atmosphere

MDPI

Article

SAL Method Applied in Grid Forecasting Product Verification with Three-Source Fusion Product

Debin Su [1,2], Jinhua Zhong [1,2], Yunong Xu [1,2], Linghui Lv [1,2], Honglan Liu [3], Xingang Fan [4], Lin Han [5] and Fuzeng Wang [1,2,*]

[1] College of Electronic Engineering, Chengdu University of Information Technology, Chengdu 610225, China; sudebin@cuit.edu.cn (D.S.); zhongjh@cuit.edu.cn (J.Z.); 3210305004@stu.cuit.edu.cn (Y.X.); 3210307020@stu.cuit.edu.cn (L.L.)
[2] Key Laboratory of Atmospheric Sounding, Chengdu 610225, China
[3] Zhangye Meteorological Bureau, Zhangye 734000, China; gszylhl@126.com
[4] Department of Earth, Environmental, and Atmospheric Sciences, Western Kentucky University, Bowling Green, KY 42101, USA; xingang.fan@wku.edu
[5] College of Atmospheric Sciences, Chengdu University of Information Technology, Chengdu 610225, China; hanlin@cuit.edu.cn
* Correspondence: wangfz@cuit.edu.cn

Abstract: Quantitative precipitation forecast (QPF) verification stands out as one of the most formidable endeavors in the realm of forecast verification. Traditional verification methods are not suitable for high-resolution forecasting products in some cases. Therefore, the SAL (structure, amplitude and location) method was proposed as a method of object-based spatial verification that studies precipitation verification in a certain range, which is combined with factors including structure, amplitude and location of the targets. However, the setting of the precipitation threshold would affect the result of the verification. This paper presented an improved method for determining the precipitation threshold using the QPF from ECMWF, which is an ensemble forecast model and three-source fusion product that was used in China from 1 July to 31 August 2020, and then the results obtained with this method were compared with the other two traditional methods. Furthermore, the SAL and the traditional verification methods were carried out for geometric, simulated and real cases, respectively. The results showed the following: (1) The proposed method in this paper for determining the threshold was more accurate at identifying the precipitation objects. (2) The verification area size was critical for SAL calculation. If the area selected was too large, the calculated SAL value had little significance. (3) *ME* (Mean Error) could not identify the displacement between prediction and observation, while *HSS* (Heidke Skill Score) was sensitive to the displacement of the prediction field. (4) Compared with the traditional verification methods, the SAL method was more straight forward and simple, and it could give a better representation of prediction ability. Therefore, forecasters can better understand the model prediction effect and what needs to be improved.

Keywords: precipitation forecasting; verification; SAL; threshold; three-source fusion product

Citation: Su, D.; Zhong, J.; Xu, Y.; Lv, L.; Liu, H.; Fan, X.; Han, L.; Wang, F. SAL Method Applied in Grid Forecasting Product Verification with Three-Source Fusion Product. *Atmosphere* **2024**, *15*, 1366. https://doi.org/10.3390/atmos15111366

Academic Editors: Michael L. Kaplan and Jimy Dudhia

Received: 24 July 2024
Revised: 22 October 2024
Accepted: 28 October 2024
Published: 13 November 2024

1. Introduction

The verification of quantitative precipitation forecasts (QPFs) remains one of the most challenging aspects of weather forecast verification [1]. Precipitation forecasting is particularly difficult due to the complexity of precipitation processes, including the spatial and temporal variability of rainfall. Proper verification of these forecasts is crucial, as this not only evaluates model performance but also provides insights into errors, such as their sources and distributions. These insights help improve forecasting methods and models, especially in terms of precipitation location and intensity [2,3].

In recent years, with the improvement of numerical modeling and prediction techniques, the accuracy of weather prediction has improved steadily [4,5]. A good numerical

prediction verification should not only quantitatively evaluate the performance of the prediction, but also provide error information regarding the prediction and observation, such as the source and distribution of the error [6,7]. The results from the forecast verification could provide an effective reference for forecasters to improve their forecasting skills. This was especially important for the location and intensity of precipitation forecasting [8].

Traditionally, a point-to-point verification scheme is the most commonly used method for the verification of numerical predictions. The prediction of the model was usually evaluated using statistical metrics, such as root mean square error (*RMSE*), bias (BIAS) or probability of detection (POD), false alarm ratio (FAR) and frequency bias indicator (*FBI*) scores. The *RMSE* and BIAS were sensitive to discontinuities, clutter and boundary values, while the others were sensitive to the location of events [9,10]. The point-to-point method works well in low-resolution model products, but for high-resolution model products, it may underestimate the prediction accuracy. For example, point-to-point verification often showed lower scores for precipitation and had a complex structure, and its scale was usually less than 100 km. Although the results of this method are quantitative and objective, it was very sensitive to the location and did not fully account for the complexity of precipitation in terms of its location, intensity and structure [11]. In the high-resolution numerical model's predicted precipitation, the issue of misses and false alarms was particularly serious, which was known as the "double penalty" issue [8]. This method also lacked direct verification of the precipitation area shape, structure and location, which was exactly the problem that needs attention when using the model predictions. Given the importance and the difficulty of this, many spatial precipitation verification methods have been proposed.

The spatial verification techniques could generally be divided into four categories, in which the techniques of neighborhood and scale separation could be described as filtering methods, and the other two techniques of feature-based and field deformation could be broadly thought of as displacement methods [12].

The most commonly used neighborhood methods include Upscaling [13], Minimum coverage [14], Fractions skill score [15], etc., which have proved to be effective in improving the "double penalty" phenomenon to some extent. One kind of scale decomposition technology, intensity scale [16], was specifically useful for rainfall verification, which provided a way of evaluating the forecast skill as a function of precipitation rate and spatial scale, but the calculation process was more complicated.

The displacement method includes the field deformation method and object-based feature verification [12]. The difference between the two methods was that the object-based approaches analyze each feature separately, whereas the field deformation approaches analyze the entire field.

The latest field decomposition method includes Keil and Craig's improved method of FQM (forecast quality measure), called DAS (displacement and amplitude score) [17], which allowed false alarms to be handled correctly, whereas the FQM does not [17]. Similarly, this kind of method comprises a large amount of calculation and can only give an overall evaluation. At present, the commonly used object-oriented feature verification methods include MODE (method for object-based diagnostic evaluation) [18,19], SAL (structure, amplitude and location) [20], CRA (continuous rain area) [8], etc. These methods pay attention to the differences in structure and location between prediction and observation fields in a limited identification area. Some studies [21–25] showed that the CRA method has high computational complexity, and the determination of relevant parameters of the MODE method formula was highly subject [18,19,26], while the SAL method has more scientific results, and the calculation process is relatively simple. SAL is a spatial verification method developed by Wernli et al. [20] to assess high-resolution NWP precipitation forecasts. The method evaluates three primary attributes of both observed and simulated precipitation: (1) the amplitude A, which refers to the total precipitation across the domain; (2) the location L, indicating where the precipitation occurs; and (3) the structure S, defining the size and shape of the precipitation areas. Precipitation areas, also referred to as features, are defined as contiguous regions where precipitation exceeds a specified threshold.

SAL spatial verification technology [20] was applied to the verification of the precipitation forecast in the Elbe River basin in Germany. This method belonged to the object-based spatial verification method to study precipitation targets in a certain region. According to its connectivity, the precipitation area in this region was divided into different precipitation objects, which then could be combined to calculate the structure (S), amplitude (A) and location (L) of the precipitation objects, and the value ranges for these components were [−2,2], [−2,2] and [0,2], respectively.

