Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Enhancing the Robustness of Temperature Simulations in India through a Bias-Corrected Multi-Model Ensemble Framework

Avijit Paul

engr.avijit.paul@gmail.com

Central Institute of Technology Kokrajhar

Monomoy Goswami

Central Institute of Technology Kokrajhar

Short Report

Keywords: CMIP6, Multi-model ensemble, Temperature variability, Granger–Ramanathan Averaging, India

Posted Date: April 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9094575/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Theoretical and Applied Climatology on June 15th, 2026. See the published version at https://doi.org/10.1007/s00704-026-06362-0.

Enhancing the Robustness of Temperature Simulations in India through a Bias-Corrected Multi-Model Ensemble Framework

Avijit Paul* and Monomoy Goswami

Department of Civil Engineering, Central Institute of Technology Kokrajhar, Kokrajhar, India

Email: engr.avijit.paul@gmail.com; monomoy99@hotmail.com

*Corresponding author

Abstract

Reliable regional-scale temperature projections are essential for climate impact assessment and strategizing remedial measures for future. The complex physiography, which is the characteristics of mainland India, introduces large uncertainties in Global Climate Model (GCM) outputs for temperature analysis over the vast expanse of this geographical entity. This study presents a framework for reducing the uncertainties of GCM outputs by applying a multi-model ensemble (MME) technique using CMIP6 Bias Corrected GCM outputs. Monthly maximum, mean, and minimum temperatures for the period 1965–2014 were evaluated against ERA5 reanalysis data. Five Multi-Model Ensemble (MME) techniques—Equal Weight Averaging, Bates–Granger Averaging, Granger–Ramanathan Averaging (GRA), Mallows Model Averaging, and Bayesian Model Averaging—were systematically assessed using k-fold cross-validation and full-sample calibration at multiple spatial resolutions. Results show that GRA consistently outperforms other techniques by minimizing mean absolute error and maximizing the index of agreement across all temperature variables. The proposed framework enhances the robustness of temperature simulations and provides a strong foundation for future climate projections under different emission scenarios.

Keywords: CMIP6; Multi-model ensemble; Temperature variability; Granger–Ramanathan Averaging; India

1. Introduction

Temperature is one of the fundamental climate variables that serves as an indicator of climate change across the globe, and governs the environment, agriculture, water resources, energy demand etc. (Frederick and Major 1997) in local scales. Modelling the temperature variable is therefore of utmost importance for understanding future values of temperature and its trends, predicting future scenarios of local climates, and assessing likely impacts of the future climate on various sectors. Using the outputs from such a model, policy makers can develop effective adaptation and mitigation strategies to ensure sustainable development and well-being for a country's population.

Different GCMs developed under 6th phase of a Coupled Model Intercomparison Project (CMIP6), has been a global scientific effort that provides a large collection of climate model simulations that are in use to study and predict how the Earth's climate might change in the future (Tebaldi et al. 2021). Different models can give different simulated outputs thereby resulting in uncertainties (Jose and Dwarakish 2020). These uncertainties in climate model projections stem from several factors that include natural variability of climate systems and differences in how models simulate complex processes like cloud formation and natural feedback mechanisms (Ceppi et al. 2017). While the GCMs under CMIP6 offer a much better understanding of global climate trends, there is still work to be done towards improving regional-level forecasts that are crucial for countries like India because of its vulnerability to climate change and because of the need for detailed information for planning and decision-making.

Multi model ensembles (MMEs) aims at combining outputs from GCMs, by assigning weights to different models depending on their individual performances in simulating reference data. Thus, MMEs attempt to enhance the reliability of projections by reducing systematic and random errors of the individual GCMs and influences of model's outputs that are outliers, and consequently highlight consistent trends across simulations thereby bringing in essential improvements for reliably conducting climate risk evaluations and informing evidence-based adaptation for policy development. While many researchers address climate modelling uncertainty using single procedures, combining multiple methods such as bias correction (BC) of outputs and MMEs yields more consistent results (Patel et al. 2025). This study adopts an integrated framework utilizing these combined approaches of the BC developed previously by authors in (Paul and Goswami 2025), and devising of an efficient MME technique. In doing so, the authors used Equal Weight Averaging (EWA) (Shamseldin et al. 1997), Bates-Granger averaging (BGA) (Bates and Granger 1969), Granger-Ramanathan averaging (GRA) (Granger and Ramanathan 1984) and Mallows model averaging (MMA) (Hansen 2007) for finding the weights of MMEs. Whereas these methods were widely used in hydrologic applications (Najafi and Moradkhani 2016) and in combining soil property raster (Wang et al. 2022), these methods were not used in devising MMEs of GCM outputs in climate modelling. Besides the above four, the authors also used Bayesian Model Averaging (BMA) (Hoeting et al. 1999) as an additional MME technique, this technique being a well-established method of reducing biases and enhancing the accuracies of ensemble projections (Talukder et al. 2025) in climate modelling.

