

Spatial Spread-Skill Relationship in Terms of Agreement Scales for Precipitation Forecasts in a Convection-Allowing Ensemble

Xi Chen^{1,2}, Huiling Yuan^{1,2,*} and Ming Xue^{1,3}

1. *Key Laboratory of Mesoscale Severe Weather/Ministry of Education and School of Atmospheric Sciences, Nanjing University, Nanjing, China*

2. *Joint Center for Atmospheric Radar Research of CMA/NJU, China*

3. *School of Meteorology, and Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma*

Submit to Quarterly Journal of Royal Meteorological Society

Accepted October , 2017

*Corresponding author:

Prof. Huiling Yuan

Email: yuanhl@nju.edu.cn

ABSTRACT

Verification of precipitation is one of the major issues in evaluating numerical weather prediction. In this study, a recently developed neighbourhood-based method in terms of agreement scales is applied to characterize scale-dependent spatial spread-skill relationship of precipitation forecasts in a 3-km convection-allowing ensemble prediction system (EPS) over the Yangtze-Huaihe river basin of China. Thirty cases during the Meiyu season of 2013 are classified into two weather regimes, large coverage (LC) and small coverage (SC) based on the precipitation fractional coverage. Overall, precipitation distributions for these two weather regimes are reasonably forecast by the EPS. The results show that the spatial spread-skill relationship depends highly on the weather regime. The spatial spread-skill relationship under SC is poorer and shows more diurnal variations compared to that under LC. In addition, this paper extends the neighbourhood-based method to investigate the relative influence of precipitation intensity and placement on the spatial spread-skill relationship. With increasing precipitation threshold, the relative impact of precipitation intensity on the relationship gradually decreases, and the influence of precipitation placement becomes dominant.

Key Words: convection-allowing ensemble; precipitation forecast; agreement scale; spatial spread-skill relationship

2. Introduction

Accurate precipitation forecasts have been particularly challenging for numerical weather prediction models. In last decade, as the convection-allowing models have been developed rapidly (Kawabata et al., 2007; Baldauf et al., 2011; Seity et al., 2011), convection-allowing ensemble prediction systems (EPSs), which explicitly depict convection and take the uncertainties from many sources into account, show advantages to improve precipitation forecasts (Roberts and Lean, 2008; Clark et al., 2010; Schwartz et al., 2015). The convection-allowing EPSs have been operational or quasi-operational at many forecasting centers, such as MOGREPS-UK with 2.2-km resolution (Tennant, 2015), AROME-EPS with 2.5-km resolution (Vie et al., 2011; Bouttier et al., 2012; Nuissier et al., 2012; Bouttier et al., 2016), COSMO-DE-EPS with 2.8-km resolution (Peralta et al., 2012; Harnisch and Keil, 2015) and NCAR's experimental real-time convection-allowing EPS with 4-km resolution (Schwartz et al., 2015). These convection-allowing EPSs provide important guidance in operational weather forecasting (Iyer et al., 2016), especially when predicting the high impact weather such as extreme heavy rain. However, the complexity of nonlinear error growth and the predictability of moist convection on convective scale (Hohenegger and Schar, 2007; Zhang et al., 2007; Clark et al., 2010; Melhauser and Zhang, 2012) pose great challenges for the improvements and assessment of precipitation forecasts in a convection-allowing EPS. Therefore, verification of precipitation forecasts becomes one of the major issues in evaluating convection-allowing EPSs. This study aims to verify the short-range precipitation forecasts in a convection-allowing EPS, focusing on the scale-dependent spatial spread-skill relationship using a new neighbourhood-based method (Dey et al., 2016a, b).

Traditionally, the spread-skill relationship is used to depict the agreement between the ensemble spread (standard deviation between ensemble members that measuring the uncertainty of ensemble) and the expected forecast error of ensemble-mean or control forecast (Grimit and Mass, 2007). Many studies use the traditional spread-skill relationship (or spread-error relationship) to evaluate ensemble precipitation forecasts (Ebert, 2001; Martin et al., 2010; Bouttier et al., 2012; Su et al., 2014). However, the traditional metrics of forecast error, such as root-mean square error (RMSE) and mean absolute error (MAE), are not suitable for the convection-allowing EPSs due to the serious double penalty (Mittermaier et al., 2013). In addition, the traditional error and spread metrics using all statistical samples are scale-independent. In order to address these problems, some spatial verification methods (Gilleland et al., 2009), such as the Method for Object-based Diagnostic Evaluation (MODE; Davis et al., 2006, 2009), the neighbourhood-based method (Ebert 2008, 2009), the method of Contiguous Rain Area (CRA; Ebert and McBride 2000), the method of Fractions Skill Score (FSS; Roberts and Lean, 2008) and so on, have been developed for high-resolution weather forecasts. Although these methods can reduce the double penalty problem and provide more significant scale information than the traditional ones, they are proposed for verifying deterministic forecasts not for EPS. In order to extend these methods to verify the convection-allowing EPS, some new methods are further developed (e.g. Zacharov and

Rezacova, 2009; Duc et al., 2013; Dey et al., 2014).

Recently, Dey et al. (2016a, b) have developed a new neighbourhood-based method to identify agreement scales and used it to evaluate the scale-dependent spatial spread-skill relationship for ensemble precipitation forecasts in a convection-allowing EPS developed over the UK. The spatial spread-skill relationship refers to the comparison of the average agreement scale for ensemble members, denoted as $S_{ij}^{A(\overline{mm})}$, and the average agreement scale for ensemble members and observations, denoted as $S_{ij}^{A(\overline{mo})}$. Through a long-term routine model evaluation, they got useful information about ensemble performance and characteristics in summer 2013 over the UK (Dey et al., 2016b). In order to test the feasibility of this latest method under different climate conditions in different regions, this study applies this neighbourhood-based method to verify a convection-allowing EPS over the Yangtze-Huaihe river basin of China. Also, Dey et al. (2016a, b) pointed out that three factors, including precipitation intensity, precipitation placement and the distance from the precipitation, influence the spatial spread-skill relationship in the EPS. However, their paper didn't compare the relative contributions of these factors on the spatial spread-skill relationship. This study extends a new application of this latest method to further analyze the comparative importance of precipitation intensity and placement on the spatial spread-skill relationship in a convection-allowing EPS.

At the same time, the predictability and forecast skill are varied with weather regimes. Previous studies show that the forecasts under weak forcing regimes tend to produce poorer skill than that under strong forcing regimes (Keil and Craig, 2011), meanwhile the predictability and forecast skill on smaller scales are usually worse than larger-scale forecasts (Roberts, 2008; Roberts and Lean, 2008; Surcel et al., 2015). Done et al. (2006) proposed the convective adjustment time-scale (τ_c) to classify the weather forcing regime. However, many researches (Zimmer et al., 2011; Keil et al., 2014; Kuhnlein et al., 2014) investigated τ_c and found out that τ_c actually is a weak indicator to classify the precipitation forecasts in terms of predictability and forecast skill of precipitation. Surcel et al. (2016) found out that the precipitation fractional coverage is the best indicator associated with precipitation predictability among three indicators (τ_c , large-scale forcing for ascent and precipitation fractional coverage). Although Dey et al. (2016b) did not verify the spatial spread-skill relationship for different weather regimes, their study also suggested that this relationship is highly associated with precipitation fractional coverage. Therefore, in this study, the precipitation fractional coverage is selected as the classification criteria of weather regime, and the cases during the Meiyu season of 2013 over the Yangtze-Huaihe river basin are categorized to large fractional coverage (LC) and small fractional coverage (SC) groups, respectively.

The Meiyu season over the Yangtze-Huaihe river basin usually spans from mid-June to mid-July (Ding and Chan, 2005), while heavy rainfall often occurs under favorable synoptic conditions, including the Meiyu front, south westerly low-level jet, low-level vortex and subtropical high (Luo and Chen, 2015). There are several studies

using convection-allowing EPS to investigate the mechanism of precipitation during the Meiyu season over the Yangtze-Huaihe river basin. For example, Luo and Chen (2015) investigated the forecast uncertainties and physical mechanisms of a quasi-linear extreme-rain-producing mesoscale convective system (MCS) along the Meiyu front in East China using convection-permitting ensembles. Wang and Zhong (2014) implemented multi-physics ensemble simulations of a summertime heavy precipitation event and studied the influence of large-scale urbanization on precipitation predictions over the lower reaches of Yangtze River valley. Due to the expensive computer resources, most of the previous studies used convection-allowing EPS to investigate the specific weather cases. Since the Meiyu season of 2013 concentrates from 23 June to 22 July 2013 over the Yangtze-Huaihe river basin, this study evaluates daily precipitation forecasts of the total 30 cases and compares the forecast performance under the LC and SC categories, based on a convection-allowing EPS.

The outline of this article is as follows. Section 2 describes the model configurations, verification data, the classification method of weather regimes and the neighbourhood-based method. The averaged results and the comparison between the traditional and spatial spread-skill relationship of one month studies during the Meiyu season of 2013 are presented in Section 3. Section 4 provides the results about the relative influence of precipitation placement and intensity on the spatial spread-skill relationship of precipitation forecasts. Summary and conclusions are given in Section 5.

3. Data and Method

2.1 Model configuration

The convection-allowing EPS over the Yangtze-Huaihe river basin is based on the Weather Research and Forecasting (WRF) model with the grid spacing of 3-km (single domain, 620 * 498 grid points) and 50 vertical model levels (Figure 1). It is driven by the control forecast and 14 perturbed forecasts randomly selected from the NCEP Global Ensemble Forecast system (GEFS; Su et al. 2014), which output is archived with 6 hours interval at 1 degree horizontal resolution (~110 km). GEFS consists of a control forecast and 20 perturbed forecasts using the bred vector-ensemble transform with rescaling (BV-ETR) method (Wei et al., 2008) and stochastic total tendency perturbation (STTP) scheme (Hou et al., 2008). Although, there is a big gap between the resolutions of these two EPSs, it is proved as an available way to directly conduct the dynamical downscaling (Lawson and Gallus, 2016; Zhu and Xue, 2016). In order to represent the uncertainty in model physics, different combinations of physical parameter schemes (Table 1), including microphysical schemes, land surface schemes, and planetary boundary schemes, are employed to construct individual forecast members in this convection-allowing EPS by referencing to the model configuration of the NOAA Hazardous Weather Testbed 2007-2010 Spring Experiments (Coniglio et al., 2010; Clark et al., 2012) and our own tests. All the members use the same short-wave and long-wave radiation schemes accounting for their less importance and the stability

for the model integration.

Thirty cases from one month (23 June to 22 July 2013) during the Meiyu season are initialized at 0000 UTC of each day and forecasted to 36 h with the 3 h output interval, except the control run with hourly rainfall output.

2.2 Precipitation verification data

The verification data are from the NOAA CPC Morphing Technique (CMORPH) merged with the rain gauge observations in China (Shen et al., 2014), which provide official grid precipitation observations in China with the highest temporal (1 h) and spatial (0.1 degree) resolutions available. Due to the relatively low spatial resolution (~10 km) of this verification data, a compromise is made to verify the precipitation forecasts on the coarser observation grid; the 3-km precipitation forecasts are interpolated using the nearest four grid points with inverse-distance weighting onto the observation grid over the verification domain (25.55°N~38.25°N, 110.05°E~122.75°E; Figure 1).

In this study, the 24 h verification period (0300 UTC-0300 UTC of next day, corresponding to 3-27 h of forecast) is defined as an individual case, and 3 h accumulated precipitation forecasts are only evaluated during this 24 h verification period. The first 3 h forecasts are discarded to reduce the spin-up problem, and the 27-36 h forecasts are removed to avoid the overlapping of the cases.

