Journal of Civil Engineering Researchers

Journal homepage: www.journals-researchers.com

A Comparative Analysis of Ensemble NWP Models for Flood Forecasting

Saleh Aminyavari [a,*]

[a] Department of Civil Engineering, Cha.C., Islamic Azad University, Chalous, Iran

ABSTRACT

This study presents the first application of gamma quantile mapping to bias-correct ensemble precipitation forecasts from seven global NWP models (ECMWF, NCEP, UKMO, CMA, JMA, ECCC, NCMRWF) in the data-scarce Saliyan Basin, Iran. The integration of these models with advanced bias correction techniques significantly improves flood forecasting accuracy. To address systematic biases in the raw forecasts, gamma quantile mapping was applied, significantly enhancing the reliability of the precipitation inputs. These bias-corrected forecasts were then used as inputs for the GR4J hydrological model to simulate river flow and predict flood events. The study period included a major flood event in March 2019, which was used to evaluate the performance of the ensemble forecasting system. Results demonstrated that bias correction using gamma quantile mapping substantially improved the accuracy of flood forecasts, with the ECMWF and UKMO models showing the highest skill scores. The ensemble approach effectively captured the uncertainty in flood predictions, providing valuable insights for risk assessment and decision-making. This research highlights the importance of bias correction in ensemble forecasting and offers a robust framework for flood prediction in data-scarce regions. The findings have significant implications for improving flood early warning systems and mitigating flood-related damages in similar basins worldwide.

ARTICLE INFO

Received: October 13, 2025
Accepted: Novembre 01, 2025

Keywords:
Flood Forecasting
Ensemble Numerical Weather Prediction (NWP)
Gamma Quantile Mapping
Uncertainty Analysis
Iran

DOI: 10.61186/JCER.7.4.23
DOR: 20.1001.1.22516530.1399.11.4.1.1

1. Introduction

Weather forecasts play a crucial role in hydrological applications. Predicting the likelihood of floods provides an opportunity for disaster managers to plan for damage reduction, manage water distribution, and mitigate potential damages. Therefore, managers need an index of forecast uncertainty to assess the risks associated with management decisions. Today, significant advancements have been made in improving weather forecasts, one of which is numerical weather predictions (NWP). NWP values are generated by solving the governing equations of the atmosphere using three-dimensional numerical methods for various temporal and spatial scales. Improving numerical precipitation forecasts is a primary goal of forecasting centers and a major challenge for the hydrometeorological research community. The weaknesses of NWP models in accurately describing atmospheric processes, along with unavoidable random errors in solving numerical equations, result in significant

* Corresponding author. Tel.: +989112912393; e-mail: saleh1161a@gmail.com.

uncertainty in NWP forecasts. Since the data from NWP models are used as inputs for hydrological models to predict flow, the accuracy of these data affects the simulated flow results. Thus, NWP model results impose considerable uncertainty on hydrological models. Given the limitations of deterministic forecasts in atmospheric states and changes in initial conditions, ensemble forecasting methods have been developed to improve the capability of numerical and probabilistic forecasts. Ensemble forecasts involve multiple individual forecasts generated with different physical parameterizations or initial conditions [1]. One of the key issues in using ensemble precipitation forecasts in hydrological models is addressing input uncertainty. Extensive research has been conducted on the application of ensemble forecasts in hydrological applications, some of which are discussed below.

Thirel et al. (2008) evaluated the capabilities of two ensemble forecasting systems, ECMWF and PEARP, for river flow forecasting across France. The results showed that ensemble flow forecasts based on PEARP data were better for floods and small basins, while ECMWF data performed better for large basins and low flows [2]. He et al. (2010) used TIGGE meteorological data to create a flood warning system for the upper Huai River basin in China. A grand ensemble was created from five centers with equal weight coefficients. Precipitation was classified into low, medium, and high categories, and the results showed that evaluation scores decreased from low to heavy precipitation, indicating that the equal weight coefficients needed further investigation [3]. Alfieri et al. (2014) evaluated the European Flood Awareness System (EFAS), which uses outputs from ECMWF and DWD models. The results indicated that model performance significantly decreased for basins smaller than 300 km² due to underestimation of runoff in mountainous areas [4]. Bennett et al. (2014) developed a continuous ensemble hydrological forecasting system (SCHEF) for nine Australian basins using NWP forecasts. NWP forecasts were post-processed using the BJP method, and the GR4H rainfall-runoff model was used for hydrological modeling. The results showed that SCHEF effectively predicted river flow, especially for a 1- to 6-day forecast horizon [5]. Zomerdijk (2015) examined the development of a flood ensemble forecasting system in Quzhou, eastern China, for a 1- to 10-day forecast horizon and evaluated flood forecasts using precipitation forecasts from four meteorological centers: ECMWF, NCEP, UKMO, and CMA. The integrated GR4J hydrological model was used for flood forecasting. Ensemble precipitation forecasts were corrected using quantile mapping, and the results showed that all models had good flood forecasting capabilities, with ECMWF performing the best and CMA the worst [6]. Thiemig et al. (2015) assessed the capabilities of the African Flood Forecasting System (AFFS) based on ECMWF forecasts. The results showed that AFFS correctly predicted 70% of floods [7]. Matsueda and Nakazawa (2015) created a rapid warning system using ensemble forecasts from UKMO, NCEP, ECMWF, and JMA models in the TIGGE database. They assessed the probability of extreme weather events and found that these models successfully predicted severe events like the 2010 Russian heatwave, the 2010 Pakistan floods, and Hurricane Sandy in 2012 [8]. Cai et al. (2019) studied uncertainty in precipitation forecasts from four TIGGE centers in China's Huai River basin during flood season. They introduced a new model using fuzzy probabilities and Bayesian theory (GPDF), finding it highly accurate, reliable, and sharp. Uncertainty rose with forecast horizon, and the model provided acceptable accuracy for flood risk analysis up to three days [9].

