



Impacts of multi-physics ensemble on heavy precipitation prediction in South Korea: focusing on the performance of ensemble mean

Han-Gyul Jin^{1,2} · Jong-Jin Baik³

Received: 16 April 2024 / Accepted: 16 May 2025

© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2025

Abstract

The multi-physics ensemble method is a widely used method to represent the prediction uncertainty arising from model errors and has great potential to improve precipitation forecasts. This study performs multi-physics ensemble simulations of ten heavy precipitation cases in South Korea and evaluates the performance of the ensemble mean. The multi-physics ensemble is generated from 27 different combinations of cloud microphysics, planetary boundary layer, and radiation schemes. In the prediction of 24-h accumulated precipitation amount, the ensemble is underdispersive for most cases, indicating that the prediction uncertainty is only partly represented by the ensemble. The overall performance of the ensemble mean is better than that of any individual ensemble member. No individual ensemble member consistently shows good performance for every case. The root-mean-square error (RMSE) of the ensemble-mean prediction is smaller than the average RMSE of individual ensemble members for every case. The relative difference in RMSE exhibits a strong positive correlation with the spread-error ratio. In comparison with subset ensembles, the total ensemble shows the most stable performance. Among the three types of physics parameterization in the multi-physics suite, the cloud microphysics parameterization contributes the most and the radiation parameterization contributes the least to the ensemble spread and the ensemble-mean performance. Each physics parameterization scheme has tendencies to predict cloud and precipitation properties to be larger or smaller than those predicted by other schemes, which stem from differences in the choices of individual process parameterizations and the physics parameter values used. This systematic difference contributes to the ensemble spread of the multi-physics ensemble, which is a key factor for the ensemble-mean performance.

1 Introduction

Providing quantitatively better precipitation forecasts has been a long goal of numerical weather prediction, but still precipitation forecasts show high uncertainty due to the fact that various physical processes such as radiation processes, boundary layer processes, and cloud microphysical processes along with their complicated interactions are involved in precipitation processes. The uncertainty of numerical

precipitation prediction originates from the errors in the initial conditions that amplify with model integration due to the chaotic nature of the nonlinear governing equations (Lorenz 1965), from the errors in the model's representations of physical processes with physical parameterizations and numerical schemes (Tribbia and Baumhefner 1988), and from the errors in the lateral boundary conditions if a regional model is used.

Ensemble prediction is a widely used method to deal with the prediction uncertainty and has great potential to improve weather prediction (Romine et al. 2014). In an ensemble prediction system, the prediction uncertainty from initial and lateral boundary condition errors can be represented by perturbations on the initial and lateral boundary conditions (e.g., Houtekamer and Derome 1995; Wang and Bishop 2003). To account for the prediction uncertainty from model errors, many different ensemble methods have been developed. Buizza et al. (1999) developed an ensemble method called the stochastically perturbed parameterization tendencies (SPPT) scheme, which perturbs the tendency of physical

Responsible Editor: Silvia Trini Castelli.

✉ Jong-Jin Baik
jjbaik@snu.ac.kr

¹ Department of Atmospheric Sciences, Pusan National University, Busan 46241, South Korea

² Institute for Future Earth, Pusan National University, Busan 46241, South Korea

³ School of Earth and Environmental Sciences, Seoul National University, Seoul 08826, South Korea

processes by multiplying random noise that has spatiotemporal correlations. This method considers the limitation of physical parameterizations that only grid-mean properties are used to calculate the physics tendencies and their uncertainties by unresolved subgrid-scale processes are not considered. Another ensemble method using stochastic perturbation, the stochastic kinetic energy backscatter (SKEB) scheme (Mason and Thomson 1992; Shutts 2005), perturbs the streamfunction and potential temperature tendencies to represent the uncertain effects of scale interactions between unresolved and resolved processes.

Besides the stochastic perturbation methods, another popular branch of ensemble prediction that represents model errors is the multi-physics ensemble method (e.g., Stensrud et al. 2000; Meng and Zhang 2007; García-Ortega et al. 2017; Gaudet et al. 2021). In this method, each ensemble member is assigned a different combination of physics parameterization schemes. For example, multiple cumulus parameterization schemes, planetary boundary layer (PBL) schemes, and cloud microphysics schemes can be used for the multi-physics ensemble prediction of precipitation. Using several different parameterizations for the same physical process, the uncertainty in the representation of the physical process can be taken into account.

Multi-physics ensembles have some known theoretical limitations that may affect the ensemble performance, such as the clustering of ensemble members and their distinguishability due to systematic similarities and differences among physics parameterization schemes. They also involve some practical challenges, such as substantial efforts required to establish a multi-physics ensemble system and potential compatibility issues between selected physics parameterization schemes. Despite these drawbacks, multi-physics ensembles can introduce significant diversity into the ensemble prediction, given that ensembles accounting only for the initial and lateral boundary condition errors are often under-dispersive and insufficient in representing the prediction uncertainty (Buizza et al. 2005; Romine et al. 2014). This increased diversity in ensemble prediction can contribute to improvement of prediction performance (Jankov et al. 2017).

In precipitation forecasting, ensemble prediction can provide not only probabilistic forecasts but also ensemble-derived representative forecasts that often outperform single-member forecasts. A commonly used approach is the ensemble mean. When applied to discontinuous fields such as short-term accumulated precipitation amount, the ensemble mean may overly smooth out precipitation features, expanding the overall area of precipitation and diminishing precipitation maxima (e.g., Du and Zhou 2011; Ancell 2013). While several alternative methods have been proposed to address this limitation, including probability matching (Ebert 2001; Clark 2017), the ensemble mean still remains one of the most straightforward

methods. This study targets the performance of ensemble mean in predicting summertime heavy precipitation in South Korea, which is typically long-lasting with a wide spatial distribution, conditions under which fair performance of ensemble mean may be expected.

The superiority of the ensemble mean over individual ensemble members has been experienced by many studies (e.g., Du et al. 1997; Toth and Kalnay 1997). Surcel et al. (2014) stated that a major reason for the superiority of the ensemble mean is that the unreliable features on which many of the ensemble members disagree are filtered out while taking the ensemble mean. Christiansen (2018) theoretically showed that when the observation and the ensemble members are drawn independently from the same distribution, the ensemble mean outperforms individual ensemble members in terms of distance measures in high-dimensional spaces (e.g., the root-mean-square error over numerous grid points) and the error of the ensemble mean is almost always 30% smaller than the median error of individual ensemble members. This was attributed to the non-intuitive geometric properties of high-dimensional spaces. Christiansen (2019) extended the assumption of the previous work to state that the ensemble members are drawn independently from the same distribution but the observation is not, allowing for the possibility of bias between the entire ensemble and the observation. He derived an analytical expression for the rank of the ensemble mean in comparison with ensemble members and demonstrated the superiority of the ensemble mean.

For multi-physics ensembles, it is not likely that the ensemble members are drawn independently from the same distribution, because the ensemble members in a multi-physics ensemble are systematically different from each other (Xu et al. 2020). Therefore, the superiority of the ensemble mean may not be guaranteed, which calls for evaluation studies on the ensemble-mean performance of multi-physics ensembles that provide evidence to determine if a multi-physics ensemble could be a promising method to improve particular forecasts in particular regions. This study aims to evaluate the ensemble-mean performance of a multi-physics ensemble on the prediction of heavy precipitation cases in South Korea, which is expected to contribute to the regional precipitation forecast.

This paper is organized as follows: the configuration of the multi-physics ensemble simulations in this study and the evaluation metrics are introduced in Sect. 2. In Sect. 3, the performance of the multi-physics ensemble mean on precipitation prediction is evaluated, in comparison with individual ensemble members and subset ensembles. Section 4 describes how the ensemble spread in the multi-physics ensemble is generated. Summary and conclusions are given in Sect. 5.

2 Methodology

2.1 Multi-physics parameterization scheme suite

The multi-physics parameterization scheme suite used for the ensemble consists of three cloud microphysics schemes, three PBL schemes, and three radiation schemes available in the Weather Research and Forecasting (WRF) model (Skamarock et al. 2019). One theoretical disadvantage of multi-physics ensembles is that physics parameterization schemes do not compose a consistent distribution: they are often closely related to each other (Jankov et al. 2017). Although this is an inherent limitation of multi-physics ensembles and cannot be completely avoided, we carefully selected physics parameterization schemes that have distinctive features from each other and that are not expected to significantly underperform than others to partly mitigate this limitation. It is noted that previous studies on the simulations of summertime precipitation associated with the East Asian summer monsoon have identified different schemes as performing best, depending on the study (e.g., Song and Sohn 2018; Guo et al. 2019; Jo et al. 2023). Therefore, there can be many reasonable alternatives for this multi-physics parameterization scheme suite. For the cloud microphysics schemes, the predicted particle properties (P3) scheme (Morrison and Milbrandt 2015), the Thompson aerosol-aware scheme (Thompson and Eidhammer 2014), and the WRF double moment 6-class (WDM6) scheme (Lim and Hong 2010) are considered, and hereafter they are called MP, MT, and MW, respectively. The three schemes substantially differ from each other in the assumptions of particle size distributions and terminal velocity relations as well as in the parameterizations of each microphysical process. It is also noted that MP stands apart from the other schemes by representing ice hydrometeors without predefined categories and allowing free evolution of their several physical properties.

There are various other cloud microphysics schemes available in the WRF model, but many of them have a large similarity with one of the selected schemes and are therefore excluded to alleviate clustering of ensemble members. For example, MW shares many parameters and formulations with the WRF single moment and double moment schemes with different number of hydrometeor classes (e.g., WSM3, WSM5, WSM6, WSM7, WDM5, WDM7). MP shares its warm microphysics parameterizations with the Morrison scheme (Morrison et al. 2005) and the ice-spheroids habit model with aspect-ratio evolution (ISHMAEL) scheme (Jensen et al. 2017). Some schemes such as the National Severe Storms Laboratory (NSSL) scheme (Mansell et al. 2010) do not share many features

with the selected schemes, but they are not included in the ensemble due to limitations in computational resources.

For the PBL schemes, the Yonsei University (YSU) scheme (Hong et al. 2006), the Mellor–Yamada–Nakanishi–Niino (MYNN) level 2.5 scheme (Nakanishi and Niino 2006), and the University of Washington (UW) scheme (Bretherton and Park 2009) are considered, and hereafter they are called PY, PM, and PU, respectively. For each PBL scheme, a surface layer scheme is paired. The revised MM5 surface layer scheme (Jiménez et al. 2012) is paired with PY and PU, and the MYNN surface layer scheme is paired with PM. PY is a nonlocal scheme that employs the K-profile method, while PM and PU are local schemes with turbulent kinetic energy closure. PM considers an improved mixing length scale based on large-eddy simulation data, and PU employs an explicit entrainment parameterization at the PBL top. For the radiation schemes (for both shortwave and longwave radiation), the rapid radiative transfer model for general circulation models (RRTMG) scheme (Iacono et al. 2008), the Community Atmosphere Model (CAM) scheme (Collins et al. 2004), and the new Goddard scheme (Chou and Suarez 1999; Chou et al. 2001) are considered, and hereafter they are called RR, RC, and RG, respectively. The three schemes exhibit differences in the division of shortwave spectrum, the treatment of optical properties, and the prescribed compositions of ozone, trace gases, and aerosols.

The compatibility among the physics parameterization schemes used in this study is examined, and it is confirmed that, in all scheme combinations, no required input variable is missing for the operation of any scheme. Some schemes are able to better represent interactions between different physical processes by more effectively utilizing the output from other schemes. For example, RR and RG have an advantage over RC in representing cloud-radiation interaction, as they utilize radiative effective radii of hydrometeors provided by cloud microphysics schemes rather than using prescribed values. Note that all three cloud microphysics schemes used in this study provide radiative effective radii that are computed consistently with their respective microphysical assumptions.