2.3 Classification of weather regimes

The precipitation fractional coverage is defined as the proportion of the verification domain with hourly rainfall rates larger than 0.1 mm h^{-1} . When both median and mean values of precipitation fractional coverage of hourly observations during a 24 h verification period are less than 0.1, that day is categorized as the SC day and otherwise the LC day. By using the coverage threshold of 0.1, the ratio of the SC and LC cases is similar to the percentage of the two categories found in previous studies (Zimmer et al 2011; Surcel et al. 2016).

According to the boxplot of the hourly precipitation coverage of observations (Figure 2a), 8 of 30 cases are classified as SC cases, while the remaining 22 cases are LC cases. Taking the control forecast as an example, Figure 2b shows scatter plot of observed fractional coverages of hourly precipitation versus control forecast coverages; it shows a good agreement between forecasts and observations, with a correlation of 0.86, indicating that the model is predicting generally correct regimes in terms of precipitation fractional coverage.

2.4 Location-dependent agreement scales

The neighbourhood-based method (Dey et al., 2016a, b) has been developed to quantify the location-dependent spatial agreement scales between two precipitation fields. The calculation of the agreement scales was described in Dey et al (2016a, b) and is repeated as follows:

The distance D_{ij}^S of one grid point (i, j) in the domain at the agreement scale S between two precipitation fields $f1$ and $f2$ is defined as:

$$D_{ij}^S = \begin{cases} \frac{(f1_{ij}^S - f2_{ij}^S)^2}{(f1_{ij}^S)^2 + (f2_{ij}^S)^2} & \text{if } f1_{ij}^S > 0 \text{ or } f2_{ij}^S > 0 \\ 1 & \text{if } f1_{ij}^S = 0 \text{ and } f2_{ij}^S = 0 \end{cases}$$

where $f1_{ij}^S$ and $f2_{ij}^S$ are the area averaged precipitation within a squared grid box centered upon the point (i, j), which has the side length of $2*S+1$. The values of the agreement scale S are varied from 0 to S_{lim} , until $f1_{ij}^S$ and $f2_{ij}^S$ are deemed sufficiently similar when the criterion $D_{ij}^S \leq D_{crit,ij}^S$ is met, where $D_{crit,ij}^S = \alpha + (1 - \alpha) \frac{S}{S_{lim}}$. More details about this method can refer to Dey et al. (2016a, b). In this study, $\alpha = 0.25$ and $S_{lim} = 30$ are used, considering the size of the domain and the scales of precipitation fields. Other values of S_{lim} , such as 40 and 60, are also tested and produce similar results (see supplementary material Figure S2). After the above calculation of each point in the verification domain between two precipitation fields, the agreement scales can be mapped. By its definition, small (large) spatial agreement scales correspond to high (low) spatial agreement.

This study applies the neighbourhood-based method to evaluate the location-dependent spatial agreement between the pairs of ensemble members or between any individual ensemble forecast and the corresponding observation. To assess the general performance and filter the noise in the individual maps of agreement scales (Dey et al. 2016a, b), the verification focuses to analyze the average scales from all paired fields. For an EPS with N members (in this paper, $N=14$), the averaged result (denoted as $S_{ij}^{A(\overline{mm})}$) of the pairs between ensemble members is generated from the $N(N-1)/2$ agreement scale maps, and the averaged result (denoted as $S_{ij}^{A(\overline{mo})}$) of the pairs between individual ensemble members and the observations is the average of the N agreement scale maps. In addition, interpretation of agreement scales depends on the verification grid spacing, while similar agreement scales on coarser grids represent coarser spatial scale information than that on finer grids.

In addition, the ‘‘moving binned scatter’’ (Dey et al., 2016a, b) method is applied to compare the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ in this paper. To avoid the noisy scatter plot between $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$, it is necessary to plot the mean values of $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ for each moving bin based on the values of $S_{ij}^{A(\overline{mm})}$. For each bin, the mean value of $S_{ij}^{A(\overline{mm})}$ over those points whose values of agreement scales fall into this bin range is calculated and plotted on the x-axis, then the corresponding value of y-axis is the mean $S_{ij}^{A(\overline{mo})}$ over these same points. After all bins are calculated, a line of mean $S_{ij}^{A(\overline{mm})}$ against mean $S_{ij}^{A(\overline{mo})}$ can be plotted. If the curve falls above (below) the diagonal, the ensemble is said to be spatially underdispersive (overdispersive) in terms of agreement scales. After examination (see supplementary material Figure S3), the bin size of 6 grid points is used in this study to retain enough statistical samples and sufficient scale-dependent information in each bin (Dey et al., 2016a, b).

3. Meiyu season averaged results

This section investigates the agreement scales $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ and the

spatial spread-skill relationship (relationship between $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$). The spatial and diurnal characteristics of the spatial spread-skill relationship are studied, and the advantages of the new neighbourhood-based method (Dey et al., 2016a, b) are highlighted by comparing with traditional spread-skill relationship (RMSE and spread).

Dey et al. (2016b) demonstrated the necessity to test the effect of systematic intensity biases on the verification of spatial spread-skill relationship. Thus, this study also examined the effect of systematic biases by multiplying the observations with a factor of 1.5 (approximately the maximum 3 h intensity bias averaged between the observations and ensemble members; Figure S1). Consistent with previous results (Dey et al., 2016b), systematic biases show insignificant influence on the spatial spread-skill relationship.

3.1 Spatial characteristics

Figure 3 displays the spatial distributions of 3 h accumulated precipitation averaged over the 3-27 h lead times from observations and control forecasts under two different weather regimes. Under LC, precipitation distributions are often controlled by the Meiyu front at large-scale circulations and the averaged precipitation is widely spread throughout the whole verification domain (Figure 3a and 3c). Because of the shift of the subtropical high and the Meiyu front, most of the precipitation events are east-west or northeast-southwest orientated, forming three major rainbands in the domain (Figure 3a and 3c). One rainband is located in the northern area extending from coastal Shandong ($\sim 120^\circ$ E and 37° N) to mountainous Shanxi ($\sim 112^\circ$ E and 36° N) province, one is in the central plain area from Anhui ($\sim 117^\circ$ E and 31° N) to Hubei ($\sim 114^\circ$ E and 30° N), and the last one in the southern hilly area from Eastern Jiangxi ($\sim 117^\circ$ E and 29° N) to Western Jiangxi ($\sim 115^\circ$ E and 28° N). In contrast, under SC, precipitation observations and the control forecasts (Figure 3b and 3d) are more intense and localized along heterogeneous and steep terrains, such as the mountainous areas in the northwest domain ($\sim 111^\circ$ E and 37° N), Shandong peninsular ($\sim 118^\circ$ E and 36° N) and the coastal areas ($\sim 122^\circ$ E and 35° N) in the northeast domain. In general, the control forecasts of the convection-allowing EPS under two different weather regimes reasonably reproduce the observed precipitation distributions, but both show the displacement errors of rainfall centers. Under LC, the averaged rain intensities of control forecasts (Figure 3c) are overestimated, especially in Jiangxi province ($\sim 116^\circ$ E and 29° N) in the middle domain. Compared to the observations under SC (Figure 3b), the control forecasts generate more isolated and spurious rain cells (Figure 3d), but underforecast the maxima of larger precipitation cells (7.33 mm in Figure 3b vs. 3.30 mm in Figure 3d in the northeast domain).

Figure 4 shows the spatial distributions of $S_{ij}^{A(\overline{mm})}$ averaged over the 3-27 h lead times under LC and SC, which are similar to the spatial distributions of $S_{ij}^{A(\overline{mo})}$ under the two regimes (see supplementary material Figure S5). Overall, the agreement scales under LC are lower than that under SC (2.75 less on domain average), which is consistent with the conclusion that the larger precipitation fractional coverage is

associated with smaller agreement scales (higher spatial agreement) (Dey et al., 2016a, b). The smallest agreement scales locate close to the rainfall centers, indicating more confident about the location of precipitation than other regions. The minimum agreement scale is only 10.1 grid points (around the neighbourhood length of 212 km) under LC and 11.6 grid points (~242 km) under SC. This means the EPS is not confident about the location of precipitation at the scales smaller than these grid scales. In the research of Dey et al. (2016b), the minimum agreement scale of seasonal averaged is 12 grid points, which indicates much smaller physical scales (~ 55 km) for the high-resolution MOGREPS-UK with the 2.2-km resolution. As mentioned in Section 2.4, although the minimum agreement scale of this study is smaller than that of Dey et al. (2016b), more spatial scale information is lost.

3.2 Diurnal characteristics

Section 3.1 showed spatial characteristics of precipitation and its agreement scales from the convection-allowing EPS. This section focuses on investigating the diurnal variations of domain averaged precipitation amounts, and comparing spatial spread-skill relationship with traditional spread-skill relationship (RMSE and spread) for 22 cases under LC and 8 cases under SC.

Figure 5 shows the diurnal variations of domain averaged precipitation amounts under the two weather regimes. Overall, the domain averaged precipitation amounts under LC (Figure 5a) are larger than that under SC (Figure 5b). Based on the observations during 0300 UTC to 0300 UTC cycle (corresponding to 3-27 h lead times), two rainfall peaks under LC occur at 0900 UTC and 0000 UTC, respectively, while there is only one peak at 2100 UTC for SC. Under LC, the ensemble mean forecasts reproduce the two peaks successfully (Figure 5a), but largely overestimate the first rainfall peak and underestimate the second one. For the precipitation forecasts under SC, the ensemble mean forecasts generate a spurious peak at the 9 h lead time and a 3 h time lag for the 2100 UTC observed peak. The ensemble mean forecasts show obvious overestimations for the specific period during the afternoon and evening (0600 UTC - 1500 UTC; corresponding to 6-15 h lead times, 1400 LST – 2300 LST) under both regimes and slight underestimations during the early morning 2100 UTC – 0300 UTC (corresponding to 21-27 h lead times, 0500 LST – 1100 LST) for the LC cases. The overestimation (underestimation) of the precipitation on a diurnal cycle may be related to that the physics parameterization schemes (Li et al., 2009; Katragkou et al., 2015) in the convection-allowing model tend to overestimate (underestimate) the afternoon and evening deep convection (shallow convection in the early morning).

The diurnal variations of domain averaged RMSE, spread and the traditional spread-skill relationship (spread/RMSE) under two weather regimes (Figure 6a and 6b) show the similar patterns with the domain averaged precipitation amounts (Figure 5). Although RMSE and spread under LC are larger than that under SC, there are no obvious differences between the diurnal variations of traditional spread-skill relationship under two weather regimes. Figure 6c shows the diurnal variations of domain averaged agreement scales of $S_{ij}^{A(mm)}$ and $S_{ij}^{A(mo)}$. In general, they are negatively correlated with the domain averaged precipitation amounts, RMSE and

spread. Same as the previous study (Dey et al., 2016b), the forecasts under SC show larger domain averaged agreement scales than LC, revealing that the smaller fractional coverage of precipitation is associated with the poorer spatial agreement. Under both regimes, the domain averaged agreement scales of $S_{ij}^{A(\overline{mm})}$ are smaller than that corresponding scales of $S_{ij}^{A(\overline{mo})}$ at the same lead time. In order to quantitatively compare the differences of forecast performance for domain averaged scales under LC and SC, the diurnal variations of the ratios between the domain averaged agreement scales of $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ are presented in Figure 6d. Different with the traditional spread-skill relationship (Figure 6b), the spatial spread-skill relationship (Figure 6d) under the two regimes are always less than 1 (underdispersion). The ratio under SC fluctuates more with the lead time than that under LC (Figure 6d), indicating that the spatial spread-skill relationship under SC shows higher variations on the diurnal cycle. The ratio under SC reveals the worst spatial spread-skill relationship (Figure 6d) at the 18 h lead time, but the traditional spread-skill relationship under SC (Figure 6b) is nearly perfect (~ 1) at the same time.