Zhang et al. (2020) demonstrated that advanced bias correction methods, such as gamma quantile mapping, can significantly enhance the accuracy of precipitation forecasts [10]. Similarly, Li et al. (2021) investigated the impact of high-resolution ensemble NWP models on flood forecasting in small basins and found that these models can effectively reduce uncertainty [11]. In another study, Wang et al. (2022) combined ensemble NWP models with machine learning techniques for flood prediction and observed a notable improvement in forecast accuracy [12]. Additionally, Kumar et al. (2023) examined the influence of input data quality on flood forecasting accuracy and highlighted the critical role of bias correction in improving results [13]. Martinez et al. (2021) conducted an uncertainty analysis in flood forecasting using ensemble NWP models and advanced statistical methods, showing that these approaches can effectively manage uncertainty [14]. Finally, Lee et al. (2023) explored the impact of regional characteristics on the performance of flood forecasting methods and emphasized the importance of selecting appropriate techniques for each region [15].

In evaluating TIGGE ensemble precipitation forecasts for Iran, Aminyavari et al. (2018) examined forecasts from three centers (ECMWF, UKMO, and NCEP) in eight different precipitation groups across Iran. The results showed that ECMWF performed better in most regions, UKMO in mountainous areas, and NCEP along the Persian Gulf coast [16]. Aminyavari et al. (2018) post-processed TIGGE forecast data in the Beshar basin using a combination of quantile mapping and Bayesian averaging. They concluded that the forecasting skill of the models for the Beshar basin improved, and the VR histogram obtained from each model showed a uniform distribution. The combined BMA forecast had higher skill than individual models [17]. Aminyavari et al. (2019) evaluated NWP models (ECMWF, UKMO, NCEP) and GPM satellite for 2019 floods in Iran. Satellite estimates were more accurate

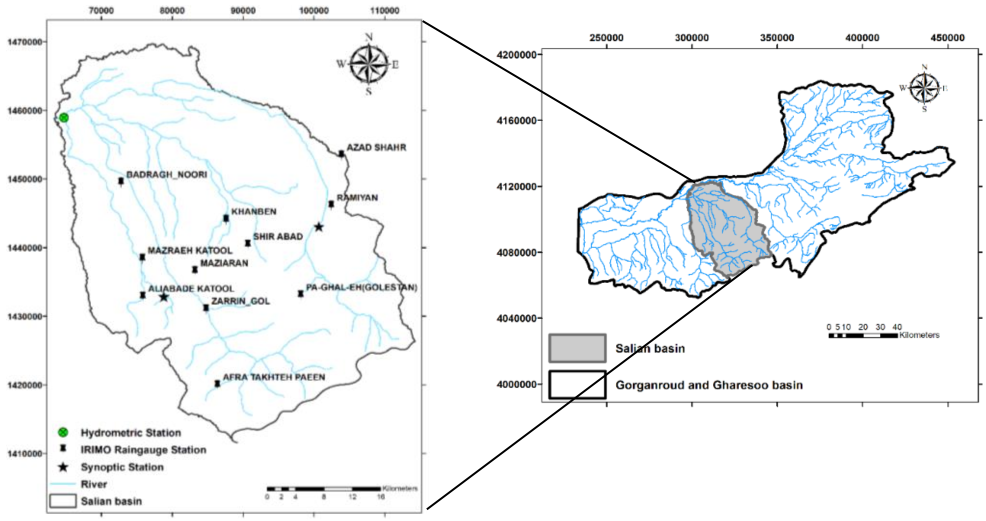

Figure 1. Layout of rain gauge of Research Basin

Table1.

Specifications of seven NWP models in TIGGE database

Center	No. of Ensemble Members	Horizontal Resolution Archived	Forecast Length (day)	Initial Perturbation Method
ECMWF	50	N320(~0.28°) N160(~0.56°)	0-10 10-15	EDA-SVINI
NCEP	20	1.0°×1.0°	0-16	BV-ETR
UKMO	17	0.83°×0.56°	0-15	ETKF
CMA	14	0.56°×0.56°	0-10	BV
JMA	26	1.25°×1.25°	0-11	SV
ECCC	20	1°×1°	0-16	EKF
NCMRWF	11	0.25°×0.25°	0-10	ETKF

for precipitation amount, UKMO excelled in spatial distribution, NCEP's performance decreased with higher thresholds, and ECMWF had better POD and lower false alarms at specific thresholds [18]. Hoghoughinia et al. 2024 evaluated three post-processing methods—Quantile Mapping (QM), Support Vector Machine (SVM), and Random Forest (RF)—applied to ECMWF precipitation forecasts over Iran. The RF method significantly improved forecast accuracy, particularly in regions with higher precipitation rates, demonstrating the importance of post-processing for enhancing flood forecasting and management [19].

Despite advances in ensemble forecasting, the combined use of seven NWP models with gamma quantile mapping remains unexplored, particularly in regions prone to flash floods due to complex topography and sparse data.

In this study, ensemble precipitation forecasts from seven centers (UKMO, ECMWF, NCEP, ECCC, JMA, NCMRWF, and CMA) were extracted from the TIGGE database for the Saliyan basin and bias-corrected using gamma quantile mapping. The bias-corrected ensemble precipitation forecasts were then used as inputs for the G4RJ rainfall-runoff model. In terms of innovation, no research has been conducted on bias-correcting ensemble precipitation forecasts from seven NWP models using gamma quantile mapping and analyzing uncertainty in ensemble flood forecasts for the Saliyan basin in 2019.

2. Materials and Methods

2.1. Study Basin Characteristics and Forecast Data

The main river of the Saliyan basin is the Saliyan Tappeh or Habib Eshan River. This river is formed by the confluence of the Qarasu Ramian, Siah Jub, Zarringol, and Kabul Val rivers and flows into the Gorgan River near the village of Habib Eshan after the Gorgan Dam. The basin area up to the Baghe Saliyan hydrometric station is approximately 1800 km². Eleven rain gauge stations from the Meteorological Organization, whose characteristics and locations in the basin are shown in Figure 1 based on UTM coordinates, were selected for precipitation.

The Baghe Saliyan hydrometric station was chosen for observed discharge, and the Aliabad Katul synoptic station was selected for temperature. The precipitation data from the 11 rain gauge stations and the Saliyan hydrometric station were tested for trends using the Mann-Kendall test. The results showed that p-values for all stations were above 0.05, indicating no trend in the observed data, and the Sen's slope was approximately zero for all stations. The Pettitt non-parametric test was also applied to test the homogeneity of the 11 rain gauge stations, and no significant changes were observed in the precipitation time series.