The determination of precipitation objects was the basis for the accurate calculation of SAL. Therefore, an appropriate threshold had a great impact on the calculation results. Previous studies had shown that there were still some defects in the determination of threshold; for example, the threshold determined by the precipitation body determination scheme proposed by Wernli [20] was too large, and some of the main precipitation information was ignored. Therefore, this paper will propose an improved SAL method and compare this method with the original one and the traditional verification method as well, so as to provide more accurate and objective error information for forecasters.

The data and the study area are presented in Section 2, including the SAL verification method in detail and the improved strategy for threshold selection. Section 3 presents the experimental results under geometric, simulated and real precipitation scenarios, comparing these results with traditional verification methods. Finally, Section 4 summarizes the main findings of this study and discusses potential future research directions.

2. Data and Methodology

2.1. Data Description

The data used were taken from the 24 h precipitation forecast results of the ECMWF ERA5 ensemble forecast model (0.5° × 0.5°) from 1 July to 31 August 2020, and the region around China (70~140° E, 15~55° N) was selected as the analysis area. The parameter of total precipitation of ERA5 was the accumulated liquid and frozen water, comprising rain and snow. It was the sum of large-scale precipitation and convective precipitation and did not include fog, dew or the precipitation that evaporates in the atmosphere before it lands on the surface of the Earth.

The gridded observation, developed by the National Meteorological Information Center [27], was a three-source fusion precipitation product with a spatial resolution of 5 km and temporal resolution of 1 h, which used the Bayesian fusion method to ingest the regional ground rainfall observation (more than 30,000 stations in China), the high-resolution QPE product from radar and satellite-retrieved precipitation.

Then, the EC precipitation field and the observation field were interpolated into a specified resolution as 0.25° for analysis. This is done to ensure that the number of grid points in the observational field matches that of the EC precipitation field. Secondly, normalization is performed, which means unifying the length of effective precipitation values in both the observational field and the EC precipitation field. For grid points without observational precipitation data, the model forecast values are set as missing, thereby unifying the meaningful data range of both observations and forecasts. This ensures that all data participate in the verification in the same format and within the same non-missing data range. The numerical forecast and the observational field are shown in Figure 1.

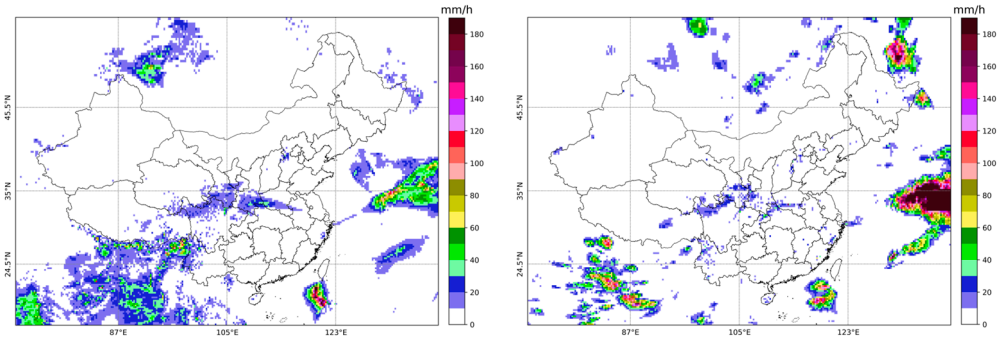

Figure 1. Precipitation accumulation over 24 h. from (**a**) numerical forecast and (**b**) observation at 08:00 on 14 July 2020, UTC.

2.2. Methods

2.2.1. SAL Metrics

SAL is a kind of object-oriented verification method. Based on the distribution, and according to certain criteria, the precipitation within the forecasted and observed areas is divided into different precipitation components (main objects and individual objects). Then, the forecast of the main/individual precipitation objects is verified based on three key forecast elements: amplitude (A) or intensity, location (L) and structure (S). Among these, the A value reflects the deviation between the forecasted and observed main/individual precipitation objects; the L value typically measures the distance between the centroids of the forecasted and observed main precipitation objects, as well as the distance between the centroid of each individual object and the center of the main object; the S value comprehensively reflects the deviation of central values and sizes of all individual objects between the forecasted and observed precipitation. The final verification results are three values for S, A and L, with ranges of [−2,2], [−2,2] and [0,2], respectively. When all three values become 0, it indicates the best forecast performance. This method needed to identify precipitation objects and an area D (referring to the area that was connected and the grid point with values that were bigger than the threshold), and then to describe the amplitude, location and structure of all precipitation objects. Similar to the CRA method, three components were given to represent objects of interest, but compared with the CRA method, it did not need to match each precipitation object in the region one-to-one, and the operation was relatively simple. This method can not only avoid the "double penalty" phenomenon, but also enable forecasters to understand the essential differences between forecasting and observation in terms of structure, location and amplitude.

a. Structure (s)

$$S = \frac{V(R_{mod}) - V(R_{obs})}{0.5[V(R_{mod}) + V(R_{obs})]} \quad (1)$$

where $V(R) = \frac{\sum_{n=1}^{m} R_n v_n}{\sum_{n=1}^{m} R_n}$, $V_n = \sum_{(i,j)\in R_n} R_{i,\,j} / R_n^{max} = R_n / R_n^{max}$.

$R_{i,j}$ represents the precipitation at grid point (i, j), R_n^{max} represents the maximum precipitation within object n, R_n represents the total precipitation within object n, Vn is the ratio of the total precipitation to the maximum in the nth object, and $V(R)$ is the weighted average of Vn. When S is negative, it means that the region of predicted precipitation is relatively small or the precipitation center value is significantly larger than the surrounding areas, or the two conditions exist at the same time, and vice versa.

b. Amplitude

$$A = \frac{D(R_{mod}) - D(R_{obs})}{0.5[D(R_{mod}) + D(R_{obs})]} \quad (2)$$

where $D(R)$ represents the average value of all grid points in the selected area (R_{mod} represents the prediction field and R_{obs} represents the observation field). A is in the range of $[-2,2]$, $A = 0$ indicates that the prediction effect was optimal, $A > 0$ shows that the prediction amplitude was bigger than the observation, and $A < 0$ indicates that the prediction amplitude was lower than the observation.

c. Location

This component consists of two parts, i.e., $L = L_1 + L_2$. L_1 represents the normalized result of the distance between the gravity centers of the observation and the prediction field:

$$L_1 = \frac{|x(R_{mod}) - x(R_{obs})|}{d} \tag{3}$$

where $x(R_{mod})$, $x(R_{obs})$ represents the gravity center of prediction, the observation field in region D, and d represents the maximum distance between valid precipitation grid points in region D. The range of L_1 was [0,1], $L_1 = 0$ indicates that the gravity center of prediction is overlapped with the observation field, and the prediction is perfect. However, even if the precipitation fields are completely different, the gravity centers may be similar. For this reason, another parameter L_2 was introduced to account for the average distance between the gravity center of the overall objects and each individual precipitation, as follows:

$$L_2 = 2 * \frac{|r(R_{mod}) - r(R_{obs})|}{d} \tag{4}$$

where $r = \frac{\sum_{n=1}^{m} R_n \cdot |x - x_n|}{\sum_{n=1}^{m} R_n} R_n = \sum_{(i,j) \in R_n} R_{i,j}$.

m is the number of individuals in the region, and R_n is the total precipitation of the nth individual, which is the weighted average of the distance between each individual object and overall objects. The significance of r was that the larger the total precipitation of the individual, the farther it was from the gravity center of overall precipitation objects. The value of L_2 is in the range of [0,1], and the L value is in the range of [0,2]. Therefore, we can obtain a perfect match when both L_1 and L_2 are equal to 0. The closer L is to 0, the better the prediction is.