To further investigate the contradictory results (Figure 6b and 6d) between the spatial and traditional spread-skill relationship, Figure 7 presents the spatial distributions of precipitation amounts, RMSE, spread, and agreement scales averaged over all cases under SC at the 18 h lead time. Obviously, the forecasts for heavier precipitation amounts (bold black contours in Figure 7a-f) show larger forecast errors (Figure 7a), increased spread (Figure 7b), and higher agreement (Figure 7d and 7e). In terms of the differences between the spread and RMSE (Figure 7c), the EPS is severely underdispersive in the precipitation areas and overdispersive in relatively large dry regions especially around rain cells. Thus, the domain averaged difference between the spread and RMSE is quite small (0.14), which leads to the best (~ 1) traditional spread-skill relationship (*i.e.*, the ratio of spread and RMSE) at the 18 h lead time (Figure 6b). On contrary, Figure 7f shows underdispersive forecasts overspreading the verification area, with extremely insufficient spread in the rain cells. Overall, both the traditional and spatial spread-skill relationship metrics (Figure 7c and 7f) indicate that the EPS lacks spread in the rainfall regions over the northwest mountainous areas and northeast coastal areas. However, the traditional spread-skill relationship only measures the static information of point-wise error and spread (Figure 7c) and provides spurious perfect spread-skill relationship (Figure 6b), while the spatial spread-skill relationship is capable of reflecting the correct spatial scale information in the neighbourhood space (Figure 7f) and thus obtaining consistent total statistics (Figure 6d).

In addition, more scale information about the spatial spread-skill relationship is obtained by using the “moving binned scatter” method (Dey et al., 2016a, b). Figure 8a shows the spatial spread-skill relationship over the 3-27 h lead times under LC and SC. As the curves situate above the diagonal, the agreement scales below 20 grid points indicate the underdispersive spread-skill relationship under both regimes, and more severe underdispersion for that under SC. For different lead times, the spread-skill relationship shows slight diurnal variations under LC (Figure 8b), while it shows much larger diurnal variations under SC (Figure 8c) with the worst skill at the 18 h lead time. This is consistent with the results shown in Figure 6d. In order to examine whether the differences under the two regimes are caused by different sample sizes, 8 cases are

randomly selected under LC (see supplementary material Figure S4) and similar conclusion can be drawn. Therefore, the differences of the spatial spread-skill relationship under LC and SC are mainly attributed to different predictabilities of these two weather regimes.

4. Relative influence of precipitation intensity and placement

In Section 3, the agreement scales ($S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$) and spatial spread-skill relationship were investigated, and compared with the traditional spread-skill relationship during the Meiyu season of 2013. Dey et al. (2016a, b) demonstrated that $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ are influenced by three factors, including the placement of precipitation, bias in precipitation intensity, and distance from the precipitation. However, their paper has not discussed the relative importance of these factors. For the precipitation grids (the grids whose rain rates from any ensemble members or the observations are above a specific precipitation threshold), the precipitation placement and intensity are the main impact factors on the spatial spread-skill relationship. In this section, we extend a new application of this neighbourhood-based method to further investigate the relative influence of precipitation intensity and placement on the frequency distributions of agreement scales and spatial spread-skill relationship.

Before calculating the agreement scales, the rain rates below a specific precipitation threshold are processed as zero values and the ones above the threshold remain unchanged to provide the first type of precipitation field (hereinafter, the threshold raw field). Similarly, the original rainfall amounts below the threshold are truncated to zero values and the amounts above the threshold are set to one, which provides the second type of precipitation field (hereinafter, the threshold binary field). In this study, the agreement scales $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ are calculated and compared for the threshold raw and binary fields, respectively. The differences between the agreement scales of these two types of precipitation fields are caused by the bias in precipitation intensity alone, because the threshold raw fields and threshold binary fields are the same except for the precipitation intensity above the selected threshold.

4.1 Relative influence on frequency distributions of agreement scales

The frequency distributions of $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ (Figure 9) for all precipitation grids are presented for the two types of precipitation fields at four precipitation threshold (0.1, 2, 7 and 15 mm 3h⁻¹) over the 3-27 h lead times. For the threshold raw precipitation fields, most of the $S_{ij}^{A(\overline{mm})}$ range from 5 to 10 grid points (~ 110 to 210 km), and the distributions are similar for different precipitation thresholds under both weather regimes. In terms of $S_{ij}^{A(\overline{mo})}$, the proportions of agreement scales larger than 25 grid points obviously increase with the increasing threshold, especially under SC. These results demonstrate that the spatial agreement between ensemble members is high and less affected by weather regimes and precipitation thresholds. However, with the increasing threshold, the spatial agreement between ensemble members and observations gradually decreases, especially for the SC regime. This is the reason why the spatial spread-skill relationship becomes more and more underdispersive with the increasing precipitation threshold (see supplementary material

Figure S6), which can be inferred from both types of precipitation fields.

Comparing the frequency distributions of $S_{ij}^{A(\overline{mm})}$ or $S_{ij}^{A(\overline{mo})}$ for the threshold raw and threshold binary precipitation fields, the frequencies of agreement scales with lower values change the most. By removing the bias of precipitation intensity, the frequencies of small agreement scales for the threshold binary precipitation fields are greatly enhanced than the threshold raw ones. With the increasing agreement scale and precipitation threshold, the frequency differences of $S_{ij}^{A(\overline{mm})}$ or $S_{ij}^{A(\overline{mo})}$ between these two types of precipitation fields steadily decrease, indicating that the precipitation intensity plays a more important role on the agreement scales for smaller rainfall thresholds. Indeed, the EPS can reasonably predict the spatial coverage of major rainbands during the Meiyu season, and hence the intensity becomes relatively important for the lighter precipitation (Figure 3).

4.2 Relative influence on the spatial spread-skill relationship

This section aims to investigate the relative influence of precipitation intensity on the spatial spread-skill relationship. Figure 10 shows the moving bin scatter plots of all precipitation grids for the threshold raw and binary precipitation fields at four precipitation thresholds. For the 2, 7, 15 mm 3h⁻¹ thresholds, the spatial spread-skill relationship is generally underdispersive. By contrast, the forecasts exhibit mixed overdispersion and underdispersion at the lowest threshold (0.1 mm 3h⁻¹). With the increasing threshold, the spatial agreement for both regimes decreases, suggesting that the EPS tends to provide precipitation forecasts with more insufficient spread for higher precipitation intensities. The spatial agreement under LC is generally higher than that under SC for all thresholds. Under the same weather regime at a specified threshold, the spatial agreement for threshold binary precipitation fields is higher than the corresponding threshold raw precipitation fields, because the influences of precipitation intensity bias are eliminated for the threshold binary precipitation fields. With the increasing threshold, the scale differences between these two types of precipitation fields become smaller and smaller, indicating the impact of precipitation intensity on the spatial spread-skill relationship reduces.

In order to quantitatively analyze the relative influence of precipitation intensity on the spatial spread-skill relationship, the relative skill difference percentages between the threshold binary and raw precipitation fields are defined as $s^i = \frac{s_b^i - s_r^i}{s_r^i}$, where s_b^i and s_r^i denote the skill (the ratio between averaged $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ within the i^{th} moving bin) for the threshold binary and threshold raw fields, respectively.

Figure 11 shows that skill differences between the two types of precipitation fields decrease with the increasing agreement scale and precipitation thresholds. For light (0.1mm 3h⁻¹; Figure 11a) and moderate (2mm 3h⁻¹; Figure 11b) rain, the influence of precipitation intensity under SC is larger than that under LC, especially for the smaller agreement scales. At larger agreement scales, there are no obvious differences between the two types of precipitation fields, in particular for higher rainfall thresholds (Figure 11c and 11d). With increasing agreement scales (relatively larger area), removing the

precipitation intensity bias above higher thresholds may cause negative skill difference percentage (Figure 11b-11d), because spurious rainfall cells may reduce dry intensity bias of heavy rainfall in raw precipitation forecasts and relatively increase the skill in the threshold raw field. Overall, the impacts of precipitation intensity on the spatial spread-skill relationship decrease with the increasing threshold, while the precipitation placement becomes more dominant to influence the spatial spread-skill relationship for heavy precipitation events.

4.3 Two cases studies

In addition, two cases are analyzed to provide a more intuitive understanding of the relative influence of precipitation intensity and placement on the spatial spread-skill relationship. Case 1 (initialized at 0000 UTC 27 June, 2013) and case 2 (initialized at 0000 UTC 7 July, 2013) are selected to represent LC and SC events, respectively. The relatively stationary rainband in case 1 is controlled by the Meiyu front that locates at the south of the verification domain, and it reaches the maximum rain intensity at the 27 h lead time (Figure 12). In case 2, the precipitation is mainly caused by the MCSs moving from the southwest to the northeast, and it develops to the strongest cell at the 12 h lead time (Figure 13).

For the LC case, the forecasted precipitation rainband (shaded area) over both thresholds is generally consistent with the observation (blue contours). The small agreement scales of $S_{ij}^{A(mm)}$ for the threshold raw fields (Figure 12a and 12d) are confined within the precipitation grids, indicating that the EPS is very confident about the location of this rainband controlled by the Meiyu front. For example, the smallest scale is 1.2 grid points (~ 34 km) at the $0.1 \text{ mm } 3\text{h}^{-1}$ threshold (Figure 12a). The positive values of the differences ($S_{ij}^{A(mm)} - S_{ij}^{A(mo)}$) reveal an overdispersion of the spatial spread-skill relationship, especially for larger agreement scales (Figure 12a and 12d), mixed with the underdispersion (negative values) around the rainband center with smaller agreement scales at both thresholds (Figure 12b and 12e). As shown in the Sections 4.1 and 4.2, the larger the differences of $S_{ij}^{A(mm)}$ between the threshold raw and binary fields, the stronger the relative impact of precipitation intensity on agreement scales. Apparently, the LC case demonstrates that the relative influence of precipitation intensity decreases with the increasing precipitation threshold (Figure 12c and 12f).

Compared to the LC case, the coverage and displacement biases are much larger in the SC case, while multiple rain cells develop and shift fast with MCSs. Although the EPS is also confident about the location of most rain cells with small agreement scales for this SC case (Figure 13a and 13d), the underdispersion (Figure 13 b and 13e) of the spatial spread-skill relationship for the SC case is more severe than that for the LC case at both thresholds. The statistics for different lead times (Figure 8b and 8c) also suggests that the spatial spread-skill relationship is less underdispersive and more reasonable under LC than that under SC. Similar to the LC case, the precipitation placement gradually becomes a dominant factor to affect the spatial spread-skill relationship at higher precipitation thresholds (Figure 13c and 13f).

5. Summary and conclusions

In this study, a new neighbourhood-based method in terms of agreement scale (Dey et al., 2016a, b) is applied to verify the spatial spread-skill relationship of precipitation forecasts in a convection-allowing EPS. The precipitation forecasts during the Meiyu season from 23 June to 22 July 2013 are produced by this EPS with a 3-km grid spacing over the Yangtze-Huaihe river basin of China. Since the forecast skill highly depends on weather regime, precipitation fractional coverage over the verification domain is used to classify the total 30 cases into two categories: the large-coverage (LC) and small-coverage (SC) regimes. This study also proposes a new application of this new neighbourhood-based method, which further investigates the relative influence and importance of precipitation placement and intensity on the spatial spread-skill relationship of precipitation forecasts.

The verification results show that for the 30 days considered here, the spatial spread-skill relationships under both LC and SC are underdispersive, except that for the light precipitation at the $0.1 \text{ mm } 3\text{h}^{-1}$ threshold. This underdispersion deteriorates with increasing precipitation threshold, suggesting more difficulties in predicting the location of rain for higher precipitation thresholds. Also, the spatial spread-skill relationship under SC is poorer and has more diurnal variations compared to that under LC. In addition, this study compares the traditional and spatial spread-skill relationship. With the traditional scale-independent approach, spread-skill relationship is evaluated over the entire verification domain, scale-dependent behaviors that are common with localized precipitation cannot be revealed. The scale-dependent neighbourhood-based method examined in this paper provides a remedy.