2. Study Area and Datasets

2.1 Study Area

In this study, the Indian mainland, extending over a 3.287 million km² area between latitudes 6°N and 36°N and longitudes 68°E and 98°E approximately serves as the focus area for evaluation of model

performance. Geographically, India ranks as the seventh-largest nation globally that has the Himalayas from the north to the northeast, the Bay of Bengal forming the eastern coastline, the Arabian Sea forming the western coastline, and the Indian Ocean bounding the south.

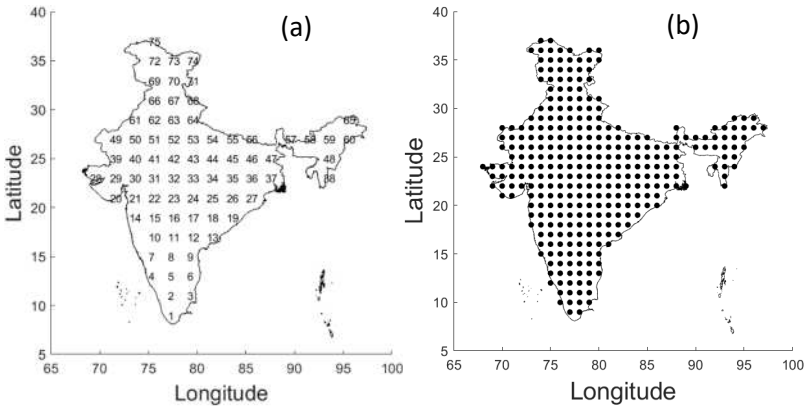


Figure 1 Mainland India showing (a) 75 grid points at 2°× 2° resolution and (b) 316 grid points at 1°× 1° resolution

In order to compare the performance of different MME techniques considered in this study by applying k-fold cross-validation for training and validation of these techniques, i.e. models, 75 grid points covering the mainland India at 2° × 2° resolution were considered as shown in Figure 1(a). Additionally, a relatively large subset of 316 grid points at 1° × 1° resolution as shown in Figure 1(b) was considered for comparing the performance of these models using full-sample calibration. The selections of the subsets were made from the considerations of less computational load and short processing time.

2.2 Datasets

In order to devise, analyse and evaluate different MME techniques of EWA, BGA, MMA, GRA and BMA, outputs from 17 GCMs under CMIP6 were used. A list of these GCMs is provided in Table A-1 as an Annexure. The European Centre for Medium Range Weather Forecasting (ECMWF)'s fifth-generation reanalysis product (ERA5) was employed as the reference dataset for the three temperature variables over the study area during the period considered from 1965 to 2014 as these data were found to perform well in several earlier temperature related studies (Pang et al. 2025).

3. Methodology

Initially, all GCM outputs are bias-corrected following the methodology established by (Paul and Goswami 2025) and followed by the application of the MME technique as discussed below.

3.1 Multi Model Ensemble Techniques

The primary objective of applying an MME technique is to assign weights to a selection of component models so that the combined weighted outcome provides a more accurate prediction of the observed system behaviour than any single model, i.e. component, within the ensemble. Accordingly, an MME is devised by adopting a generalized formulation given in equation (1):

$$Y_t = \sum_{i=1}^{k} W_i X_{i,t} \quad \ldots (1)$$

where, Y= (Y_1,Y_2, …Y_n) is a column vector of the ensembled data at a particular grid point with 'n' number of monthly data. For example, for a dataset of 50 years, n=600.

W=(W_1,W_2, …W_k) is a column vector of the weights of 'K' number of components, i.e. GCM models. In this study K=17.

X is matrix of (n x k) dimensions of individual GCM data points of a month t=(1,2,…n) at a grid point. Following are the details of the selected procedures for devising MMEs in this study:

Equal Weight Averaging (EWA)

In this method, all models are assigned equal weights, which can be calculated mathematically as:

$$W_{EWA} = \frac{1}{k} \quad \ldots (2)$$

Where, W_{EWA} is the estimated weight of each individual GCM by EWA method.

Bates-Granger averaging (BGA)

The BGA approach assign weights by minimizing the root mean squared error between the GCM Outputs and reference data, and is calculated as:

$$W_{BGA,k} = \frac{\frac{1}{\lambda_k}}{\sum_{k=1}^{K} \lambda_k^2} \quad \ldots (3)$$

where $W_{BGA,i}$ is the estimated weight of each individual GCM by BGA method and λ_i is Root Mean Square Error of the corresponding GCM.

