Journal of
Water & Climate Change

IWA PUBLISHING

Journal of Water and Climate Change Vol 14 No 7, 2122 doi: 10.2166/wcc.2023.426

An integrated method for extended-range prediction of heavy Precipitation process in the flood season over Hunan Province based on S2S models

Chengmin Mao[a,b,c,d], Jianming Zhang[a,b], Yuxing Zeng[a], Hui Zhao[a], Jiadong Peng[a,b,d,e,*], Yihao Tang[a] and Shaofeng Peng[f,g]

[a] Climate Center of Hunan Province, Changsha, China
[b] Yueyang National Climatological Observatory, Yueyang, Hunan, China
[c] Huaihua Meteorological Bureau, Huaihua, Hunan, China
[d] Key Laboratory of Hunan Province for Meteorological Disaster Prevention and Mitigation, Changsha, China
[e] Turpan Meteorological Bureau, Turpan, Xinjiang, China
[f] Hunan Academy of Forestry, Changsha, China
[g] Shennongguo Oil Eco-Agriculture Development Co. Ltd, Hengyang, Hunan, China
*Corresponding author. E-mail: 14865976@qq.com

YZ, 0000-0002-3142-7037; JP, 0000-0003-2709-8899

ABSTRACT

Floods in the middle reaches of Yangtze River threaten millions of people and cause casualties and economic losses. Yet, the prediction of floods especially on the sub-seasonal scale in this region is still challenging. To better predict the floods during the flood season (from April to August) in Hunan Province, the models from the China Meteorological Administration (CMA), the European Centre for Medium-Range Weather Forecasts (ECMWF) and the National Centers for Environmental Prediction (NCEP) that participated in the sub-seasonal to seasonal (S2S) prediction project were chosen to evaluate their extended-range (the next 11–30 days) prediction skills for heavy precipitation. The original prediction score of single model (original score), the score of single model using optimal threshold of heavy precipitation (adjusted score) and the score of multi-model integration (integrated score) were calculated by the scoring rules for heavy precipitation process. The results show that the integrated score in the extended-range is 75.1, which is 10.3 and 6.9 higher than the average scores of original models and adjusted method, respectively. The false alarm (missing) rate of the integrated method is 5.4% (33.9%), which is 8.6% (4.7%) and 2.9% (3.3%) smaller than the average rates of original models and adjusted method, respectively.

Key words: heavy precipitation process, Hunan Province, multi-model integration, optimal threshold, sub-seasonal to seasonal model

HIGHLIGHTS

- The sub-seasonal to seasonal model data established by the World Meteorological Organization are used.
- The method of multi-mode integration is used to effectively improve the prediction score of the heavy precipitation process.
- On the basis of multi-mode integration, the empty rate and the missing rate of the heavy precipitation process are effectively reduced by adjusting the threshold of heavy precipitation.

1. INTRODUCTION

The sub-seasonal to seasonal (S2S) prediction research project is jointly established by the World Weather Research Program (WWRP) and World Climate Research Program (WCRP) under the World Meteorological Organization (WMO). The sub-seasonal prediction, which has multiple time scales, is the key interface between the weather forecast and seasonal prediction. Applying the numerical weather forecast to short-term climate prediction can bridge the gap between medium-range weather forecasts and seasonal predictions (WMO 2013).

In recent years, benefiting from the continuous developments of models, assimilation technology, and ensemble methods, the prediction skills of the S2S models have been remarkably improved. For example, the forecast lengths are 44 and 46 days for the National Center for Environment Prediction (NCEP) and the European Centre for Medium-Range Weather Forecasts (ECMWF), respectively, indicating a higher prediction skill of the latter (CMA 2018). China Meteorological Administration

(CMA) actively participates in the international cooperation of the S2S prediction project, launches China's sub-seasonal timescale forecast research (Qi & Rong 2014), successively develops several versions of climate system models, and gradually realizes their operational application (Wu *et al.* 2010, 2013, 2014; Wu 2012; Hu *et al.* 2020), which is helpful to improve the capability of global forecast service (Hu *et al.* 2019). At present, the CMA model has a forecast length of up to 60 days.

Scholars have comprehensively evaluated the prediction skills for precipitation and temperature (Tian *et al.* 2017), cold surges (Li *et al.* 2017), climate change process in summer of East Asia (Liang & Lin 2017), and summer monsoon anomaly in Australia (Marshall & Hendon 2015) on the sub-seasonal timescale. In addition, Bombardi *et al.* (2017) assessed the sub-seasonal predictability of three global S2S models on the start and end dates of rainy seasons in South America, East Asia, and northern Australia. Zhao *et al.* (2016) comprehensively evaluated the performance of the BCC_CSM1.2 S2S model. These studies have revealed that the S2S models have certain predictability, but they are still difficult to perform skilled predictions (Lin *et al.* 2019).