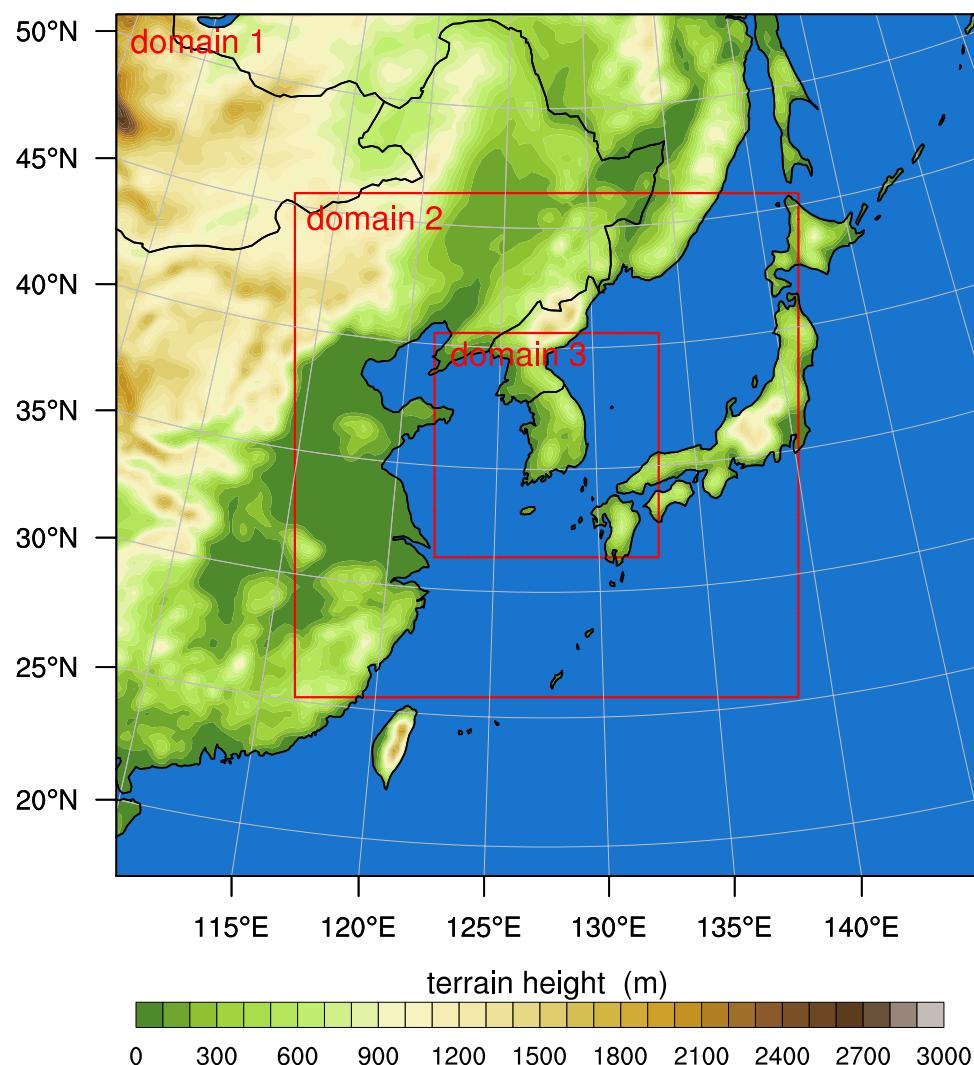
2.2 Experimental design

Multi-physics ensemble simulations are performed for ten heavy precipitation cases in South Korea with the WRF model version 4.3.3. The multi-physics ensemble consists of 27 multi-physics ensemble members, each corresponding to a unique combination of cloud microphysics, PBL, and radiation schemes. As described in subsect. 2.1, three cloud microphysics schemes, three PBL schemes, and three radiation schemes are selected, leading to a total of 27 multi-physics ensemble members. Each multi-physics ensemble member is obtained by averaging the results

(e.g., precipitation amount) of three simulations which are initialized using three slightly different initial conditions. The first initial condition is taken directly from a reanalysis dataset (introduced below). The other two initial conditions are generated by adding random perturbations within $[-0.3, 0.3 \text{ K}]$ on potential temperature to every column in the first initial condition. These three initial conditions are shared by all 27 multi-physics ensemble members for each case so that the differences among multi-physics ensemble members are solely attributed to the different choices of physics parameterization schemes. As a result, 81 simulations (27 multi-physics ensemble members \times 3 initial conditions per member) are conducted for each case, and a total of 810 simulations are conducted for the ten cases selected in this study. Hereafter, the term “multi-physics ensemble member” is simply referred to as “ensemble member”.

The simulations are initialized at 12 LST on the previous day and run for 36 h until 24 LST of the target day. The first 12 h is regarded as the spin-up time and the simulated 24-h accumulated precipitation amount on the target day is mainly analyzed. Three nested domains where the horizontal grid spacing is 27, 9, and 3 km and the number of horizontal grids is 144×144 , 252×252 , and 336×336 , respectively, are used for the simulations (Fig. 1). Vertically 44 layers whose depth increases with height from 50 m up to 800 m are considered. The model top is 50 hPa. The Kain–Fritsch cumulus parameterization scheme (Kain 2004) and the unified Noah land surface model (Tewari et al. 2004) are used for all simulations. The cumulus parameterization is turned off for the innermost domain. For the boundary conditions for the outermost domain and the initial conditions, the European Centre for Medium-range Weather Forecasts Reanalysis version 5 (Hersbach et al. 2020) with a temporal resolution of 1 h and a spatial resolution of $0.25^\circ \times 0.25^\circ$ is used.

Fig. 1 Three nested domains with terrain height



2.3 Case description

The ten precipitation cases simulated in this study correspond to the ten different days with the highest daily precipitation amounts in South Korea during June, July, and August of 2020. Here, the daily precipitation amount is the 24-h accumulated precipitation amount from 00 to 24 LST. The maximum observed daily precipitation amount exceeds 130 mm for all ten cases, which is sufficient to classify them as heavy precipitation cases. The dates of the ten cases are given in Table 1. The observed daily precipitation amount averaged over 549 rain gauge points in South Korea operated by the Korea Meteorological Administration ranges from 30.9 to 72.2 mm. Figure 2 shows the synoptic condition of each case at 12 LST. Cases 2, 4, 5, and 6, which occurred during late June–late July, are caused by developing extratropical cyclones with upper-level troughs located at their west, indicating baroclinicity. The extratropical cyclones are accompanied by strong midlevel upward motions on their front side and active low-level moisture transport at their southeast and move eastward, inducing heavy precipitation while passing South Korea. Cases 1, 3, 7, 8, 9, and 10, which occurred during late July–mid August, are caused by quasi-stationary monsoon rainbands formed along the path of strong low-level moisture transport at the boundary of the western North Pacific subtropical high. In most of these cases, the low-level jet over South Korea is coupled with an upper-level jet at its north, enhancing upward motions at the south of the upper-level jet entrance through a secondary circulation (Shin and Lee 2005). The upward motions are relatively weak compared to those in the cases associated with extratropical cyclones, but the quasi-stationary monsoon rainbands stay in South Korea for a long time, accumulating a large amount of precipitation. More detailed synoptic analysis on the heavy precipitation cases in 2020 summer in South Korea can be found in Park et al. (2021a).

Table 1 Date, station-averaged observed and ensemble-mean 24-h accumulated precipitation amounts, correlation coefficient and RMSE calculated by comparing the ensemble-mean and observed 24-h accumulated precipitation amounts, and ensemble spread for each case

Case	Date	Observed 24-h precipitation amount (mm)	Ensemble-mean 24-h precipitation amount (mm)	R of ensemble mean	RMSE of ensemble mean (mm)	Ensemble spread (mm)
1	08/08	55.3	51.7	0.78	39.0	43.7
2	07/23	67.5	58.8	0.58	32.6	24.6
3	08/07	38.8	36.4	0.60	46.4	29.0
4	07/10	34.8	47.2	0.65	39.2	26.2
5	07/13	72.2	73.5	0.86	29.9	19.6
6	06/29	39.0	46.5	0.77	22.9	13.6
7	08/10	32.5	23.8	0.32	29.2	15.1
8	07/29	30.9	19.2	0.37	32.9	16.1
9	08/11	34.0	19.3	0.42	33.8	16.2
10	08/06	36.3	27.6	0.37	33.0	13.6

All dates refer to the year 2020

2.4 Evaluation metrics

Partly considering random errors from initial condition uncertainty, each member of multi-physics ensemble, which corresponds to a unique combination of physics parameterization schemes, is obtained by averaging the three different simulations initialized using three different initial conditions. The precipitation predictions of the obtained multi-physics ensemble members and their mean are evaluated in comparison with the rain gauge data observed at 549 rain gauge points in South Korea. As evaluation metrics, the root-mean-square error (RMSE) and the Pearson correlation coefficient (R) are used, which are given as

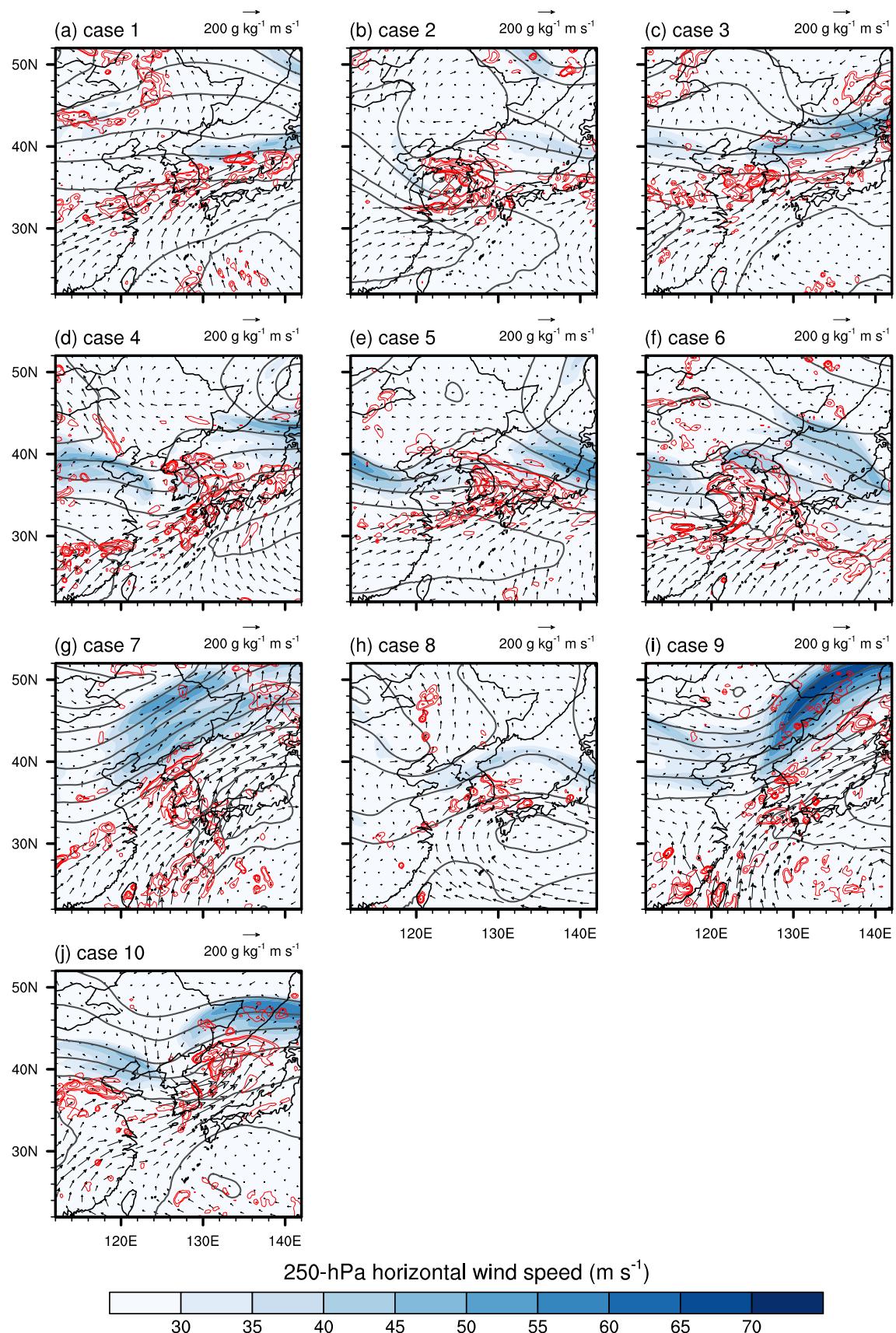
$$\text{RMSE} = \sqrt{\frac{1}{M} \sum_{m=1}^M (f_m - o_m)^2}, \quad (1)$$