Granger-Ramanathan averaging (GRA)

The GRA method applies the ordinary least squares (OLS) technique to determine the weight of each GCM, thereby reducing the sum of squared errors (SSE). The weights for each GCM are computed using the following equation:

$$W_{GRA} = (G^T G)^{-1} G^T R \quad \ldots (4)$$

where G is the matrix of (t x n) dimension, t is number GCM outputs in months, G^T is the transpose matrix of G and R is the matrix of (t x 1) dimension of reference data

Mallows model averaging (MMA)

Hjort and Claeskens,(2003) adopted a frequentist approach to model ensemble against the more common Bayesian approach, the main concept of this approach is that no single model can universally be considered the best, rather, the notion of an optimal model is subjective and depends on the specific objective under consideration. Extending this approach Hansen, (2007), demonstrated that combining models using Mallows' criterion asymptotically produces forecasts with the lowest achievable mean squared error and the mathematical proof of this approach is provided by Wan et al., (2010) which assumes a linear regression framework. The weights are calculated by minimizing Mallows' criterion which is the penalised sum of squared residuals as objective function mentioned below:

$$C_n = \sum_{t=1}^{n} (Y_t - X_t W)^2 + 2\alpha^2 \sum_{k=1}^{K} W_k p_k \quad \ldots (5)$$

where, C_n is the p_k is the number of parameters of the k^{th} GCM (in this study number of parameters of all the GCMs have been considered equal since exact number of parameters used in the GCMs is not available in the literature); α^2 is with smallest root mean square error among the GCMs (Diks and Vrugt 2010), since equal number of parameter is assumed for each GCM. The weights of the GCMs according to MMA that belong to the real space are calculated by minimising the above equation using Differential Evolution Adaptive Metropolis (DREAM) and adaptive Markov chain Monte Carlo (MCMC) algorithm (Vrugt 2016).

$$W_{MMA} = \arg\min C_n \quad \ldots (6)$$

Bayesian Model Averaging (BMA)

The BMA techniques explicitly accounts for the uncertainty associated with each GCM's output (Raftery et al. 2005) while assigning weights. This technique is widely used in hydrology, climate studies and meteorology (Talukder et al. 2025). BMA estimates the posterior distribution of forecasting variables by weighting each model's posterior distribution according to its probabilistic likelihood. Models that demonstrate superior predictive performance are assigned higher weights, while those

performing poorly receive lower ones. Therefore, this method assigns weights to all candidate models based on their statistical evidence, which reflects both their independence and predictive capability. Weights are calculated by the following formula:

$$W_{BMA,k} = p(X_k|R) = \frac{p(R|X_k)p(X_k)}{\sum_{k=1}^{K} p(R|X_k)p(X_k)} \quad \ldots (7)$$

where, $p(X_k|R)$ is the posterior probability of GCM model 'X' in comparison with referenced data (R) set which measures the accuracy of the model prediction. $p(X_k)$ is the prior probability of the model. $p(R|X_k)$ is the marginal likelihood of the GCM model which means the probability of the reference data, if the model is assumed to be true model.

3.2 Training & Validation

For training and validation of MME techniques, i.e. models, used in this study, k-fold cross-validation scheme adopted. In k-fold cross-validation, dataset is divided into k equal parts. The process involves k training–validation cycles, where in each cycle, a distinct fold is set aside for validation, and the remaining (k – 1) folds are used for model training (Refaeilzadeh et al. 2009). In this study, each set containing 50 years of data of the three temperature variables T_{max}, T_{mean} and T_{min} was segmented into five folds of 10 years each. In this scheme, the first fold comprises data of the 10 years 1965-1974 for validation, and the remaining 40 years of data being used for model training. Accordingly, the consideration of five folds ensures that at least 80% data is used for training and the remaining 20% is used for validation for each training-validation cycle of a model. The folds also ensure that the data for model training are drawn from different time periods rather than on a continuous chronological sequence, and every record in the dataset is used for model training in at least one iteration.

At each grid point of the 75 grid-point subset at 2° × 2° resolution, in each cycle of the k-fold cross-validation scheme comprising sequentialized data of four folds for training each of the five selected MME techniques, all 17 GCMs were considered as component models, and the weights of these component models in simulating the sequentialized reference data of the folds were estimated for each MME. The weights thus estimated were then applied to the remaining fold, designated for validation, to compute the MME outputs. The selected performance assessment indices, namely MAE and IoA, were then evaluated for each MME technique using the MME outputs and the corresponding reference data for the fold for validation for each of the three temperature variables. This process was repeated, until all the folds were used for training and validation. The relative performances of the different MME techniques in simulating reference data were evaluated by studying the values of two performance assessment indices as described in the following subsection.

Noting from analysis that the levels of performance of two MME techniques in k-fold cross-validation were comparable, full-sample calibration involving the whole 50 years of historic data were considered

for identifying the best-performing MME technique. This procedure involved a single run of each MME technique, unlike k training-validation cycles, and hence runs, in the case of k-fold cross-validation, for evaluating the performance of the selected MME techniques, and was applied on the subset of 316 grid points at $1° \times 1°$ resolution.

3.3 Model performance assessment indices

Assessing the performance of 5 different Multi model ensemble methodologies and their consistency with reference data is essential for reliable climate change projections. Suitable performance assessment indices provide a comprehensive measure of how effectively MMEs reproduce observed climate patterns.

