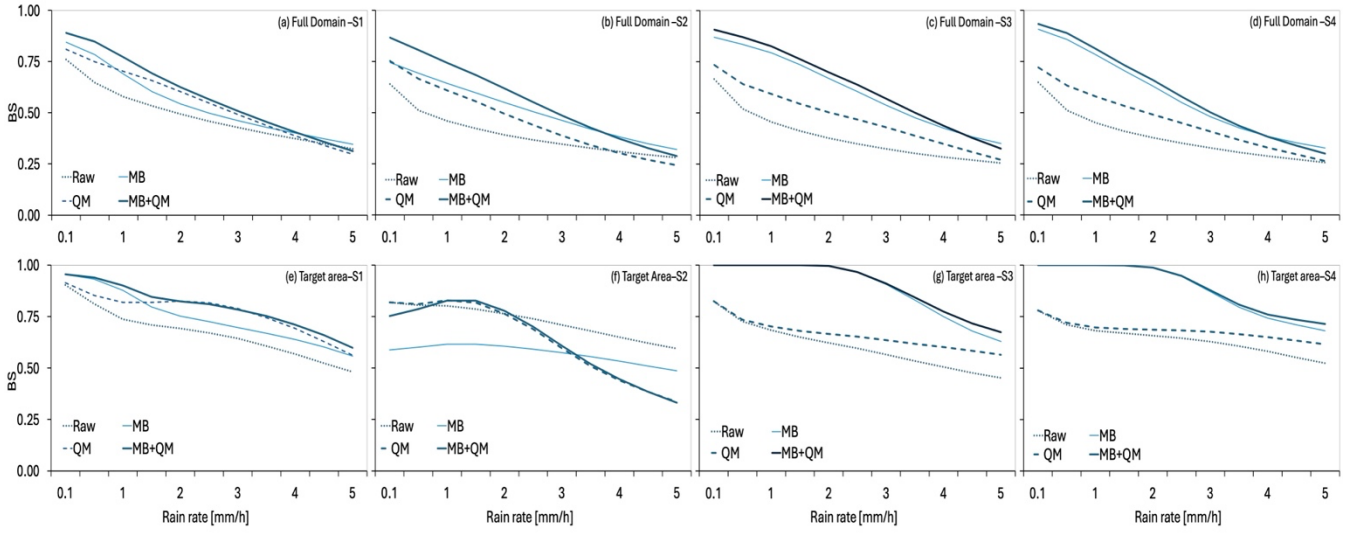


Figure 8: Average grid-wise Bias score (BS) for full domain (first row) and target area (second row) for each ensemble set as indicated on top right corner in each panel.

We presented the ETS in Fig. 9. Across the full domain, all raw and QM-only simulations produce higher ETS than MB- or MB+QM-corrected fields. S1 raw has the highest ETS (~0.3), followed by S4 slightly below 0.2, while S2 and S3 remain near or below 0.1. After applying MB, ETS drops below 0.1 for all ensembles at low thresholds. With increasing thresholds, S1 and S4 gradually recover and converge toward their raw ETS curves, reaching slightly below 0.1 around the 5 mm h^{-1} threshold. S2 and S3 remain consistently low, with ETS staying below 0.1 across all thresholds. Within the rainband corridor, similar behaviour appears for S1, S3, and S4: at thresholds below ~2 mm h^{-1}, MB and MB+QM sharply reduce ETS to values near zero. In S1, the raw ETS reaches ~0.6, and QM slightly enhances ETS for low-to-moderate thresholds

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

(improvement of ~0.2). MB+QM catches up around 2.5 mm h^{-1} and then exceeds the raw ETS at higher thresholds, maintaining ETS > 0.3 at 5 mm h^{-1}. S4 also shows improved ETS relative to raw at higher thresholds (ETS ≈ 0.2). S3 follows the same upward trend at higher thresholds, though the magnitude of improvement is small. S2 behaves differently, where QM and MB+QM significantly improve ETS only at low thresholds (<2.5 mm h^{-1}), but ETS degrades relative to raw at higher thresholds.