Ensemble precipitation forecast data from seven centers (UKMO, ECMWF, NCEP, ECCC, JMA, NCMRWF, and CMA) were extracted from the TIGGE database (https://apps.ecmwf.int/datasets/data/tigge/levtype=sfc/type=pf/) for the Saliyan basin with a resolution of 50 km. Their specifications are listed in Table 1.

Since forecast values are located at the center of the forecast grid points, which are 50 km apart and differ from the spatial coordinates of the rain gauge stations, these data need to be interpolated to the locations of the ground observation stations for accurate evaluation. Various interpolation methods exist but based on successful experiences in similar studies [20, 21, 22], the inverse distance weighting (IDW) method was used. IDW is a non-linear interpolation method that uses a weighted average of forecast values near the target station. In this study, four grid points around each station were selected, weighted based on their diagonal distance to the target station, and used in the IDW formula to calculate the forecast precipitation at the selected station. This step was performed for all seven models over the study years for each day with all ensemble members.

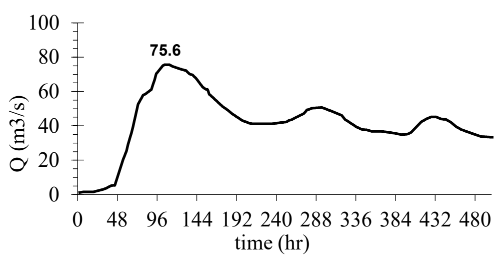

Figure 2. Flood hydrograph of Research Basin

On March 17, 2019, heavy rainfall in the northern provinces of Iran, including Golestan, Mazandaran, and North Khorasan, caused flooding. Many factors contributed to the flooding, but the most significant were soil erosion due to excessive forest exploitation, continuous rainfall, and the release of water from dams, leading to flooding in the two northern provinces of Iran March 2019. In this study, most sub-basins of the Gorgan River were examined, and the Saliyan sub-basin was selected for evaluating the performance of numerical models in flood forecasting due to the availability of statistical data and calibration results. Figure 2 shows the flood hydrograph for this basin. The hydrograph starts on March 17, 2019, with a peak flow of 75.6 m^3/s occurring on April 1, 2019, lasting for six hours (8 AM to 2 PM). The total flood volume was 75.32 million cubic meters. According to the Golstan Regional Water Authority, the total precipitation in the basin was approximately 250 mm, which, based on statistical analysis, was unprecedented in the past 50 years. The training period (October 2018–March 2019) was chosen to encompass seasonal variability in precipitation, ensuring robust calibration of the gamma function.

2.2. Study Basin Characteristics and Forecast Data

Most ensemble precipitation forecasts have bias errors that need correction. Bias can be unconditional (systematic) or conditional. Systematic bias refers to the difference between the mean forecasts and observations over the study period, which is not influenced by the user, while conditional bias is based on thresholds set by the user. The goal of bias correction is to correct systematic bias errors. Gamma quantile mapping was selected for its ability to handle zero-inflated, skewed precipitation distributions (Piani et al., 2010), unlike methods assuming normality (e.g., linear scaling). This is critical for the Saliyan Basin, where 80% of annual rainfall occurs in sporadic, high-intensity events [23]. The gamma distribution is the best fit for this type of data [24]. To perform bias correction, the cumulative distribution function (CDF) of observations and forecasts is first extracted based on the gamma distribution. Then, using the following formula, quantiles (Qn) from the forecast CDF are extracted, and new precipitation values are calculated from the observed CDF based on the obtained quantiles.

$$BC_{fcst}=CDF_{obs}^{-1}(CDF_{fcst}(Fcst))=CDF_{obs}^{-1}\left(Q_n\right) \quad (1)$$

The gamma quantile mapping method was implemented using the hyfo package [25] in R. The training period for bias correction was from October 1, 2018, to March 15, 2019, and forecasts from March 15, 2019, to April 15, 2019, were bias-corrected based on the fitted function from the training period. This process was performed separately for each ensemble member of each numerical model. In other words, bias correction was performed 158 times (50+20+17+14+26+20+11) for each day based on the third column of Table 1 (number of ensemble members).

2.3. Flow forecasting

The GR4J integrated continuous conceptual rainfall-runoff model was used to predict the flow of the Baghe Saliyan River during the study years. The model was implemented using the airGR package in R [26]. This model was selected based on reputable research [6, 7, 17] in flood forecasting. The main inputs for the model are precipitation, temperature, and potential evapotranspiration, which must be averaged over the basin. The Thiessen polygon method was used to calculate the average values of the required inputs for the Saliyan basin. First, observed precipitation and forecasts from all seven models were interpolated to the 11 selected stations in the basin using the Thiessen polygon method. This was done separately for each ensemble member for the 11 stations. After preparing the precipitation, discharge, and potential evapotranspiration data for the Saliyan basin, these values were merged based on common days, resulting in 5170 days from December 31, 2001, to August 22, 2017, available for calculating optimal parameters. A one-year warm-up period was set before the calibration period for model initialization. Thus, the period from December 31, 2001, to December 15, 2002 (some days lacked data), was used as the warm-up period, from December 16, 2002, to May 4, 2014, as the calibration period, and from May 5, 2014, to August 22, 2017, as the validation period. The optimal parameters obtained from the model were then used to predict floods using numerical precipitation forecasts from the seven centers from March 17, 2019, to April 15, 2019. The GR4J model has four parameters for flow forecasting, which must be optimized during calibration and tested during validation. Parameter (X1) represents the maximum soil moisture capacity of the basin. Soil moisture acts like a reservoir that is filled with precipitation and empties with potential evapotranspiration. Parameter (X2) indicates the influence of groundwater on the routing reservoir. A positive value indicates groundwater inflow into the routing reservoir, while a negative value indicates a decrease in the routing reservoir height and inflow into groundwater. A negative value suggests that some precipitation in the basin enters groundwater. Parameter (X3) represents the capacity of the routing reservoir, and parameter (X4) is the base time of the unit hydrograph for routing. In this study, four optimization methods were used to calculate the optimal parameters during calibration: the Michel optimization algorithm (available in GR4J) [27], the differential evolution algorithm (DE), the improved particle swarm optimization algorithm (PSO), and the memetic algorithm with local search (MA-LS).