$$R = \frac{\sum_{m=1}^M (f_m - \mu_f)(o_m - \mu_o)}{\sqrt{\sum_{m=1}^M (f_m - \mu_f)^2} \sqrt{\sum_{m=1}^M (o_m - \mu_o)^2}}, \quad (2)$$

where M is the number of rain gauge points used for evaluation, m is the rain gauge point index, f is the forecast value, and o is the observation value. μ_f and μ_o indicate the spatial means of the forecast and observation values, respectively. Observational errors are not considered in this study.

The ensemble spread at a point s_m is represented by the sample standard deviation as

$$s_m = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (f_{m,n} - \bar{f}_m)^2}, \quad (3)$$



◀Fig. 2 Synoptic fields at 12 LST for the ten cases. The arrows represent 850-hPa water vapor flux; the red contours indicate 500-hPa upward velocity (starting from -0.25 Pa s^{-1} and doubling at each interval); the grey contours indicate 250-hPa geopotential height (in 80-m intervals); the blue shades represent 250-hPa horizontal wind speed

where N is the number of ensemble members, n is the ensemble member index, and the overbar refers to the ensemble mean. The ensemble spread of an ensemble prediction s is calculated by first averaging the squared ensemble spread over rain gauge points and then taking the square root, following the method of Fortin et al. (2014):

$$s = \sqrt{\frac{1}{M} \sum_{m=1}^M \frac{1}{N-1} \sum_{n=1}^N (f_{m,n} - \bar{f}_m)^2}. \quad (4)$$

Note that all the ensemble spreads given in this study indicate the multi-physics ensemble spread, which means that $f_{m,n}$ in Eqs. (3) and (4) is the forecast value of a single multi-physics ensemble member (the average of three simulations with different initial conditions).

3 Performance of the multi-physics ensemble

3.1 Ensemble prediction of the ten precipitation cases

The comparison of the ensemble-mean prediction of 24-h accumulated precipitation amount with the observation for the ten precipitation cases as well as the ensemble spread is presented in Figs. 3 and 4. For cases 1–7, the observed precipitation pattern for each case is overall well represented by the ensemble mean. For each of these cases, the ensemble mean successfully represents the observed heavy precipitation area although it is slightly mislocated and the precipitation amount is to some extent over or underestimated. For example, for case 1, the west–east oriented band-like pattern of precipitation over the southern part of South Korea is well represented by the ensemble mean, but the precipitation in the western (eastern) part of the band is underestimated (overestimated) by the ensemble mean and the western part of the band in the ensemble mean expands further north compared to the observation. For cases 8–10, the ensemble mean shows a relatively poor performance in reproducing the observed precipitation pattern. For example, for case 8, the ensemble mean fails to reproduce the precipitation in the southwestern part of South Korea, and for case 10, it totally misses the main precipitation area in the northern part of South Korea and exaggerates the precipitation in the

southwestern part of the country. The ensemble spread is overall relatively large for cases 1–4, and large ensemble spread appears within the observed main precipitation area for these cases. This indicates that for cases 1–4, there is large uncertainty stemming from physical parameterizations and they play important roles in precipitation prediction, which implies that multi-physics ensemble prediction could be especially useful for these cases.

The performance of ensemble mean in comparison with the observation and the ensemble spread for each case are summarized in Table 1. The ensemble mean overestimates the station-averaged 24-h accumulated precipitation amount for three cases (cases 4–6) and underestimates it for the other seven cases. Cases 1–7, for which the observed precipitation pattern is overall well represented by the ensemble mean (Figs. 3 and 4), show moderate or high correlation coefficients between the ensemble mean and the observation except for case 7, which shows a markedly low correlation coefficient. Cases 8–10 also show low correlation coefficients. In terms of the ensemble spread, cases 1–4 exhibit relatively large ensemble spreads compared to cases 5–10.

Comparison between the RMSE of ensemble mean and the ensemble spread shows whether the prediction uncertainty is appropriately represented by the ensemble. Figure 5 shows the ratio of the ensemble spread to the RMSE of ensemble mean (hereafter, the spread-error ratio). A spread-error ratio close to the unity suggests that the ensemble spread well represents the prediction uncertainty. A spread-error ratio significantly smaller than the unity suggests that the ensemble is underdispersive, meaning that the ensemble is too confident and does not sufficiently account for model errors. On the other hand, a spread-error ratio significantly larger than the unity suggests that the ensemble is overdispersive, which means that the ensemble is too uncertain about its prediction, potentially reducing the usefulness of prediction.

According to the spread-error ratio, case 1 shows the best agreement between the RMSE and ensemble spread among the ten cases, followed by case 2. For nine of the ten cases (cases 2–10), the ensemble is underdispersive, indicating that the prediction uncertainty is only partly represented by the ensemble. The overall underdispersiveness in short-time forecasts has also been reported for other types of ensemble systems as well, such as ensemble predictions using initial condition perturbations, the SPPT scheme, the SKEB scheme, and some combinations of these (e.g., Berner et al. 2011; Jankov et al. 2017).

From Figs. 3, 4, and 5 and Table 1, the ten precipitation cases can be characterized as follows: Cases 1 and 2 are characterized by relatively large ensemble spreads which are located within the observed main precipitation area and better representation of prediction uncertainty by the ensemble compared to the other cases. Cases 3 and 4

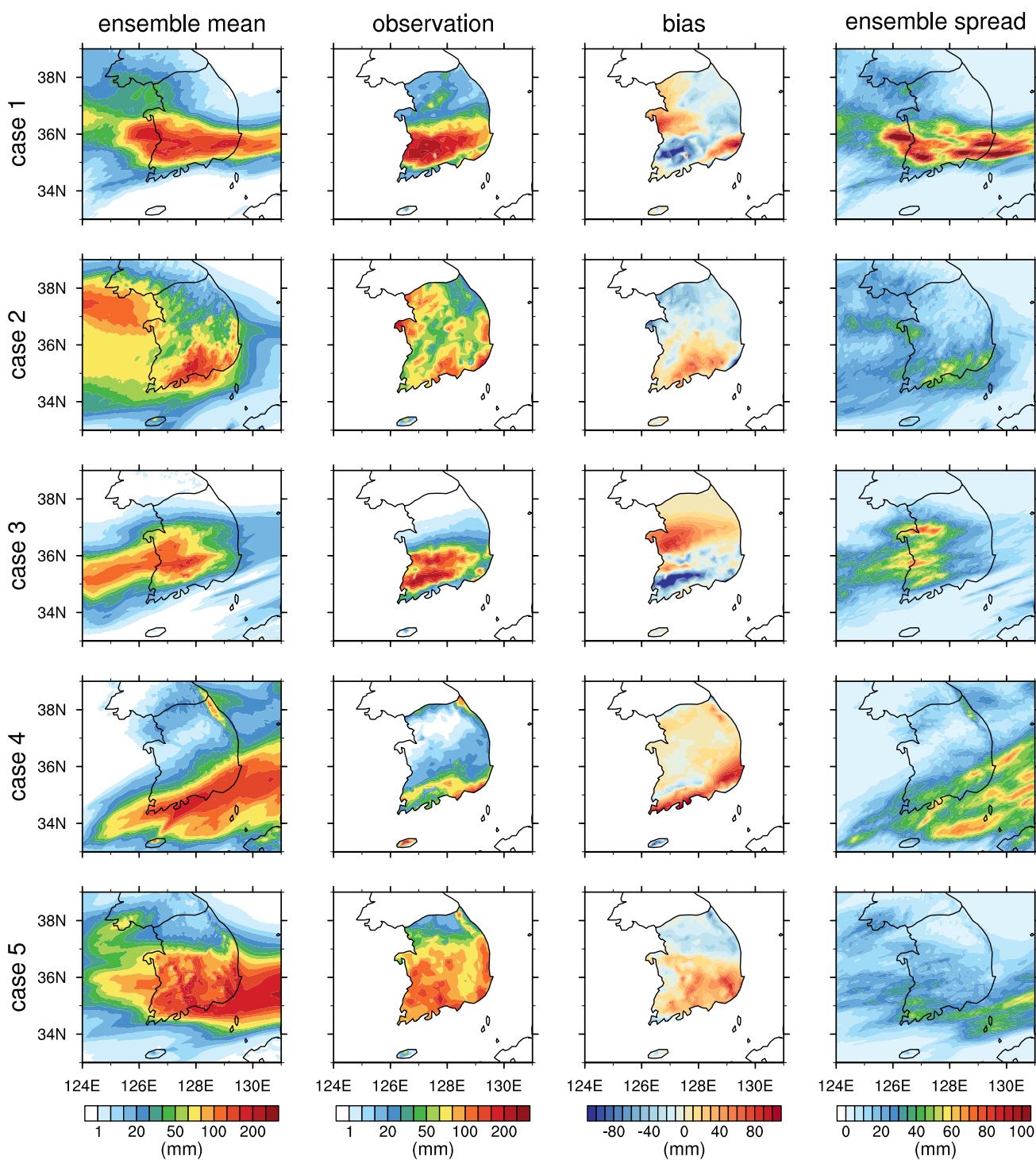


Fig. 3 Ensemble-mean (first column) and observed (second column) 24-h accumulated precipitation amounts, the bias of the ensemble mean from the observation (third column), and the ensemble spread (fourth column) for cases 1–5

also show relatively large ensemble spreads located within the observed main precipitation area, but the ensemble is more underdispersive compared to case 2. Cases 5 and 6 are characterized by relatively small ensemble spreads, underdispersiveness of the ensemble, and high correlation

coefficients between the ensemble mean and observation, which indicates that the precipitation is well predicted for these cases but the use of multi-physics ensemble makes only limited contribution to this. Cases 7–10 are characterized by relatively small ensemble spreads, even stronger