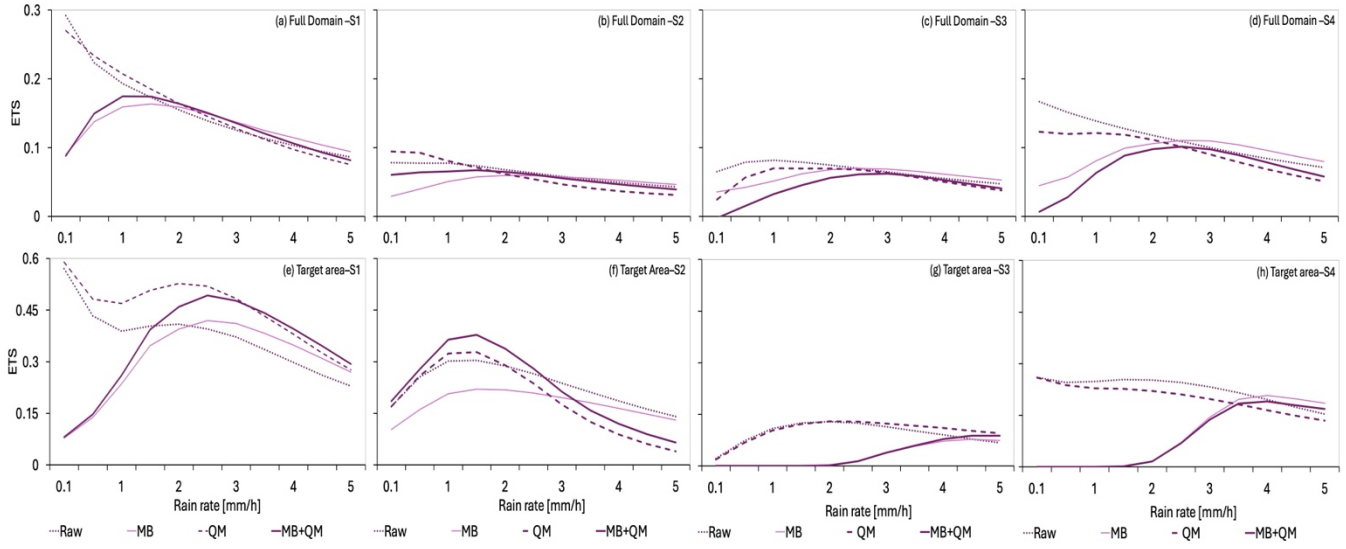


Figure 9: Average grid-wise equitable threat score (ETS) for full domain (first row) and target area (second row) for each ensemble set as indicated on top right corner in each panel.

4 Discussion

Bias correction substantially modifies the spatial and statistical characteristics of the ensemble rainfall forecasts, but its effectiveness depends strongly on the nature of the dominant error in each ensemble. The contrasting behavior between the coarse S1 ensemble and the finer S2–S4 ensembles confirms that higher spatial resolution alone does not guarantee improved forecast skill for *senjō-kōsuitai*. Although finer grids can resolve convective cores and sharper moisture gradients, they also introduce heightened sensitivity to initial perturbations and displacement errors, an effect amplified when ensemble size is limited. With only 50 members, S2–S4 lack sufficient spread to represent positional uncertainty, leading to overly extensive and intense rainbands compared to the observed data (Figs. 3–4). Although this limitation reduces the effectiveness of post-processing for S2 and S3, S4 improves more consistently because its longer time integration produces a more coherent rainfall structure. As a result, the remaining errors in S4 are mainly amplitude-related rather than positional, allowing MB+QM to correct them more effectively.

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

4.1 Impact of Bias Correction on Rainfall Intensity and Spatial Structure

Bias correction improves the intensity distribution for all ensembles, but the degree of improvement depends on whether the dominant error is amplitude-related or structural. S1 benefits the most from MB+QM because its errors are largely amplitude-related, while its spatial structure is already realistic. As shown in Fig. 3, the orientation and placement of the rainband in S1 closely match JMA-R, with a similar southwest–northeast alignment and a comparable distribution of areas exceeding 100 mm of accumulated rainfall. The main discrepancy lies in the magnitude of the intensities rather than in the geometry of the band. Therefore, MB+QM can effectively correct S1 by scaling the rainfall toward the observed distribution without needing to repair shape or positional errors.