At the same time, some studies (Rashid *et al.* 2011; Kim *et al.* 2014; Neena *et al.* 2014; Vitart 2014) have shown that the qualitative and quantitative prediction skills of the S2S models for precipitation in the extended range vary with the forecast month. For example, qualitative prediction skills are higher in the flood season, especially in midsummer, whereas quantitative prediction skills are the lowest in midsummer. From the perspective of qualitative prediction skills on a sub-seasonal timescale, models of the CMA and BOM (Bureau of Meteorology Australia) perform the best in predicting extreme precipitation in May and June to September, respectively (Pang *et al.* 2021).

Located in the Subtropical Monsoon Climate Zone of Central China, Hunan Province has experienced frequent rainfall in the flood season, which is easy to cause natural disasters such as mountain torrents, mudslides, and urban waterlogging. In addition, there are one to three concentration periods for rainfall in Hunan Province during the flood season every year. During these periods, the short interval of precipitation processes and large rainfall intensity will cause basin floods easily. Sub-seasonal to seasonal prediction can extend the forecast length of the heavy rainfall process, which can help the government departments to switch from passive response to active mitigation of disasters. It is also crucial to improve the meteorological service capacity and build the first line of defence in disaster prevention and mitigation. Since the S2S prediction is performed on the global scale, there is still a certain degree of deviation in the local-scale prediction, such as in Hunan, especially for the insufficiency of prediction stability by the single model. Thus, based on the products of S2S models from the CMA, ECMWF, and NCEP, the extended range prediction skills of the heavy precipitation process in the flood season over Hunan Province were evaluated. On this basis, an adjusted method for single-model prediction and an integration method for multi-model predictions were established in this study, which has important scientific and social application values.

The remainder of this article is organized as follows. Section 2 introduces the observation and prediction data of precipitation, the definition and scoring rules of the heavy precipitation process, and the adjusted and integration methods for the S2S models. Section 3 presents the assessments of original models, adjusted and integration methods. Finally, the conclusions and discussion are presented in Section 4.

2. DATA AND METHODS

2.1. Precipitation data

The S2S models used in this study are from the CMA, ECMWF, and NCEP. Table 1 lists the model details, including forecast length, resolution, number of ensemble members, and forecast frequency. The CMA S2S model has a forecast length of 60 days and a horizontal resolution of about $0.5° \times 0.5°$, with four ensemble members and a forecast frequency of one time per day. The ECMWF S2S model has a forecast length of 46 days and a horizontal resolution of 16 km within 0–15 days and 32 km after 15 days, which has 51 ensemble members and makes forecasts twice in 1 week. The NCEPS2S model has a forecast length of 44 days and a horizontal resolution of about $1° \times 1°$, with 16 ensemble members and the same forecast frequency as the CMA model. The prediction skills of MJO for the aforementioned three models are 24, 23–27, and 20 days, respectively (Peng *et al.* 2021).

The forecast data of daily precipitation output by the aforementioned three models were used in this study, which covers the period from April to August during 2017–2020. Meanwhile, the observational data of daily precipitation during the same period from 96 national meteorological observation stations in Hunan Province were also used for the assessment of prediction performance (shown in Figure 1).

Table 1 | Information of three sub-seasonal to seasonal (S2S) models

Name	Forecast length	Horizontal resolution	Forecast frequency	Ensemble size	Prediction skills of MJO
CMA	60 days	~0.5° × 0.5°	1/day	4	24 days
ECMWF	46 days	16 km (≤15 d) 32 km(>15 d)	2/week	51	23–27 days
NCEP	44 days	~1° × 1°	1/day	16	20 days

Note: ~ means the resolution is an approximate value.