2.2.2. Threshold Determination Scheme

It was critical and skillful to determine the objects of interest for a reasonable verification according to a specific rainfall threshold. A bigger threshold would reduce the size of precipitation objects, while a smaller threshold would introduce more weak precipitation areas [28,29]. The key to the SAL method is the identification of precipitation areas and individual precipitation entities; thus, the determination of the precipitation threshold is particularly important [30,31]. Therefore, an appropriate threshold is a key step to determine the precipitation area and calculate SAL accurately [32]. Wernli initially took 1/15 of the maximum precipitation in the verification area as the threshold; this value is an experienced value, i.e., it is found to be appropriate through testing. When the maximum precipitation has extreme or abnormal values, the threshold will be too big. This kind of threshold determination scheme (simply abbreviated to TS1) has obvious disadvantages, as shown in Figure 2a.

The precipitation events were verified in the Yangtze River by Ying Gong [23], which has been improved on the basis of Wernli. The precipitation value at each point has been put in order. A threshold can be determined then by multiplying the value at the index which is at the 95th percentile by 1/15. This method is marked as TS2, as shown in Figure 2b. But this approach suffers from significant drawbacks that there were too many smaller precipitation values and the determined threshold would be significantly smaller, which could lead to the introduction of many small and dispersed weak precipitation areas [33].

Because of the defects of the two methods mentioned above, an improved method has been proposed. Firstly, we put the precipitation values in sequence order. The grid points with precipitation values of 0 in the convective precipitation area account for the majority.

Smaller precipitation values will introduce many small and dispersed weak precipitation areas. Here, we remove the grid with precipitation less than 0.1 mm. Then, a threshold could be determined by multiplying the value at the index which was at the 95th percentile by 1/15. The grid point with values that are bigger than this threshold was the object in the area of interest, which is marked as TS3, as shown in Figure 2c.

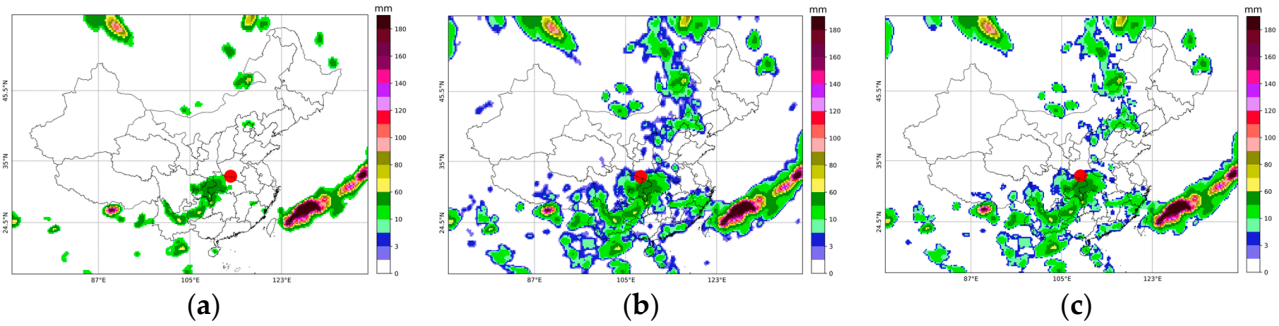

Figure 2. Precipitation objects determined by (**a**) TS1, (**b**) TS2 and (**c**) TS3.

Figure 2 shows the precipitation objects (accumulated rainfall in 24 h) determined by three threshold schemes using the ECMWF ERA5 model data around China at 08:00 on 1 July 2020 (UTC) in (a) to (c).

The three schemes were used to calculate the accumulated precipitation threshold in 24 h from 08:00, 1 July to 31 August 2020. The results (calculated every 24 h and averaged every four days) are shown in Figure 3. It could be observed that the threshold determined by TS1 was much bigger than the others, and precipitation objects with potential increasement would be lost. The threshold determined by TS2 was the smallest, and weak precipitation areas were reserved. Using TS3, more reasonable precipitation areas could be extracted, which not only retained the features with severe rainfall that need to be focused on, but also removed the weak sporadic precipitation, as shown in Figure 3.

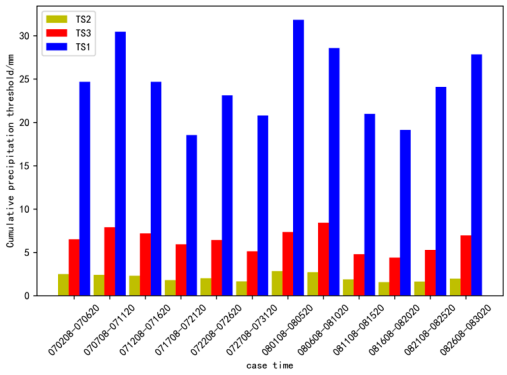

Figure 3. Precipitation threshold determined by three threshold schemes from 1 July to 31 August 2020.

The identified precipitation objects were similar by thresholds calculated by TS2 and TS3, which could well reflect the shape of the precipitation area. However, TS2 easily introduce small and dispersed precipitation areas, while the precipitation objects determined by TS1 were not reasonable. The forecasting results (accumulated rainfall in 24 h), which were calculated by S, A and L from ECMWF at 08:00 2 July 2020, are shown in Figure 4, in which parameter A from the three schemes was consistent, the S and L calculated by TS1 were negative and slightly larger, respectively, while the S and L calculated by TS2 and TS3 differ little, indicating that the results of these two methods were similar. Therefore, the threshold calculated by TS3 was selected for analysis.

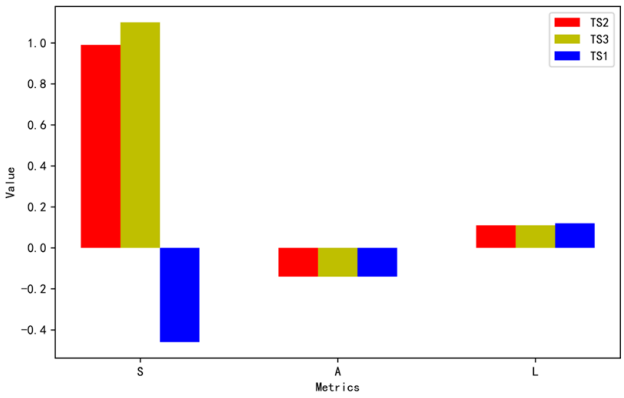

Figure 4. *S*, *A* and *L* calculated by three schemes.

2.2.3. Traditional Verification Metrics

Mean Error (*ME*)

The average error was defined as the average value of the difference between all grid points of observation and prediction:

$$ME = \frac{1}{N}\sum_{(i,j)\in D}[R_{mod}(i,j) - R_{obs}(i,j)] \quad (5)$$

where N is the number of grid points in region D, and $R(i,j)$ is the value of grid points for observation or prediction. When *ME* was 0, it indicated perfect prediction. Positive or negative *ME* indicated overestimation or underestimation.

Root Mean Square Error (*RMSE*)

Root mean square error was defined as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{(i,j)\in D}[R_{mod}(i,j) - R_{obs}(i,j)]^2} \quad (6)$$

The root mean square error was positive, and its value was 0 in a perfect prediction.

Frequency Bias Index (*FBI*)

The *FBI* was based on a contingency table, which corresponds to the ratio of the predicted number to the observed number:

$$FBI = (a+b)/(a+c) \quad (7)$$

where a is hit, b is false alarm and c is miss. The *FBI* range is 0 to infinity; 1 indicated perfect prediction, and greater than or less than 1 indicated overestimation or underestimation.

Heidke Skill Score (*HSS*)

The *HSS* was also based on a contingency table:

$$HSS = 2(ad - bc)/[(a+c)(c+d) + (a+b)(b+d)] \quad (8)$$

where d is correct negative, *HSS* range is [0,1], 1 is the perfect score, 0 means no skill and *HSS* represents the proportion of correct prediction after eliminating the correct prediction due to random opportunities.