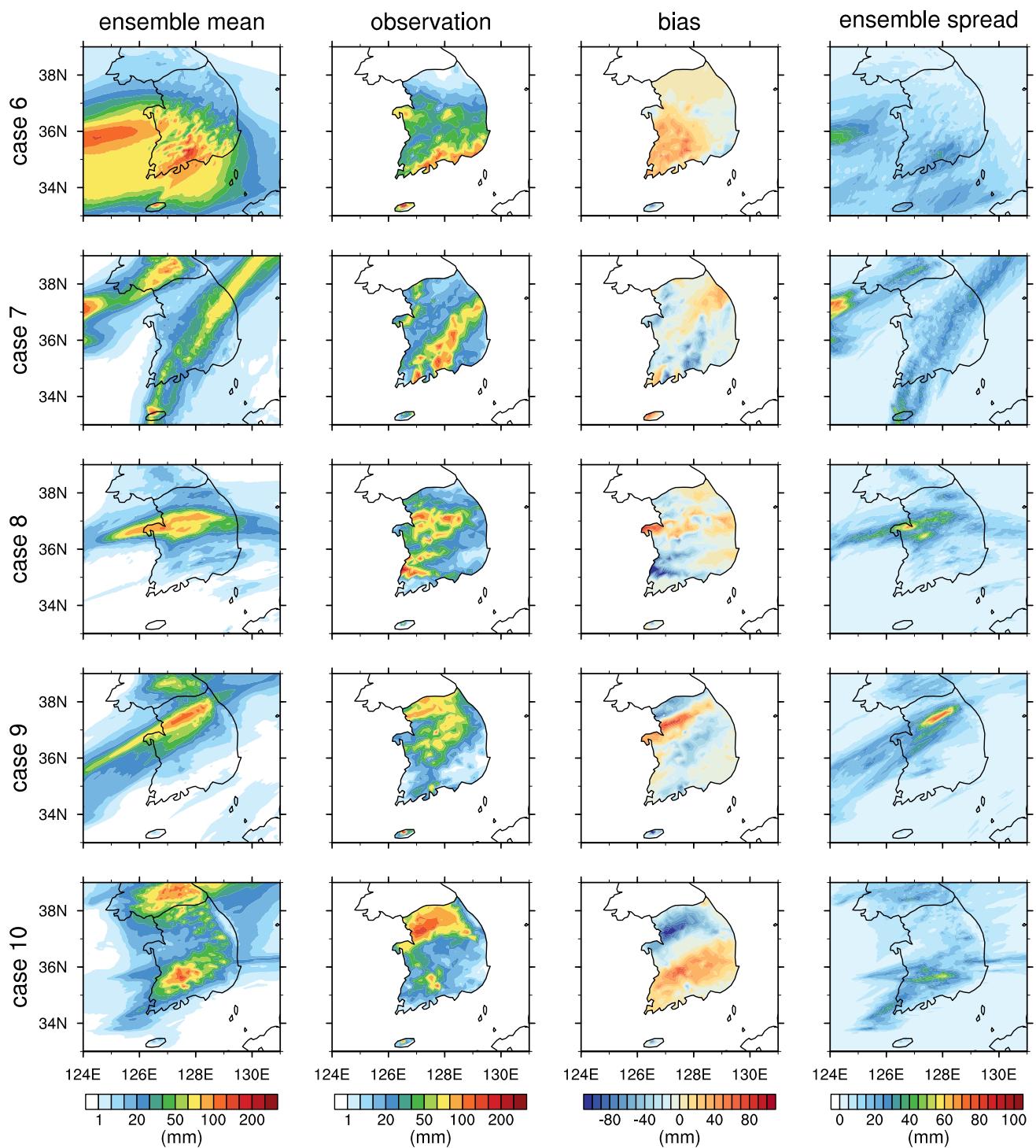


Fig. 4 As Fig. 3 but for cases 6–10

underdispersiveness of the ensemble, and low correlation coefficients between the ensemble mean and observation, which indicates that the precipitation prediction is unsuccessful and that a large part of the prediction uncertainty for these cases may have originated from factors other than the physics parameterizations. For these cases where the

multi-physics ensemble represents only a limited portion of the prediction uncertainty, uncertainties in initial and boundary conditions and model dynamics, along with uncertainties from unresolved features of the system, could account for a large portion of the total prediction

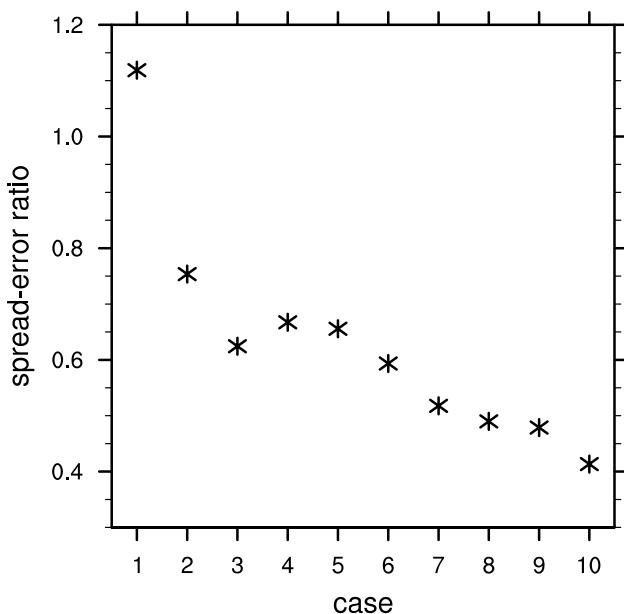


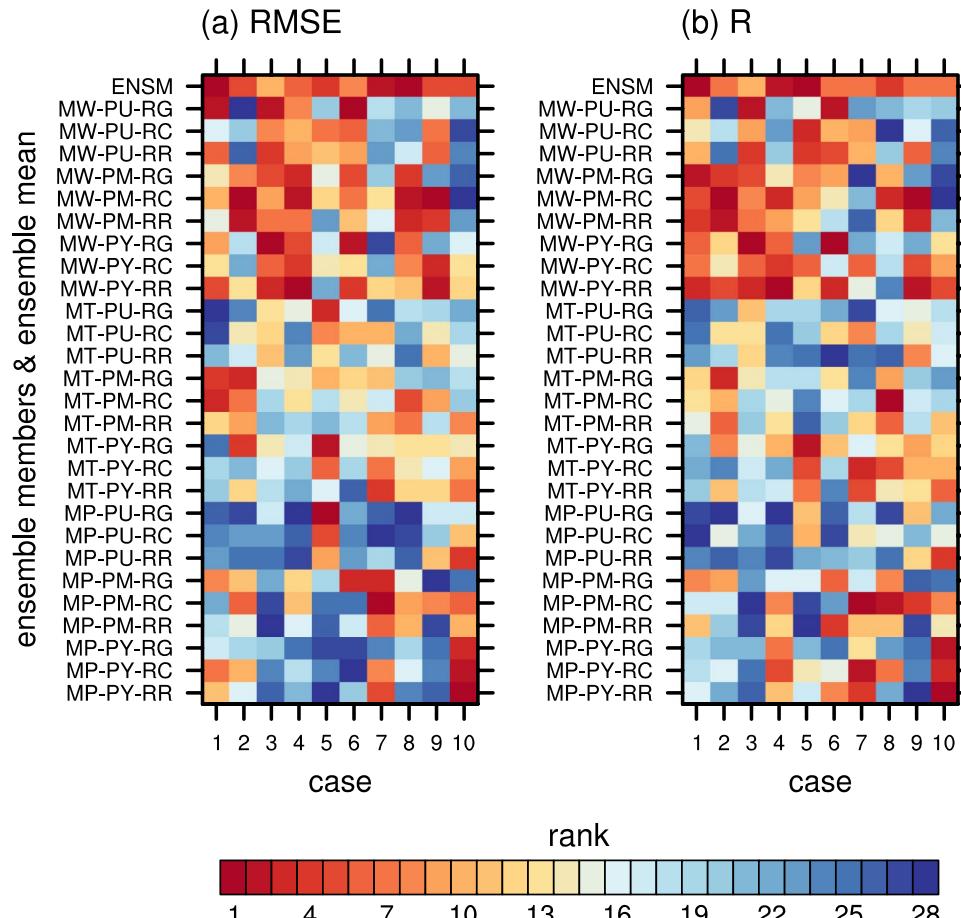
Fig. 5 Ratio of the ensemble spread to the RMSE of ensemble mean for the ten cases

uncertainty, which may be better addressed by using other types of ensemble systems.

3.2 Evaluation of the ensemble mean in comparison with individual ensemble members

The performance of the multi-physics ensemble mean is evaluated in comparison with those of individual multi-physics ensemble members. Total 28 forecasts, that is, the 27 ensemble members and the ensemble mean, are ranked by the RMSE and correlation coefficient calculated by comparing the forecasted and observed 24-h accumulated precipitation amount (Fig. 6). Note that for RMSE, the highest rank (i.e., the first place) corresponds to the smallest value, whereas for correlation coefficient, it corresponds to the largest value. In terms of the RMSE, the ensemble mean shows better overall performance than any of the individual ensemble members. The rank of the ensemble mean is the first place for two cases, within the top five for seven cases, and within the top ten for all cases. Paired t-tests indicate that the RMSE of the ensemble mean is statistically significantly smaller than that of 18 individual ensemble members as well as the average RMSE of 27 individual ensemble members. No individual

Fig. 6 Ranks of each of the ensemble members (MX-PX-RX) and the ensemble mean (ENSM) in terms of the **a** RMSE and **b** correlation coefficient obtained by comparing the forecasted and observed 24-h accumulated precipitation amounts for the ten cases



ensemble member consistently shows good performance (having a rank higher than the median rank) for every case. For example, MW-PM-RC, which shows the overall best performance among the individual ensemble members, has a rank better than the median rank for nine cases but takes the very last place for case 10. Compared to this best-performing individual ensemble member, the ensemble mean shows smaller RMSE for six cases. It is noteworthy that the superiority of the ensemble mean over the individual ensemble members does not seem to significantly depend on the magnitude of the ensemble spread, the spread-error ratio, and the correlation coefficient between the ensemble mean and the observation because the ensemble mean shows a high rank not only for cases 1 and 2 but also for cases 7–10. There does exist some variations in the rank of the ensemble mean, but it is rather associated with the difference in performance among individual physics parameterization schemes for each case. In terms of the correlation coefficient, the rank of the ensemble mean is the first place for two cases, within the top five for four cases, and within the top ten for all cases (Fig. 6b). As in Fig. 6a, only the ensemble mean shows a rank better than the median rank for all cases.

For each case, the performances of individual physics parameterization schemes are compared to each other by means of box plots in Fig. 7. Each box (including the whiskers) for a physics parameterization scheme consists of nine multi-physics ensemble members that use the scheme. The boxes for different schemes largely overlap with each other for most cases, which indicates that the impact of a single scheme on the prediction performance is usually limited. Overall, the cloud microphysics schemes exhibit narrower overlaps than the PBL and radiation schemes, indicating that they tend to have a greater control over precipitation prediction performance. For some cases (e.g., cases 3 and 4), the boxes for different cloud microphysics schemes are notably separated, suggesting that the choice of cloud microphysics scheme can sometimes play a decisive role in determining precipitation prediction performance.

For cases where the ensemble mean takes the top rank (cases 1 and 8), there is no individual physics parameterization scheme that performs overwhelmingly better than the others. In contrast, for cases where the rank of the ensemble mean is not within the top five (cases 3, 4, and 6), there is a physics parameterization scheme that performs substantially better than the others for each case. For example, for cases 3 and 4, nine out of the best ten forecasts are done by ensemble members with MW. For case 6, seven out of the best ten forecasts are done by ensemble members with MW. This indicates that for a case where a physics parameterization scheme significantly outperforms other schemes, the multi-physics ensemble may not provide a satisfactory ensemble-mean forecast.