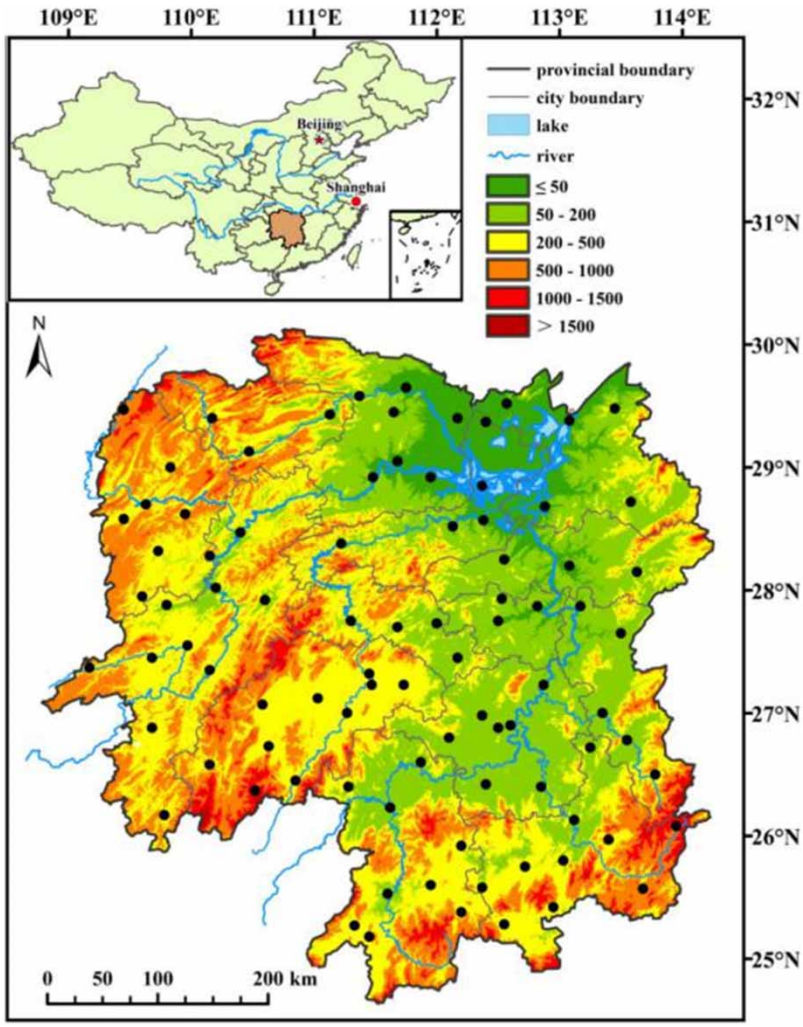

Figure 1 | Geographical distribution of the observation stations.

2.2. Methodology

The gridded daily precipitation from model predictions was interpolated to 96 national stations in Hunan Province by the bilinear method. Then the predicted precipitation was matched with the observation from April to August during 2017–2020 to calculate the prediction score of the heavy precipitation process to evaluate the prediction skills of the three models.

The standard of moderate rain in Hunan is that the precipitation exceeds 10 mm in a day, which will have a certain impact on traffic, agriculture, and so on. Therefore, the daily precipitation at a single station above 10 mm is considered heavy precipitation in this study. The heavy precipitation process is determined by the standard that there are at least six stations in the

whole province having heavy precipitation in a single day. The heavy precipitation process can be classified as a short process and a long process. The short process refers to the process with a duration of 2 days or less, and the long process lasts for 3 days or more.

2.2.1. Scoring rules for predictions of heavy precipitation process

At present, the meteorological department in China has not yet established a national standard for the evaluation of extended range prediction of the heavy precipitation process. Due to the actual needs of the local vocational work, we developed the scoring rules in this article based on actual service needs and practical experiences in Hunan. The following are specific rules.

For short precipitation processes, the statistical period is from the day before the first day of the process to the day after the last day. If heavy precipitation is observed at six stations or more on any single day in the statistical period, it is considered that the process is correctly predicted; otherwise, it is a false alarm prediction. For the prediction process that lasts only one day, the score is 100 for correct prediction, and the score is 0 for false alarm prediction.

When the precipitation process is predicted to be 2 days, if the observed heavy precipitation process is completed within the model forecast period, the prediction score is recorded as 100. If only part of the observed process occurs in the model forecast period, the prediction is scored as 70. If no heavy precipitation process is observed during the model forecast period, the score is marked as 0, indicating a false alarm prediction.

For a long precipitation process, the statistical period is the same as the forecast period. In this period, if the number of observed heavy precipitation days is greater than or equal to 1/3 of the forecast period, the prediction is considered correct, and the score is recorded as 100; otherwise, it is a false alarm prediction with the score being 0.

The days outside the statistical periods are regarded as the days without predicted heavy precipitation. These days, if the heavy precipitation process is observed, it is regarded as a missing prediction.

The prediction score of the heavy precipitation process, P_{r0}, is calculated as follows:

$$P_{r0} = \frac{\sum_{i=1}^{r} P_i}{N + N_f}, \tag{1}$$

where P_i is the score of each correct prediction, r is the number of heavy precipitation process, which is correctly predicted, N_f is the number of missing predictions, and N is the number of predicted processes.

For the different forecast leading times, the average score can be calculated as follows:

$$P_{r0}(\text{ARI\_AVG}) = \frac{P_{rA1-5} + P_{rA6-10} + P_{rA11-15}}{3} \tag{2}$$

where P_{rA1-5} is the prediction score at a short leading time of 1–5 days, P_{rA6-10} is the prediction score at a medium leading time of 6–10 days, and $P_{rA11-15}$ is the prediction score at long leading times of 11–15 days, which are all calculated by Equation (1). $P_{r0}(ARI\_AVG)$ is the average prediction score of a single model.

It should be specially noted that if there are multiple prediction samples in different forecast leading times for a certain forecast future period, the average scores of all samples during the corresponding forecast future period were taken as the values of P_{rA1-5}, P_{rA6-10}, and $P_{rA11-15}$. The original prediction score of a single model (or original score for short) was calculated by Equations (1) and (2).