3. Experimental Results

3.1. Geometric Cases

By simplifying complex precipitation distributions into geometric shapes (such as ellipses), it becomes easier to isolate and study specific properties such as amplitude, structure and location. This helps in breaking down complex systems into more manageable components for analysis. Figure 5 shows four idealized precipitation objects, which were labeled as G0 to G3. The objects in dark gray denote more intense precipitation, and the surrounding light gray indicates weak precipitation. Compared to object G0, the amplitude and structure of G1 and G2 were almost equal. The center of G1 was located 40 grid points east of G0, with only a difference in displacement. The center of G2 and G3 were shifted 40 grid points to the right from G0, and they had a larger rainfall area and different aspect ratio. Details of these objects are shown in Table 1. x, y represent the number of grid points from the center of the ellipse to the horizontal and vertical axis, respectively. Table a and b represent the major and minor axes of the ellipse, respectively.

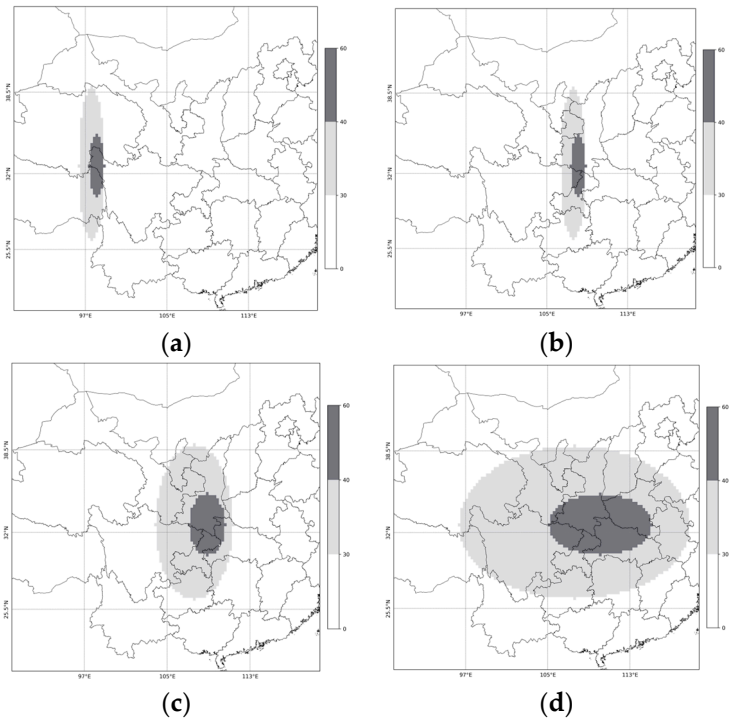

Figure 5. The simulated four types of elliptical precipitation areas: (**a**) G0; (**b**) G1; (**c**) G2; (**d**) G3.

Table 1. Several types of ellipse parameters defined.

Precipitation Object	x	y	a	b	Aspect Ratio a/b
G0	50	30	25	5	5
G1	50	70	25	5	5
G2	50	70	25	15	1.67
G3	50	70	25	45	0.56

Table 2 showed SAL of several elliptical precipitation areas and the values of traditional verification metrics *ME*, *RMSE*, *HSS* and *FBI*. In the comparison of G0 and G1, it is assumed that G0 was the observation field and G1 was the prediction field, their shape and amplitude were equal and there was eastward displacement relative to the observation field, resulting in S and A being 0 and L being 0.266. In the traditional verification score, *ME* was 0 because the positive and negative offset each other. Therefore, *ME* could not identify the displacement of prediction and observation. In addition, the *FBI* was close to 1, indicating that the forecast was perfect, but there were differences in positions. The score of *HSS* was

equal to 0, which indicated no skill. Therefore, the *FBI* and *HSS* metrics were not enough to represent their prediction effect. This case showed that there was diversity in quality assessment when using traditional verification metrics, that is, the prediction was described as perfect by the *ME* and *FBI* metrics, and although *RMSE* was not perfect, its value was very small, while *HSS* indicated no skill. On the contrary, SAL provides more valuable information, that is, the amplitude and structure of model prediction had a perfect score of 0, and there was a certain displacement.

Table 2. SAL and traditional *ME*, *RMSE*, *HSS* and *FBI* for idealized elliptical precipitation objects.

	S	***A***	***L***	***ME***	***RMSE***	***HSS***	***FBI***
G0 vs. G1	0	0	0.266	0	8.268	−0.030	1
G0 vs. G2	1	1	0.258	2.080	11.616	−0.050	3.047
G0 vs. G3	1.60	1.600	0.261	8.308	17.345	0.048	9.166
G2 vs. G3	1	1	0.004	2.077	8.137	0.469	3.047

For the comparison of G0 and G2, and G0 and G3 (in both cases, G0 was taken as the observation field), the predicted precipitation area moved eastward relative to the observation, and there was a difference in shape. This feature was captured by SAL, resulting in $L > 0$ and $S > 0$. The traditional verification metrics tend to have poor prediction effects, especially G0 and G3, with the largest being *ME*, *RMSE* and *FBI*, and *HSS* tends to be 0. However, to some extent, the forecast effect was not as bad as indicated in the traditional test scores. The reason for the poor scores of G0 and G3 evaluated by traditional verification metrics was that metrics such as *ME*, *RMSE* and *FBI* were very sensitive to the area of precipitation. If there was no overlap between the forecast and the observation field, the scores calculated by traditional methods were poor. According to the evaluation of SAL, G0 and G1 had the worst prediction in terms of location, while G0 and G2, and G0 and G3 had the worst prediction in terms of amplitude and structure.

In the comparison of G2 and G3, as we can see, component *L* and *HSS* had the best ideal values relative to the other four comparisons. This was due to the center of gravity being closed. Their *S* and *A* parameters indicate that G3 overestimates the size and amplitude of the precipitation area, which was consistent with the results obtained by *ME*, *RMSE* and *FBI*. When there was no displacement error or the displacement error was small, the traditional scores were more accurate. When the displacement error was large, there was a contradiction between the scores of the traditional methods and SAL methods. Therefore, the *HSS* metric was very sensitive to the displacement of the prediction field, and when the observation and prediction overlapped, the *HSS* value was positive; otherwise, it was negative.

3.2. Simulation Cases

In this paper, the 24 h cumulative precipitation reported from 08:00 on 19 July 2020 is selected as the actual field, marked as F0, as shown in Figure 6a. Three different treatments are made for the actual field as the prediction field: F1 (move F0 to the southeast), F2 (move F0 to the southeast and multiply it by 1.5 times: F2 = 1.5 × F1) and F3 (move F0 to the southeast and reduce each grid value by 1.27 mm; if the grid value is negative, it is 0, i.e., F3 = max [0, F1–1.27 mm]). The three simulated precipitation fields are shown in Figure 6b–d, in which the red outline represents the area determined by TS3. This paper will compare the results of these three artificial simulation prediction fields.

This showed that the precipitation areas identified by F0, F1 and F2 were identical, because F1 was obtained by F0 displacement, while F2 was obtained by displacement and linear transformation. However, some precipitation areas would be moved out of the study border during displacement. Therefore, although the precipitation distribution was the same, the values of S and A were not 0. The results of SAL and the traditional inspection index of F0 and F1, F0 and F2, and F0 and F3, respectively, are shown in Table 3.