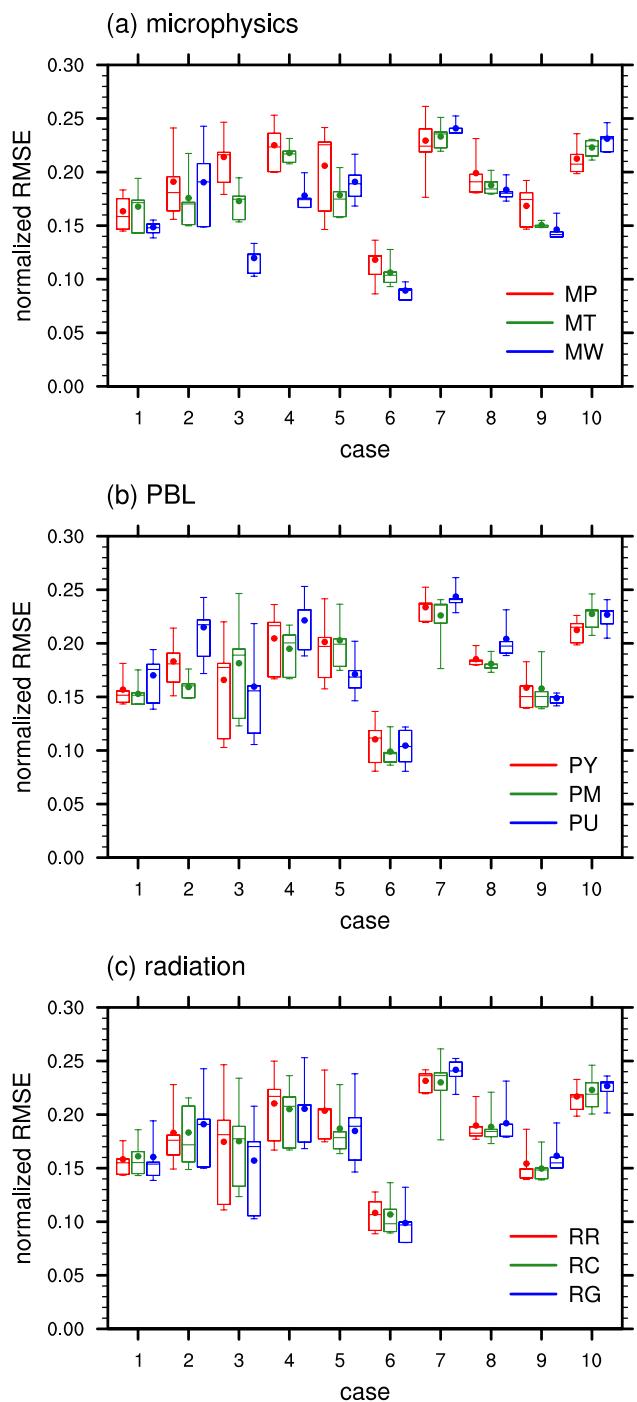


Fig. 7 Box plots of RMSE normalized by the observed maximum precipitation amount for the ensemble members with **a** cloud microphysics, **b** PBL, or **c** radiation scheme for the ten cases. Each box consists of nine ensemble members. The lower boundary, middle line, and upper boundary of each box indicate the lower quartile, median, and upper quartile, respectively. The lower and upper whiskers indicate the minimum and maximum, respectively, and the dot indicates the mean

In Fig. 6, it is shown that for each case, there exist a few ensemble members that perform as successfully as the ensemble mean. However, it is almost impossible to identify the best-performing combination of physics parameterization schemes for a case in advance of simulating the case because the best-performing combination completely changes with the case. For this reason, it is assumed that the typical performance of a single-member prediction can be represented by the average RMSE of individual ensemble members, and the expected improvement of performance brought by the ensemble-mean prediction in comparison with the typical performance of a single-member prediction is evaluated. Figure 8 shows the relative decrease in RMSE from the average RMSE of individual ensemble members to the ensemble-mean RMSE (i.e., the RMSE of the ensemble mean) for each case and its relation to the spread-error ratio and the ensemble spread. For all cases, the ensemble-mean RMSE is smaller than the average RMSE of individual ensemble members, which shows the superiority of the ensemble-mean prediction using the multi-physics ensemble over a single-member prediction. The relative decrease in RMSE ranges from 7% for case 10 to 33% for case 1. The relative decrease in RMSE is almost perfectly correlated with the spread-error ratio ($R = 1.00$). In order to expect a greater improvement of prediction performance, the multi-physics ensemble should represent a large part of prediction uncertainty. This is not surprising because the following equality holds:

$$\begin{aligned} & \frac{1}{N} \sum_{n=1}^N \frac{1}{M} \sum_{m=1}^M (f_{m,n} - o_m)^2 \\ &= \frac{1}{M} \sum_{m=1}^M \frac{1}{N} \sum_{n=1}^N (f_{m,n} - \bar{f}_m)^2 \\ &+ \frac{1}{M} \sum_{m=1}^M (\bar{f}_m - o_m)^2. \end{aligned} \quad (5)$$

The left-hand side is the mean squared error averaged over the individual ensemble members, the first term in the right-hand side is equal to the squared ensemble spread multiplied by $(N - 1)/N$, and the second term in the right-hand side is the mean squared error of the ensemble mean. Given the same average RMSE of individual ensemble members, a larger ensemble spread yields a smaller ensemble-mean RMSE. The relative decrease in RMSE is also strongly correlated with the ensemble spread ($R = 0.92$). For the heavy precipitation cases in South Korea, a larger ensemble spread is needed for a multi-physics ensemble prediction to be more effective.

3.3 Evaluation of the total ensemble mean in comparison with subset ensemble means

In subsect. 3.2, it was shown that there is no physics parameterization scheme that consistently performs well for every

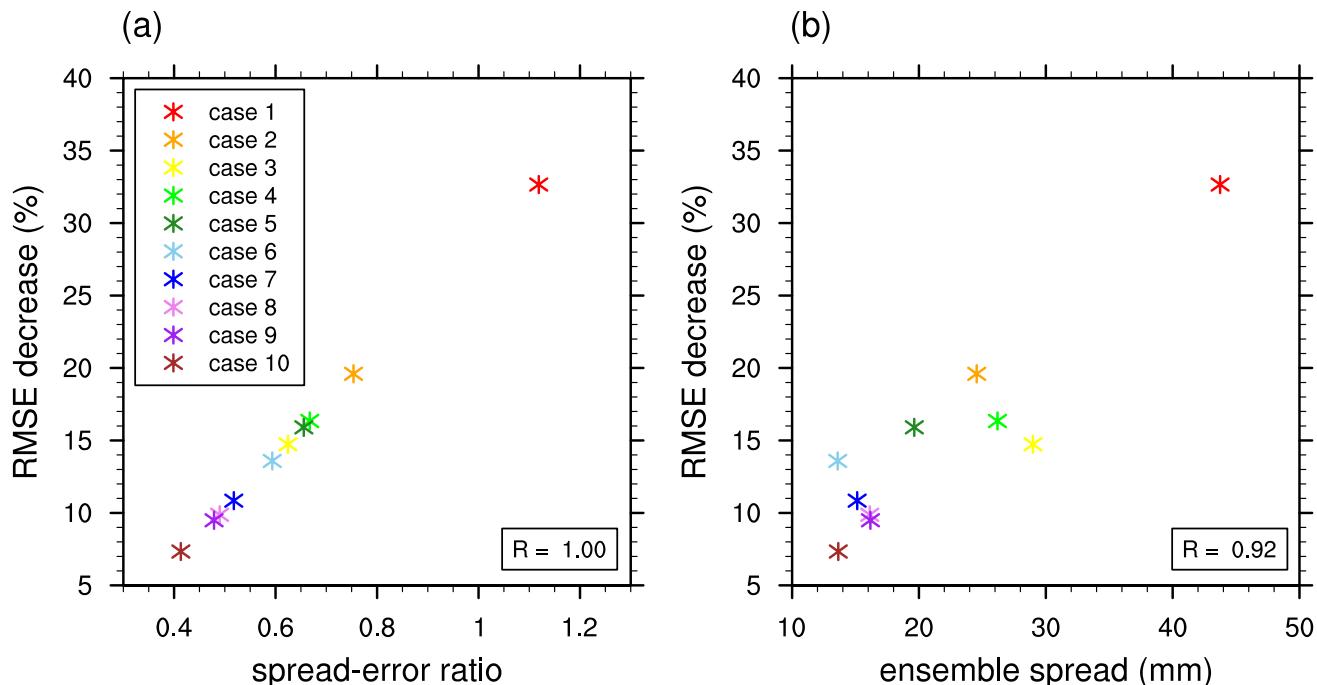


Fig. 8 **a** Scatter plot of the spread-error ratio and the relative decrease in RMSE from the average RMSE of individual ensemble members to the ensemble-mean RMSE. **b** Same as **a** but for the ensemble spread instead of the spread-error ratio

case. However, there do exist certain physics parameterization schemes that overall perform better than the others. For example, in Fig. 7, the ensemble members with MW (PM) shows a smaller mean RMSE than those with the other microphysics (PBL) schemes for six cases. On the contrary, there also exist certain physics parameterization schemes whose performances are not as good as the others. The ensemble members with MP (PU) shows a larger mean RMSE than those with the other microphysics (PBL) schemes for seven (five) cases.

The existence of a physics parameterization scheme that significantly outperforms the others tends to reduce the superiority of the multi-physics ensemble-mean prediction over the single-member prediction with a properly selected combination of physics parameterization schemes (Fig. 6). This gives a motivation to examine if the performance of the multi-physics ensemble mean is improved when only the ensemble members with an overall well-performing scheme are included in the ensemble or when the ensemble members with an overall poorly-performing scheme are excluded from the ensemble. Figure 9a shows the ranks of the total ensemble mean in terms of the RMSE in comparison with the subset ensemble means which are obtained using ensemble members that share the same microphysics scheme (e.g., MW subset ensemble) or those that use one of two specific microphysics schemes (e.g., MT/MW subset ensemble). The same is done in Figs. 9b, c but for PBL schemes and radiation schemes, respectively. The performance of the subset ensembles with one microphysics scheme is overall relatively poor than that of the subset ensembles with two

microphysics schemes and that of the total ensemble. The MW subset ensemble, which is the overall best-performing subset ensemble among the subset ensembles with one microphysics scheme, takes the top rank for cases 3, 4, and 6, but it also takes the bottom rank for cases 2, 7, and 10, showing a large variation in its performance depending on the case. The subset ensembles with two microphysics schemes show relatively stable performance, never taking the bottom rank for any case. For any case, a subset ensemble with two microphysics schemes always shows RMSE that is either intermediate between those of the subset ensembles with one of the two microphysics schemes or smaller than both of those. For example, the MT/MW ensemble takes a higher rank than both of the MT and MW subset ensembles for cases 1, 5, 7, 8, and 9 and an intermediate rank between the MT and MW subset ensembles for cases 2, 3, 4, 6, and 10. This proposition also holds for PBL schemes (Fig. 9b) and for radiation schemes (Fig. 9c).