2.2.2. Adjusted method for single-model prediction

Since there are some systematic errors in the model prediction, the heavy precipitation prediction was adjusted by changing the threshold value of identifying heavy precipitation. The specific procedures are as follows: (1) take the threshold values from 6 to 15 mm with an interval of 1 mm to determine the heavy precipitation day; (2) calculate the prediction score under each threshold value; (3) select the threshold value with the highest prediction score as the optimal threshold to determine the heavy precipitation process (hereinafter referred to as optimal threshold method). Finally, the adjusted prediction scores of each model (or adjusted scores for short) at different forecast leading times for certain forecast future periods are calculated by Equations (1) and (2).

It should be noted that analysis has been conducted on the range of threshold. If the threshold is less than 5 mm, it will increase the false alarm rate. On the contrary, if it exceeds 16 mm, it will increase the missing rate.

2.2.3. Integration method for multi-model predictions

To determine whether a certain day is a heavy precipitation day by the results of multiple models, an integration method was established by defining the coefficient P_d as follows:

$$P_d = \frac{\frac{W_{\mathrm{CMA1}}}{N_{1-5}}\sum_{i=1}^{5}P_i + \frac{W_{\mathrm{CMA2}}}{N_{6-10}}\sum_{i=6}^{10}P_i + \frac{W_{\mathrm{CMA3}}}{N_{11-15}}\sum_{i=11}^{15}P_i + \frac{W_{\mathrm{NCEP1}}}{N_{1-5}}\sum_{i=1}^{5}P_i + \frac{W_{\mathrm{NCEP2}}}{N_{6-10}}\sum_{i=6}^{10}P_i + \frac{W_{\mathrm{NCEP3}}}{N_{11-15}}\sum_{i=11}^{15}P_i + \frac{W_{\mathrm{ECMWF1}}}{N_{1-5}}\sum_{i=1}^{5}P_i + \frac{W_{\mathrm{ECMWF2}}}{N_{6-10}}\sum_{i=6}^{10}P_i + \frac{W_{\mathrm{ECMWF3}}}{N_{11-15}}\sum_{i=11}^{15}P_i}{W_{\mathrm{CMA1}} + W_{\mathrm{CMA2}} + W_{\mathrm{CMA3}} + W_{\mathrm{NCEP1}} + W_{\mathrm{NCEP2}} + W_{\mathrm{NCEP3}} + W_{\mathrm{ECMWF1}} + W_{\mathrm{ECMWF2}} + W_{\mathrm{ECMWF3}}}, \quad (3)$$

where P_i is the coefficient of heavy precipitation day predicted by a single model using the optimal threshold method, and the whole value is 1 for heavy precipitation day and 0 for nonheavy precipitation day. N_{1-5} is the number of prediction samples for each model at short forecast leading times (1–5 days), N_{6-10} is the number of prediction samples at medium forecast leading times (6–10 days), and N_{11-15} is the number of prediction samples at long forecast leading times (11–15 days).

W_{CMA1}, W_{CMA2}, and W_{CMA3}; W_{NCEP1}, W_{NCEP2}, and W_{NCEP3}; and W_{ECMWF1}, W_{ECMWF2}, and W_{ECMWF3} represent the weight coefficients of the CMA, NCEP, and ECMWF S2S models at different forecast leading times(1–5, 6–10, and 11–15 days), respectively, which were determined by the model's prediction skill for heavy precipitation process.

When the P_d of a certain day is equal to or larger than 0.5, the day is determined as an integrated heavy precipitation day; otherwise, there is no heavy precipitation on that day. Then, the prediction scores of multi-model integration (or integrated scores for short) at different forecast future times are calculated by Equation (1).

2.2.4. False alarm and missing rates

As described in Section 2.2.1, the false alarm prediction refers to the situation that there is a heavy precipitation process in prediction when no heavy precipitation process is found in observation. The missing prediction refers to the situation that there is a heavy precipitation process in observation but no heavy precipitation in prediction. Therefore, the false alarm rate is the ratio of the number of false alarm predictions to the total number of predicted heavy precipitation processes, and the missing prediction rate refers to the number of missing predictions divided by the total number of observed processes.

3. RESULTS

3.1. Assessment of single model

3.1.1. Original scores of single-model prediction

As shown in Figure 2, the original scores of the CMA, ECMWF, and NCEP models are relatively high in May and June, with the averages of the three models being 73.2 and 80.5, respectively. This may be related to the relatively concentrated

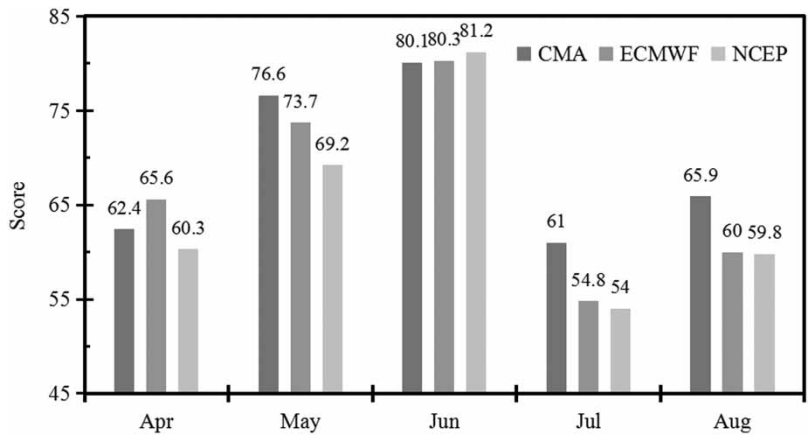

Figure 2 | Original scores of each model from April to August.

precipitation in Hunan during the 2 months. The scores in April, July, and August are relatively low, with the lowest score being 56.6 in July.