In contrast, S2–S4 exhibit structural issues—such as an east–west-tilted rainband and overly wide convective cores—that cannot be resolved through amplitude-based post-processing alone. Despite these structural limitations, S4 shows substantial improvement because its longer integration time produces a more organized and physically consistent rainfall field; the remaining errors are dominated by intensity overestimation, which MB+QM can effectively suppress.

These findings are consistent with previous studies (Hamill et al., 2017; Yang et al., 2020), which emphasize that bias correction performs best when the training period is sufficiently long to sample the full spectrum of model–observation relationships. In this study, the combination of large ensemble size (50 members) and high temporal resolution effectively increases the number of training samples even within a short time window, allowing the correction to remain feasible despite limited data. This also explains why S4, with longer integration time produces a more coherent and physically organized rainfall field, benefits more from MB+QM than S2 and S3. However, the short duration of the training dataset remains a key limitation, as the corrected fields presented here are valid primarily for this specific event, and broader operational or multi-event applications would still require a longer training period to ensure robust and generalizable bias relationships..

The misalignment in the raw high-resolution ensembles (S2–S4) is evident in both the probability and RMSE maps. For example, in Fig. 5, raw S4 produces a high-probability region that is incorrectly oriented relative to the observed distribution of heavy rainfall. After applying MB+QM, the probability field becomes substantially more realistic, and the area with >0.9 probability of exceeding 5 mm h^{-1} aligns closely with the observed heavy-rainfall footprint. S1 and especially S2 show broad regions of negative probability differences relative to S4, indicating systematic underprediction of the localized high-probability core. RMSE patterns confirm these tendencies: S1 maintains comparatively low RMSE across Kyushu, while S2 exhibits widespread RMSE exceeding 50 mm due to exaggerated rainband expansion. MB+QM significantly reduces the high-RMSE regions in S3 and S4, enabling the corrected S4 field to approach S1's spatial accuracy while more effectively capturing the extreme-rainfall core.

All ensembles, regardless of resolution, display a persistent southwest–northeast RMSE corridor that mirrors the actual propagation path of the *senjō-kōsuitai*. This feature arises from displacement and timing errors: slight offsets in a narrow moving band produce a continuous streak of RMSE along its trajectory. Because MB+QM only modifies intensities, not

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

spatial alignment or timing, this error corridor remains even after post-processing. Its presence in all ensembles highlights that **misalignment, not intensity bias, is the dominant source of forecast error** for this event.

4.2 Interpretation of BS and ETS Behavior

The BS patterns show that categorical performance is governed more by spatial realism than by intensity accuracy alone, consistent with the spatial-skill findings of Gilleland et al. (2009) and Skok & Roberts (2016). The general decline of BS with increasing thresholds indicates systematic underforecasting of high-intensity grid cells. High-resolution ensembles (S3–S4) achieve BS values close to 1 in the target region at moderate thresholds, reflecting their superior ability to resolve convective cores (Schwartz et al., 2009). However, their strong sensitivity to initial conditions broadens the intensity distribution, leading to increasing underestimation at higher thresholds, a known effect of convection-permitting models with limited ensemble size (Clark et al., 2016; Raynaud and Bouttier, 2017).

The distinctive behavior of S2, in which BS decreases after correction, indicates that its raw rainfall fields were spatially inflated. QM suppresses these excessive rainband extensions, reducing false alarms but simultaneously reducing hits—particularly in the narrow target corridor where spatial alignment is critical. This pattern was not apparent in the full domain but becomes clear when focusing on the corridor, illustrating that bias correction can paradoxically reduce skill when structural errors dominate the ensemble, which is in agreement with findings from (Scheuerer and Hamill, 2015).