For model performance assessment, an error-based index, namely Mean Absolute Error (MAE), given by equation (8) and its descriptive statistics were used.

$$MAE = \frac{1}{n}\sum|t_r - t_g| \quad \ldots (8)$$

Here n is the length of data, and t_r and t_g are the reference and GCM output data respectively.

Additionally, modified index of agreement (IoA) as given by equation (9) was used as a measure of relative error in which the 'errors and differences are given their appropriate weighting, not inflated by their squared value' (Legates and McCabe Jr. 1999).

$$IoA = 1 - \frac{\sum|t_r - t_g|}{\sum(|t_g - \bar{t_r}| + |t_r - \bar{t_r}|)} \quad \ldots (9)$$

Here, n, t_r and t_g are as defined for equation (8), and $\bar{t_r}$ is the mean of the n values of t_r.

4. Results and Discussions

Having trained each of the five MME techniques by 5-fold cross-validation by using data of each of the three temperature variables at each of the 75 grid points, the relative performances of these techniques in validation were examined using the performance assessment indices, MAE and IoA. The results of this analysis are presented in combined violin and histogram plots in Figure 2.

Each plot in Figure 2 shows the location of the central tendency in terms of median of each performance assessment index for each validation fold as a small circle. It may be seen from this figure that the violin distributions for GRA and MMA are notably short and are centred on low MAE and high IoA values relative to those for other three MME techniques, thereby indicating both small spread of error and superior central tendency in model performance across folds and variables. This pattern of the values of performance assessment indices is most pronounced for T_{max} and T_{mean} where GRA consistently attains the lowest median MAE and highest median IoA values in all five folds. For T_{min}, the differences in values of the performance assessment indices among methods are smaller; yet GRA maintains a

modest advantage in terms of reduced spread and consistently high median IoA for each fold. Thus, it emerges from cross-validation phase that GRA produces more robust and reliable ensemble.