For the average score of each model from April to August, the score of the CMA model is 69.2, which ranks the first among the three, followed by 66.9 for the ECMWF model and 64.9 for the NCEP model.

3.1.2. Adjusted scores of single-model prediction

To eliminate the systematic errors of the CMA, ECMWF, and NCEP models, different thresholds were selected to calculate the prediction scores of heavy precipitation process during April to August for each model (shown in Tables 2–4). It can be found that the optimal thresholds from April to August are 7, 7, 6, 14, and 6 mm for the CMA model (shown in Table 2), while they are 8, 7, 6, 6, and 6 mm for the ECMWF model (shown in Table 3) and 6 mm, 6 mm, 9, 6 and 6 mm for the NCEP model (shown in Table 4), respectively.

In summary, the predicted precipitation is systematically lower than the observation, resulting in optimal thresholds being mostly lower than the original threshold of the observed heavy precipitation process (10 mm). With the increase of the threshold, the score shows a general downward trend.

Table 2 | Scores of the CMA model under different thresholds

Month \ Threshold (mm)	6	7	8	9	10	11	12	13	14	15
4	64.48	**64.5***	63.34	63.14	62.37	63.7	62.16	61.7	57.56	57.03
5	79.15	**79.92***	79.13	77.38	76.56	75.65	74.85	74.23	71.23	69.27
6	**81.92***	80.8	79.85	79.9	80.07	80.35	79.77	77.99	77.39	76.34
7	62.64	61.31	60.99	61.14	60.98	61.27	61.44	62.65	**63.23***	62.26
8	**72.24***	69.47	68.24	67.63	65.9	66.75	66.31	65.86	63.49	61.44

Table 3 | Scores of the ECMWF model under different thresholds

Month \ Threshold (mm)	6	7	8	9	10	11	12	13	14	15
4	66.58	65	**67.11***	66.97	65.6	66.18	66.55	65.63	66.32	67.01
5	77.32	**77.67***	74.18	73.23	73.69	75.5	74.18	75.01	73.3	71.31
6	**84.18***	82.51	82.14	81.21	80.33	79.41	78.74	78.84	76.53	76.9
7	**58.32***	56.62	53.38	53.18	54.81	53.43	52.82	53.76	53.76	52.85
8	**65.53***	61.23	60.3	60.79	60	62.64	56.06	50.79	55.37	55.15

Table 4 | Scores of the NCEP model under different thresholds

Month \ Threshold (mm)	6	7	8	9	10	11	12	13	14	15
4	**61.73***	60.28	60.68	60.02	60.28	58.83	57.26	56.75	56.25	55.51
5	**74.94***	71.4	70.94	70.91	69.2	68.14	68.54	67.49	65.99	65.08
6	81.36	82.28	82.61	**82.63***	81.18	76.92	79.03	76.32	74.84	75.05
7	**60.01***	59.97	57.52	58.06	54.04	54.63	53.31	51.69	52.35	50.72
8	**65.77***	64.78	62.72	61.24	59.84	59.29	59.8	56.33	54.17	51.92

Table 5 | Original and adjusted scores of different models in different forecast periods for the future

Model \ Future period	1–10 days	11–20 days	21–30 days	Average of extended range
Original CMA	75.7	67.8	64.0	65.9
Adjusted CMA	71.5	71.1	72.5	71.8
Original ECMWF	69.2	66.1	65.2	65.6
Adjusted ECMWF	72.6	65.0	65.2	65.1
Original NCEP	69.2	67.4	58.2	62.8
Adjusted NCEP	69.4	69.2	66.5	67.8

Based on the optimal thresholds, the predictions of the heavy precipitation process of CMA, ECMWF, and NCEP models were adjusted. Table 5 shows that the adjusted scores of the CMA and NCEP models are higher than the original scores in both forecasts of 11–20 and 21–30 days for the future. For the whole extended range of 11–30 days, both the average adjusted scores of the CMA and NCEP models are larger than the original scores, for 5.9 and 5.0, respectively. However, the average adjusted score of the ECMWF model slightly decreases by about 0.5 in the whole extended range. This is because the prediction skills of ECMWF in the next 1–10 days are also considered when determining the optimal threshold. After the threshold is determined, the prediction skills of ECMWF in the next 1–10 days are improved, but decreased in the next 11–30 days. In general, the skills in the next 1–30 days are still improved.