For calibration and validation of the GR4J rainfall-runoff model and calculation of optimal parameters, daily observed data from the Saliyan basin, including discharge, precipitation, and temperature from 2002 to 2017, were used. Optimal parameters were calculated using four optimization methods. The RMSE index was used as the optimization criterion. The optimal parameters for all four methods are shown in Table 2. As evident from the table, the parameters were nearly identical across all methods.

The simulated discharges with optimal parameters were evaluated based on RMSE, bias in mean and standard deviation of simulated and observed discharges, the Nash-Sutcliffe efficiency (NSE), and the correlation coefficient (R2) for both calibration and validation periods, as shown in Table 3. The results were relatively good for the validation period.

Table 2.
Optimal parameters obtained in 4 optimization methods

Model	X1 (mm)	X2 (mm/d)	X3 (mm)	X4 (d)
airGR	391.937	-19.52	122.983	2.48
DE	387.927	-19.802	123.833	2.484
PSO	388.205	-19.763	123.727	2.482
MA-LS	387.258	-19.769	123.594	2.509

Table 3.
Streamflow simulation results with observational data on calibration and validation periods

	RMSE (mm/day)	$BIAS_{sd}$	$BIAS_{mean}$	NSC	R^2
Calibration	0.17	0.93	0.8	0.76	0.83
Validation	0.20	0.51	0.5	0.53	0.55

3. Results

In this section, all numerical precipitation forecasts from the seven models with all ensemble members were bias-corrected using gamma quantile mapping. Seventy percent of the data were used for training, and the remaining 30% were bias-corrected based on the fitted function from the training period. Based on the optimal parameters obtained in the previous section, flow forecasts were performed using raw and bias-corrected numerical precipitation model inputs. Figure 3 shows the flow forecast results for the CMA, ECMWF, JMA, NCEP, UKMO, ECCC, and NCMRWF models in both raw and bias-corrected states. As evident from the figures, the CMA model performed poorly in the raw state but improved after bias correction, although it still slightly underestimated the flood volume. The ECMWF model performed better in the raw state compared to other models and excelled after bias correction The JMA model performed poorly in the raw state and overestimated the flood volume after bias correction, likely due to inherent regional biases in its convective parameterization, which struggles to resolve orographic precipitation in the Saliyan Basin's steep topography (see Figure 1). The UKMO model performed relatively well in the raw state and excelled after bias correction. The ECCC model had moderate performance in the raw state and slightly underestimated the flood volume after bias correction, but all ensemble members improved. The NCMRWF model performed poorly in the raw state but improved after bias correction, although it still slightly underestimated the flow.

For a better comparison of the numerical models' performance in ensemble flow forecasting after bias