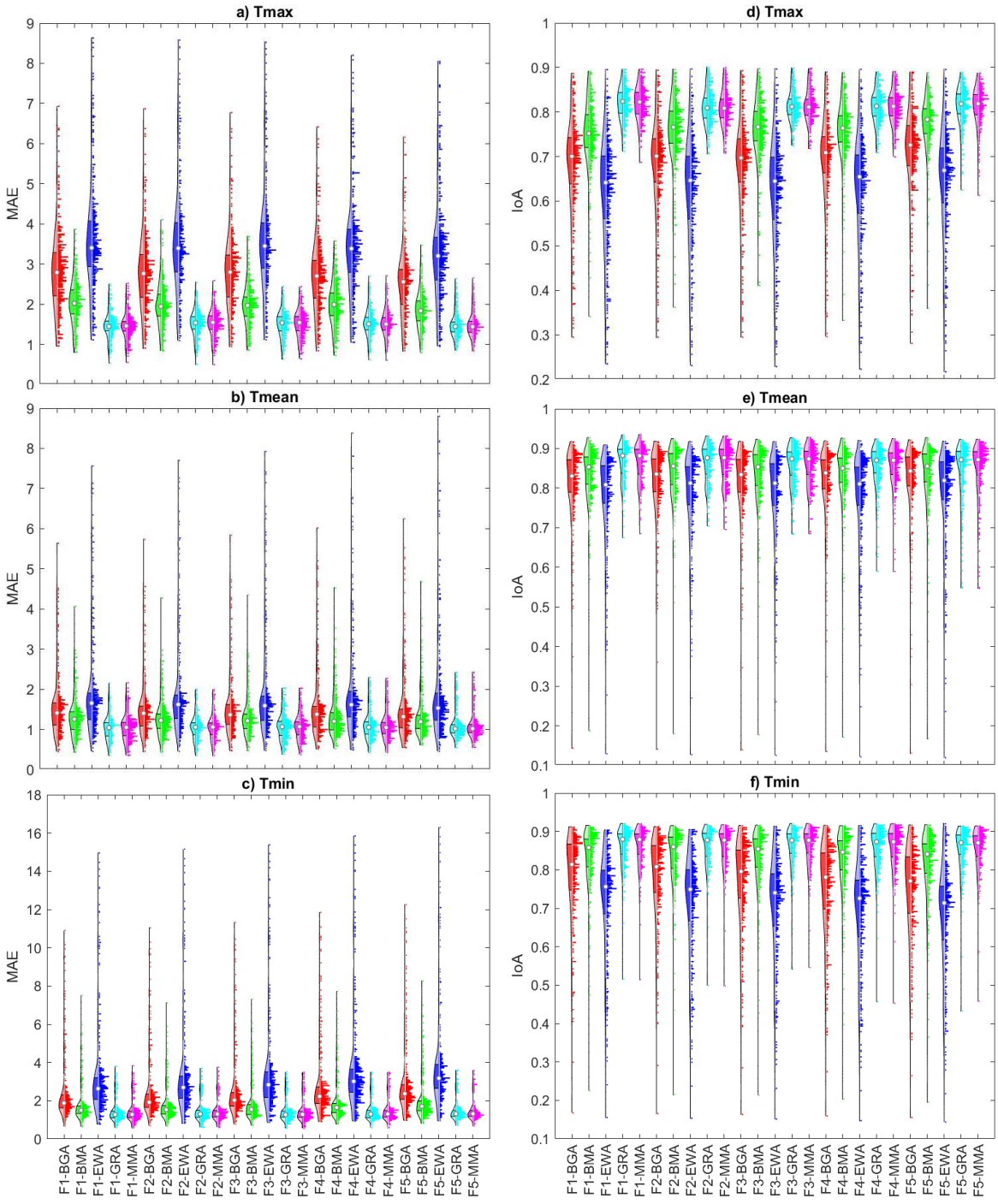


Figure 2 Violin (on the left) and histogram (on the right) plots of MAE and IoE produced in validation by each of the five selected MME techniques in each of the five validation folds for (a) & (d) T_{max} , (b) & (e) T_{mean} and (c) & (f) T_{min} respectively with F1, F2, F3, F4 and F5 indicating the first, second, third, fourth and fifth validation folds respectively.

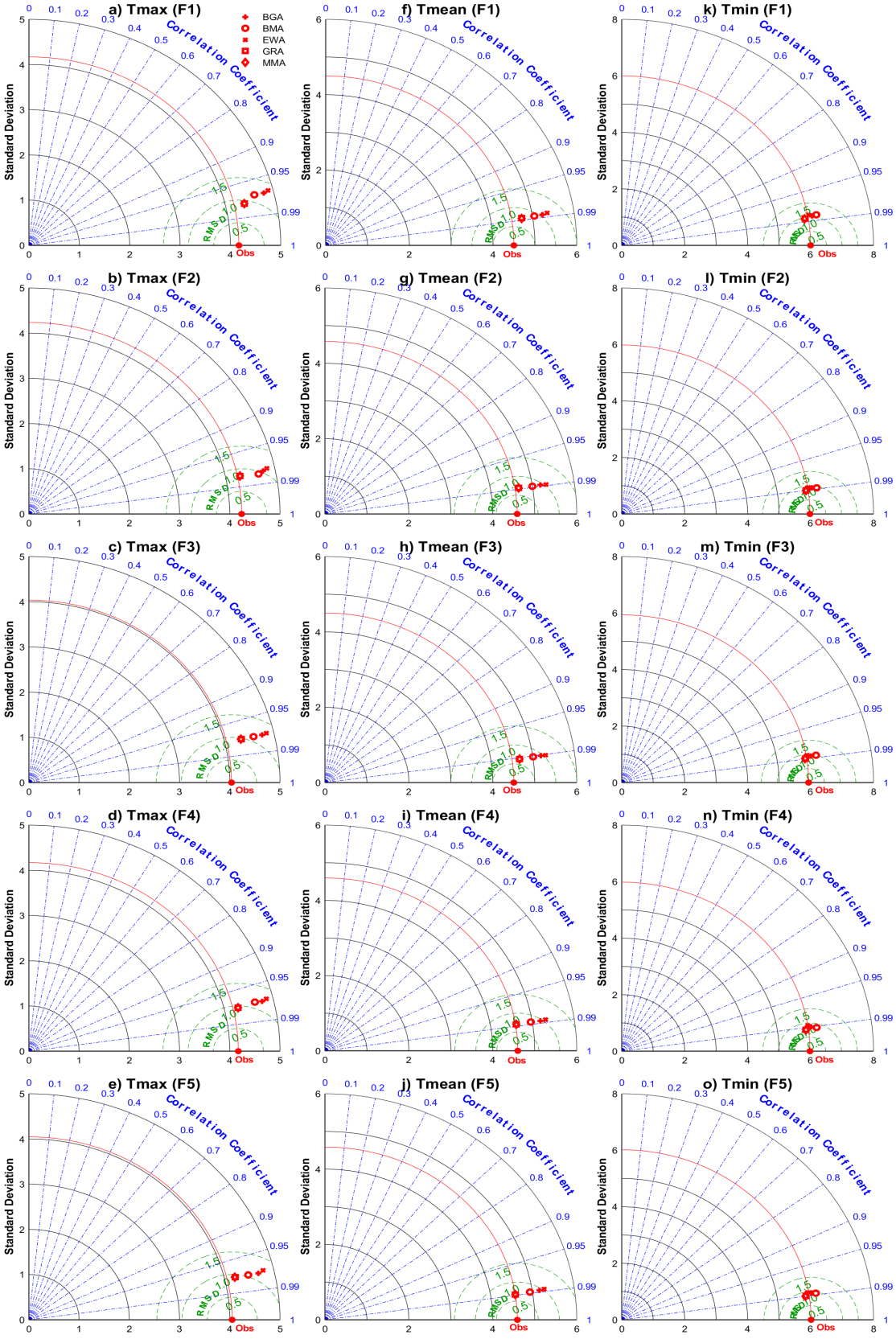


Figure 3 Taylor Diagrams of the MME techniques designated by legends on the plot on the top left for (a) to (e) T_{max}, (f) to (j) T_{mean} and (k) to (o) T_{min} validation folds indicated by F1, F2, F3, F4 and F5 respectively