ETS provides a complementary measure of categorical skill and is even more sensitive to spatial alignment (Gilleland et al., 2009). In both the full and target domains, raw forecasts consistently exhibit higher ETS than their MB- or MB+QM-corrected counterparts at lower thresholds. This decline arises because MB correction reduces ensemble spread and suppresses fine-scale variability needed to match observed rain/no-rain patterns. At low thresholds, spatial distribution rather than intensity governs performance, meaning post-processing can increase both false alarms and misses.

At higher thresholds, however, the dominant error shifts from positional misalignment to intensity realism. In this regime, MB+QM improves ETS for S1, S3, and especially S4 by removing systematic overestimation while retaining the primary convective cores, consistent with the behavior described by Scheuerer and Hamill (2015). Because heavy-rain areas are spatially compact, the influence of displacement uncertainty is reduced. Consequently, S1 and S4 outperform their raw versions above ~4 mm h^{-1}, with S4 reaching ETS values of ~0.2 and S1 reaching ~0.3.

S2 again exhibits unique behavior: its ETS improves only at low thresholds, where suppressing raw over-intensity reduces false alarms. At higher thresholds, its structural misalignment becomes the dominant limitation, and QM's suppression of extremes increases missed detections, causing ETS to fall below the raw version.

Taken together, the ETS results highlight two key points: (1) fine spatial resolution increases the potential to detect high-intensity extremes, but (2) insufficient ensemble size amplifies displacement noise and structural uncertainty, reducing categorical performance. Bias correction enhances ETS only when intensity error, not spatial misalignment, is the dominant source of forecast deficiency.

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

4.3 Implications for Ensemble Configuration and Downstream Applications

These findings suggest that no single ensemble configuration is sufficient for *senjō-kōsuitai* forecasting. Large-member coarse ensembles (e.g., S1) offer stable spatial patterns, while high-resolution ensembles (e.g., S4) provide sharper intensity details but are sensitive to displacement errors.

For downstream applications, such as flood modeling or targeted weather-modification experiments (Hiraga et al., 2025; Yang et al., 2020), a practical strategy is to use the coarse ensemble to define the robust spatial structure and then refine local features using selected high-resolution members. This combined workflow offers a better balance between spatial reliability and dynamical detail than using either system alone.

As the ensemble simulation in this study was originally done for weather-modification experiments, the implications of these findings are particularly important. Because the goal of cloud seeding or rainfall suppression experiments is to detect artificially induced changes in rainfall, any bias in the model baseline directly affects the interpretation of modification effects. Without post-processing, high-resolution ensembles such as S3 and S4 tend to exaggerate rainfall amounts and produce overly broad rainbands, while coarse ensembles like S1 may underestimate localized extremes. Either condition can produce false conclusions: an overpredicting model may falsely suggest that seeding "increased" rainfall when the bias originates from model physics, while an underpredicting model may mask genuine modification effects. The consistent overgrowth and structural distortion seen in S2–S4 before correction illustrate this risk.

As we presented in here, MB+QM reduces amplitude biases, but its effectiveness depends on the availability of an observational baseline. In this study, JMA radar provides a reliable reference, but such a baseline is not available during controlled modification trials, where the goal is to predict counterfactual (unmodified) rainfall. Using an operational radar climatology for post-processing in such cases may inadvertently remove or distort the very perturbations that the experiment aims to measure. Therefore, weather-modification forecasting requires caution: post-processing must correct systematic model bias without forcing the ensemble toward a climatology that erases the signal of the intervention. A dual-ensemble approach—using a large-member coarse ensemble to define the unmodified background and a high-resolution ensemble to simulate seeded scenarios—offers a pathway that minimizes these risks.

5 Conclusion

This study evaluated high-resolution ensemble rainfall forecasts of the August 2021 Kyushu extreme rainfall, characterized by multiple linear rainbands (*senjō kōsui-tai*), using RMSE, probability maps, ETS, and BS metrics. Post-processing via Mean Bias (MB) and Quantile Mapping (QM) was applied to assess how ensemble size and resolution influence forecast skill.