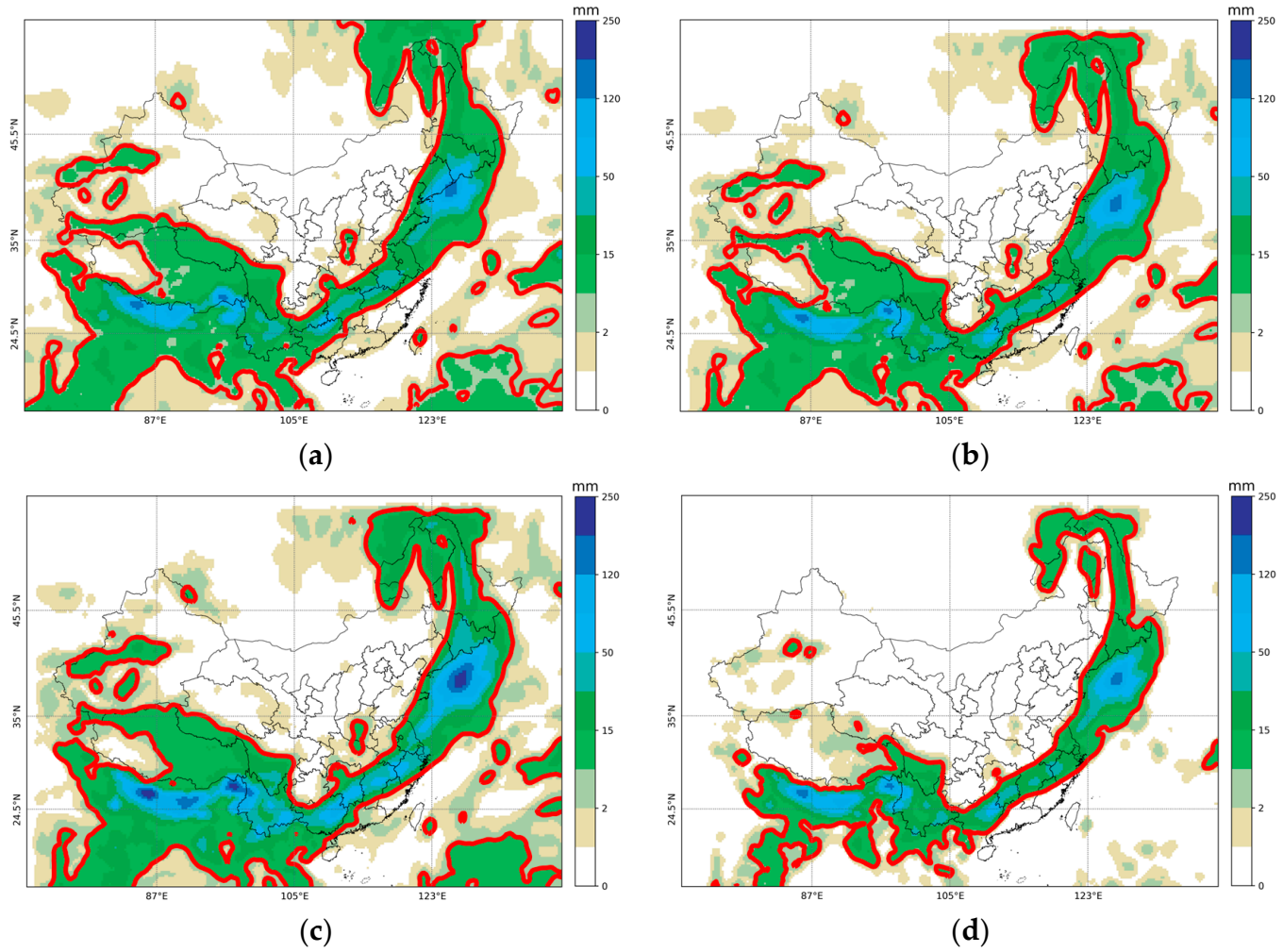

Figure 6. Simulated precipitation field: (**a**) F0; (**b**) F1; (**c**) F2; (**d**) F3.

Table 3. SAL value of simulated precipitation area and traditional *ME*, *RMSE*, *HSS* and *FBI*.

	S	***A***	***L***	***ME***	***RMSE***	***HSS***	***FBI***
F0 vs. F1	0.044	−0.045	0.030	−0.235	15.983	0.515	0.900
F0 vs. F2	0.956	0.353	0.213	2.316	21.047	0.513	0.879
F0 vs. F3	−0.219	−0.406	0.071	−1.821	15.929	0.368	0.481

In the comparison of F0 and F3, F3 reduces each grid point by 1.27 mm compared with F0, so the value of A becomes negative, which was −0.406. The structural component S was also negative, which was due to the obvious reduction in precipitation areas when the value of each grid point decreased to 1.27 mm. The phenomenon that the grid value decreases R0 and the precipitation areas decrease significantly showed that the center of the precipitation area comprised strong precipitation, while the surrounding area was widely distributed with weak precipitation, and the precipitation intensity changed significantly in a certain area. This kind of precipitation was usually convective precipitation, while the precipitation with uniform distribution and insignificant change in a certain area was usually stratiform precipitation, as shown in Figure 7.

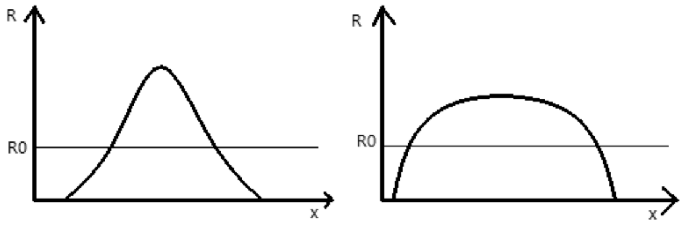

Figure 7. Distribution characteristics of convective precipitation and stratiform precipitation; R0 is the precipitation threshold.

3.3. Real Cases

In this section, the 24 h cumulative precipitation areas in three time periods were selected for analysis, at 08:00 on 23 July 2020, 08:00 on 24 July 2020 and 20:00 on 25 August 2020. The main purpose of this study was to analyze the importance of region size selection when calculating SAL.

Firstly, SAL was applied to the whole China region. Each observation and prediction field has a main precipitation area and multiple small precipitation areas. The SAL values and traditional verification scores of the overall area (Figure 8), main area and local area (Figure 9) were calculated. The results are shown in Table 4. They showed that the forecast at 20:00 on 25 August 2020 had the smallest absolute values of *S*, *A* and *L*, and the effect was the best compared with the other two. For the overall prediction, the *S* was positive, indicating that the predicted precipitation area was greater than the actual precipitation area. In the main area, the *S* was also positive, because the maximum values in the overall and main precipitation areas were the same, resulting in the same precipitation threshold, so the identified precipitation areas were also similar. In the local area, the *S* value becomes negative, because it could be concluded that the small precipitation area had little effect on the overall *S* value.

In the overall, main and local areas, the main difference was the change in the *L* value. With the decrease in the area, the *L* value gradually increased, which was caused by the decrease in d in Formula (2).

Therefore, it was more meaningful to apply SAL to a limited area, because there were many different precipitation systems in a large overall area, and their formation mechanisms were different. If SAL calculation was carried out in a large range, it was of little significance.

Table 4. SAL values and traditional *ME*, *RMSE*, *HSS* and *FBI* values of precipitation areas in actual cases.