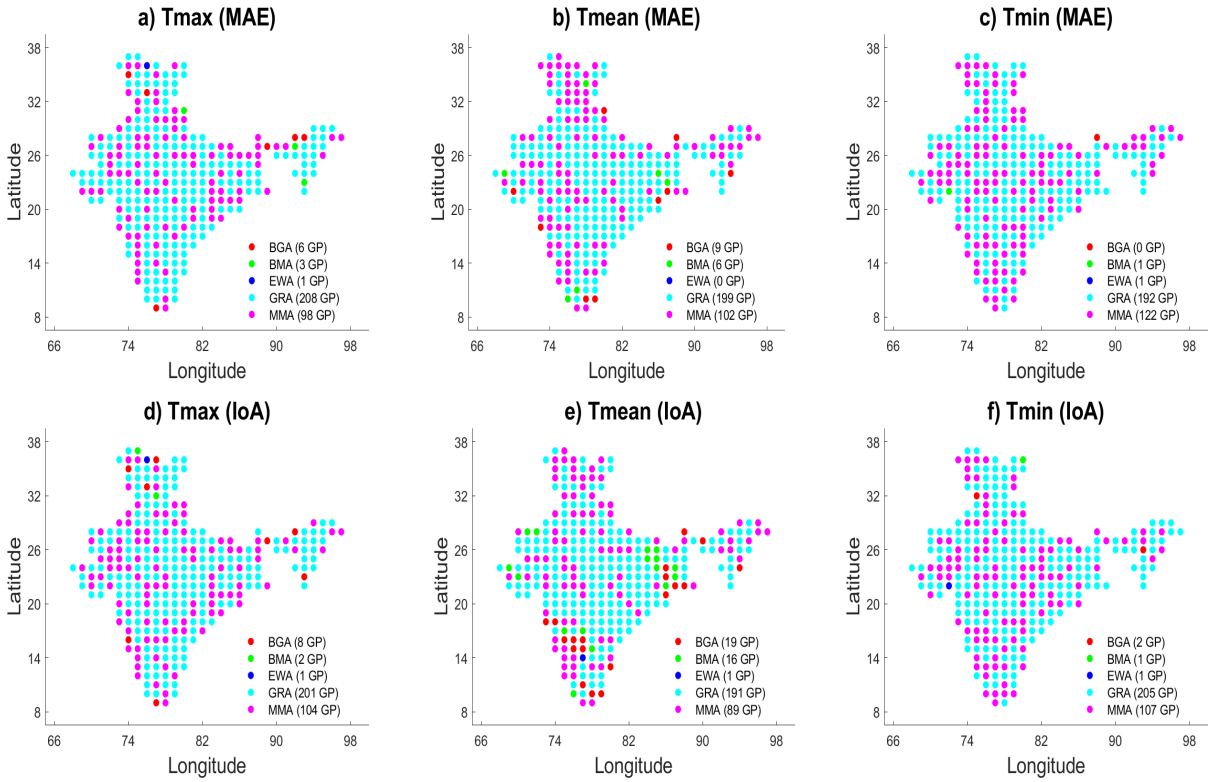


Figure 4 Values of MAE and IoA for the five MME techniques at 316 Grid Points (GPs) at 1°×1° spatial resolution for (a) & (d) T_{max}, (b) & (e) T_{mean} and (c) & (d) T_{min} respectively. (Note: the bracketed number alongside the legend of each MME technique indicates the number of GPs in which an MME performed the best)

In order to evaluate a spatially aggregated measure of performance, the MME techniques were applied on the data of each temperature variable by flattening, i.e. rearranging data of the two-dimensional grid points into a one-dimensional vector, i.e. a single series. Taylor diagrams were then produced as shown in Figure 3 by plotting the Correlation Coefficient (CC), centred Root Mean Square Error (cRMSE) and Standard Deviation (SD) between the reference data and the simulated outputs from each MME technique for each of the five validation folds for each of the three temperature variables.

It may be observed from the Taylor Diagrams in Figure 3 that all MME techniques for all folds and all temperature variables achieve CC values between 0.95 and 0.99 and cRMSE values closely packed around 1°C. However, in terms of SD values, the performances of both GRA and MMA techniques are comparable, and are relatively close to the corresponding values of the reference data for each temperature variable in comparison with other three techniques. The closeness of the SD values for GRA and MMA are particularly consistently pronounced in the case of the second of the five cross-validation folds being considered as the validation fold as may be seen from the plots numbered (b), (g) and (l) of Figure 3.