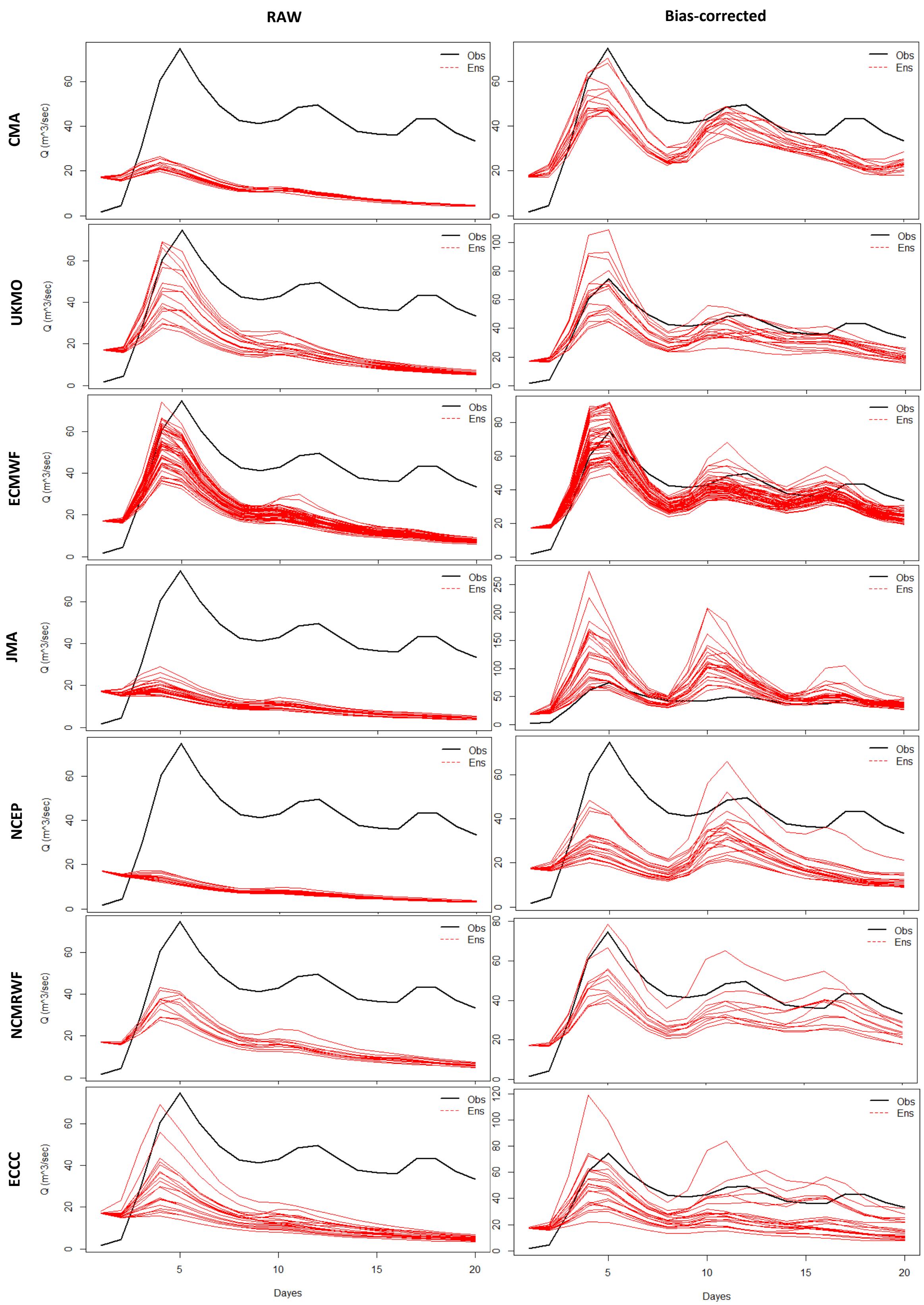

Figure 3. Ensemble flow forecast in two raw and bias-corrected modes of models

correction versus the raw state, box plots of the NSE values for the ensemble members of each numerical model are shown in Figure 4. As evident, the NSE values in the bias-corrected state were mostly above 0.5, indicating good model performance after bias correction. The ECMWF and UKMO models performed exceptionally well after bias correction, with nearly 50% of ensemble members achieving NSE values above 0.7. This superior performance likely stems from their higher spatial resolution (N320/0.28° for ECMWF and 0.83°×0.56° for UKMO; see Table 1) and advanced perturbation methods (EDA-SVINI for ECMWF and ETKF for UKMO), which enable better capture of localized rainfall dynamics in the Saliyan Basin. For the JMA model, although some ensemble members had very poor NSE values (as low as -2.5), most members had NSE values between 0.2 and 0.6.

Figure 5 shows the results of probabilistic flow forecasting evaluation for the seven numerical models using the continuous ranked probability score (CRPS) in both raw and bias-corrected states. As evident, all models performed poorly in the raw state and were weaker than the reference forecasts. After bias correction, the CRPS scores of all models improved significantly, with the UKMO and ECMWF models achieving positive scores. The cumulative distribution function (CDF) of the flow forecasts after bias correction became closer to the observed CDF.

Figure 6 shows the results of probabilistic flow forecasting evaluation for the seven models using the Brier skill score (BSS) in both raw and bias-corrected states. As evident, all models performed poorly in the raw state and were weaker than the reference forecasts. After bias correction, the BSS scores of all models, especially JMA, increased, indicating that the numerical models better predicted the probability of floods after bias correction. Additionally, the increase in BSS scores indicates that after bias correction, the uncertainty of all seven numerical models decreased, and their resolution increased.

Rank histograms were used to evaluate the performance of probabilistic flow forecasts from the seven numerical ensemble precipitation models. Figure 7 shows the rank histograms for the seven models in both raw and bias-corrected states.

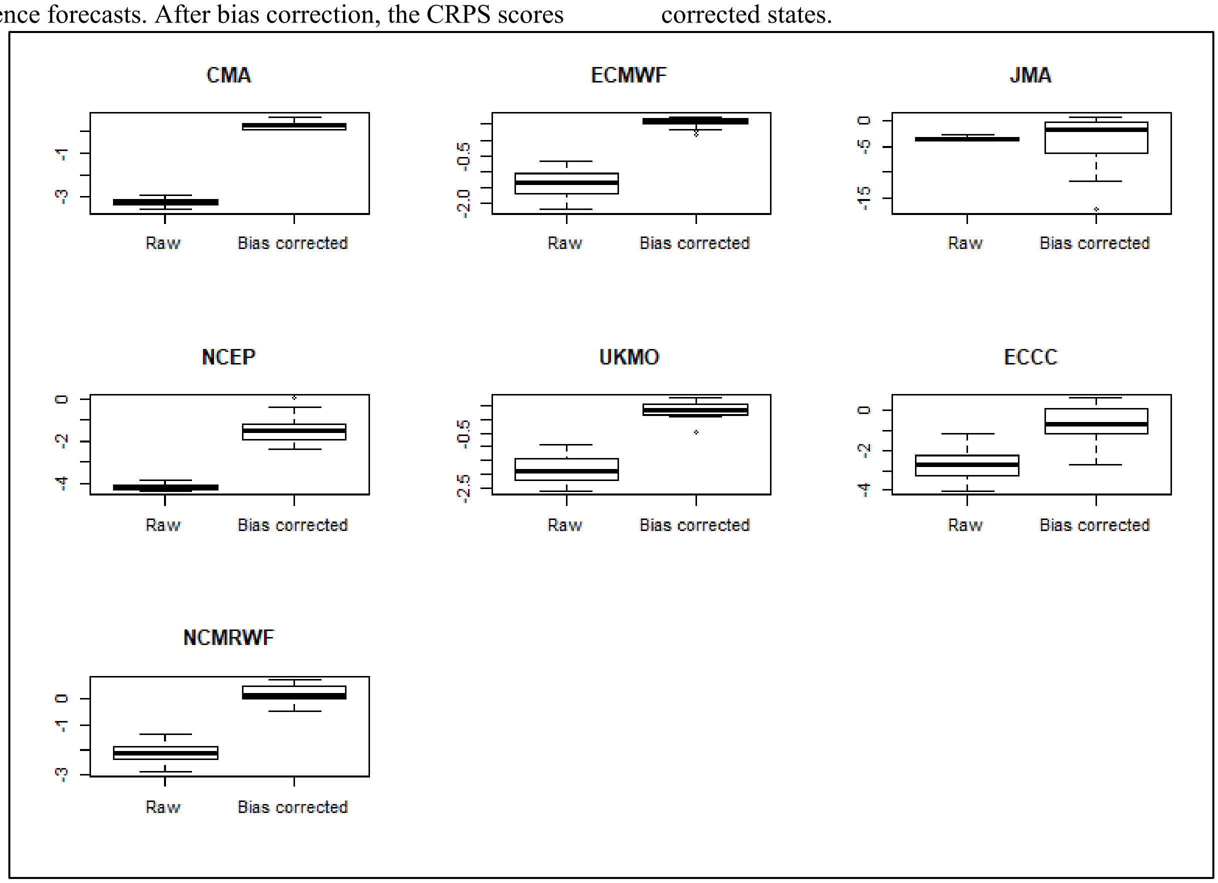

Figure 4. Boxplot for ensemble flow forecast in two raw and bias-corrected states for the Nash-Sutcliffe criterion