The results demonstrate that ensemble size and resolution jointly determine forecast reliability. Coarser-resolution S1, with 100 members, produced relatively stable spatial patterns, and MB+QM correction improved rainfall representation, despite persistent underestimation of some extreme values produced in finer resolution model. Finer-resolution ensembles (S2–S4)

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

captured localized peaks better but exhibited higher variability and overestimation, particularly in S2, which were partly corrected by MB+QM. S4, with MB+QM applied, achieved the closest match to observed rainfall, illustrating that high-resolution ensembles can effectively capture extreme intensity when coupled with robust post-processing.

Evaluation across RMSE, probability maps, ETS, and BS reveals consistent trends: MB+QM improves magnitude and probability representation, reduces RMSE hotspots, and enhances spatial coherence, but cannot fully compensate for displacement errors or structural misrepresentation in smaller ensembles. Notably, a sufficiently large ensemble (S1) can achieve comparable post-processed skill to finer ensembles for general spatial patterns, highlighting the critical role of ensemble size in operational forecasting under computational constraints.

Overall, these findings underscore the nuanced interplay between resolution, ensemble size, and post-processing. Ensemble design must balance member number, grid spacing, and temporal resolution to reliably capture both spatial distribution and intensity of extreme rainfall. MB+QM correction proves essential for improving forecast reliability, particularly in high-resolution ensembles, but care must be taken to recognize persistent under- or overestimation. The results provide a practical foundation for both operational extreme rainfall prediction and future weather-modification assessments, illustrating that **accuracy is not solely determined by resolution—ensemble size and intelligent post-processing are equally decisive.**

Code and data availability

The JMA-R data are publicly available at [https://database.rish.kyoto-u.ac.jp/arch/jmadata/data/jma-radar/synthetic/original/], and AMeDAS data can be accessed at [https://www.data.jma.go.jp/stats/etrn/index.php.] Codes for downloading the JMA-R data, as well as for statistical analysis and plotting, are available at [https://github.com/irasyarif1906/ensemble-post-processing.git] upon request. ERA5 reanalysis data were obtained from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [https://cds.climate.copernicus.eu/].The ensemble simulation data are not publicly available because they are still under development for their primary research purpose; access can be granted upon request, subject to JST approval.

Author contributions

MS and HA conceptualized the paper. MS wrote and prepared the manuscript, conducted the pre- and post-processing of the ensemble simulation results, statistical analysis and visualization of all the results in this paper, HH conducted the pre-analysis and visualization of ensemble simulation results, KY conducted all the ensemble simulation by SCALE-RM. KY and HA supervised the work. All authors contributed equally to the revision and internal review process.

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

Competing interests

The authors declare that they have no competing interests.

Disclaimer

All results, interpretations, and conclusions presented in this paper are those of the authors. The publisher and affiliated institutions are not responsible for any errors or opinions expressed.

Acknowledgements

The authors gratefully acknowledge the computational resources provided by the Fugaku supercomputer at the RIKEN Center for Computational Science (R-CCS). We thank the Japan Meteorological Agency (JMA) for providing the C-band Radar-corrected gauge and AMeDAS observational data. ERA5 reanalysis data were obtained from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS), and we acknowledge Copernicus for making these data publicly available.

Financial support

This study was supported by the Japan Science and Technology Agency (JST) Moonshot R&D Program (JPMJMS2389). The authors gratefully acknowledge the computational resources provided by the Fugaku supercomputer at the RIKEN Center for Computational Science (R-CCS).

References

Clark, P., Roberts, N., Lean, H., Ballard, S. P., and Charlton-Perez, C.: Convection-permitting models: A step-change in rainfall forecasting, Meteorological Applications, 23, 165–181, https://doi.org/10.1002/MET.1538;CTYPE:STRING:JOURNAL, 2016.

Gascón, E., Lavers, D., Hamill, T. M., Richardson, D. S., Bouallègue, Z. B., Leutbecher, M., and Pappenberger, F.: Statistical postprocessing of dual-resolution ensemble precipitation forecasts across Europe, Quarterly Journal of the Royal Meteorological Society, 145, 3218–3235, https://doi.org/10.1002/QJ.3615;JOURNAL:JOURNAL:1477870X;ISSUE:ISSUE:DOI, 2019.