3.2. Assessment of the multi-model integration

According to the optimal threshold method, the scores of each model in predicting the heavy precipitation process in the next 1–30 days for different forecast leading times were analysed. It is found that the prediction score increases with the decreasing of forecast leading time. The prediction at the forecast leading times of 1–5 days has the highest average score of 70.2, followed by 69.5 at 6–10 days and 67.8 at 11–15 days. For different models, the CMA model has the highest prediction score, with an average of 71.6, followed by the NCEP model with an average score of 68.4, and the ECMWF model has the lowest average score of 67.5 (shown in Table 6).

Considering the prediction skills of different models, the weight coefficients of models during the process of multi-model integration were determined by the prediction scores. Specifically, higher prediction scores and smaller forecast leading times correspond to larger weight coefficients. Therefore, the coefficients of the CMA, NCEP, and ECMWF models and the forecast leading times of 1–5, 6–10, and 11–15 days were both assigned with 1, 0.8, and 0.6, respectively. The weight coefficient for a certain model at certain forecasting leading time was obtained by multiplying the two corresponding coefficients (shown in Table 7).

Based on the adjusted scores in Table 6 and the weight coefficients in Table 7, the integrated scores of the heavy precipitation process at different forecast future times were calculated by Equations (1)–(3). The results show that the integrated scores at the forecast of 11–20 days, 21–30 days, and the whole extended range for the future are 77.4, 72.8, and 75.1, respectively, which is higher than the original and adjusted scores (shown in Figure 3).

Table 6 | Adjusted scores of different models at different forecast leading times

Model	1–5 days	6–10 days	11–15 days	Average
CMA	70.87	72.44	71.54	71.6
NCEP	70.61	66.99	67.52	68.4
ECMWF	69.16	69.07	64.26	67.5
Average	70.2	69.5	67.8	69.2

Table 7 | Weight coefficients of different models at different forecast leading times

Model / Forecast leading time	CMA	NCEP	ECMWF
1–5 days	1	0.8	0.6
6–10 days	0.8	0.64	0.48
11–15 days	0.6	0.48	0.36

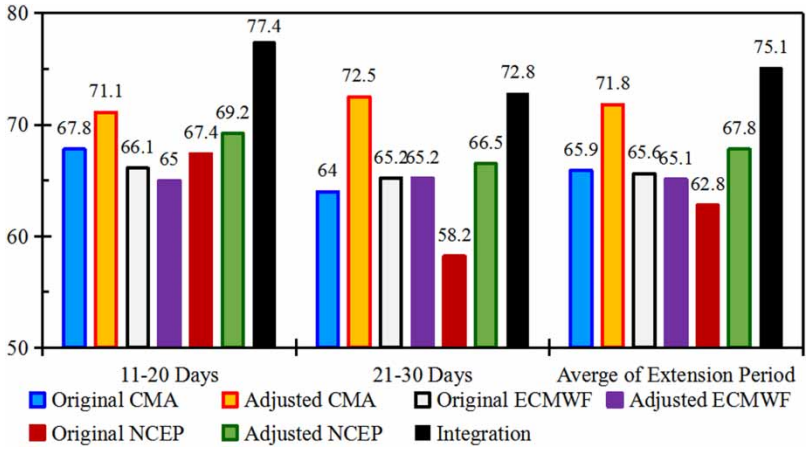

Figure 3 | Original, adjusted, and integrated scores of three models at different forecast future times.

Specifically, the original scores of the CMA, ECMWF, and NCEP models in the range of next 11–20 days are 67.8, 66.1, and 67.4, respectively, with an average of 67.1, while the adjusted scores are 71.1, 65.0, and 69.2, respectively, with an average of 68.4, slightly higher than the average of original scores. The integrated score is 77.4, which is 10.3 and 9.0 higher than the average of original and adjusted scores, respectively.

For the next 21–30 days, the original scores of the CMA, ECMWF, and NCEP models for heavy precipitation process prediction are 64.0, 65.2, and 58.2, respectively, with an average of 62.5. The adjusted scores are 72.5, 65.2, and 66.5, respectively, with an average of 68.1, higher than the average of original scores (62.5) by 5.6. The integrated score is 72.8, which is 10.3 and 4.7 higher than the averages of original and adjusted scores over the three models, respectively.

The original scores of the CMA, ECMWF, and NCEP models in the whole extended range are 65.9, 65.6, and 62.8, respectively, with an average of 64.8. The adjusted scores are 71.8, 65.1, and 67.8, respectively, with an average of 68.2, higher than the average of original scores (64.8) by 3.4. The integrated score is 75.1, which is 10.3 and 6.9 higher than the averages of original and adjusted scores over the three models, respectively.

3.3. Evaluation of the false alarm and missing rates

False alarms and missing rates are also important indicators to evaluate the prediction skills of models.