The total ensemble shows the most stable performance, taking a rank within the top four for all cases. The total ensemble of 27 members can be regarded as the combination of a subset ensemble of 18 members with two microphysics schemes (e.g., the MT/MW ensemble) and a subset ensemble of 9 members with one microphysics scheme (e.g., the MP subset ensemble). The MT/MW subset ensemble is a well-performing ensemble that takes a rank within the top two for four cases, but for cases 2, 7, and 10, it shows worse performance than the MP subset ensemble although MP is the overall worst-performing microphysics scheme among the three (Fig. 7). Including MP into the ensemble mitigates

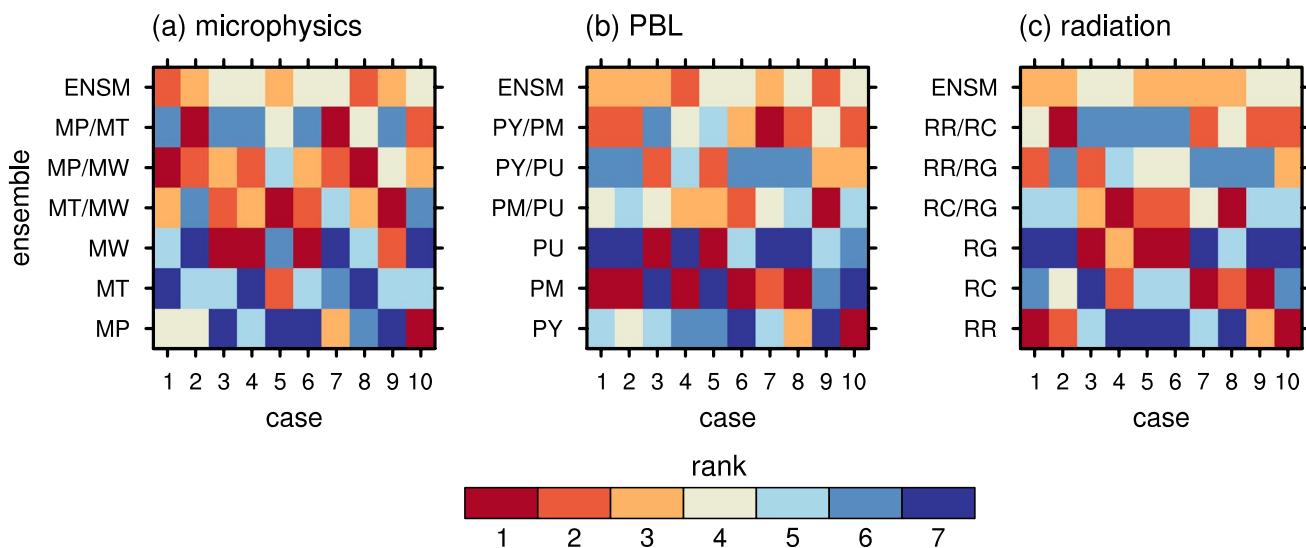


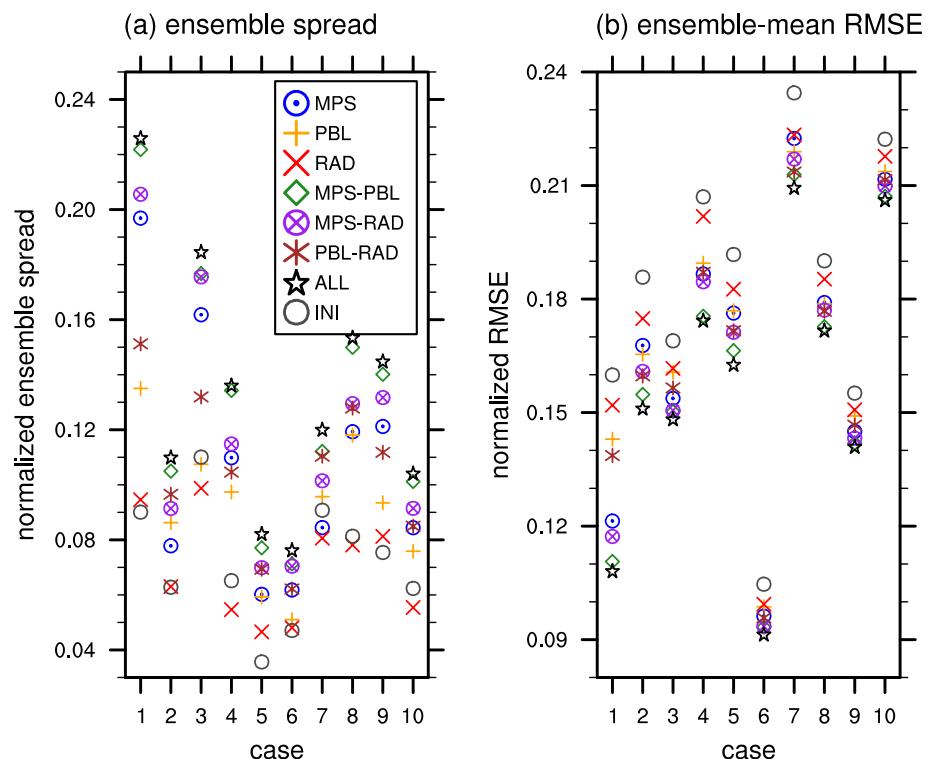
Fig. 9 Ranks of the total ensemble mean (ENSM; the mean of 27 members) and the subset ensemble means in terms of the RMSE for the 24-h accumulated precipitation amount for the ten cases. In **a**, the subset ensemble mean is obtained using ensemble members that share the same cloud microphysics scheme (MP, MT, and MW; each

is the mean of 9 members) or those that use one of two specific cloud microphysics schemes (MT/MW, MP/MW, and MP/MT; each is the mean of 18 members). The same is done in **b** but for PBL schemes and in **c** but for radiation schemes

the poor performance of the MT/MW subset ensemble for cases 2, 7, and 10. This implies that even an overall poorly-performing physics parameterization scheme can contribute to the performance of the ensemble-mean prediction, especially to the stability of the performance. On the other hand, the inclusion of MP into the MT/MW subset ensemble degrades the performance of the ensemble mean for cases 3, 4, 5, 6, and 9. Therefore, although MP does have some contributions to the performance of the total ensemble mean, the ensemble mean is expected to perform better if MP is replaced with an overall better-performing cloud microphysics scheme.

The above findings from microphysics schemes are similar to those from PBL schemes and radiation schemes. For the PBL schemes, the PU subset ensemble is the overall worst-performing one among the subset ensembles with one PBL schemes, and the PY/PM subset ensemble that excludes PU from the total ensemble shows overall good performance taking a rank within the top two for five cases. The PY/PM subset ensemble shows poor performance for cases 3 and 5, and the poor performance for the two cases is mitigated in the total ensemble by including PU. For the radiation schemes, the RC/RG subset ensemble that excludes RR from the total ensemble shows overall good performance taking a rank within the top two for four cases, but for cases 1, 2, 9, and 10, it shows poor performance. The poor performance for the four cases are mitigated in the total ensemble by including RR.

Fig. 10 **a** Ensemble spread and **b** ensemble-mean RMSE averaged over cloud microphysics ensembles (MPS), PBL ensembles (PBL), radiation ensembles (RAD), cloud microphysics–PBL ensembles (MPS–PBL), cloud microphysics–radiation ensembles (MPS–RAD), and PBL–radiation ensembles (PBL–RAD) and those for the total ensemble (ALL) for the ten cases. Those averaged for initial condition ensembles (INI) are also presented for comparison. For each case, the ensemble spread is normalized by the maximum precipitation amount in the ensemble mean and the ensemble-mean RMSE is normalized by the observed maximum precipitation amount



The multi-physics ensemble used in this study consists of three physics parameterization types, that is, the cloud microphysics parameterization, the PBL parameterization, and the radiation parameterization. The contribution of each physics parameterization type to the performance of the ensemble mean is examined by comparing the ensemble spread and ensemble-mean RMSE of subset ensembles with one or two physics parameterization types (Fig. 10). These subset ensembles and the total ensemble have different numbers of ensemble members according to the number of physics parameterization types included. Because the purpose of this analysis is to see how the ensemble spread brought by each physics parameterization type contributes to the performance of the total ensemble mean, the ensemble spread in Fig. 10a is calculated slightly differently from Eq. (4) by multiplying $[(N-1)/N]^{1/2}$ to Eq. (4). The modified ensemble spread is directly related to the ensemble-mean RMSE by Eq. (5). For comparison, the ensemble spread and ensemble-mean RMSE of the initial condition ensemble are also provided in Fig. 10. Among the subset ensembles with one physics parameterization type, the cloud microphysics ensemble shows the largest ensemble spread for eight cases while the radiation ensemble shows the smallest ensemble spread for all cases. The ensemble spread of the radiation ensemble is overall comparable to that of the initial condition ensemble. This suggests that in heavy precipitation prediction, the radiation schemes are likely to produce only a limited amount of ensemble spread, which is comparable in magnitude to the spread representing the randomness

stemming from initial condition uncertainty. The PBL ensemble shows the largest ensemble spread for two cases and the intermediate ensemble spread for eight cases. The ensemble spread of the cloud microphysics ensemble is even larger than that of the PBL–radiation ensemble for four cases, which reflects the importance of cloud micro-physics schemes in the ensemble prediction of precipitation. Among the subset ensembles with two physics parameterization types, the cloud microphysics–PBL ensemble shows the largest ensemble spread for all cases while the PBL–radiation ensemble shows the smallest ensemble spread for eight cases. For most cases, the ensemble spread of the cloud microphysics–PBL ensemble is close to that of the total ensemble, implying that radiation schemes do not contribute much to the ensemble spread of the total ensemble. The total ensemble shows the largest ensemble spread for all cases.

For the cases examined in this study, a relatively large ensemble spread leads to a relatively small ensemble-mean RMSE (Fig. 8b). For all cases except case 8, the order of the subset ensembles and total ensemble for the ensemble spread in Fig. 10a is exactly reversed for the ensemble-mean RMSE in Fig. 10b. Therefore, the cloud microphysics ensemble and the cloud microphysics–PBL ensemble show the best ensemble-mean performance among the subset ensembles with one and two physics parameterization types, respectively. The total ensemble shows the smallest ensemble-mean RMSE for all cases, and to this performance of the total ensemble mean, the contribution of the cloud microphysics parameterization is largest while that of the radiation parameterization is smallest.

4 Systematic generation of the ensemble spread in the multi-physics ensemble

In Sect. 3, it was shown that the performance of the multi-physics ensemble-mean prediction relative to a single-member prediction is strongly related to the ensemble spread. In this section, we examine how the ensemble spread of the multi-physics ensemble is generated. In the multi-physics ensemble, different physical parameterizations used in different ensemble members cause systematic differences in cloud and precipitation properties. Each ensemble member can have a tendency to predict certain precipitation properties to be larger or smaller than those predicted by other ensemble members, according to its unique way of representing physical processes. Because the unique representation of physical processes of each ensemble member stems from the combination of physics parameterization schemes, each physics parameterization scheme's tendency in the prediction of cloud and precipitation properties is examined.