Gilleland, E., Ahijevych, D., Brown, B. G., Casati, B., and Ebert, E. E.: Intercomparison of Spatial Forecast Verification Methods, Weather Forecast, 24, 1416–1430, https://doi.org/10.1175/2009WAF2222269.1, 2009.

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

Golian, S. and Murphy, C.: Evaluating Bias-Correction Methods for Seasonal Dynamical Precipitation Forecasts, J Hydrometeorol, 23, 1350–1363, https://doi.org/10.1175/JHM-D-22-0049.1, 2022.

Hamill, T. M., Engle, E., Myrick, D., Peroutka, M., Finan, C., and Scheuerer, M.: The U.S. National Blend of Models for Statistical Postprocessing of Probability of Precipitation and Deterministic Precipitation Amount, Mon Weather Rev, 145, 3441–3463, https://doi.org/10.1175/MWR-D-16-0331.1, 2017.

Henzi, A., Ziegel, J. F., and Gneiting, T.: Isotonic distributional regression, J R Stat Soc Series B Stat Methodol, 83, 963–993, https://doi.org/10.1111/RSSB.12450;JOURNAL:JOURNAL:2517617X;WEBSITE:WEBSITE:RSS;WGROUP:STRING:PUBLICATION, 2021.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/QJ.3803, 2020.

Hiraga, Y. and Meza, J.: Extreme precipitation modeling and Probable Maximum Precipitation (PMP) estimation in Chile, J Hydrol Reg Stud, 58, 102274, https://doi.org/10.1016/J.EJRH.2025.102274, 2025.

Hiraga, Y., Mbugua, J. M., Kotsuki, S., Suzuki, Y., Chen, S.-H., Hamada, A., Yasunaga, K., and Funatomi, T.: Numerical Experiments of Cloud Seeding for Mitigating Localization of Heavy Rainfall: A Case Study of Mesoscale Convective System in Japan, EGUsphere, 1–29, https://doi.org/10.5194/EGUSPHERE-2025-3524, 2025.

Holloway, C. E. and Neelin, D. J.: Moisture Vertical Structure, Column Water Vapor, and Tropical Deep Convection, J Atmos Sci, 66, 1665–1683, https://doi.org/10.1175/2008JAS2806.1, 2009.

Ilaboya, I. R., Onaiwu, O., Okonkwo, N. H., and Oranyelu, O. B.: Enhancing Climate Forecast Precision in Specified Nigerian Regions Through Statistical Downscaling and Bias Correction Techniques, NIPES - Journal of Science and Technology Research, 6, 98–110, https://doi.org/10.5281/ZENODO.10971717, 2024.

Kobayashi, T., Shirai, S., and Kitadate, S.: AMeDAS: Supporting Mitigation and Minimization of Weather-related Disasters, FUJITSU Sci. Tech. J, 53, 53–61, 2017.

Lehner, F., Nadeem, I., and Formayer, H.: Evaluating skills and issues of quantile-based bias adjustment for climate change scenarios, Adv Stat Climatol Meteorol Oceanogr, 9, 29–44, https://doi.org/10.5194/ASCMO-9-29-2023, 2023.

Lenderink, G., Ban, N., Brisson, E., Berthou, S., Cortés-Hernández, V. E., Kendon, E., Fowler, H. J., and De Vries, H.: Are dependencies of extreme rainfall on humidity more reliable in convection-permitting climate models?, Hydrol Earth Syst Sci, 29, 1201–1220, https://doi.org/10.5194/HESS-29-1201-2025, 2025.

Makihara, Y.: Algorithms for precipitation nowcasting focused on detailed analysis using radar and raingauge data, Technical Report of Meteorological Research Institute, 39, 63–111, 2000.

Nishizawa, S., Tomita, H., and Scale, T.: Detailed formulation of SCALE-RM, 2024.