This study further investigates the relative influence of precipitation intensity and placement on agreement scales and spatial spread-skill relationship. By transforming the precipitation fields to the threshold raw and binary fields, the neighbourhood-based method is able to compare the relative impact of the precipitation intensity and placement on the spatial spread-skill relationship. The results indicate that the effect of precipitation intensity on the spatial spread-skill relationship concentrates on the lower agreement scales (closer to the center of precipitation), and decreases with the increasing precipitation threshold and agreement scale. The relative skill difference percentages (Figure 11) and two case studies (Figure 12 and 13) reveal that the precipitation placement gradually dominates the spatial spread-skill relationship with the increasing precipitation threshold. In particular, the precipitation intensity has much less influence on the spatial spread-skill relationship under SC at higher thresholds, consistent with the increasing difficulty of precipitation forecasting for local heavy rain. The impact of precipitation placement on the spatial spread-skill relationship is more important than precipitation intensity, especially for heavy precipitation and the SC cases.

Although this study analyze the relative influence of precipitation intensity and placement quantitatively, other impact factors such as precipitation structure are also worthy of investigations in the future. There are other limitations in this study. The resolution of the merged CMORPH precipitation data is limited to 0.1 degree ($\sim 10 \text{ km}$)

which is the best gridded data currently available in China. Because of that, spatial spread-skill relationship for scales below the 10 km cannot be examined. Higher-resolution data are desirable. In addition, many studies (Wang et al., 2014; Johnson and Wang, 2016) have emphasized the importance of initial small-scale information on the forecasts in a convection-allowing EPS. While it is simple to drive initial conditions of a convection-allowing EPS by downscaling a global ensemble (Peralta et al., 2012; Zhu and Xue, 2016), small-scale initial perturbations are missing from the global ensemble. This is perhaps another reason for the relatively poor spatial spread-skill relationship under SC. In future versions of our convection-allowing EPS, small-scale initial perturbations will be implemented and the spatial spread-skill relationship will be evaluated to assess the benefit of the improved initial perturbations.

Acknowledgements: This study received support from National Basic Research Program of China (973 Program) (2013CB430106), the R&D Special Fund for Public Welfare Industry (Meteorology) (GYHY201206005), and Natural Science Foundation of China (41675109, 41175087). We also thank the support of Jiangsu Collaborative Innovation Center for Climate Change, China, and High Performance Computing Center of Nanjing University.

Supporting Information

As the supplement to the main manuscript, we provide the following materials:

1. The effect of systematic intensity biases (Figure S1), S_{lim} (Figure S2) and moving bin size (Figure S3) on the spatial spread-skill relationship.
2. The spatial spread-skill relationship of 3 h accumulated precipitation at different lead times for 8 cases randomly selected from the 22 LC cases (Figure S4).
3. The spatial distribution of $S_{ij}^{A(\overline{m}\overline{o})}$ (Figure S5), compared with $S_{ij}^{A(\overline{m}\overline{m})}$ (Figure 4 in the manuscript).
4. The spatial spread-skill relationship at different precipitation thresholds (Figure S6).

References:

- Baldauf, M., A. Seifert, J. Forstner, D. Majewski, M. Raschendorfer, and T. Reinhardt. 2011. Operational Convection-allowing Numerical Weather Prediction with the COSMO Model: Description and Sensitivities. *Mon. Weather Rev.* **139**: 3887-3905. doi: 10.1175/mwr-d-10-05013.1.
- Bouttier, F., B. Vie, O. Nuissier, and L. Raynaud, 2012. Impact of Stochastic Physics in a Convection-Permitting Ensemble. *Mon. Weather Rev.* **140**: 3706-3721. doi: 10.1175/MWR-D-12-00031.1.
- Bouttier, F., L. Raynaud, O. Nuissier, and B. Ménétrier. 2016. Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. *Q. J. R. Meteorol. Soc.* **142**: 390-403. doi: 10.1002/qj.2622.
- Clark, A. J., and Coauthors. 2012. An overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bull. Amer. Meteor. Soc.* **93**: 55–74. doi: 10.1175/BAMS-D-11-00040.1.
- Clark, A. J., W. A. Gallus, M. Xue, and F. Y. Kong. 2010. Convection-Allowing and Convection-Parameterizing Ensemble Forecasts of a Mesoscale Convective Vortex and Associated Severe Weather Environment. *Wea. Forecasting*. **25**: 1052-1081. doi: 10.1175/2010waf2222390.1.
- Coniglio, M. C., K. L. Elmore, J. S. Kain, S. J. Weiss, M. Xue, and M. L. Weisman. 2010. Evaluation of WRF model output for severe-weather forecasting from the 2008 NOAA Hazardous Weather Testbed Spring Experiment. *Wea. Forecasting*. **25**: 408–427. doi: 10.1175/2009WAF2222258.1.
- Davis, C. A., B. G. Brown, and R. G. Bullock. 2006. Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.* **134**: 1772–1784. doi: 10.1175/MWR3145.1.
- Davis, C. A., B. G. Brown, R. G. Bullock, and J. Halley Gotway. 2009. The Method for Object-based Diagnostic Evaluation (MODE) applied to WRF forecasts from the 2005 Spring Program. *Wea. Forecasting*. **24**: 1252–1267. doi: 10.1175/2009WAF2222241.1.
- Dey SR, Leoncini G, Roberts NM, Plant RS, Migliorini S. 2014. A spatial view of ensemble spread in convection-permitting ensembles. *Mon. Weather Rev.* **142**: 4091–4107. doi: 10.1175/MWR-D-14-00172.1.
- Dey, S. R. A., N. M. Roberts, R. S. Plant, and S. Migliorini. 2016a. A new method for the characterization and verification of local spatial predictability for convection-allowing ensembles. *Q. J. R. Meteorol. Soc.* **142**: 1982-1996. doi: 10.1002/qj.2792.
- Dey, S. R. A., R. S. Plant, N. M. Roberts, and S. Migliorini. 2016b. Assessing spatial precipitation uncertainties in a convection-allowing ensemble. *Q. J. R. Meteorol. Soc.* **142**: 2935–2948. doi: 10.1002/qj.2893.
- Ding, Y., and J. C. Chan. 2005. The East Asian summer monsoon: An overview, *Meteorol. Atmos. Phys.* **89(1–4)**: 117–142. doi: 10.1007/s00703-005-0125-z.
- Done, J. M., G. C. Craig, S. L. Gray, P. A. Clark, and M. E. B. Gray. 2006. Mesoscale simulations of organized convection: Importance of convective equilibrium. *Q. J. R. Meteorol. Soc.* **132**: 737-756. doi: 10.1256/qj.04.84.
- Duc L, Saito K, Seko H. 2013. Spatial–temporal fractions verification for high-resolution ensemble forecasts. *Tellus A* **65**: 1–21. doi: 10.3402/tellusa.v65i0.18171.
- Ebert, E. E. 2001. Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Weather Rev.* **129**: 2461-2480. doi: 10.1175/1520-0493(2001)129<2461:aoapms>2.0.co;2.
- Ebert, E. E. 2008. Fuzzy verification of high-resolution gridded forecasts: A review and proposed

- framework. *Meteorol. Appl.* **15**: 51–64. doi: 10.1002/met.25.
- Ebert, E. E. 2009. Neighborhood verification: A strategy for rewarding close forecasts. *Wea. Forecasting*. **24**: 1498–1510. doi: 10.1175/2009WAF2222251.1.
- Ebert, E. E., and J. L. McBride. 2000. Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol.* **239**: 179–202. doi: 10.1016/S0022-1694(00)00343-7.
- Gilleland E, Ahijevych D, Brown BG, Casati B, Ebert EE. 2009. Intercomparison of spatial forecast verification methods. *Weather and Forecasting* **24**: 1416–1430. doi: 10.1175/2009WAF2222269.1.
- Grimit, E. P., and C. F. Mass. 2007. Measuring the ensemble spread-error relationship with a probabilistic approach: Stochastic ensemble results. *Mon. Weather Rev.* **135**: 203–221. doi: 10.1175/mwr3262.1.
- Harnisch, F., and C. Keil. 2015. Initial Conditions for Convection-allowing Ensemble Forecasting Provided by Ensemble Data Assimilation. *Mon. Weather Rev.* **143**: 1583–1600. doi: 10.1175/mwr-d-14-00209.1.
- Hohenegger, C., and C. Schar. 2007. Atmospheric predictability at synoptic versus cloud-resolving scales. *Bull. Amer. Meteor. Soc.* **88**: 1783–1793. doi: 10.1175/bams-88-11-1783.
- Hou, D., Z. Toth, Y. Zhu, and W. Yang. 2008. Impact of a stochastic perturbation scheme on global ensemble forecast. Preprint, *the 19th AMS conference on probability and statistics*, Am. Meteorol. Soc., New Orleans, Louisiana.
- Iyer, E., A. Clark, M. Xue, and F. Kong. 2016. A Comparison of 36–60-h Precipitation Forecasts from Convection-Allowing and Convection-Parameterizing Ensembles. *Wea. Forecasting*. **31**: 647–661. doi: 10.1175/WAF-D-15-0143.1.
- Johnson, A., and X. Wang. 2016. A Study of Multiscale Initial Condition Perturbation Methods for Convection-Permitting Ensemble Forecasts. *Mon. Weather Rev.* **144**: 2579–2604. doi: 10.1175/mwr-d-16-0056.1.
- Katragkou, E., et al. 2015. Regional climate hindcast simulations within EURO-CORDEX: Evaluation of a WRF multi-physics ensemble. *Geosci. Model Dev.* **8**: 603–618. doi: 10.5194/gmd-8-603-2015.
- Kawabata, T., and Coauthors. 2007. An assimilation and forecasting experiment of the nerima heavy rainfall with a cloud-resolving nonhydrostatic 4-dimensional variational data assimilation system. *J. Meteorol. Soc. Jpn.* **85**: 255–276. doi: 10.2151/jmsj.85.255.
- Keil, C., and G. C. Craig. 2011. Regime-dependent forecast uncertainty of convective precipitation. *Meteorol. Z.* **20**: 145–151. doi: 10.1127/0941-2948/2011/0219.
- Keil, C., F. Heinlein, and G. C. Craig. 2014. The convective adjustment time-scale as indicator of predictability of convective precipitation. *Q. J. R. Meteorol. Soc.* **140**: 480–490. doi: 10.1002/qj.2143.
- Kuhnlein, C., C. Keil, G. C. Craig, and C. Gebhardt. 2014. The impact of downscaled initial condition perturbations on convection-allowing ensemble forecasts of precipitation. *Q. J. R. Meteorol. Soc.* **140**: 1552–1562. doi: 10.1002/Qj.2238.
- Lawson, J., and W. A. Gallus, Jr. 2016. On Contrasting Ensemble Simulations of Two Great Plains Bow Echoes. *Wea. Forecasting*. **31**: 787–810. doi: 10.1175/waf-d-15-0060.1.
- Li, G., Y. Wang, K.-H. Lee, Y. Diao, and R. Zhang. 2009. Impacts of aerosols on the development and precipitation of a mesoscale squall line. *J. Geophys. Res.* **114**: D17205. doi: 10.1029/2008JD011581.
- Luo, Y. L., and Y. R. X. Chen. 2015. Investigation of the predictability and physical mechanisms of an extreme-rainfall-producing mesoscale convective system along the Meiyu front in East China: An ensemble approach. *J. Geophys. Res.-Atmos.* **120**: 10593–10618. doi: 10.1002/2015jd023584.
- Martin, M. L., D.Santos-Munoz, F.Valero, and A. Morata. 2010. Evaluation of an ensemble precipitation