Noting that the performance of both the GRA and the MMA techniques of MME are comparable in most of the tests of cross-validation on both gridded and flattened data sets, efforts were next made to find if distinction between the best-performing techniqes could be found by considering full-sample calibration. Accordingly, the performance indices MAE and IoA for each of the five MME techniques applied at each of the 316 grid points across mainland India at $1° \times 1°$ spatial resolution for each of the three temperature variables were evaluated. Each plot in Figure 4 shows these 316 grid points acrosss mainland India by dots with each dot shown with a unique colour representing an MME technique yielding best performance in terms of either the least MAE or the largest IoA values at a grid point. The number of grid points of a particular colour corresponding to a best-performing MME technique is also entered alongside the legend of the MME on each plot. It may be found from this figure that the GRA techniques perfomed best amongst all five MME techniques in most locations for all three variables in terms of both performance assessment indices. It is also evident that the GRA technique performed much better in the Central India region both for T_{max} and T_{mean}, while MMA teachnque, which is the second best performing teachnique, performed relatively better in the case of T_{mean}, generally in the northern and sourthern regions of India. Overall, it can be concluded that the GRA technique generally outperforms the other MME techniques when applied to three temperature variables across mainland India. The above observation about the GRA performning best amongst different MME techiques is also found in other studies, albeit in different fields of application. (Nand et al., (2025) found that GRA demonstrated superior performance in minimizing variance in crop simulations by effectively incorporating variance reduction techniques. Kumar et al., (2015) found GRA as being the most effective MME technique for integrating simulated river discharge outputs from eight hydrological models. The use of GRA offers a significant advantage over BMA and MMA, as it provides a straightforward approach for determining weights. In contrast, identifying optimal weights in BMA and MMA involves more complex and computationally demanding procedures, such as the DREAM and MCMC algorithms. Although an MME provides an opportunity to learn from errors across different individual component models and enhance overall model performance, certain individual models may still surpass the ensemble mean or median in accuracy (Kothari et al. 2022).

5. Conclusions

This study comprehensively assessed the efficiency of different MME techniques towards improving the accuracies of simulations and projections of three temperature variables over mainland India using outputs from 17 selected GCMs of CMIP6 phase. The evaluation incorporated rigorous k-fold cross-validation and multiple spatial resolutions to ensure robust model performance assessment across the T_{max}, T_{mean} and T_{min} variables. Among the five ensemble techniques tested, the GRA technique consistently demonstrated superior skill in minimizing MAE and enhancing the IoA across nearly all grid points covering mainland India, and all folds of k-fold cross-validation procedure, over a long 50-

year period 1965-2014 of data. The GRA's methodological simplicity stemming from its direct and deterministic approach to the determination of weights provided a notable computational advantage over more complex methods, namely BMA and MMA, which rely on iterative algorithms like DREAM and MCMC. The study reaffirmed findings from earlier works that GRA effectively reduces variance within ensemble simulations while maintaining stability and consistency across temporal and spatial scales.

Author Contributions

The first author conceptualized, analysed, and prepared graphical outputs of the research. The second author suggested improvements on the methodology and presentation of results, and prepared the manuscript.

Conflict of Interest Statement

The authors declare no conflict of interest.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Funding

This research received no funding from external sources.

Data Availability Statement

Raw simulated outputs of GCMs of the CMIP6 phase are available at https://esgf-node.llnl.gov/projects/cmip6/

Acknowledgement

The authors are thankful to the anonymous reviewers for their valuable suggestions and insightful comments.

ORCID

Avijit Paul: 0000-0001-9388-9571

Monomoy Goswami: 0000-0003-1380-6276

References

Arsenault R, Gatien P, Renaud B, et al (2015) A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. Journal of Hydrology 529:754–767. https://doi.org/10.1016/j.jhydrol.2015.09.001

Bates JM, Granger CWJ (1969) The Combination of Forecasts. OR 20:451–468. https://doi.org/10.2307/3008764

Ceppi P, Brient F, Zelinka MD, Hartmann DL (2017) Cloud feedback mechanisms and their representation in global climate models. WIREs Climate Change 8:e465. https://doi.org/10.1002/wcc.465

Diks CGH, Vrugt JA (2010) Comparison of point forecast accuracy of model averaging methods in hydrologic applications. Stoch Environ Res Risk Assess 24:809–820. https://doi.org/10.1007/s00477-010-0378-z

Frederick KD, Major DC (1997) Climate Change and Water Resources. Climatic Change 37:7–23. https://doi.org/10.1023/A:1005336924908

Granger CWJ, Ramanathan R (1984) Improved methods of combining forecasts. Journal of Forecasting 3:197–204. https://doi.org/10.1002/for.3980030207