Onishi, R., Matsuda, K., and Takahashi, K.: Lagrangian Tracking Simulation of Droplet Growth in Turbulence–Turbulence Enhancement of Autoconversion Rate*, J Atmos Sci, 72, 2591–2607, https://doi.org/10.1175/JAS-D-14-0292.1, 2015.

Pierce, D. W., Cayan, D. R., Maurer, E. P., Abatzoglou, J. T., and Hegewisch, K. C.: Improved Bias Correction Techniques for Hydrological Simulations of Climate Change, J Hydrometeorol, 16, 2421–2442, https://doi.org/10.1175/JHM-D-14-0236.1, 2015.

Raynaud, L. and Bouttier, F.: The impact of horizontal resolution and ensemble size for convective-scale probabilistic forecasts, Quarterly Journal of the Royal Meteorological Society, 143, 3037–3047, https://doi.org/10.1002/QJ.3159;WGROUP:STRING:PUBLICATION, 2017.

https://doi.org/10.5194/egusphere-2026-1052
Preprint. Discussion started: 4 March 2026

Sato, M., Kodama, Y., Tsuji, M., and Odajima, T.: Co-Design and System for the Supercomputer "Fugaku," IEEE Micro, 42, 26–34, https://doi.org/10.1109/MM.2021.3136882, 2022.

Sato, Y., Nishizawa, S., Yashiro, H., Miyamoto, Y., Kajikawa, Y., and Tomita, H.: Impacts of cloud microphysics on trade wind cumulus: which cloud microphysics processes contribute to the diversity in a large eddy simulation?, Prog Earth Planet Sci, 2, 1–16, https://doi.org/10.1186/S40645-015-0053-6/FIGURES/11, 2015.

Scherrer, S. C., de Valk, C., Begert, M., Gubler, S., Kotlarski, S., and Croci-Maspoli, M.: Estimating trends and the current climate mean in a changing climate, Clim Serv, 33, 100428, https://doi.org/10.1016/J.CLISER.2023.100428, 2024.

Scheuerer, M. and Hamill, T. M.: Statistical Postprocessing of Ensemble Precipitation Forecasts by Fitting Censored, Shifted Gamma Distributions, Mon Weather Rev, 143, 4578–4596, https://doi.org/10.1175/MWR-D-15-0061.1, 2015.

Schwartz, C. S., Kain, J. S., Weiss, S. J., Xue, M., Bright, D. R., Kong, F., Thomas, K. W., Levit, J. J., and Coniglio, M. C.: Next-Day Convection-Allowing WRF Model Guidance: A Second Look at 2-km versus 4-km Grid Spacing, Mon Weather Rev, 137, 3351–3372, https://doi.org/10.1175/2009MWR2924.1, 2009.

Skok, G. and Roberts, N.: Analysis of Fractions Skill Score properties for random precipitation fields and ECMWF forecasts, Quarterly Journal of the Royal Meteorological Society, 142, 2599–2610, https://doi.org/10.1002/QJ.2849;REQUESTEDJOURNAL:JOURNAL:1477870X;PAGE:STRING:ARTICLE/CHAPTER, 2016.

Takuto, K., Satoshi, S., and Satoshi, K.: AMeDAS: Supporting Mitigation and Minimization of Weather-related Disasters, Fujitsu Scientific & Technical Journal, 53, 53–61, 2017.

Taylor, J., Amemiya, A., Honda, T., Maejima, Y., and Miyoshi, T.: Predictability of the July 2020 Heavy Rainfall with the SCALE-LETKF, Scientific Online Letters on the Atmosphere, 17, 48–56, https://doi.org/10.2151/sola.2021-008, 2021.

Tomita, H.: New Microphysical Schemes with Five and Six Categories by Diagnostic Generation of Cloud Ice, Journal of the Meteorological Society of Japan. Ser. II, 86A, 121–142, https://doi.org/10.2151/JMSJ.86A.121, 2008.

Yang, C., Yuan, H., and Su, X.: Bias correction of ensemble precipitation forecasts in the improvement of summer streamflow prediction skill, J Hydrol (Amst), 588, 124955, https://doi.org/10.1016/J.JHYDROL.2020.124955, 2020a.