- prediction system over the Western Mediterranean area. *Atmos. Res.* **98**: 163-175. doi: 10.1016/j.atmosres.2010.07.002.
- Melhauser, C., and F. Q. Zhang. 2012. Practical and Intrinsic Predictability of Severe and Convective Weather at the Mesoscales. *J. Atmos. Sci.* **69**: 3350-3371. doi: 10.1175/jas-d-11-0315.1.
- Mittermaier, M., N. Roberts, and S. A. Thompson. 2013. A long-term assessment of precipitation forecast skill using the Fractions Skill Score. *Meteorol. Appl.* **20**: 176-186. doi: 10.1002/met.296.
- Nuissier, O., B. Joly, B. Vie, and V. Ducrocq. 2012. Uncertainty of lateral boundary conditions in a convection-permitting ensemble: a strategy of selection for Mediterranean heavy precipitation events. *Nat. Hazards Earth Syst. Sci.* **12**: 2993-3011. doi: 10.5194/nhess-12-2993-2012.
- Peralta, C., Z. Ben Bouallegue, S. E. Theis, C. Gebhardt, and M. Buchhold. 2012. Accounting for initial condition uncertainties in COSMO-DE-EPS. *J. Geophys. Res.-Atmos.* **117**: D07108. doi: 10.1029/2011jd016581.
- Roberts, N. 2008. Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteorol. Appl.*, **15**: 163-169. doi: 10.1002/met.57.
- Roberts, N. M., and H. W. Lean. 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Weather Rev.* **136**: 78-97. doi: 10.1175/2007mwr2123.1.
- Schwartz, C. S., G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman. 2015. NCAR's Experimental Real-Time Convection-Allowing Ensemble Prediction System. *Wea. Forecasting.* **30**: 1645-1654. doi: 10.1175/waf-d-15-0103.1.
- Seity, Y., and Coauthors. 2011. The AROME-France Convection-allowing Operational Model. *Mon. Weather Rev.* **139**: 976-991. doi: 10.1175/2010mwr3425.1.
- Shen, Y., Zhao, P., Pan, Y., Yu, J. 2014. A high spatiotemporal gauge-satellite merged precipitation analysis over China. *J. Geophys. Res.-Atmos.* **119**: 3063-3075. doi: 10.1002/2013jd020686.
- Su, X., H. Yuan, Y. Zhu, Y. Luo, and Y. Wang. 2014. Evaluation of TIGGE ensemble predictions of Northern Hemisphere summer precipitation during 2008-2012. *J. Geophys. Res.-Atmos.* **119**: 7292-7310. doi: 10.1002/2014jd021733.
- Surcel, M., I. Zawadzki, and M. K. Yau. 2015. A Study on the Scale Dependence of the Predictability of Precipitation Patterns. *J. Atmos. Sci.* **72**: 216-235. doi: 10.1175/JAS-D-14-0071.1.
- , 2016. The Case-to-Case Variability of the Predictability of Precipitation by a Storm-Scale Ensemble Forecasting System. *Mon. Weather Rev.* **144**: 193-212. doi: 10.1175/mwr-d-15-0232.1.
- Tennant, W. 2015. Improving initial condition perturbations for MOGREPS-UK. *Q. J. R. Meteorol. Soc.* **141**: 2324-2336. doi: 10.1002/qj.2524.
- Vie, B., O. Nuissier, and V. Ducrocq. 2011. Cloud-Resolving Ensemble Simulations of Mediterranean Heavy Precipitating Events: Uncertainty on Initial Conditions and Lateral Boundary Conditions. *Mon. Weather Rev.* **139**: 403-423. doi: 10.1175/2010mwr3487.1.
- Wang, H., and Z. Zhong. 2014. Ensemble simulations to investigate the impact of large-scale urbanization on precipitation in the lower reaches of Yangtze River Valley, China. *Q. J. R. Meteorol. Soc.* **140**: 258-266. doi: 10.1002/qj.2125.
- Wang, Y., M. Bellus, J. F. Geleyn, X. L. Ma, W. H. Tian, and F. Weidle. 2014. A New Method for Generating Initial Condition Perturbations in a Regional Ensemble Prediction System: Blending. *Mon. Weather Rev.* **142**: 2043-2059. doi: 10.1175/Mwr-D-12-00354.1.
- Wei, M., Z. Toth, R. Wobus, and Y. Zhu. 2008. Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus.* **60A**: 62-79.

doi:10.1111/j.1600-0870.2007.00273.x.

- Zacharov P, Rezacova D. 2009. Using the fractions skill score to assess the relationship between an ensemble QPF spread and skill. **94**: 684–693. doi: 10.1016/j.atmosres.2009.03.004.
- Zhang, F. Q., N. F. Bei, R. Rotunno, C. Snyder, and C. C. Epifanio. 2007. Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *J. Atmos. Sci.* **64**: 3579-3594. doi: 10.1175/jas4028.1.
- Zhu, K., and M. Xue. 2016. Evaluation of WRF-based convection-permitting multi-physics ensemble forecasts over China for an extreme rainfall event on 21 July 2012 in Beijing. *Adv. Atmos. Sci.* **33**: 1240-1258. doi: 10.1007/s00376-016-6202-z.
- Zimmer, M., G. C. Craig, C. Keil, and H. Wernli. 2011. Classification of precipitation events with a convective response timescale and their forecasting characteristics. *Geophys Res Lett.* **38**: L05802. doi: 10.1029/2010GL046199.

Table 1. Model configurations of the convective-scale EPS with different microphysical, land surface and planetary boundary layer schemes.

Member	MP	LSM	PBL
Gec00	Thompson	Noah	MYJ
Gep01	Lin	RUC	YSU
Gep02	M-Y	NOAH	QNSE
Gep03	Morrison	NOAH	MYNN2
Gep04	WDM6	RUC	ACM2
Gep05	Lin	NOAH	MYJ
Gep06	M-Y	NOAH	YSU
Gep07	Morrison	RUC	QNSE
Gep08	WDM6	NOAH	MYNN2
Gep09	Thompson	RUC	ACM2
Gep10	M-Y	RUC	MYJ
Gep11	Morrison	NOAH	YSU
Gep12	WDM6	NOAH	QNSE
Gep13	Thompson	RUC	MYNN2
Gep14	Lin	RUC	ACM2

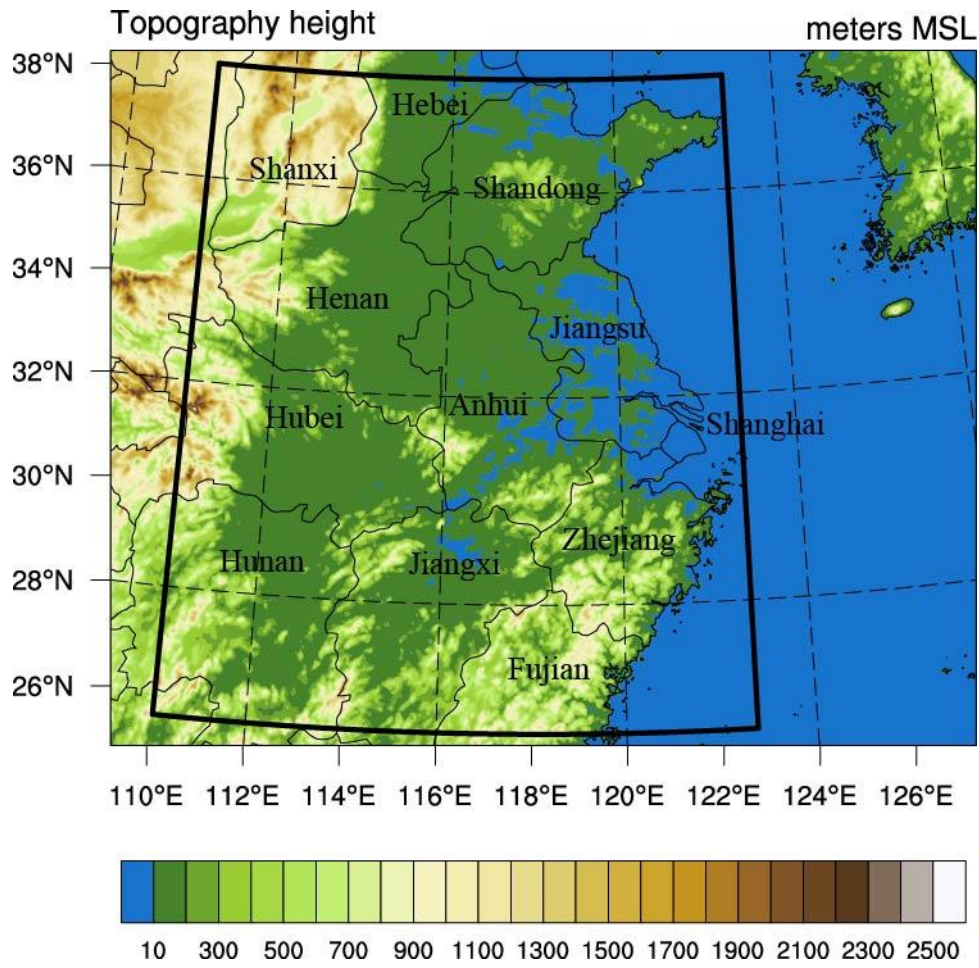


Figure 1. Domain of the convective-scale EPS with coastlines, province boundaries and topography heights, and the verification domain of precipitation forecasts (inside thick solid lines).

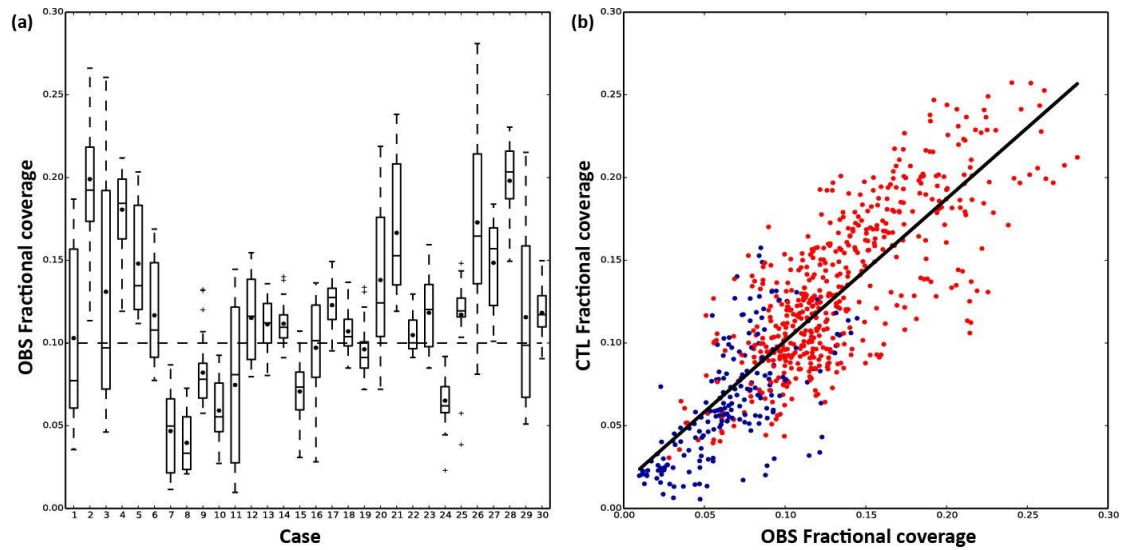


Figure 2. (a) Boxplot of hourly precipitation coverage of the merged CMORPH observations for 30 cases (the dot stands for the mean fractional coverage of each case),

and (b) scatterplot of hourly precipitation coverage of the merged CMORPH observations vs. the control forecasts for the SC (blue) and LC cases (red).