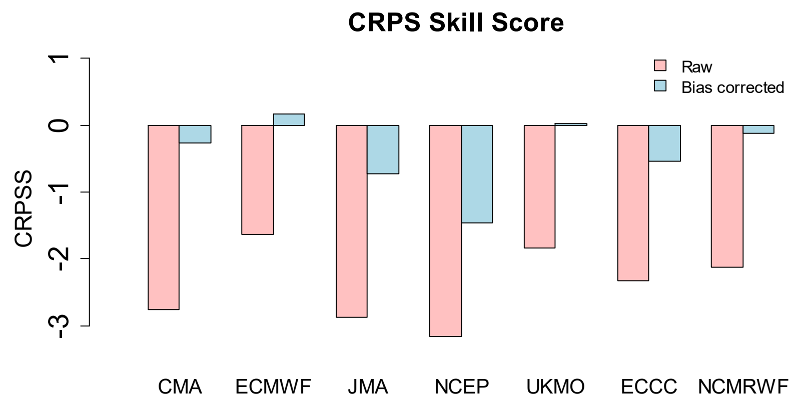

Figure 5. Results of ensemble flow forecasting evaluation of models with CRPS skill score

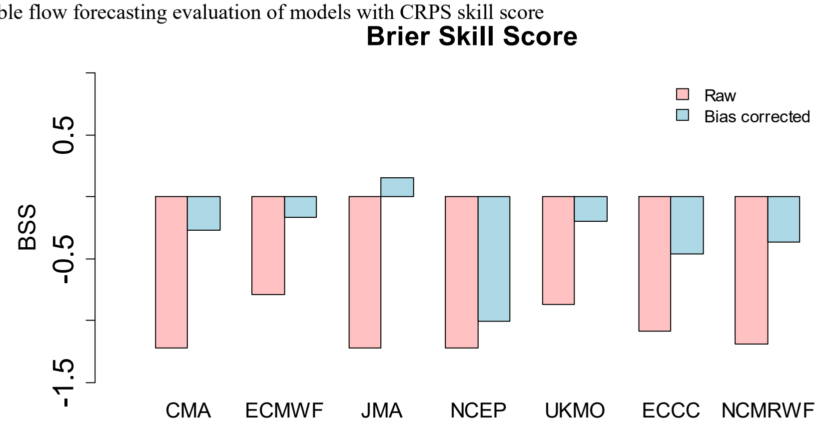

Figure 6. Results of ensemble flow forecasting evaluation of models with Brier skill score

The NCMRWF model has 11 ensemble members. After sorting the ensemble members in ascending order for each day, most observed discharges were higher than the ensemble forecasts, indicating that the model underestimated in the raw state. After bias correction, some observed discharges fell within the higher ensemble members, indicating improved flow forecasting. The CMA model has 14 ensemble members. Similar to the previous model, no observed discharges fell within the ensemble members in the raw state, but after bias correction, a few observed discharges fell within the higher ensemble members, although the model still underestimated flood discharges. The UKMO model has 17 ensemble members. In the raw state, a few observed discharges fell within the ensemble members. In the bias-corrected state, most of the 20 observed discharges fell within the ensemble members, indicating that the model effectively identified uncertainties. The ECCC model has 20 ensemble members. In the raw state, 4 observed discharges fell within the ensemble members, and in the bias-corrected state, 15 observed discharges fell within the ensemble members, indicating good flood forecasting performance. The NCEP model also has 20 ensemble members but captured the fewest observed discharges in both raw and bias-corrected states compared to other models. The JMA model has 26 ensemble members and performed poorly in the raw state, with almost all observed discharges falling into higher ranks. In contrast, in the bias-corrected state, most observed discharges fell within the lower ranks, indicating overestimation. Finally, the ECMWF model has 50 ensemble members. In the raw state, 4 observed discharges fell within the ensemble members, and in the bias-corrected state, 15 observed discharges fell within the ensemble members. The ECMWF model was the only one with observed discharges uniformly distributed among the ensemble members, indicating better performance compared to other models.

4. Discussion

The results demonstrate that high-resolution models like ECMWF/UKMO are particularly effective in topographically complex regions like northern Iran, where localized rainfall dynamics dominate flood risks. This aligns with Lee et al. (2023), who emphasized region-specific model selection for flood forecasting. The application of gamma quantile mapping effectively reduced systematic biases in the raw ensemble precipitation forecasts, leading to more reliable inputs for