	Region	***S***	***A***	***L***	***ME***	***RMSE***	***HSS***	***FBI***
2020072408	overall	0.216	−0.125	0.237	−1.632	17.548	0.328	1.125
2020082520	overall	0.181	−0.043	0.124	−0.896	18.823	0.549	1.265
2020072308	overall	0.196	−0.145	0.275	−1.418	17.417	0.416	1.227
2020072408	main	0.105	−0.015	0.330	−1.289	19.968	0.340	1.407
2020082520	main	0.459	−0.463	0.181	−4.950	24.680	0.608	1.006
2020072308	main	0.178	0.094	0.276	−1.647	21.379	0.408	1.313
2020072408	local	−0.866	−1.027	0.338	−5.157	15.081	0.362	0.518
2020082520	local	−0.188	0.324	0.190	1.343	6.194	0.462	1.153
2020072308	local	−0.863	−1.292	0.284	−11.614	23.096	0.429	0.443
2020083120	local	0.673	−0.073	0.215	0.976	36.620	0.156	1.643
2020082908	local	−1.150	−0.683	0.138	−14.139	39.592	0.298	0.454

Figure 10 shows the prediction results of two models with the maximum *RMSE*. The *RMSE* had reached more than 30. However, the prediction quality of the two prediction results was very different. The difference in prediction quality can be seen intuitively from Figure 10. At 20:00 on 31 August 2020, there was a precipitation area in the southeast coastal

area of China. It can be seen from Table 4 that it was mainly the difference in structure (*S*) and location (*L*). It can be seen from Figure 10 that the prediction field was much larger than the observation field, the center of gravity of the two precipitation fields also had a large displacement and the prediction center of gravity moves to the southeast, compared to the observation. At 08:00 on 29 August 2020, the error mainly comes from the structure (*S*) and strength (*A*). The prediction field was smaller than the observation field, and the amplitude was weaker than the observation field. Therefore, although the two had similar *RMSE*s, the error sources were different. In the first example, the "double punishment" phenomenon was caused by the precipitation target displacement, while in the second example, the *RMSE* value was too large due to the underestimation of precipitation intensity, which highlights the importance of SAL evaluation and was more meaningful than traditional verification metrics.

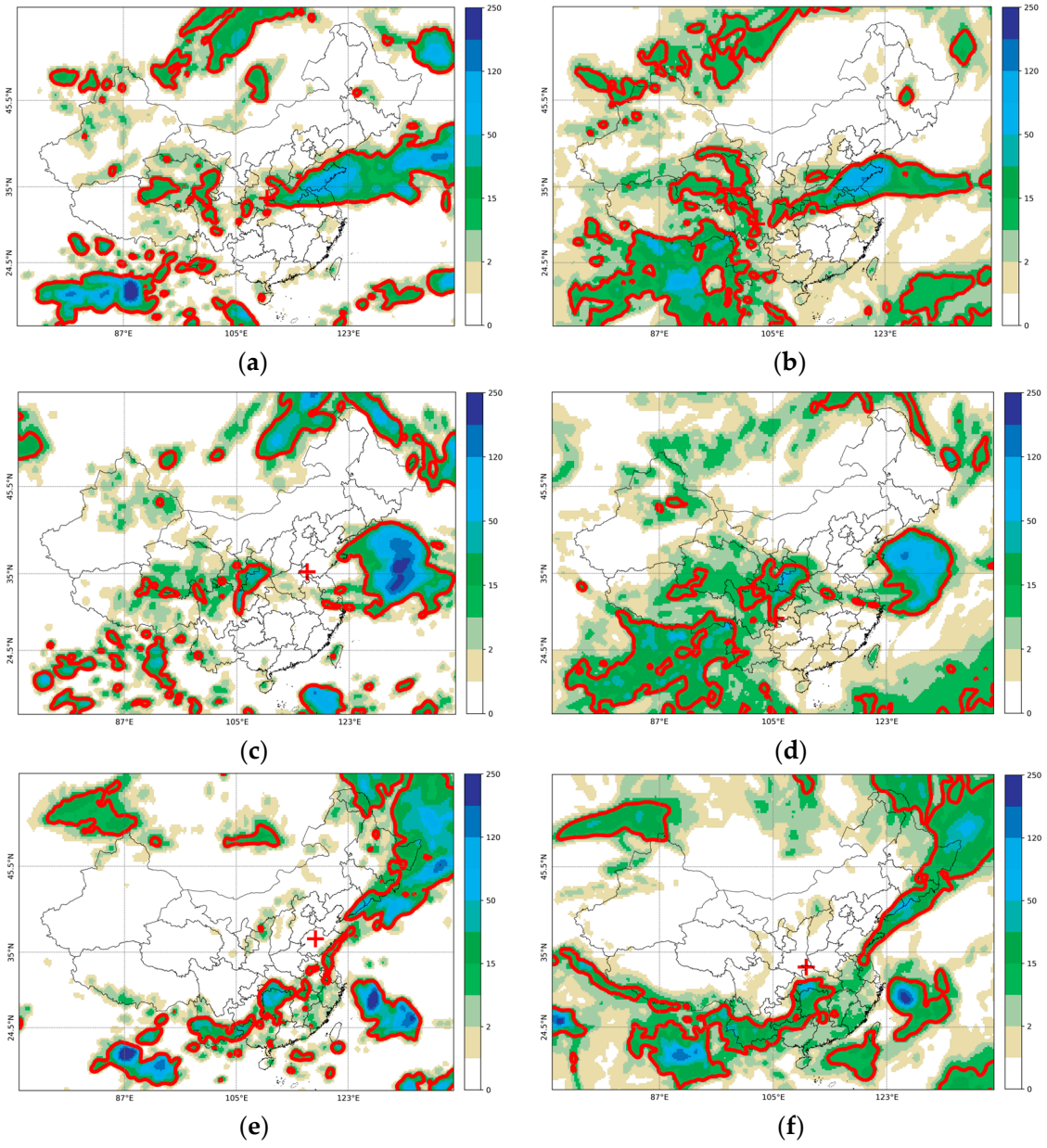

Figure 8. Cumulative precipitation over 24 h. at 08:00 on 23 July (**a**), 08:00 on 24 July (**c**) and 20:00 on 25 August 2020 (**e**). The left is the observation field (**a**,**c**,**e**), and the right is the forecast field (**b**,**d**,**f**). The red outline is the precipitation threshold, the red cross is the center of gravity of the precipitation field, and the yellow box is the main body and local area.

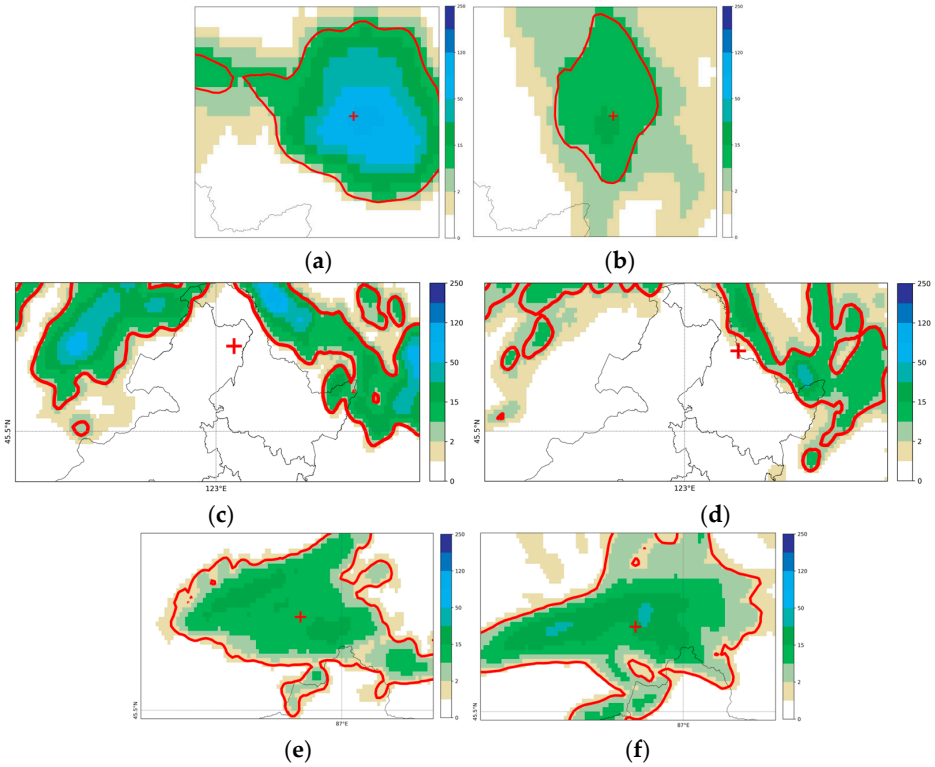

Figure 9. Accumulated precipitation in local areas at 08:00 on 23 July (**a**), 08:00 on 24 July (**c**) and 20:00 on 25 August 2020 (**e**). The left is the observation field (**a**,**c**,**e**), and the right is the forecast field (**b**,**d**,**f**). The red outline is the precipitation threshold, and the red cross is the center of gravity of the precipitation field.