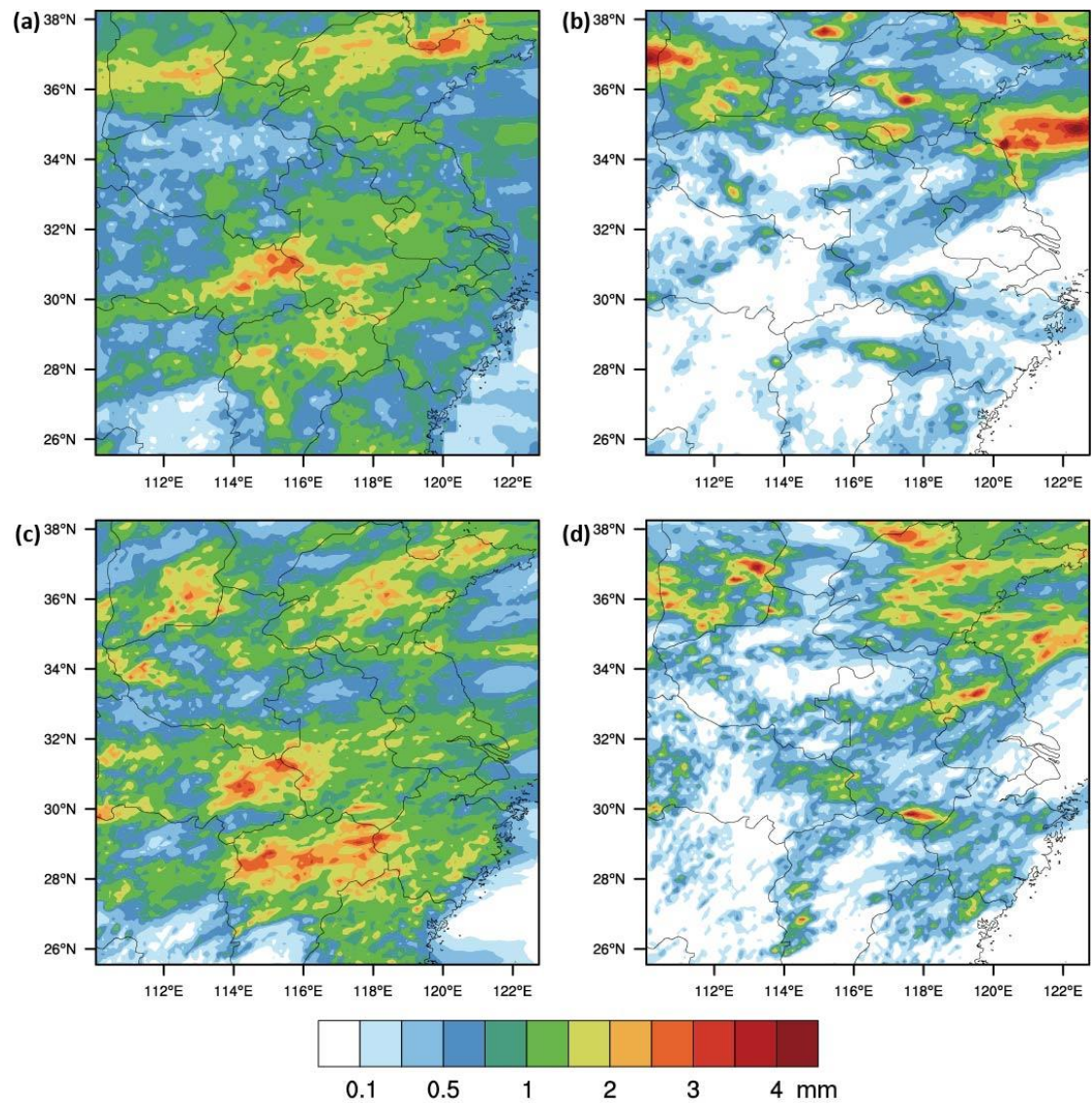


Figure 3. Spatial distributions of 3h accumulated precipitation averaged from (a, b) the merged CMORPH observations and (c, d) the control forecasts over the 3-27 h lead times for (a, c) the LC cases and (b, d) the SC cases.

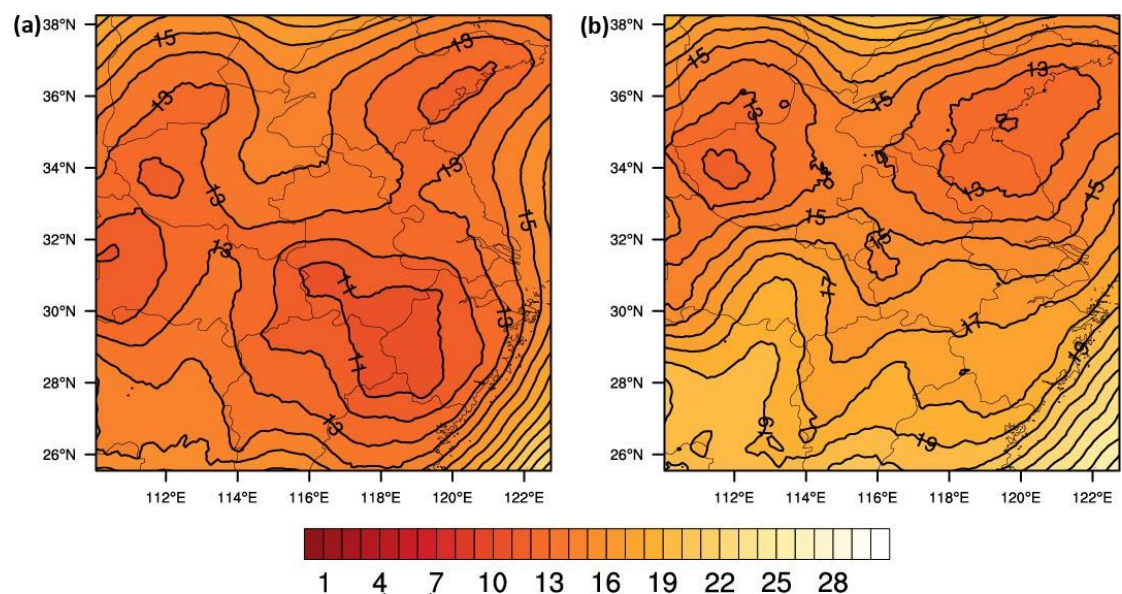


Figure 4. Scale maps of $S_{ij}^{A(mm)}$ averaged over the 3-27 h lead times for (a) the LC cases and (b) the SC cases.

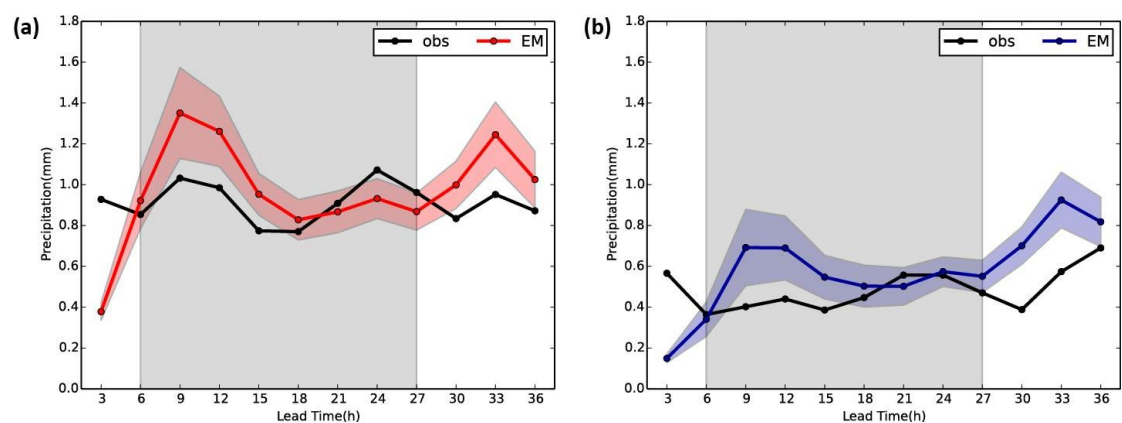


Figure 5. Diurnal variations of domain averaged 3 h accumulated precipitation observations and ensemble mean forecasts for (a) the LC cases and (b) the SC cases. The shadow stands for the standard deviation of ensemble members.

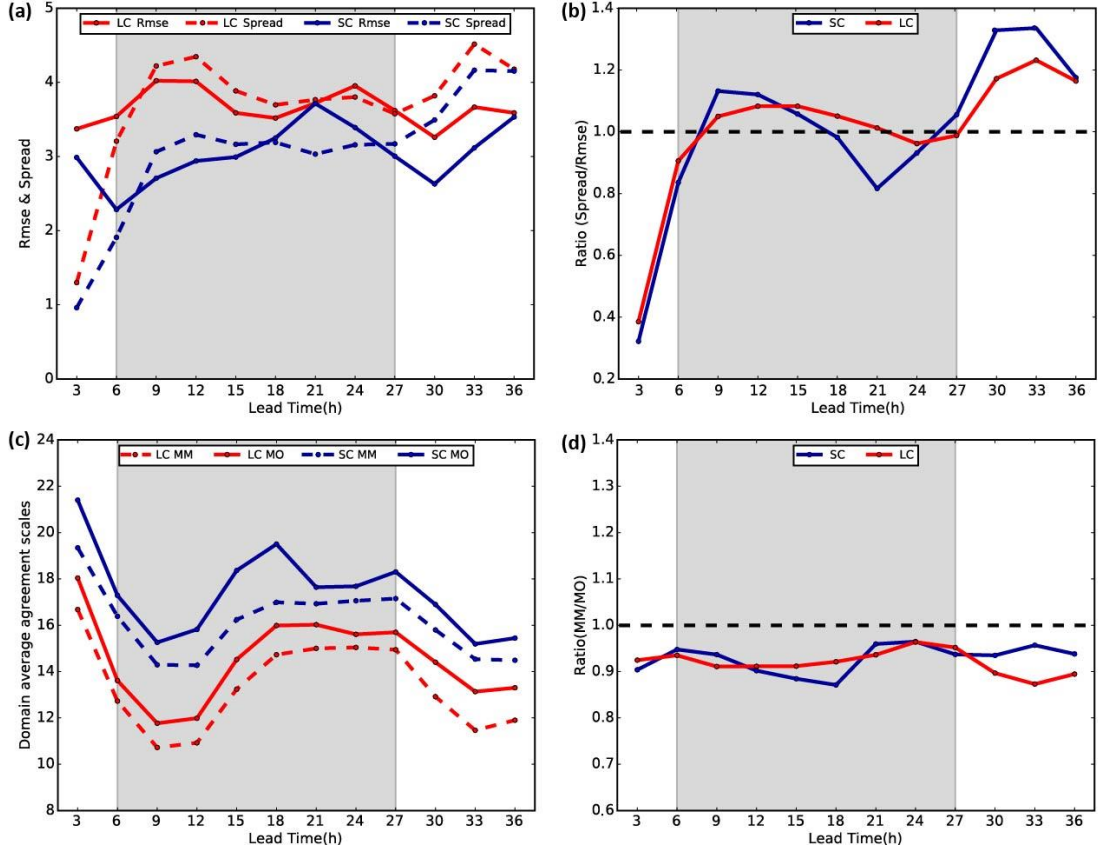


Figure 6. Diurnal variations of domain averaged (a) RMSE and spread for 3 h accumulated precipitation, (b) ratios of spread and RMSE, (c) $S_{ij}^{A(mm)}$ and $S_{ij}^{A(mo)}$, and (d) the ratios of $S_{ij}^{A(mm)}$ and $S_{ij}^{A(mo)}$ under two weather regimes.