It was found that the false alarm rates of the original CMA, ECMWF, and NCEP models are 8.0, 16.9, and 17.1%, respectively, with an average of 14.0%, while the false alarm rates of the adjusted method decrease to 6.2, 8.7, and 10.0%, with an average of 8.3%, which is 5.7% smaller than the original average rate. The false alarm rate of the integrated method is 5.4%, which decreases by 8.6 and 2.9% with respect to the original and adjusted average rates, respectively.

The missing rates of the original CMA, ECMWF, and NCEP models are 37.2, 38.7, and 39.9%, respectively, with an average of 38.6%. The missing rates of the adjusted method are 34.8, 37.9, and 39.0%, respectively, with an average of 37.2%, which is 1.4% smaller than the original average rate. The missing rate of the integration method is 33.9%, which is 4.7 and 3.3% smaller than the original and adjusted average rates, respectively (shown in Table 8).

Table 8 | The false alarm and missing rates of original models, adjusted and integration methods

Method / Rate	Original CMA	Adjusted CMA	Original ECMWF	Adjusted ECMWF	Original NCEP	Adjusted NCEP	Original average	Adjusted average	Integration
False alarm	8.0%	6.2%	16.9%	8.7%	17.1%	10.0%	14.0%	8.3%	5.4%
Missing	37.2%	34.8%	38.7%	37.9%	39.9%	39.0%	38.6%	37.2%	33.9%

4. CONCLUSIONS AND DISCUSSION

In this study, we first defined the heavy precipitation process and the scoring rules for heavy precipitation prediction according to the characteristics of precipitation in Hunan Province. Based on the forecast precipitation data during 2017–2020 from three operational models (CMA, ECMWF, and NCEP) that participated in the S2S prediction project and the observational daily precipitation data from 96 national meteorological stations in Hunan, the monthly prediction skills from April to August of the models were evaluated. According to the prediction scores under different thresholds of heavy precipitation, the optimal threshold of each model for each month was obtained to calculate the adjusted scores of different models at different forecast leading times. The weight coefficients of the models were further obtained based on the adjusted scores and were used to determine the multi-model integration method and calculate the integrated scores. Finally, by defining the false alarm prediction and missing prediction, the false alarm and missing rates of original models, and adjusted and integration methods were analysed. The main conclusions are as follows.

The original prediction scores of the heavy precipitation process for each model are relatively high during May to June and low in April, July, and August, with the highest score of 80.5 in June and the lowest score of 56.6 in July. The CMA model has the highest average score, followed by the ECMWF and NCEP models.

The adjusted scores of the CMA and NCEP models are higher than the original scores in all forecasts of 11–20 days, 21–30 days, and the whole extended range for the future, while the adjusted scores of the ECMWF model are lower. For the whole extended range of 11–30 days, the adjusted scores of both the CMA and NCEP models increase by about 5.9 and 5.0, but it slightly decreases by about 0.5 for the ECMWF model.

The integrated scores in the forecasts of 11–20 and 21–30 days for the future are 77.4 and 72.8, respectively, both higher than the original and adjusted scores of single-model prediction. The integrated score in the whole extended range is 75.1, which is 10.3 and 6.9 higher than the averages of original and adjusted scores over the three models, respectively.

The false alarm rate of the integration method is 5.4%, which is 8.6 and 2.9% smaller than the average rates of the original models and adjusted methods, respectively. Similarly, the missing rate of the integration method is 33.9%, which is 4.7 and 3.3% smaller than the average missing rates of original models and adjusted methods, respectively.

The primary purpose of this study is to establish a relatively stable method for forecasting heavy precipitation processes as far ahead as possible to provide services for the government in disaster prevention and mitigation. However, it should be pointed out that the prediction of heavy rainfall in the extended range is a very complex work. Taking this study as an example, there is still a lot to be improved as discussed in the following.

The scoring rules of the heavy precipitation process only consider the duration of the heavy precipitation process, but ignore the spatial distribution. This is because current models have low accuracy in predicting the location of heavy precipitation processes. However, it is relatively accurate to predict the time period when heavy precipitation processes occur on a larger scale. Therefore, government decision-makers are more concerned about whether there has been heavy rain or the frequency of heavy rain in the administrative region, and release corresponding warning information (such as control of reservoir levels). Anyway, this will bring great limitations and make the prediction unable to fully meet the requirement.

In determining the weight coefficients of models, although the prediction skills at different forecast leading times are considered, where smaller forecast leading times and higher prediction scores correspond to larger weight coefficients, there are still many subjective factors. It is necessary to adopt different methods to obtain the optimal scheme to calculate the weight coefficients in the future.

ACKNOWLEDGEMENTS

We thank Nanjing Hurricane Translation for reviewing the English language quality of this paper.

AUTHOR CONTRIBUTIONS

C. M. analysed the data. J. P. put forwarded the idea. C. M. and J. P. wrote the manuscript. All authors revised the manuscript.

FUNDING

This study was supported by the Support Plan of Enterprise Innovation and Entrepreneurship Team for science and technology in Hunan and the Capacity Building Project of Hunan Meteorological Bureau in 2020 (Grant No. NLJS04).