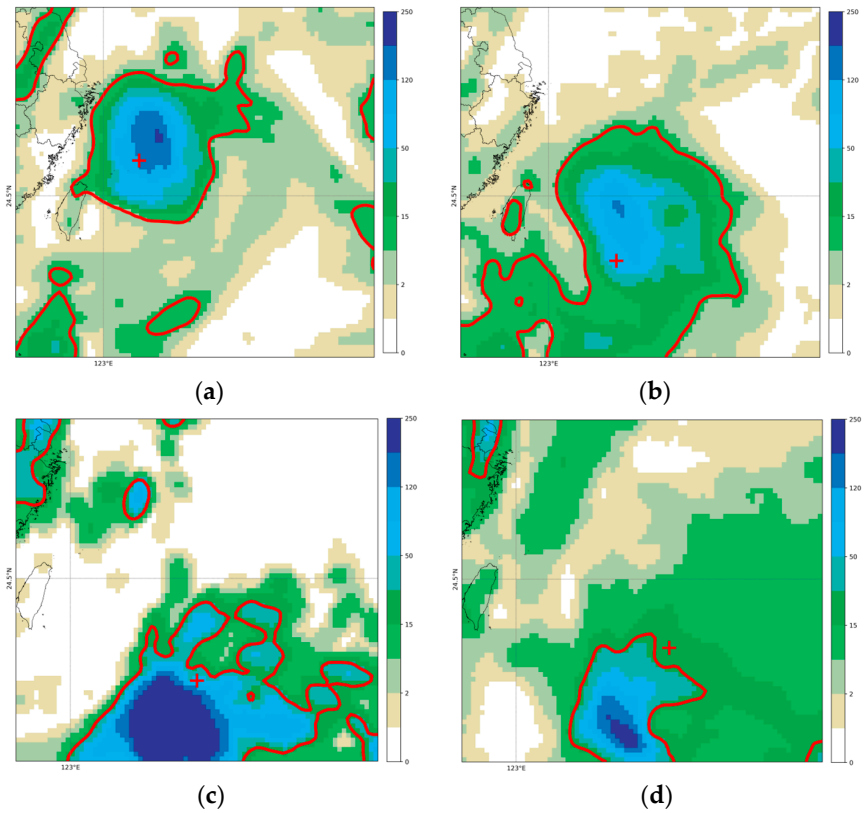

Figure 10. Cumulative precipitation in local areas at 20:00 on 31 August 2020 (**a**) and 08:00 on 29 August 2020 (**c**), with the observation field on the left (**a**,**c**) and the forecast field on the right (**b**,**d**).

4. Conclusions

This paper proposes a new precipitation target determination method based on the SAL spatial verification method, and this method was used to analyze virtual geometric cases, simulations cases and real 24 h cumulative precipitation from July to August 2020. We explored the differences and advantages between the SAL test method and traditional verification metrics. The results show that, compared to traditional verification metrics such as *ME* (Mean Error) and *RMSE* (root mean square error), the SAL method can more intuitively reflect the differences between forecasts and observations in terms of structure, intensity and location. In particular, when dealing with the "double penalty" problem in high-resolution precipitation forecasts, the SAL method demonstrates significant advantages. Compared to previous research by Wernli et al. [21], our improved threshold method can more accurately identify the main precipitation regions, thereby improving the precision of precipitation object identification. From this study, the main conclusions drawn are as follows:

(1) The key step in the SAL spatial verification method was the identification of the precipitation body. By comparing three threshold determination schemes, the threshold determination method proposed in this paper could identify the main body of precipitation more accurately and effectively.
(2) The traditional *ME* metric could not identify the displacement between prediction and observation, while the *HSS* index was very sensitive to the displacement of the prediction field, and when the observation and prediction overlap, the *HSS* value was positive; otherwise, it was negative.
(3) Structure component S could be used as a metric to judge whether precipitation is convective and stratiform precipitation. Generally speaking, when the smaller value R0 is subtracted from the grid value of the precipitation area and S changes greatly and is negative, the precipitation area is convective precipitation, and vice versa.
(4) Regional selection was very important for SAL calculation. If the regional selection is too large, there may be multiple precipitation systems in the region, and the formation mechanism may be different. The calculated SAL value had little significance.
(5) Compared with the traditional verification metrics, the SAL verification method was easy to calculate and operate, and could better reflect the model prediction ability, so forecasters could better understand the model prediction effect and what needs to be improved.
(6) Since this study only used summer data for verification, it can only reflect the superiority of the SAL method in precipitation forecasting for this season, which has certain limitations. To more accurately understand the forecasting performance in other seasons, further verification work will continue to be conducted in the future, providing more references for forecasters. Our findings also emphasize that the interpretation of SAL must be specific to the chosen threshold and field. For instance, utilizing a larger domain diminishes the impact of the *L* component since object displacements are normalized by the diagonal length of the domain.

The SAL verification method could not only quantitatively analyze the prediction performance of QPF, but also reflect the physical parameters of the model from different aspects. In terms of artificial weather influence, it was necessary to obtain the evaluation of intensity, location and operation area after operation. From this perspective, SAL will apply the effect evaluation of artificial weather modification in the future, which can not only obtain quantitative evaluation parameters, but also reflect the changes in physical parameters inside the cloud cluster.

The contribution of this study to the spatial verification of quantitative precipitation estimation (QPE) lies in proposing an improved threshold selection method that can more accurately identify the main precipitation regions and reduce the misjudgment of weak precipitation areas. By applying the SAL method to different geometric, synthetic and actual precipitation scenarios, it demonstrated its effectiveness in mitigating the "double penalty" problem in high-resolution forecasts. The study compared the performance differences

between traditional verification methods and the SAL method, revealing the limitations of traditional metrics such as *ME* (Mean Error) and *RMSE* (root mean square error) in high-resolution precipitation forecasting. It provides forecasters with a more intuitive tool to understand the performance of forecast models, especially under complex terrain and diverse precipitation structures.