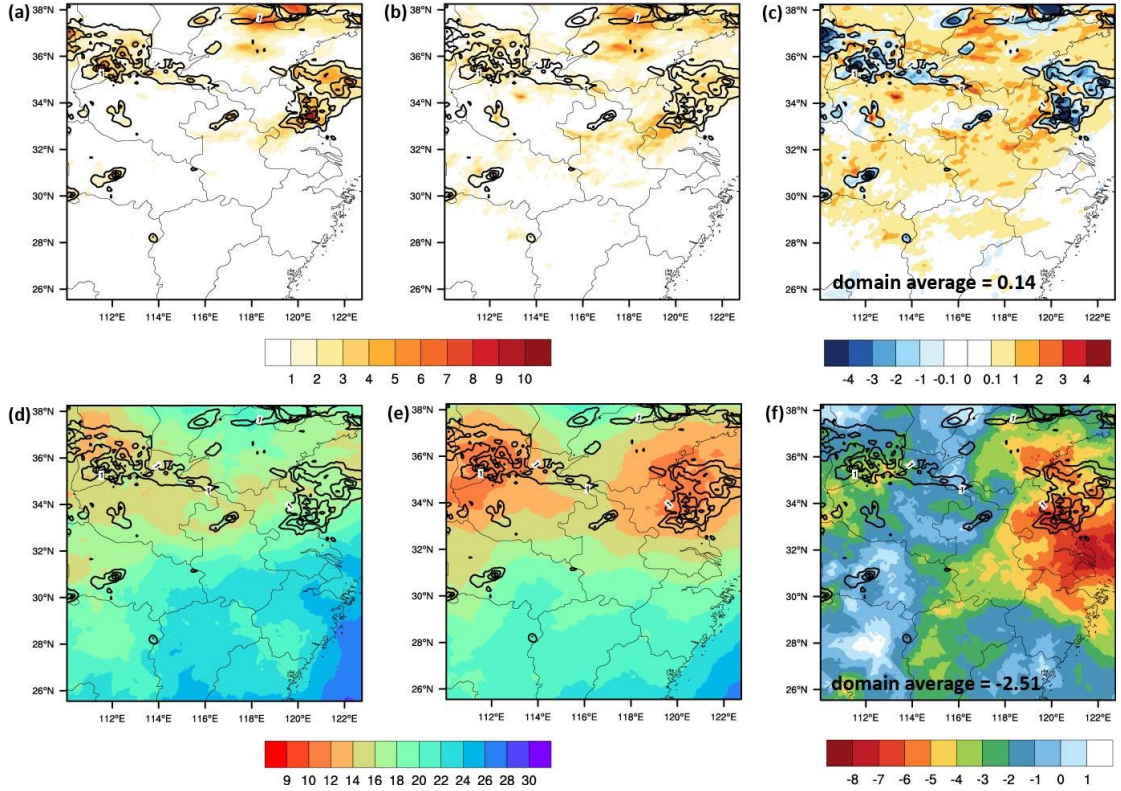


Figure 7. Spatial distributions of (a) RMSE, (b) spread, (c) the differences between spread and RMSE, (d) $S_{ij}^{A(\overline{m}\overline{o})}$, (e) $S_{ij}^{A(\overline{m}\overline{m})}$, and (f) the difference between $S_{ij}^{A(\overline{m}\overline{m})}$ and $S_{ij}^{A(\overline{m}\overline{o})}$ for 3 h accumulated precipitation averaged for the SC cases at the 18 h lead time. The bold black contours represent the observations ranging from 1 to 7 mm with the 2 mm interval.

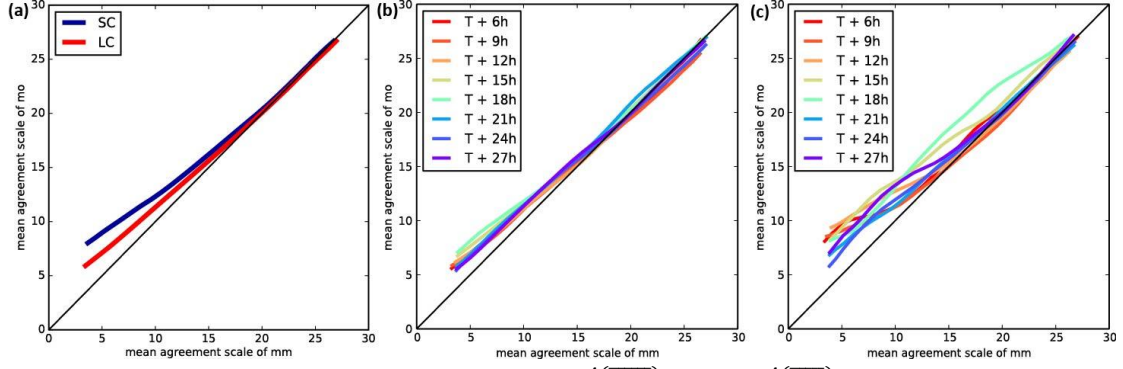


Figure 8. Moving binned scatterplots of $S_{ij}^{A(\overline{m}\overline{m})}$ and $S_{ij}^{A(\overline{m}\overline{o})}$ of 3 h accumulated precipitation for (a) the 3-27 h lead times, and different lead times under the (b) LC and (c) SC weather regimes.

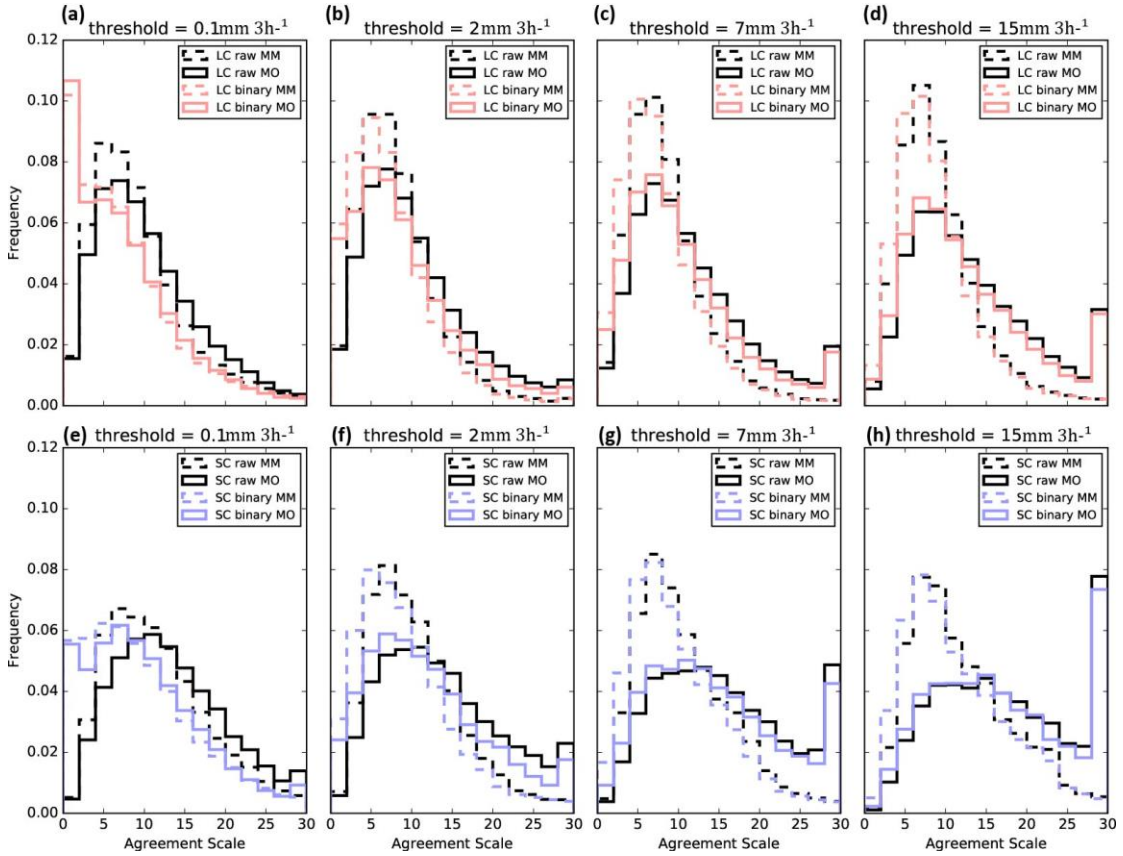


Figure 9. Frequency histograms of $S_{ij}^{A(\overline{m}\overline{o})}$ for the threshold raw (black solid) and binary (colored solid) precipitation fields, and $S_{ij}^{A(\overline{m}\overline{m})}$ for the threshold raw (black dash) and binary (colored dash) precipitation fields, over the 3-27 h lead times at the thresholds of (a), (e) 0.1 mm 3h⁻¹, (b), (f) 2 mm 3h⁻¹, (c), (g) 7 mm 3h⁻¹, and (d), (h) 15 mm 3h⁻¹ under the (a-d) LC and (e-h) SC weather regimes.

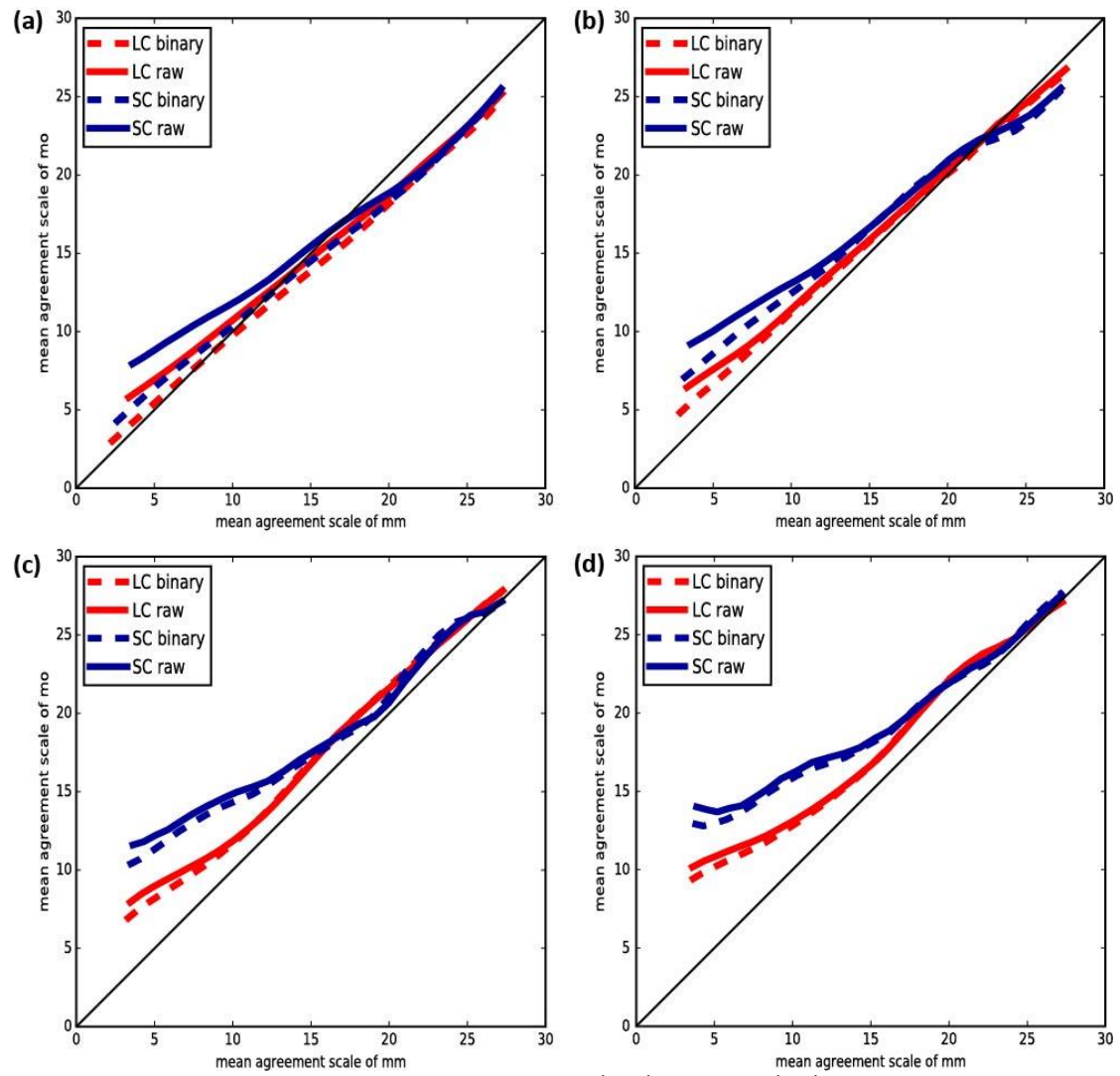


Figure 10. Moving binned scatterplots of $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ for the threshold raw and binary precipitation fields over the 3-27 h lead times at the thresholds of (a) 0.1 mm 3h⁻¹, (b) 2 mm 3h⁻¹, (c) 7 mm 3h⁻¹, and (d) 15 mm 3h⁻¹.