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Bombardi, R. J., Pegion, K. V., Kinter, J. L., Cash, B. A. & Adams, J. M. 2017 Sub-seasonal predictability of the onset and demise of the rainy season over monsoonal regions. *Frontiers in Earth Science* **5**, 14.

CMA 2018 S2S Models [EB/OL]. [2022-08-28]. Available from: http://s2s.cma.cn/Models/.

Hu, Z. G., Wei, L., Xue, F. & Zhou, Q. L. 2019 Design and implementation of the global forecast service sharing platform of the World Meteorological Center. *Meteorological Science and Technology* **47**, 581–591. (In Chinese).

Hu, X., Zhang, Z. Q., Zhang, Q. & Wang, J. 2020 Analysis and application of sub-seasonal-seasonal (S2S) forecast data from the National Meteorological Information Center. *Meteorological Science and Technology* **48**, 779–787. (In Chinese).

Kim, H. M., Webster, P. J. & Toma, V. E. 2014 Predictability and prediction skill of the MJO in two operational forecasting systems. *Journal of Climate* **27**, 5364–5378.

Li, Q. P., Yang, S., Wu, T. W. & Liu, X. W. 2017 Sub-seasonal dynamical prediction of east Asian cold surges. *Weather and Forecasting* **32**, 1675–1694.

Liang, P. & Lin, H. 2017 Sub-seasonal prediction over East Asia during boreal summer using the ECMWFCC monthly forecasting system. *Climate Dynamics* **50**, 1007–1022.

Lin, Q., Chen, J., Li, W. & Li, X. Q. 2019 Performance evaluation of S2S sub-seasonal to seasonal forecasts for global precipitation forecasting by state key laboratory of water resources and hydropower engineering science. *Water Resources Research* **8**, 547–556. (In Chinese).

Marshall, A. G. & Hendon, H. H. 2015 Sub-seasonal prediction of Australian summer monsoon anomalies. *Geophysical Research Letter* **42**, 10913–10919.

Neena, J. M., Lee, J. Y. & Waliser, D. 2014 Predictability of the Madden-Julian oscillation in the intraseasonal variability hindcast experiment. *Journal of Climate* **27**, 4531–4543.

Pang, Y. S., Qin, N. S., Liu, B., Sun, Z. X. & Yang, S. Q. 2021 Analysis on prediction skills of S2S models for extreme precipitation during flood season in Sichuan Province. *Meteor Mon.* **47**, 586–600. (In Chinese).

Peng, B., Li, X. J., Yao, Y. H. & Tang, Y. M. 2021 Predictability of the Madden-Julian Oscillation in the subseasonal-to-seasonal prediction models. *Journal of Meteorological Sciences* **41**, 339–348. (In Chinese).

Qi, Y. J. & Rong, X. Y. 2014 Application prospects and outlook of sub-seasonal-seasonal forecasting. *Advances in Meteorological Science and Technology* **24**, 74–75. (In Chinese).

Rashid, H. A., Hendon, H. H. & Wheeler, M. C. 2011 Prediction of the Madden-Julian oscillation with the POAMA dynamical prediction system. *Climate Dynamics* **36**, 649–661.

Tian, D., Wood, E. F. & Yuan, X. 2017 CFSv2-based sub-seasonal precipitation and temperature forecast skill over the contiguous United States. *Hydrology and Earth System Sciences* **21**, 1477–1490.

Vitart, F. 2014 Evolution of ECMWF sub-seasonal forecast skill scores. *Quarterly Journal of the Royal Meteorological Society* **140**, 1889–1899.

WMO 2013 Sub-seasonal to Seassonal prediction research implementation plan [EB/OL].[2022-08-28]. Available from: http://www.s2sprediction.net/file/documents_reports/S2S_Implem_plan_en.pdf

Wu, T. W. 2012 A mass-flux cumulus parameterization scheme for large-scale models: description and test with observations. *Climate Dynamics* **38**, 725–744.

Wu, T. W., Yu, R. C. & Zhang, F. 2010 The Beijing Climate Center atmospheric general circulation model: description and its performance for the present-day climate. *Climate Dynamics* **34**, 123–147.

Wu, T. W., Li, W. P. & Ji, J. J. 2013 Global carbon budgets simulated by the Beijing Climate Center climate system model for the last century. *Journal of Geophysical Research* **118**, 4326–4347.

Wu, T. W., Song, L. C., Li, W. P., Wang, Z. Z., Zhang, H. & Xin, X. G. 2014 Progress of climate system model development in Beijing Climate Center: application to climate change research. *Journal of Meteorology* **72**, 12–29. (In Chinese).

Zhao, C. B., Ren, H. L., Wu, J., Zhou, F. & Lie, Y. 2016 Comprehensive evaluation reports of BCC_CSM12 S2S. *Beijing:Laboratory for Climate Studies, China Meteorological Administration*. (In Chinese).

First received 28 September 2022; accepted in revised form 6 June 2023. Available online 19 June 2023