Hansen BE (2007) Least Squares Model Averaging. Econometrica 75:1175–1189. https://doi.org/10.1111/j.1468-0262.2007.00785.x

Hjort NL, Claeskens G (2003) Frequentist Model Average Estimators. Journal of the American Statistical Association 98:879–899. https://doi.org/10.1198/016214503000000828

Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian Model Averaging: A Tutorial. Statistical Science 14:382–401

Jose DM, Dwarakish GS (2020) Uncertainties in predicting impacts of climate change on hydrology in basin scale: a review. Arab J Geosci 13:1037. https://doi.org/10.1007/s12517-020-06071-6

Kothari K, Battisti R, Boote KJ, et al (2022) Are soybean models ready for climate change food impact assessments? European Journal of Agronomy 135:126482. https://doi.org/10.1016/j.eja.2022.126482

Kumar A, Singh R, Jena PP, et al (2015) Identification of the best multi-model combination for simulating river discharge. Journal of Hydrology 525:313–325. https://doi.org/10.1016/j.jhydrol.2015.03.060

Legates DR, McCabe Jr. GJ (1999) Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. Water Resources Research 35:233–241. https://doi.org/10.1029/1998WR900018

Malone BP, Minasny B, Odgers NP, McBratney AB (2014) Using model averaging to combine soil property rasters from legacy soil maps and from point data. Geoderma 232–234:34–44. https://doi.org/10.1016/j.geoderma.2014.04.033

Najafi MR, Moradkhani H (2016) Ensemble Combination of Seasonal Streamflow Forecasts. Journal of Hydrologic Engineering 21:04015043. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001250

Nand V, Qi Z, Ma L, et al (2025) Evaluation of multimodel averaging approaches for ensembling evapotranspiration and yield simulations from maize models. Journal of Hydrology 661:133631. https://doi.org/10.1016/j.jhydrol.2025.133631

Pang N, Kong H, Wong C, et al (2025) A CNN-Based Downscaling Model for Macau Temperature Prediction Using ERA5 Reanalysis Data. Applied Sciences 15:5321. https://doi.org/10.3390/app15105321

Patel G, Das S, Das R (2025) A Comparative Approach to Understand the Performance of CMIP6 Models for Maximum Temperature near Tropic of Cancer Using Multiple Machine Learning Ensembles. Water Resour Manage 39:3921–3936. https://doi.org/10.1007/s11269-025-04137-2

Paul A, Goswami M (2025) An Adaptive Neuro-Fuzzy Inference System (ANFIS) and a Wavelet-Integrated ANFIS (WANFIS) for Univariate Bias-Correction of GCM -Simulated Temperature. Intl Journal of Climatology e8816. https://doi.org/10.1002/joc.8816

Raftery AE, Gneiting T, Balabdaoui F, Polakowski M (2005) Using Bayesian Model Averaging to Calibrate Forecast Ensembles. https://doi.org/10.1175/MWR2906.1

Refaeilzadeh P, Tang L, Liu H (2009) Cross-Validation. In: Encyclopedia of Database Systems. Springer, Boston, MA, pp 532–538

Shamseldin AY, O'Connor KM, Liang GC (1997) Methods for combining the outputs of different rainfall–runoff models. Journal of Hydrology 197:203–229. https://doi.org/10.1016/S0022-1694(96)03259-3

Talukder A, Shaid S, Hwang S, et al (2025) Optimizing the multi-model ensemble of CMIP6 GCMs for climate simulation over Bangladesh. Sci Rep 15:11343. https://doi.org/10.1038/s41598-025-96446-0

Tebaldi C, Debeire K, Eyring V, et al (2021) Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6. Earth System Dynamics 12:253–293. https://doi.org/10.5194/esd-12-253-2021

Vrugt JA (2016) Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. Environmental Modelling & Software 75:273–316. https://doi.org/10.1016/j.envsoft.2015.08.013

Wan ATK, Zhang X, Zou G (2010) Least squares model averaging by Mallows criterion. Journal of Econometrics 156:277–283. https://doi.org/10.1016/j.jeconom.2009.10.030

Wang J, Zhao X, Deuss KE, et al (2022) Proximal and remote sensor data fusion for 3D imaging of infertile and acidic soil. Geoderma 424:115972. https://doi.org/10.1016/j.geoderma.2022.115972

ANEXURES

Annexure A-1 List of the GCMs of CMIP6 considered for developing an MME

Sl. No.	Model name
1	ACCESS-CM2
2	ACCESS-ESM1-5
3	AWI-CM-1-1-MR
4	CMCC-ESM2
5	CAS-ESM2-0
6	EC-Earth3-Veg
7	EC-Earth3-Veg-LR
8	EC-Earth3
9	FGOALS-g3
10	GFDL-ESM4
11	INM-CM4-8
12	INM-CM5-0
13	IPSL-CM6A-LR
14	MIROC6
15	MPI-ESM1-2-HR
16	MPI-ESM1-2-LR
17	MRI-ESM2-0