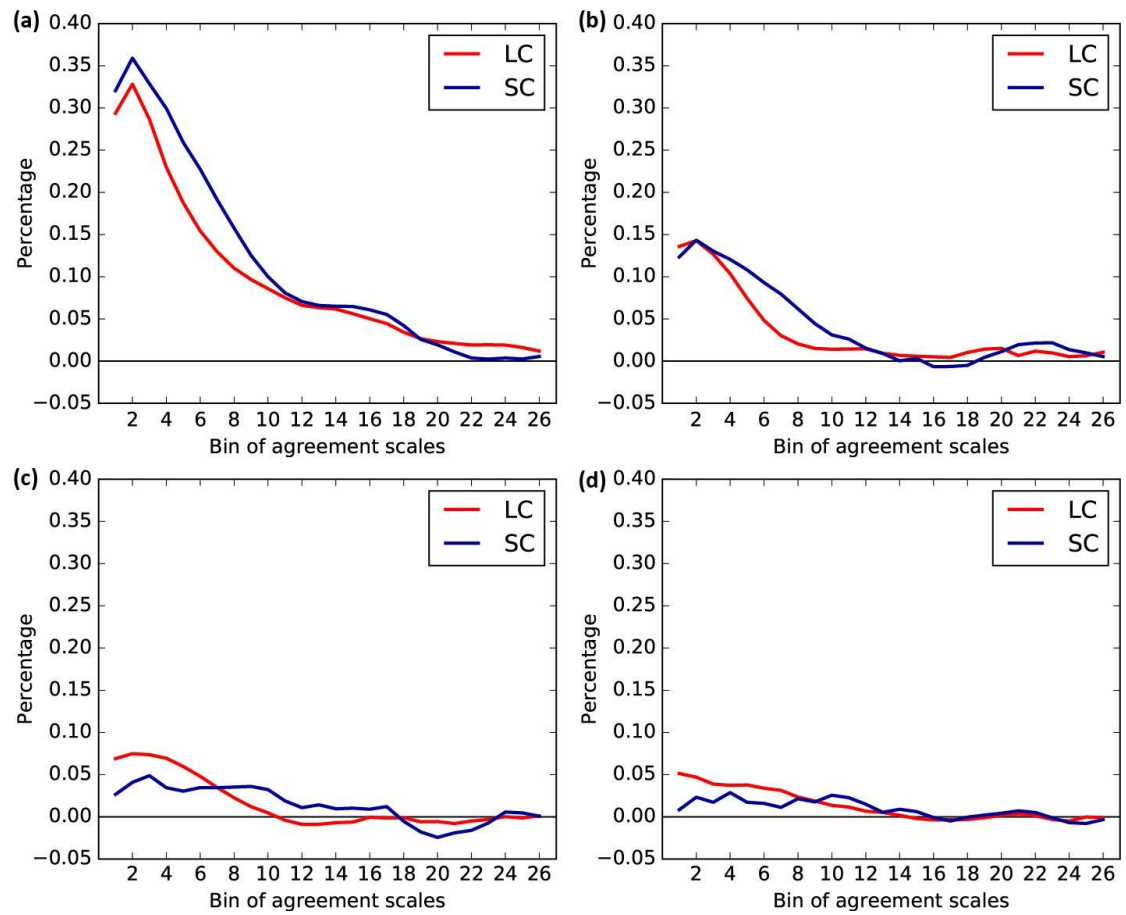


Figure 11. Relative skill difference percentages between the threshold binary and raw precipitation fields over the 3-27 h lead times at the thresholds of (a) 0.1 mm 3h⁻¹, (b) 2 mm 3h⁻¹, (c) 7 mm 3h⁻¹, and (d) 15 mm 3h⁻¹.

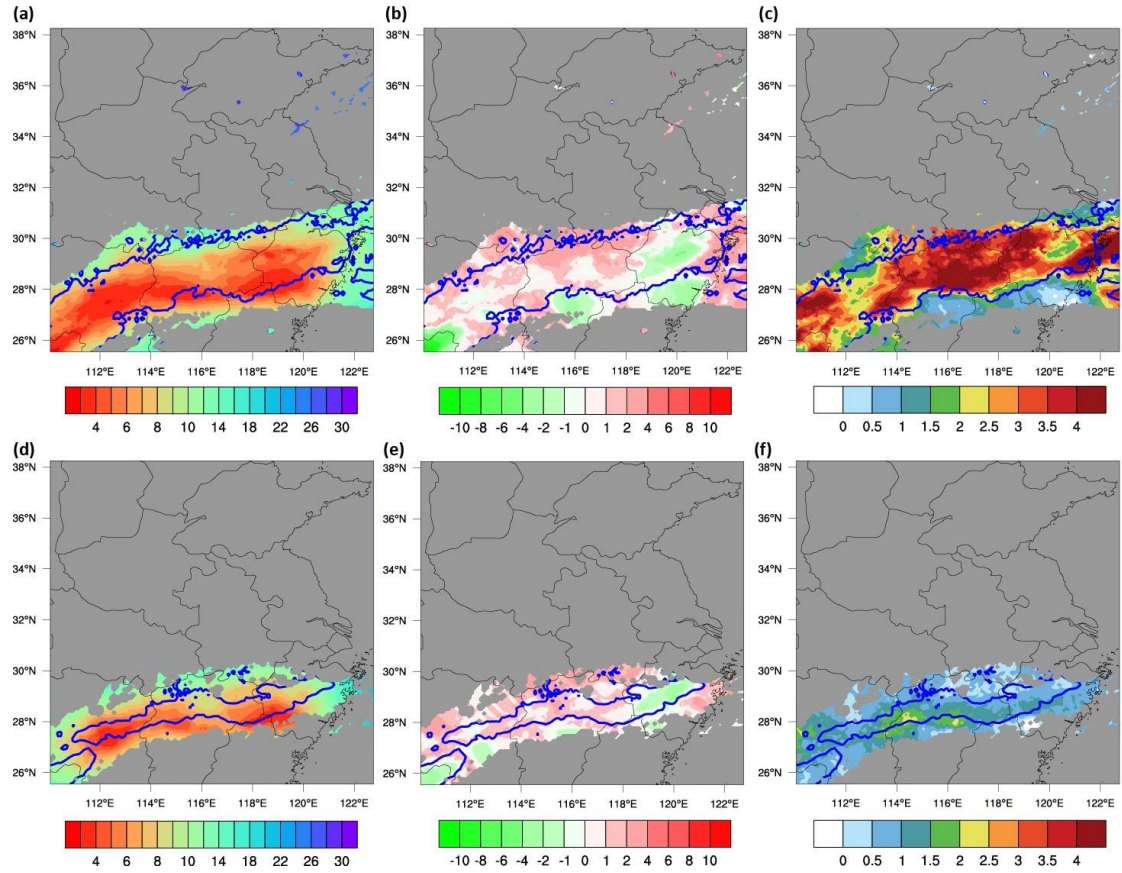


Figure 12. For case 1 (initialized at 0000 UTC 27 June, 2013) at the 27 h lead time, spatial distributions of (a, d) $S_{ij}^{A(\overline{mm})}$ and (b, e) $S_{ij}^{A(\overline{mm})} - S_{ij}^{A(\overline{m\sigma})}$ for the threshold raw fields over the thresholds of (a-c) 0.1 mm 3h⁻¹ and (d-f) 7 mm 3h⁻¹, and (c, f) the differences of $S_{ij}^{A(\overline{mm})}$ between the threshold raw and binary fields (raw-binary). The gray areas are the region below the precipitation threshold. The blue contours represent the observed precipitation threshold.

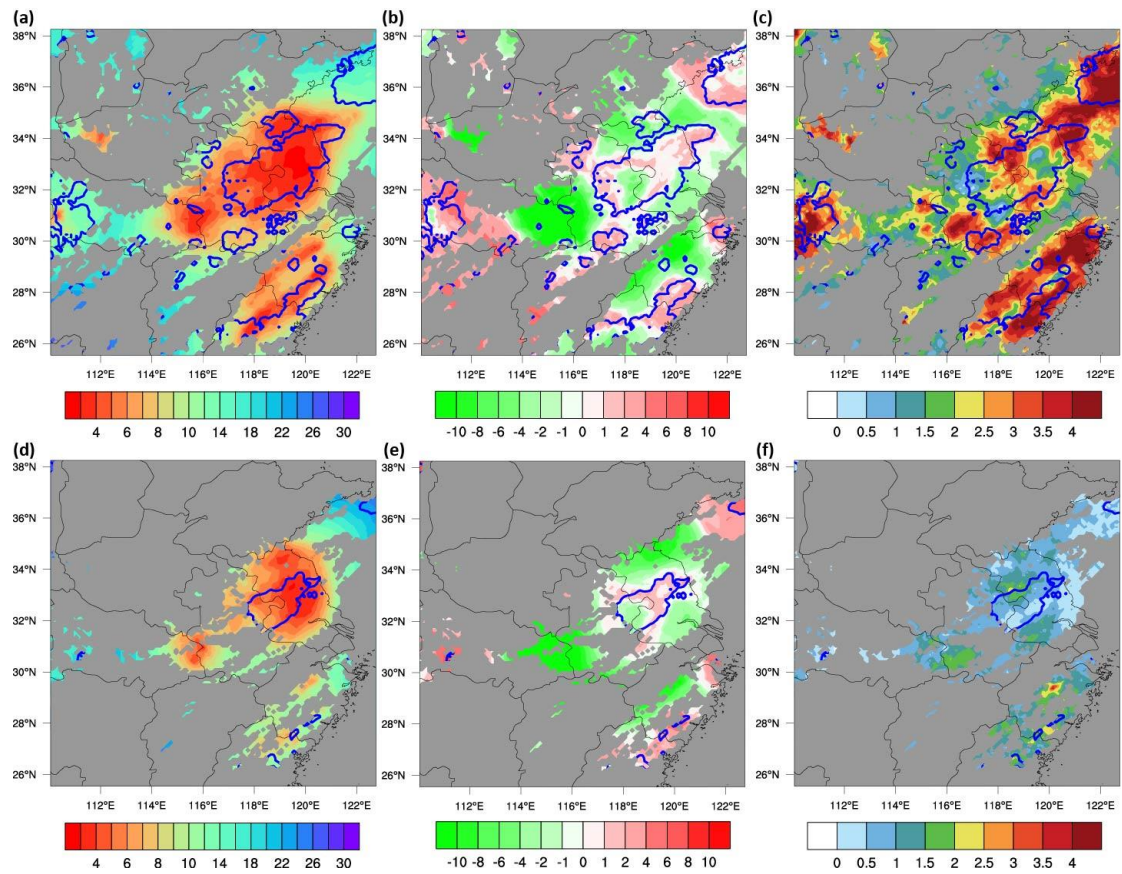


Figure 13. Same as Figure 12, but for case 2 (initialized at 0000 UTC 7 July, 2013) at the 12 h lead time.