Figure 11 shows the 24-h accumulated precipitation amount averaged over the ensemble members with each

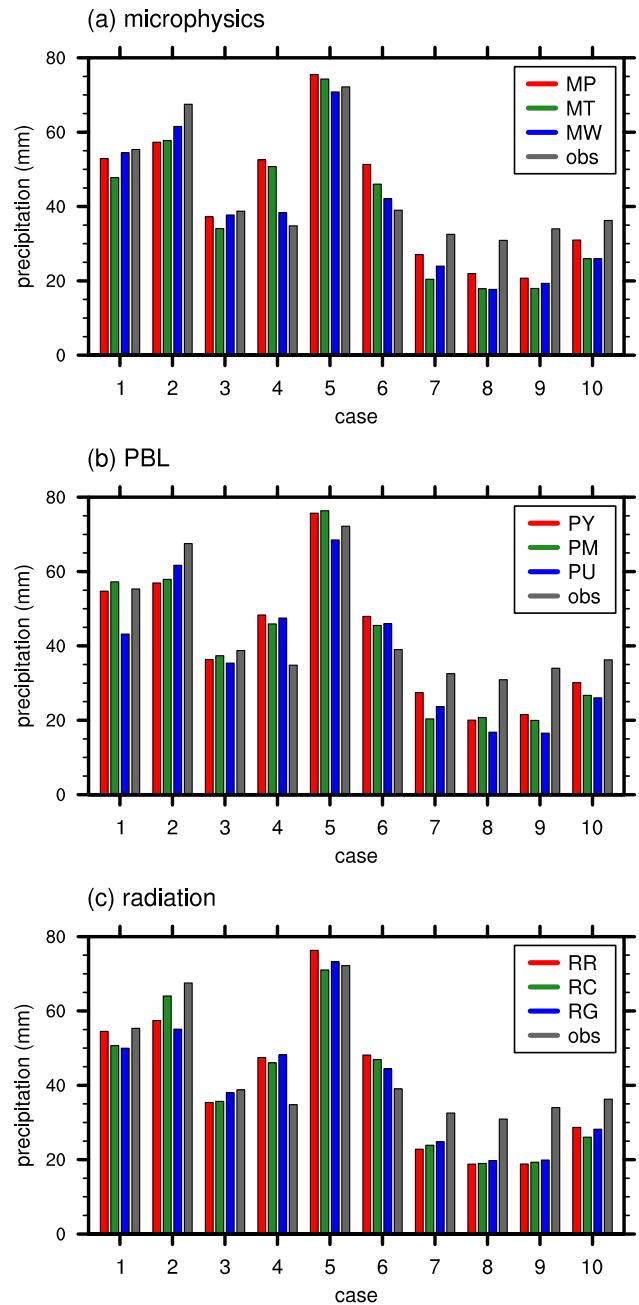


Fig. 11 24-h accumulated precipitation amount averaged over the ensemble members with each **a** cloud microphysics, **b** PBL, or **c** radiation scheme and the observation

physics parameterization scheme. The ensemble members with MP tend to show a larger accumulated precipitation amount than those with MT (for nine cases) and those with MW (for seven cases). However, it seems that this tendency does not induce a tendency of the cloud microphysics schemes in the prediction accuracy. The cloud microphysics scheme which gives the accumulated precipitation amount that is closest to the observation is MW for cases 1–6, but for cases 1–3 the ensemble members with MW show the

largest precipitation amount while for cases 4–6 they show the smallest precipitation amount. Among the PBL schemes, the use of PY tends to result in a relatively large amount of accumulated precipitation compared to when PU is used (for nine cases). However, the PBL scheme that gives the best prediction of accumulated precipitation amount can be either PY (for cases 1, 7, 9, and 10) or PU (for cases 2 and 5). For the radiation schemes, the ensemble members with RG tends to show a larger amount of accumulated precipitation than those with RC (for seven cases), but the radiation scheme that gives the best prediction of accumulated precipitation amount can be either RG (for cases 3, 5, 6, 7, 8, and 9) or RC (for cases 2 and 4).

Differences in cloud properties among different physics parameterization schemes which can be associated with the systematic differences in precipitation prediction are investigated. Figure 12 shows the 24-h mean cloud water path, rainwater path, ice water path, and total water path averaged over the ensemble members with each cloud microphysics scheme. The ensemble members with MW show the largest cloud water path and rainwater path and show smaller ice water path than the ensemble members with MT,

consistently for all cases. On the other hand, the ensemble members with MT show the intermediate cloud water path and the largest ice water path for all cases and the smallest rainwater path for nine cases. The ensemble members with MP show the smallest cloud water path and total water path for all cases and the intermediate rainwater path and the smallest ice water path for nine cases. For the same precipitation cases, MW tends to predict a relatively large amount of liquid hydrometeors, which would result in producing precipitation with more of warm-rain characteristics, while MT tends to predict relatively a large amount of ice hydrometeors, which would result in producing precipitation with more of cold-rain characteristics and producing more upper-level clouds. This finding aligns with Song and Lim (2022), who evaluated the performance of eleven cloud microphysics schemes in simulations of summertime heavy rainfall over the Korean Peninsula and found that MT produces a significantly larger amount of upper-level snow than MW. MP tends to predict relatively small amount of hydrometeors in the air but relatively large amount of surface precipitation (Fig. 11a), showing high precipitation efficiency.

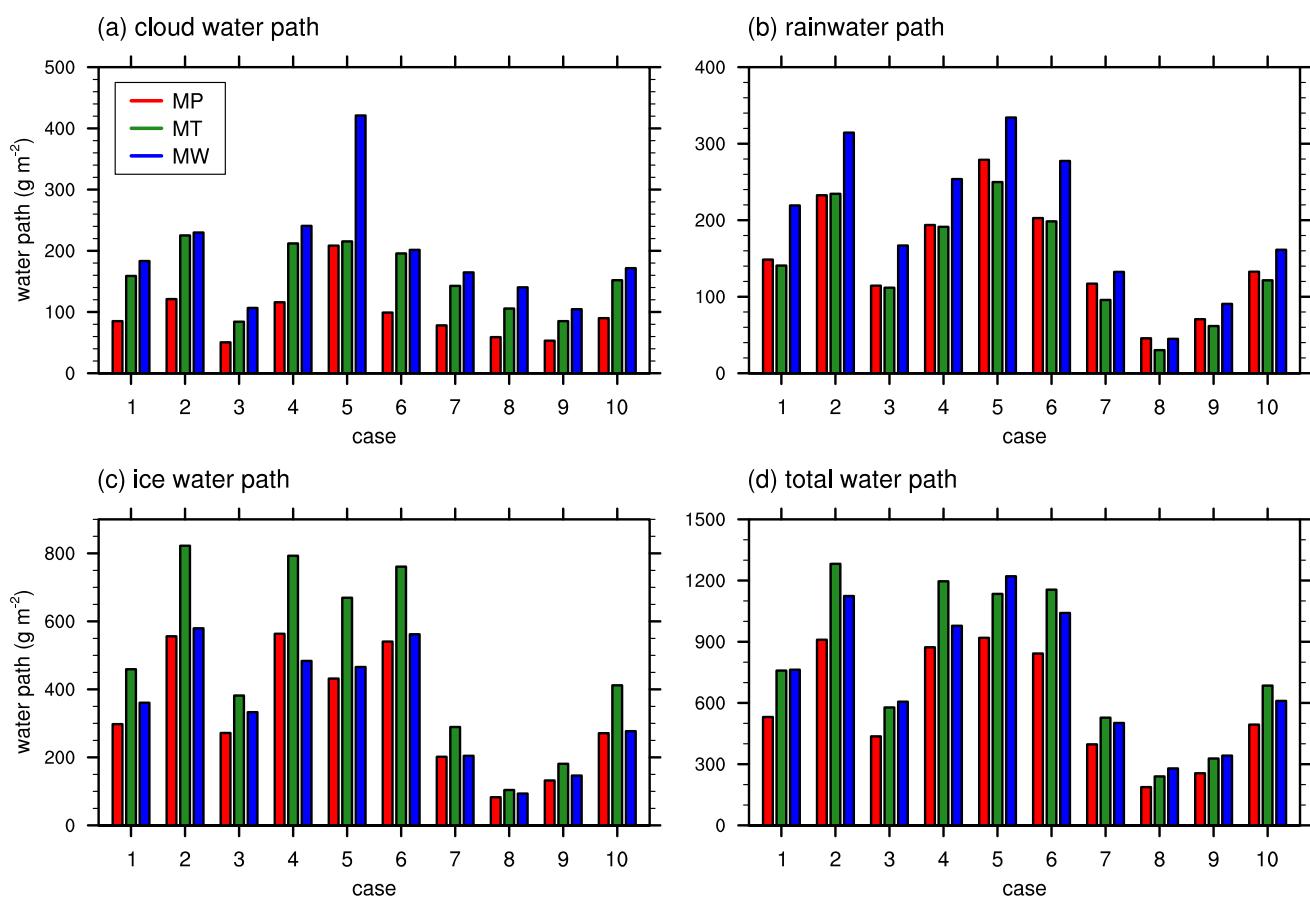


Fig. 12 24-h mean **a** cloud water path, **b** rainwater path, **c** ice water path, and **d** total (liquid and ice) water path averaged over the ensemble members with each cloud microphysics scheme

The above analyses indicate that each physics parameterization scheme does have a tendency to predict cloud and precipitation properties to be larger or smaller than those predicted by other schemes, which results in systematic generation of the ensemble spread of precipitation amount. This means that the ensemble members of a multi-physics ensemble are not randomly drawn from a distribution and there may exist specific distribution and arrangement of ensemble members for a given multi-physics ensemble. In other words, the ensemble members of a multi-physics ensemble are “distinguishable” from each other. Because the “indistinguishability” among ensemble members is one of the theoretical bases for the superiority of the ensemble-mean prediction over a single-member prediction (Christiansen 2019), multi-physics ensembles can be accused of not properly representing the model uncertainty in precipitation prediction (Xu et al. 2020). Despite this limitation, in this study, the multi-physics ensemble-mean prediction consistently shows its superiority over single-member predictions. This suggests that the distinguishable distribution of ensemble members does not necessarily lead to unsatisfactory performance of ensemble-mean prediction. If the physics parameterization schemes comprising the ensemble are chosen carefully, a multi-physics ensemble has the potential for strong ensemble-mean performance by providing a significant ensemble spread related to key physical processes.

5 Summary and conclusions

In this study, we evaluated the ensemble-mean performance of a multi-physics ensemble on the prediction of heavy precipitation cases in South Korea and investigated how the ensemble spread of the multi-physics ensemble is generated. Ten heavy precipitation cases were simulated with 27 different combinations of physics parameterization schemes and the ensemble predictions of the 24-h accumulated precipitation amount were evaluated. For most cases, the ensemble is underdispersive, indicating that the prediction uncertainty is only partly represented by the ensemble. The overall performance of the ensemble mean is better than that of any individual ensemble member. No individual ensemble member consistently shows good performance for every case. The relative decrease in RMSE from the average RMSE of individual ensemble members to the ensemble-mean RMSE is strongly positively correlated with the spread-error ratio and the ensemble spread. In comparison with subset ensembles, the total ensemble shows the most stable performance. Among the three types of physics parameterization in the multi-physics suite, the cloud microphysics parameterization contributes the most and the radiation parameterization contributes the least to the ensemble spread and the

ensemble-mean performance. The ensemble spread of the multi-physics ensemble, which is a key factor for the ensemble-mean performance, is generated systematically due to the fact that each physics parameterization scheme has a tendency to predict cloud and precipitation properties to be larger or smaller than those predicted by the other schemes.

The results in this study provide some practical helps to quantitative precipitation forecasting with a multi-physics ensemble in South Korea. From a practical point of view, it may be reasonable to include only the cloud microphysics schemes and PBL schemes in the multi-physics ensemble because of the relatively small contribution of radiation schemes to the ensemble-mean performance. In order to improve the performance of the ensemble mean, MP, PU, and RR that show overall poor performance on the prediction of heavy precipitation in South Korea may be replaced with other schemes which show overall better performance and increase the ensemble spread, although these schemes have some contributions to the stability of the ensemble-mean performance.

In this study, not all available cloud microphysics, PBL, and radiation schemes for precipitation prediction were tested. Due to computational resource limitations and the strong similarity among certain schemes, only three schemes were selected for each type. As a result, the simulation results do not represent the optimal performance that can be achieved using a multi-physics ensemble. With a more extensive study utilizing greater computational resources, there is potential for further improvement in multi-physics ensemble prediction. Despite these limitations, the results highlight the potential of the multi-physics ensemble for heavy precipitation prediction and provide preliminary guidance for enhancing the ensemble-mean performance.