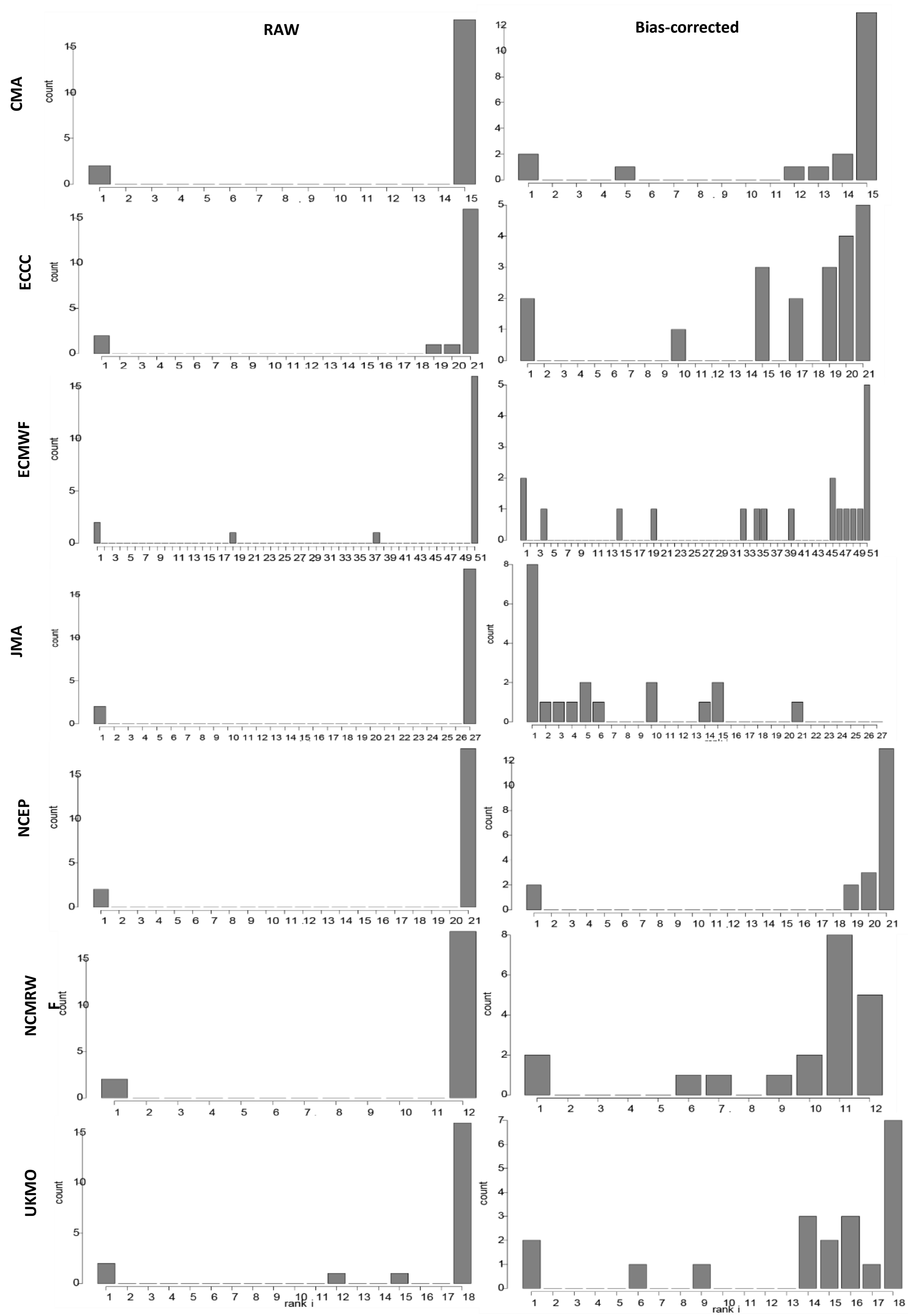

Figure 7. Rank histogram of probabilistic streamflow forecasts with raw numerical forecasts and bias-corrected precipitation for 7 models

the GR4J hydrological model. This improvement was particularly evident in the performance of the ECMWF and UKMO models, which consistently showed higher skill scores in predicting flood events. The ensemble approach also provided valuable insights into the uncertainty associated with flood forecasts, enabling better risk assessment and decision-making for flood management. The findings align with recent studies that have highlighted the importance of bias correction in ensemble forecasting. For instance, Zhang et al. (2020) [10] and Li et al. (2021) [11] emphasized the role of advanced bias correction methods in enhancing the accuracy of precipitation forecasts, especially in data-scarce regions. Similarly, the improved performance of the ECMWF and UKMO models observed in this study is consistent with the results of Wang et al. (2022) [12], who demonstrated the effectiveness of high-resolution ensemble NWP models in flood prediction. However, the relatively poor performance of the NCEP and JMA models, even after bias correction, suggests that the quality of initial conditions and model physics may play a critical role in the accuracy of ensemble forecasts. This observation is supported by Kumar et al. (2023) [13], who found that input data quality significantly influences flood forecasting outcomes. One of the key contributions of this study is the application of gamma quantile mapping to multiple ensemble NWP models, which has not been extensively explored in previous research. The results indicate that this method can effectively address systematic biases and improve the reliability of flood forecasts, particularly in regions with complex hydrological and meteorological conditions, such as the Saliyan Basin. However, the study also highlights the need for further research to optimize bias correction techniques and explore their applicability in different climatic and hydrological contexts.

5. Conclusion

This study investigated the potential of ensemble NWP models combined with gamma quantile mapping for improving flood forecasting in the Saliyan Basin, Iran. The results demonstrated that bias correction using gamma quantile mapping significantly enhanced the accuracy of precipitation forecasts, leading to more reliable flood predictions. The ECMWF and UKMO models emerged as the top-performing models, while the NCEP and JMA models showed relatively weaker performance, even after bias correction. The ensemble approach effectively captured the uncertainty in flood forecasts, providing valuable insights for flood risk management and decision-making. The findings of this study have important implications for improving flood early warning systems, particularly in data-scarce regions. By integrating ensemble NWP models with advanced bias correction techniques, it is possible to reduce the uncertainty associated with flood forecasts and enhance the reliability of hydrological predictions. Future research should focus on optimizing bias correction methods, exploring the use of machine learning techniques for post-processing ensemble forecasts, and evaluating the performance of these methods in different hydrological and climatic settings. Overall, this study contributes to the growing body of knowledge on flood forecasting and provides a robust framework for improving flood risk management in similar basins worldwide.

Acknowledgements

The author would like to thank the Islamic Azad University, Chalous Branch, for providing the necessary infrastructure and support to conduct this research. Appreciation is also extended to the Iranian Meteorological Organization and the Golestan Regional Water Authority for providing access to essential hydrological and meteorological data.

References

[1] Swinbank, R., M. Kyouda, P. Buchanan, L. Froude, T.M. Hamill, T.D. Hewson, et al. "The TIGGE project and its achievements." Bulletin of the American Meteorological Society 97.1 (2016): 49–67. https://doi.org/10.1175/BAMS-D-13-00191.1

[2] Thirel, G., F. Rousset-Regimbeau, E. Martin, and F. Habets. "On the impact of short-range meteorological forecasts for ensemble streamflow predictions." Journal of Hydrometeorology 9.6 (2008): 1301–1317. https://doi.org/10.1175/2008JHM951.1

[3] He, Y., F. Wetterhall, H. Bao, H. Cloke, Z. Li, F. Pappenberger, et al. "Ensemble forecasting using TIGGE for the July–September 2008 floods in the Upper Huai catchment: a case study." Atmospheric Science Letters 11.2 (2010): 132–138. https://doi.org/10.1002/asl.267

[4] Alfieri, L., F. Pappenberger, F. Wetterhall, T. Haiden, D. Richardson, and P. Salamon. "Evaluation of ensemble streamflow predictions in Europe." Journal of Hydrology 517 (2014): 913–922. https://doi.org/10.1016/j.jhydrol.2014.06.035

[5] Bennett, J.C., D.E. Robertson, D.L. Shrestha, Q.J. Wang, D. Enever, P. Hapuarachchi, and N.K. Tuteja. "A System for Continuous Hydrological Ensemble Forecasting (SCHEF) to lead times of 9 days." Journal of Hydrology 519 (2014): 2832–2846. https://doi.org/10.1016/j.jhydrol.2014.07.034

[6] Zomerdijk, L. "Performance of multi-model ensemble combinations for flood forecasting." Master Thesis; University of Twente: Netherlands (2015).