Author Contributions: Conceptualization, F.W. and D.S.; methodology, D.S.; formal analysis, J.Z. and Y.X.; investigation, L.L.; data curation, L.L.; writing—original draft preparation, J.Z. and Y.X.; writing—review and editing, H.L. and X.F.; funding acquisition, L.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant Nos. 42075001) and the National Key R&D Program of China (Grant Nos. 2018YFC1506605), the Sichuan Natural Science Foundation project "Quantitative Assessment of Climate Impact of Northwest Sichuan Ecological Demonstration Zone on Surrounding Regions" (2022 NSFSC0208), Ningxia Natural Science Foundation Project (2023AAC02088) "Research on Intelligent Monitoring and Evaluation Technology of Wind Disaster Indicators for Facility Greenhouses", the 2023 Science and Technology Plan Project of Liangshan Prefecture (23ZDYF0182) "Research on Forest Fire Risk Early Warning Technology and Development of Fire Risk Meteorological Level Prediction System in Liangshan Region" and the Sichuan Province 2024 Science and Technology Plan Project (24ZDYF1673).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author/s.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ahijevych, D.; Brown, B.; Ebert, E. Application of Spatial Verification Methods to Idealized and NWP-Gridded Precipitation Forecasts. *Weather Forecast.* **2009**, *24*, 1485–1497. [CrossRef]
2. Alam, A.; Ahmed, B.; Sammonds, P.; Kamal, A.M. Applying rainfall threshold estimates and frequency ratio model for landslide hazard assessment in the coastal mountain setting of South Asia. *Nat. Hazards Res.* **2023**, *3*, 531–545. [CrossRef]
3. Brunner, M.I.; Swain, D.L.; Wood, R.R.; Willkofer, F.; Done, J.M.; Gilleland, E.; Ludwig, R. An extremeness threshold determines the regional response of floods to changes in rainfall extremes. *Commun. Earth Environ.* **2021**, *2*, 173. [CrossRef]
4. Casati, B.; Ross, G.; Stephenson, D.B. A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorol. Appl.* **2004**, *11*, 141–154. [CrossRef]
5. Collier, C.G.; Krzysztofowicz, R. Quantitative Precipitation Forecasting. *J. Hydrol.* **2001**, *239*, 1–2. [CrossRef]
6. Csima, G.; Ghelli, A. On the use of the intensity-scale verification technique to assess operational precipitation forecasts. *Meteorol. Appl.* **2010**, *15*, 145–154. [CrossRef]
7. Damrath, U.; Wetterdienst, D. Verification against precipitation observations of a high-density network–what did we learn? In Proceedings of the International Verification Methods Workshop, Montreal, QC, Canada, 15–17 September 2004.
8. Davis, C.; Brown, B.; Bullock, R. Object-Based Verification of Precipitation Forecasts. Part I: Methodology and Application to Mesoscale Rain Areas. *Mon. Weather Rev.* **2006**, *134*, 1772–1784. [CrossRef]
9. Davis, C.; Brown, B.; Bullock, R. Object-Based Verification of Precipitation Forecasts. Part II: Application to Convective Rain Systems. *Mon. Weather Rev.* **2006**, *134*, 1785–1795. [CrossRef]
10. Davis, C.A.; Brown, B.G.; Bullock, R.; Halley-Gotway, J. The Method for Object-Based Diagnostic Evaluation (MODE) Applied to Numerical Forecasts from the 2005 NSSL/SPC Spring Program. *Weather Forecast.* **2009**, *24*, 1252–1267. [CrossRef]
11. Ebert, E.E. Fuzzy verification of high-resolution gridded forecasts: A review and proposed framework. *Meteorol. Appl.* **2010**, *15*, 51–64. [CrossRef]
12. Ebert, E.E.; Gallus, W.A. Toward Better Understanding of the Contiguous Rain Area (CRA) Method for Spatial Forecast Verification. *Weather Forecast.* **2009**, *24*, 1401–1415. [CrossRef]
13. Ebert, E.E.; McBride, J.L. Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol.* **2000**, *239*, 179–202. [CrossRef]
14. Gilleland, E.; Jevych, D.A.; Brown, B.G. Intercomparison of spatial forecast verification methods. *Weather Forecast.* **2009**, *24*, 1416–1430. [CrossRef]
15. Gonzalez, F.C.; Cavacanti, M.D.; Nahas Ribeiro, W.; de Mendonça, M.B.; Haddad, A.N. A systematic review on rainfall thresholds for landslides occurrence. *Heliyon* **2023**, *10*, e23247. [CrossRef] [PubMed]

16. Gong, Y. The Explanation and Application of SAL Quantitative Verification for Precipitation Forecasts. *Torrential Rain Disasters* **2010**, *29*, 51–57.
17. Grams, J.S.; Gallus, W.A.; Koch, S.E.; Wharton, L.S.; Loughe, A.; Ebert, E.E. The Use of a Modified Ebert McBride Technique to Evaluate Mesoscale Model QPF as a Function of Convective System Morphology during IHOP 2002. *Weather Forecast.* **2006**, *21*, 288–306. [CrossRef]
18. Ho, J.Y.; Liu, C.H.; Chen, W.B.; Chang, C.-H.; Lee, K.T. Using ensemble quantitative precipitation forecast for rainfall-induced shallow landslide predictions. *Geosci. Lett.* **2022**, *9*, 22. [CrossRef]
19. Jolliffe, I.T.; Stephenson, D.B. Forecast Verification: A Practitioner's Guide in Atmospheric Science, 2nd ed.John Wiley & Sons: Chichester, UK, 2003.
20. Keil, C.; Craig, G.C. A displacement-based error measure applied in a regional ensemble forecasting system. *Mon. Weather Rev.* **2007**, *135*, 3248–3259. [CrossRef]
21. Keil, C.; Craig, G.C. A displacement and amplitude score employing an optical flow technique. *Weather Forecast.* **2009**, *24*, 1297–1308. [CrossRef]
22. Korolev, V.Y.; Gorshenin, A.K. The probability distribution of extreme precipitation. *Dokl. Earth Sci.* **2017**, *477*, 1461–1466. [CrossRef]
23. Korolev, V.; Gorshenin, A. Probability Models and Statistical Tests for Extreme Precipitation Based on Generalized Negative Binomial Distributions. *Mathematics* **2020**, *8*, 604. [CrossRef]
24. Leonarduzzi, E.; Molnar, P. Deriving rainfall thresholds for landsliding at the regional scale: Daily and hourly resolutions, normalisation, and antecedent rainfall. *Nat. Hazards Earth Syst. Sci.* **2020**, *20*, 2905–2919. [CrossRef]
25. Liou, Y.; Chou, T.; Cheng, Y.; Teng, Y. The Improvement of Short-Term Quantitative Precipitation Forecast in Mountainous Areas by the Assimilation of Meteorological State Variables Retrieved by Multiple Doppler Radar Data. *Mon. Weather Rev.* **2024**, *152*, 1339–1355. [CrossRef]
26. Liu, J.Q.; Li, Z.L.; Wang, Q.Q. Quantitative Precipitation Forecasting Using an Improved Probability-Matching Method and Its Application to a Typhoon Event. *Atmosphere* **2021**, *12*, 1346. [CrossRef]
27. Roberts, N.M.; Lean, H.W. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Weather Rev.* **2008**, *136*, 78–97. [CrossRef]
28. Wang, C.C.; Chang, C.S.; Wang, Y.W.; Huang, C.-C.; Wang, S.-C.; Chen, Y.-S.; Tsuboki, K.; Huang, S.-Y.; Chen, S.-H.; Chuang, P.-Y.; et al. Evaluating Quantitative Precipitation Forecasts Using the 2.5 km CReSS Model for Typhoons in Taiwan: An Update through the 2015 Season. *Atmosphere* **2021**, *12*, 1501. [CrossRef]
29. Wernli, H.; Hormann, C.; Zimmer, M. Spatial forecast verification methods intercomparison project: Application of the SAL technique. *Weather Forecast.* **2009**, *24*, 1472–1484. [CrossRef]
30. Wernli, H.; Paulat, M.; Hagen, M.; Frei, C. SAL—A novel quality measure for the verification of quantitative precipitation forecasts. *Mon. Weather Rev.* **2008**, *136*, 4470–4487. [CrossRef]
31. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences: An Introduction*; Academic Press: San Diego, CA, USA, 1995.
32. Zhou, K.H.; Sun, J.S.; Zheng, Y.G.; Zhang, Y. Quantitative Precipitation Forecast Experiment Based on Basic NWP Variables Using Deep Learning. *Adv. Atmos. Sci.* **2022**, *39*, 1472–1486. [CrossRef]
33. Zepeda-Arce, J.; Foufoula-Georgiou, E.; Droegemeier, K.K. Spacetime rainfall organization and its role in validating quantitative precipitation forecasts. *J. Geophys. Res. Atmos.* **2000**, *105*, 10129–10146. [CrossRef]