This study highlights the importance of ensemble spread for the performance of the multi-physics ensemble mean in predicting heavy precipitation. The ensemble spread depends not only on the choice of physics parameterization schemes but also on the dynamic and physical characteristics of the precipitation event, such as the importance of latent heating, radiative heating, and subgrid-scale convection relative to that of grid-scale dynamic forcing on the event. Chen et al. (2022), who simulated ten summertime heavy rainfall events associated with afternoon convection in Taiwan with a cloud microphysics ensemble, reported much larger ensemble spreads (55%–161% of the observed mean) than those found in this study (27%–79%), which may be partially attributed to the less importance of grid-scale dynamic forcing in those events compared to the monsoonal precipitation events examined in this study. Park et al. (2021b) classified the synoptic patterns of summertime heavy rainfall events in South Korea into several clusters and showed that the ratio of dynamic forcing to diabatic forcing can vary significantly between these clusters. This suggests that the multi-physics

ensemble spread and the ensemble mean performance for the summertime heavy precipitation may be affected by the synoptic types of each event. Although not investigated in this study, the dependence of multi-physics ensemble spread on different precipitation events and the factors controlling it require further investigation.

Acknowledgements The authors are grateful to two anonymous reviewers for providing valuable comments on this work. This work was supported by the National Research Foundation of Korea under grant 2021R1A2C1007044. Han-Gyul Jin was supported by Global—Learning & Academic research institution for Master's-PhD students, and Postdocs (LAMP) Program of the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (No. RS-2023-00301938).

Funding This work was supported by the National Research Foundation of Korea under grants 2021R1A2C1007044 (Jong-Jin Baik) and RS-2023-00301938 (Han-Gyul Jin).

Data availability The WRF model output data used in this study are available on request from the authors.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

- Ancell BC (2013) Nonlinear characteristics of ensemble perturbation evolution and their application to forecasting high-impact events. *Weather Forecast* 28:1353–1365
- Berner J, Ha S-Y, Hacker JP, Fournier A, Snyder C (2011) Model uncertainty in a mesoscale ensemble prediction system: stochastic versus multiphysics representations. *Mon Weather Rev* 139:1972–1995
- Bretherton CS, Park S (2009) A new moist turbulence parameterization in the Community Atmosphere Model. *J Clim* 22:3422–3448
- Buizza R, Miller M, Palmer TN (1999) Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q J Roy Meteorol Soc* 125:2887–2908
- Buizza R, Houtekamer PL, Pellerin G, Toth Z, Zhu Y, Wei M (2005) A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon Weather Rev* 133:1076–1097
- Chen J-P, Tsai T-C, Tzeng M-D, Liao C-S, Kuo H-C, Hong J-S (2022) Microphysical perturbation experiments and ensemble forecasts on summertime heavy rainfall over Northern Taiwan. *Weather Forecast* 37:1641–1659
- Chou M-D, Suarez MJ (1999) A solar radiation parameterization for atmospheric studies. NASA Tech. Rep. NASA/TM-1999-104606. National Aeronautics and Space Administration Goddard Space Flight Center, Greenbelt, MD, USA
- Chou M-D, Suarez MJ, Liang X-Z, Yan M-H (2001) A thermal infrared radiation parameterization for atmospheric studies. NASA Tech. Rep. NASA/TM-2001-104606. National Aeronautics and Space Administration Goddard Space Flight Center, Greenbelt, MD, USA
- Christiansen B (2018) Ensemble averaging and the curse of dimensionality. *J Clim* 31:1587–1596
- Christiansen B (2019) Analysis of ensemble mean forecasts: the blessings of high dimensionality. *Mon Weather Rev* 147:1699–1712
- Clark AJ (2017) Generation of ensemble mean precipitation forecasts from convection-allowing ensembles. *Weather Forecast* 32:1569–1583
- Collins WD, Rasch PJ, Boville BA, Hack JJ, McCaa JR, Williamson DL, Kiehl JT, Briegleb B, Bitz C, Lin S-J, Zhang M, Dai Y (2004) Description of the NCAR Community Atmosphere Model (CAM3.0). NCAR Tech. Note NCAR/TN-464+STR. National Center for Atmospheric Research, Boulder, CO, USA
- Du J, Zhou B (2011) A dynamical performance-ranking method for predicting individual ensemble member performance and its application to ensemble averaging. *Mon Weather Rev* 139:3284–3303
- Du J, Mullen SL, Sanders F (1997) Short-range ensemble forecasting of quantitative precipitation. *Mon Weather Rev* 125:2427–2459
- Ebert EE (2001) Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon Weather Rev* 129:2461–2480
- Fortin V, Abaza M, Anctil F, Turcotte R (2014) Why should ensemble spread match the RMSE of the ensemble mean? *J Hydrometeorol* 15:1708–1713
- García-Ortega E, Lorenzana J, Merino A, Fernández-González S, López L, Sánchez JL (2017) Performance of multi-physics ensembles in convective precipitation events over northeastern Spain. *Atmos Res* 190:55–67
- Gaudet LC, Sulia KJ, Tsai T-C, Chen J-P, Blair JP (2021) Assessment of a microphysical ensemble used to investigate the OWLeS IOP4 lake-effect storm. *J Atmos Sci* 78:1607–1628
- Guo Z, Fang J, Sun X, Yang Y, Tang J (2019) Sensitivity of summer precipitation simulation to microphysics parameterization over eastern China: convection-permitting regional climate simulation. *J Geophys Res Atmos* 124:9183–9204
- Hersbach H, Bell B, Berrisford P et al (2020) The ERA5 global reanalysis. *Q J R Meteorol Soc* 146:1999–2049
- Hong S-Y, Noh Y, Dudhia J (2006) A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon Weather Rev* 134:2318–2341
- Houtekamer PL, Derome J (1995) Methods for ensemble prediction. *Mon Weather Rev* 123:2181–2196
- Iacono MJ, Delamere JS, Mlawer EJ, Shephard MW, Clough SA, Collins WD (2008) Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models. *J Geophys Res* 113:D13103
- Jankov I, Berner J, Beck J, Jiang H, Olson JB, Grell G, Smirnova TG, Benjamin SG, Brown JM (2017) A performance comparison between multiphysics and stochastic approaches within a North American RAP ensemble. *Mon Weather Rev* 145:1161–1179
- Jensen AA, Harrington JY, Morrison H, Milbrandt JA (2017) Predicting ice shape evolution in a bulk microphysics model. *J Atmos Sci* 74:2081–2104
- Jiménez PA, Dudhia J, González-Rouco JF, Navarro J, Montávez JP, García-Bustamante E (2012) A revised scheme for the WRF surface layer formulation. *Mon Weather Rev* 140:898–918
- Jo KR, Pak KS, Ri SN, Kim SI, Kim SR (2023) Impact of microphysics schemes on prediction of an extreme heavy rainfall event over the Democratic People's Republic of Korea: a case study using WRF model. *Meteorol Atmos Phys* 135:36
- Kain JS (2004) The Kain-Fritsch convective parameterization: an update. *J Appl Meteorol* 43:170–181
- Lim K-SS, Hong S-Y (2010) Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models. *Mon Weather Rev* 138:1587–1612
- Lorenz EN (1965) A study of the predictability of a 28-variable atmospheric model. *Tellus* 17:321–333

- Mansell ER, Ziegler CL, Bruning EC (2010) Simulated electrification of a small thunderstorm with two-moment bulk microphysics. *J Atmos Sci* 67:171–194
- Mason P, Thomson D (1992) Stochastic backscatter in large-eddy simulations of boundary layers. *J Fluid Mech* 242:51–78
- Meng Z, Zhang F (2007) Test of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: imperfect model experiments. *Mon Weather Rev* 135:1403–1423
- Morrison H, Milbrandt JA (2015) Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part I: scheme description and idealized tests. *J Atmos Sci* 72:287–311
- Morrison H, Curry JA, Khvorostyanov VI (2005) A new double-moment microphysics parameterization for application in cloud and climate models. Part I: description. *J Atmos Sci* 62:1665–1677
- Nakanishi M, Niino H (2006) An improved Mellor-Yamada level-3 model: its numerical stability and application to a regional prediction of advection fog. *Bound Layer Meteorol* 119:397–407
- Park C, Son S-W, Kim H, Ham Y-G, Kim J, Cha D-H, Chang E-C, Lee G, Kug J-S, Lee W-S, Lee Y-Y, Lee HC, Lim B (2021a) Record-breaking summer rainfall in South Korea in 2020: synoptic characteristics and the role of large-scale circulations. *Mon Weather Rev* 149:3085–3100
- Park C, Son S-W, Kim J, Chang E-C, Kim J-H, Jo E, Cha D-H, Jeong S (2021b) Diverse synoptic weather patterns of warm-season heavy rainfall events in South Korea. *Mon Weather Rev* 149:3875–3893
- Romine GS, Schwartz CS, Berner J, Fossell KR, Snyder C, Anderson JL, Weisman ML (2014) Representing forecast error in a convection-permitting ensemble system. *Mon Weather Rev* 142:4519–4541
- Shin C-S, Lee T-Y (2005) Development mechanisms for the heavy rainfalls of 6–7 August 2002 over the middle of the Korean Peninsula. *J Meteorol Soc Jpn* 83:683–709
- Shutts GJ (2005) A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Q J Roy Meteorol Soc* 131:3079–3102
- Skamarock WC, Klemp JB, Dudhia J, Gill DO, Liu Z, Berner J, Wang W, Powers JG, Duda MG, Barker DM, Huang XY (2019) A description of the Advanced Research WRF model version 4. NCAR Tech. Note NCAR/TN-556+STR. National Center for Atmospheric Research, Boulder, CO, USA
- Song HJ, Lim K-SS (2022) Evaluation of bulk microphysics parameterizations for simulating the vertical structure of heavy rainfall between Korea and the United States. *Weather Clim Extremes* 37:100490
- Song HJ, Sohn BJ (2018) An evaluation of WRF microphysics schemes for simulating the warm-type heavy rain over the Korean Peninsula. *Asia-Pac J Atmos Sci* 54:225–236
- Stensrud DJ, Bao J-W, Warner TT (2000) Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon Weather Rev* 128:2077–2107
- Surcel M, Zawadzki I, Yau MK (2014) On the filtering properties of ensemble averaging for storm-scale precipitation forecasts. *Mon Weather Rev* 142:1093–1105
- Tewari M, Chen F, Wang W, Dudhia J, LeMone MA, Mitchell K, Ek M, Gayno G, Wegiel J, Cuena RH (2004) Implementation and verification of the unified Noah land surface model in the WRF model. In: 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction, 12–16 January 2004, Seattle, WA, USA, pp 10–15.
- Thompson G, Eidhammar T (2014) A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *J Atmos Sci* 71:3636–3658
- Toth Z, Kalnay E (1997) Ensemble forecasting at NCEP and the breeding method. *Mon Weather Rev* 125:3297–3319
- Tribbia JJ, Baumhefner DP (1988) The reliability of improvements in deterministic short-range forecasts in the presence of initial state and modeling deficiencies. *Mon Weather Rev* 116:2276–2288
- Wang X, Bishop CH (2003) A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J Atmos Sci* 60:1140–1158
- Xu Z, Chen J, Jin Z, Li H, Chen F (2020) Assessment of the forecast skill of multiphysics and multistochastic methods within the GRAPES regional ensemble prediction system in the East Asian monsoon region. *Weather Forecast* 35:1145–1171

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.