[7] Thiemig, V., B. Bisselink, F. Pappenberger, and J. Thielen. "A Pan-African medium-range ensemble flood forecast system." Hydrology and Earth System Sciences 19.8 (2015): 3365–3385. https://doi.org/10.5194/hess-19-3365-2015

[8] Matsueda, M. and T. Nakazawa. "Early warning products for severe weather events derived from operational medium-range ensemble

forecasts." Meteorological Applications 22.2 (2015): 213–222. https://doi.org/10.1002/met.1456

[9] Cai, C., J. Wang, and Z. Li. "Assessment and modelling of uncertainty in precipitation forecasts from TIGGE using fuzzy probability and Bayesian theory." Journal of Hydrology 577 (2019): 123995. https://doi.org/10.1016/j.jhydrol.2019.123995

[10] Zhang, Y., X. Wang, and J. Li. "Improving precipitation forecast accuracy using gamma quantile mapping." Journal of Hydrology 585 (2020): 124789. https://doi.org/10.1016/j.jhydrol.2020.124789

[11] Li, H., Y. Chen, and Q. Zhang. "The impact of high-resolution ensemble NWP models on flood forecasting in small basins." Water Resources Research 57.5 (2021): e2020WR028945. https://doi.org/10.1029/2020WR028945

[12] Wang, L., X. Liu, and T. Zhang. "Combining ensemble NWP models with machine learning for flood prediction." Hydrology and Earth System Sciences 26.3 (2022): 1234-1245. https://doi.org/10.5194/hess-26-1234-2022

[13] Kumar, R., P. Singh, and A. Sharma. "The influence of input data quality on flood forecasting accuracy." Journal of Hydrometeorology 24.2 (2023): 567-580. https://doi.org/10.1175/JHM-D-22-0123.1

[14] Martinez, G., M. Rodriguez, and J. Lopez. "Uncertainty analysis in flood forecasting using ensemble NWP models." Natural Hazards and Earth System Sciences 21.7 (2021): 2345-2358. https://doi.org/10.5194/nhess-21-2345-2021

[15] Lee, S., H. Kim, and J. Park. "The impact of regional characteristics on flood forecasting performance." Water Resources Management 37.4 (2023): 1456-1470. https://doi.org/10.1007/s11269-023-03456-8

[16] Aminyavari, S., B. Saghafian, and M. Delavar. "Evaluation of TIGGE ensemble forecasts of precipitation in distinct climate regions in Iran." Advances in Atmospheric Sciences 35.4 (2018): 457–468. https://doi.org/10.1007/s00376-017-7063-9

[17] Aminyavari, S., B. Saghafian, and E. Sharifi. "Assessment of Precipitation Estimation from the NWP Models and Satellite Products for the Spring 2019 Severe Floods in Iran." Remote Sensing 11.23 (2019): 2741. https://doi.org/10.3390/rs11232741

[18] Aminyavari, S., B. Saghafian, and E. Sharifi. "Assessment of Precipitation Estimation from the NWP Models and Satellite Products for the Spring 2019 Severe Floods in Iran." Remote Sensing 11.23 (2019): 2741. https://doi.org/10.3390/rs11232741

[19] Hoghoughinia, K., B. Saghafian, and S. Aminyavari. "Evaluation of precipitation temporal distribution pattern of post-processed sub-daily ECMWF forecasts." Theoretical and Applied Climatology 155.8 (2024): 8401–8414. https://doi.org/10.1007/s00704-024-04923-9

[20] Abdolmanafi, A., B. Saghafian, and S. Aminyavari. "Evaluation of global ensemble prediction models for forecasting medium to heavy precipitations." Meteorology and Atmospheric Physics 133.1 (2021): 15–26. https://doi.org/10.1007/s00703-020-00772-1

[21] Saedi, A., B. Saghafian, S. Moazami, and S. Aminyavari. "Performance evaluation of sub-daily ensemble precipitation forecasts." Meteorological Applications 27.1 (2020): e1872. https://doi.org/10.1002/met.1872

[22] Hoghoughinia, K., B. Saghafian, and S. Aminyavari. "Analysis of precipitation temporal pattern of sub-daily ECMWF forecasts." Meteorology and Atmospheric Physics 134.5 (2022): 87. https://doi.org/10.1007/s00703-022-00920-7

[23] Piani, C., J.O. Haerter, and E. Coppola. "Statistical bias correction for daily precipitation in regional climate models over Europe." Theoretical and Applied Climatology 99.1–2 (2010): 187–192. https://doi.org/10.1007/s00704-009-0134-9

[24] Liu, J. and Z. Xie. "BMA probabilistic quantitative precipitation forecasting over the Huaihe Basin Using TIGGE multimodel ensemble forecasts." Monthly Weather Review 142.4 (2014): 1542–1555. https://doi.org/10.1175/MWR-D-13-00171.1

[25] Xu, Y. "Hyfo: hydrology and climate forecasting R package for data analysis and visualization." R Package Version 1.0 (2015).

[26] Coron, L., G. Thirel, O. Delaigue, C. Perrin, and V. Andréassian. "The suite of lumped GR hydrological models in an R package." Environmental Modelling & Software 94 (2017): 166–171. https://doi.org/10.1016/j.envsoft.2017.03.010

[27] Michel, C. "Hydrologie Appliquee Aux Petits Bassins Ruraux (Applied Hydrology for Small Catchments)." Internal Report; Cemagref Antony: France (1991).